EVOLUTION OF STUDENT EVALUATIONS OF COLLEGE FACULTY:

IMPLICATIONS AND CONSEQUENCES

By

José Vázquez

A dissertation submitted to the Graduate Faculty in Urban Education in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2008

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

This manuscript has been read and accepted for the Graduate Faculty in Urban Education in satisfaction of the dissertation requirements for the degree of Doctor of Philosophy

Steven M. Cahn

Date | Chair of Examining Committee

Anthony Picciano

Date | Executive Officer

Stephan Brumberg

Lynne Weikart
Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

EVOLUTION OF STUDENT EVALUATIONS OF COLLEGE FACULTY:

IMPLICATIONS AND CONSEQUENCES

By

José Vázquez

Adviser: Steven M. Cahn

Student evaluations of college faculty have been in place for about eight decades. However, these evaluations gained momentum during the 1960s and subsequently when the political climate across college campuses in the U.S. demanded major educational reforms to benefit students. The practice of having students evaluate their instructors has become a ritual at the end of each term in almost every college classroom across the nation. Student evaluations have evolved to become an important piece in judging the quality of college faculty for re-appointment, promotion, and tenure.

Using qualitative methods, I have studied forty non-tenured faculty members at two urban universities. The study shows that student evaluations have affected how instructors' teaching and grading practices. In addition, a policy of peer observation has been formulated. The consequences and implications of the evolution of student evaluations are that those evaluations have become a centerpiece in the faculty evaluation process and college teaching is not being adequately evaluated. The teaching of college faculty should be evaluated by their peers just like research is. Effective policies need to be formulated and implemented to re-conceptualize that undergraduate education's

primary function is to enhance learning and not to satisfy student demands as if higher education were a consumer item.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I: INTRODUCTION

RATIONALE FOR RESEARCH

I think there is a fundamental prima-facie reason for examining the perceptions of tenure-track, non-tenure and adjunct faculty members at a large research university. Tenure-track faculty members depend on student evaluations since those are part of the holistic evaluation system used by their department for tenure decisions. Non-tenure and adjunct faculty members are evaluated primarily on the results of student evaluations for their reappointment and/or merit pay raise. The situation of those instructors without tenure makes them more vulnerable to negative course evaluations, and the possibility of engaging in grade inflation is more likely to take place.

As colleges and universities increase the number of instructors who are temporary (non-tenure) or part-time (adjuncts), the possible relationship between job security and teaching/grading practices becomes more important. These types of instructors provide colleges and universities the opportunity to save substantial amounts of money by not providing benefits (for adjuncts) and to decrease their workforce when economic challenges emerge. The hiring of adjuncts in particular provides an enormous flexibility to the academic institution for hiring instructors for as short as a semester based on course enrollments.

Student evaluations of teaching continue to be a centerpiece in postsecondary education more than eighty years after the first documented evaluation of instructors was

developed at the University of Washington in 1924. In 2007 they continue to be used widely on college and university campuses across North America and increasingly throughout the world as sources of feedback on instructional effectiveness for a faculty member. Those evaluations serve as tools for a variety of purposes: for student course selection; as evidence of teaching effectiveness for promotion, reappointment and tenure decisions; as a criterion of academic program effectiveness; for research on evaluation of teaching; and as a constant source of heated debate in departmental and/or institutional policies.

Despite the long time since the first evaluations of college teaching were implemented, in a survey conducted in 1993 (Seldin, 1993), indicated that 86 percent of postsecondary institutions reported using them, an increase of 18 percent since 1983. In addition, Seldin reported that no other source of data gets more attention in evaluating teaching – not even grade distribution, classroom visits, or other sources of evidence about someone's teaching. Moreover, results from student ratings of instruction are used widely for more purposes than they were originally intended for. As indicated by Guthrie (1954), colleges originally set up rating systems to serve two purposes: to help administrators monitor and assess teaching quality, and to help instructors improve their own teaching. However, over time those two purposes have diverged into multiple purposes. Currently, institutions look at the result of student evaluations of teaching to assess the hiring of new faculty, in annual reviews of current faculty, for faculty reappointment, in institutional accreditation review, and for giving faculty awards and honors.

While faculty can still use the results of student evaluations to assess their own teaching, and documenting their own effectiveness inside and outside their home institution, it is also argued that instructors can change and/or modify teaching and grading practices in order to get favorable student ratings. Students can then use the ratings when choosing which courses to enroll in and in selecting teachers for awards and honors. Although it is true that many college and university instructors view favorably the use of student evaluation of teaching, many instructors aim for better and more reliable measures of their own teaching. There seems to be a sharp division between those who view student evaluations of teaching as democratic participation in evaluation an activity they are so familiar with and in which they have a high stake on and those instructors who view the student evaluations as a powerful weapon to harm those instructors who have high expectations and observe high standards of teaching thus rendering the student evaluations as a personality contest.

CONCEPTUAL FRAMEWORK

Besides questions regarding possible bias in student ratings, there are some legal issues associated with the use of student ratings. Many institutions have developed their own student evaluation of teaching forms. Those forms are usually developed by faculty and administrators (and sometimes students) grouped as a committee. The use of these forms may pose a legal liability for the institution (Arreola, 2000, p. 93). When it comes to faculty evaluation, courts have accepted faculty-determined criteria and standards of

performance provided that there is a consistent application of contractual procedures when evaluating faculty (Kaplan, 1978).

From a legal perspective, postsecondary institutions must exercise compliance with faculty evaluation procedures that are specified in the faculty contract or faculty handbook (Centra, 1999). An institution must follow due procedure for faculty evaluation and should not use criteria considered to be discriminatory such as race, gender, and sexual orientation (Braskamp and Ory, 1999). Consequently, institutions of higher education enjoy an enormous leeway regarding what to measure and how to measure when developing faculty evaluation procedures. All that is left for institutions to follow is a faculty evaluation that does not violate civil rights, and evidence that evaluation is done systematically and consistently (Arreola, p. 93).

In a review of the legal liability in all 50 states, Carr and Padgett (1992) looked at the design and use of student evaluation forms. Their main conclusion was that the primary legal issue revolves around statistical validity and reliability of the forms used. The legal liability issue emerges from using unreliable student evaluation forms for decisions regarding faculty promotion, reappointment, and tenure. The liability of the institution includes charges of discrimination and violation of individual privacy and academic freedom. As stated by Carr and Padgett (p. 69):

> "State laws provide an additional environmental element creating possible legal liability. Of particular concern are the open records laws. Unless exceptions are made, these laws provide that employment records such as salary and

> performance evaluations of public employees are public information. Some colleges and universities publish the results of student ratings of faculty. Some do not. Under open records laws, ratings would have to be made public for a single faculty member or for all faculty members if an individual or group requested this information. In this litigious society liability for publishing student evaluations of faculty members based on forms of dubious statistical validity seems rather apparent."

Out of the 50 states, 17 were found to have no exceptions to the state statutes pertaining to the release of faculty personnel records and 23 states were found to specifically exclude faculty records from open records laws. Institutions located in states that do not have exceptions to public records laws are subject to legal action dealing with the publication of faculty records, including data from unreliable and invalid student evaluation forms (Arreola, p. 94). Even in those states that specifically exclude faculty records from open records there is still a possibility of having public disclosure of faculty records if faculty personnel decisions are challenged in a court of law.

Colleges and universities must validate the design and reliability of a student evaluation form prior to the form being used. Consequently, many institutions rely on commercially available student evaluation forms to ensure the validity and reliability since those forms have undergone extensive psychometric testing during the development process. When a postsecondary institution decides to use its own in-house form, then it must be willing to engage in a one-year period of psychometric testing that ensures the

validity and reliability of such form. By doing so, the institution complies with the legal issues surrounding faculty evaluation. As Carr and Padgett indicate:

> "Civil rights legislation has provided a legal environment in which employment tests and performance evaluations must in most cases be statistically reliable and valid. That is, there must be statistical documentation that they are reliable measures, are job related and predict job performance. Otherwise, the employer's legal defense is likely to be inadequate in cases of discrimination." (p. 68)

Thus, student evaluations may be fast and inexpensive, but there are issues of fairness that continue to be disputed. Issues of objectivity or lack of are abundant in the student evaluations of college faculty literature. An "objective student evaluation of teaching" seems a mere illusion since the process by itself is subjective. In fact, the entire faculty evaluation process discussed and proposed in this study is, by definition, subjective. Having an objective evaluation system of college faculty performance implies that it would produce the same outcome regardless of who the evaluator is. This assumption is impossible to attain.

Rather than focusing on the possibility of an objective faculty evaluation system, effort must be made to reduce the subjectivity of the process. Subjectivity in the faculty evaluation process is unavoidable. One of the goals when developing and implementing an effective policy of higher education faculty evaluation is to have the impact of subjectivity reduced to a minimum or controlled. Proper guidelines for evaluation, adequate training of faculty observers, and a rigorous method for faculty evaluation

should provide a more consistent judgment of the process. One of the main issues reducing the consistency of the faculty evaluation process is the subjective interpretation of what constitutes 'good teaching'. There is no consensus as to what constitutes teaching effectiveness or how to measure it. In the pre-college educational system teaching effectiveness is often associated with learning outcomes demonstrated by students in standardized testing. Postsecondary education does not a mandated system of testing (or at least not yet) and learning outcomes are almost impossible to measure.

Another obstacle to have a less subjective faculty evaluation system is having faculty resistance. Institutional commitment is not sufficient if the faculty members disagree about how the evaluation should be carried out. The most obvious argument is that no one enjoys being evaluated. Very few (if any) faculty members would truly enjoy being constantly evaluated, observed, and then told that they need to improve their teaching performance. Faculty resistance then leads to opposition to a comprehensive evaluation system in places where their voice carries much weight. The most common reason to object to a thorough evaluation process is 'invasion of academic freedom'. Another reason deals with the underlying resentment about not being considered competent enough that a frequent evaluation process needs to be in place. There is also the fear of being evaluated by someone who is not competent enough to make a qualified assessment of someone's teaching performance. Another concerns stems from being evaluated in an area where the faculty member is not very competent. This particular concern is legitimate given that most postsecondary faculty has little or no background in pedagogy and no specific preparation was given during graduate school.

There is a current emphasis on improving college and university teaching performance. Many institutions have established centers for teaching where faculty development is emphasized and there are opportunities for junior faculty to receive professional development. Also, those centers assist graduate teaching assistants in their preparation to become future members. While this effort of paying closer attention to the quality of teaching at a particular institution is laudable, results of any significant improvement are not readily available since those centers are relatively new and longitudinal studies have not been published. Mathis (1974, p. 25) wrote an advice regarding the establishment of those faculty development centers:

> "The Center approach is successful only to the degree that Center programs and staff are responsive to the diverse, and sometimes conflicting needs of the campus. Centers should not staff themselves to reflect any one orthodoxy about teaching. The successful Center should be able to assist the faculty member who is looking for a teaching system to the same degree that it can help a faculty member with sympathetic advice. [Such Centers] ought to avoid academic evangelism as much as possible. The temptation to save natives from themselves through an aggressive program of prophylaxis, usually technological in nature, is generally nonproductive, since it involves programs for the few at the expense of the many. The natives should save themselves, and Centers should be as eclectic as possible in helping them do so."

The statement above seemed prophetic since in many of those centers the advice given is generic, a ‘one-size-fits-all’ approach to teaching. And, as mentioned by Mathis, most advice relies on some technological issue such as using more audiovisuals, for example. Effective teaching is more complex than that, and some general advice is not going to turn an ineffective instructor into an effective one by attending a few workshops or listening to some advice. Effective instruction requires a degree of ‘academic charisma’, which is hard to acquire overnight. Effective teachers are equipped with a set of pedagogical strategies that a center for teaching is not going to give anyone over a short period of time. A good teacher has to be a good communicator, is widely aware of the student to be taught, and uses effective strategies in methodology to present the material in class. Effectiveness in teaching is an acquired taste to a certain extent and requires a developmental process. Since most college instructors have not been adequately prepared in the art of teaching, many can be ineffective and remain ineffective during their entire professorial career.

The development of an effective college instructor requires an entire process of acquiring and improving teaching procedures. For that reason, formative evaluation is essential in the formation of a quality instructor. Formative evaluation can only be done by those who have the expertise in effective teaching, who could evaluate the dynamism, competency, and methodology of peers. Students lack the competence to what is of genuine benefit to their own intellectual development (Gurland, 1978, p. 81). Consequently, a system for evaluating college faculty should consist of peer evaluation for formative purposes as its cornerstone.

Although it is generally agreed that the vitality and effectiveness of higher education is of crucial importance for the future of a nation, there is still little emphasis about how to accomplish effective teaching in a higher education setting. It seems that achieving high quality teaching is not difficult to accomplish given adequate institutional policies that promote and value effective teaching. First, an institution must start by hiring those individuals who have a true vocation for teaching since teaching is a major component in the academic fabric of undergraduate institutions. Second, even those with a true vocation for teaching can benefit from the input of more experienced colleagues who have a genuine interest in perpetuating quality teaching at the departmental or institutional level. Third, effective teaching and effective evaluation of college teaching needs a supporting environment where quality teaching is valued as much as brilliant research. As academic priorities have shifted over the years, evaluation of college teaching has taken a shortcut by leaving the evaluation to the students. Students tend to be generous evaluating professors because they just lack the competence to judge effective teaching, and can be easily biased by the instructor's personality, assigned workload, or expected grade in the course. Only by effective leadership and academic vision can a system for evaluating teaching faculty can be developed and implemented. An effective evaluation system for college teaching can not only assess excellence in teaching but will also be an appropriate tool to fostering the natural inclinations of higher education faculty to excel in teaching. The criteria for teaching excellence varies widely from department to department at the institutional level and such disregard for a consistent method in evaluating teaching continues to place teaching as a secondary

activity after research, primarily at research universities. While research activity is usually judged by peers, teaching quality is judged primarily by students and when evaluated by colleagues it is done in such a way that it does not always lead to improvement of teaching. What some institutions are not assessing is that excellence in teaching leads to student learning, better courses and programs of study, and overall improvement in the teaching of a discipline at the department level and university level.

METHODOLOGY

There were two fundamental goals of this study. One component sought to collect faculty impressions on student evaluations, changes in teaching practices as a result of student evaluations, and changes in grading practices as a result of student evaluations. By obtaining personal accounts of changes implemented by non-tenured faculty members teaching and grading practices, one could then get a better sense of how grading inflation emerges (if it does), and what is the nature of some grade inflation practices. This study, coupled with additional studies supporting grade inflation as a consequence of student evaluations, could help further establish a link between course evaluations of non-tenured faculty and grade inflation . However, establishing causality between student evaluations and grade inflation was not a goal of the study; such goal is beyond the scope of this study.

The second component of the study sought to document instructors' perceptions, attitudes and beliefs about their preferred method for faculty evaluation along with the method being used in their departments. Faculty evaluation systems tend to differ greatly from department to department and from institution to institution, and more information about current practices could inform the existing literature, which in turn could be used to develop future policies regarding how college faculty should be evaluated properly. The existing discrepancy among faculty evaluations at institutions of higher education may be partly responsible for the little value placed on achieving effective college teaching.

I conducted the study at two large research universities, one public and one private. One of the institutions, university A, is located in an urban area of a major city of the northeast. Its undergraduate enrollment is approximately 19,000 students and its undergraduate faculty consists of a combination of tenure-track, non-tenure and adjuncts. The non-tenure faculty members are full-time and work on three-year contracts renewable upon completing a satisfactory performance evaluation. Adjunct faculty members work on semester-long contracts and reappointments are based upon department needs, data from student evaluations, and enrollment figures. The other institution, university B, is located in an urban area of a major city in the northern Midwest. Its undergraduate enrollment is approximately under 20,000 students and its undergraduate faculty consists mostly of tenured and tenure-track instructors. Only a small fraction of the faculty consists of adjunct instructors.

The study sample consisted of twenty instructors from each institution, recruited at random from various departments and subject categories. None of the faculty interviewed included first-year instructors. The sample used includes both men and women, and they were recruited by sending them an invitation letter by e-mail asking them to participate in a short interview about their experiences with student evaluations. All participants were offered complete anonymity. Deans from each institution were contacted prior to the study to secure their permission to contact the faculty. Deans were informed that the name of the institution and other identifying characteristics would not be used, but that in the event of publication of the study results, they would be provided with a copy of the manuscript. Initially, twenty-three faculty members from each institution were approached. The twenty instructors interviewed were chosen based upon the first twenty who responded to the email requesting their participation.

During the interview process, which consisted of about twenty minutes, I collected general impressions using the following interview protocol:

- **Attitude Towards Student Evaluations**

  Tell me your overall reaction towards student evaluations. I'm looking for any specific thought(s) regarding those evaluations.

- **Methods of Teaching**

  Have your teaching methods changed as a result of student evaluation results? If so, please describe those changes.

- **Grading System**

Have your grading system changed based upon student evaluation results? If so, what is the nature of those changes?

- **Suggestions to Improve Faculty Evaluation**
    a. Describe your ideal evaluation system
    b. Do you prefer peer evaluation over student evaluation or both? Why?
    c. If you could change the peer evaluation system used by your institution, what would you change?

The interviews were not audio taped as I felt that it would make the process more comfortable. Upon transcribing the interview notes, each faculty member received a copy of the transcribed interview for their approval. There was a strong interest among the faculty members to participate and some of the participants seemed unaware of the discrepancies in faculty evaluations among departments since they assumed that faculty evaluation is uniform across a particular institution. Many of the faculty interviewed had never taught at other institutions or only had teaching experience as graduate assistants and the only method they knew of is student evaluations. For a large group of faculty the notion of peer evaluation consisted of a classroom visit every few years. The results from these interviews are described in Chapter 3.

After the interview process, all interviews were transcribed and reviewed by each participant, and returned approved. There was some editing done to a few interviews for the purpose of conveying clarity of answers. The participants approved the editing. The interview transcripts were created using only identifying initials and the institution

(university A or B). While each participant's experience differed in substance, the overall categories were generally consistent. Each participant was very familiar with student evaluations results; however, not every participant was well informed about peer observation or had varying degrees of experience being observed in the classroom. Analysis of the survey was qualitative. Survey transcripts were reviewed to identify patterns of consistency in terms of positive or negative experiences with the results of student evaluations, peer observation, teaching practices, and grading practices.

LIMITATIONS

The limitations and challenges that I encountered as I attempted to coordinate and implement this study provided some insights into the overall attitudes of non-tenured faculty about the evaluation system in place. Throughout the interview, I was reminded by many of the instructors interviewed that "this interview is strictly confidential" or that "this statement is off the record" thus showing the lack of satisfaction with evaluation process currently in place in their department along with some fear of retribution by criticizing the evaluation used being used.. The sample size might be considered a limitation but during the interview process I encountered a wide diversity of perceptions to get a broad picture of attitudes about student evaluations, and changes in teaching and grading practices that could indicate grade inflation or lack of it.

Another limitation of the study includes the limitations of some of the responses. None of the faculty interviewed openly admitted to be engaged in grade inflation. In addition a degree of discomfort was perceived when they were asked if they had altered their grading practices as a direct result of student evaluations.

A third and major limitation deals with the limited knowledge about the best evaluation practices assessing college faculty teaching. Although it is understandable that junior faculty members are still learning about the culture of higher education, especially in a large research universe, the lack of thorough knowledge in the evaluation process indicates that there is a greater focus on research and the importance of teaching occupies a secondary role in the academic career of those individuals. At one of the institutions surveyed, those belonging to a group of 'teaching faculty'—individuals not required to engage in research but to focus instead on a large teaching load accompanied by an occasional publication, the teaching evaluation process is still vaguely defined and peer evaluation is something being currently considered. There are faculty members who have spent more than a decade teaching on renewable contracts and are still unaware of the complexity of the peer evaluation system of teaching effectiveness.

CHAPTER PLAN

In Chapter 2, I will focus on the literature review of student evaluations of teaching, including the evolutionary since they were implemented up to the most recent trend in having students post comments on a popular website. In addition, this chapter

will address peer evaluation and the difficulty underlying the concept of categorizing 'good teaching'. Hopefully, this chapter will provide an understanding of the student evaluation debate as their purpose has shifted over the years. Chapter 3 will focus on the analysis of data from the survey of non-tenured faculty at two urban universities. Survey data were grouped to provide a percentage of those faculty members who favor student evaluations for the evaluation of teaching versus those who prefer peer evaluation, which includes classroom observation, as a preferred method for evaluating faculty. In addition, an analysis of individual experiences with the method of faculty evaluation done at their institutions will be presented. For the most part, the institution of the faculty members is not mentioned because much similarity between the two institutions developed in terms of the vagueness of the methods used for faculty evaluation. In Chapter 4, the focus shifts to an expansion on the discourse of student evaluations. Issues of grading are considered along issues with peer evaluation and why peer evaluation has not become the preferred method for evaluating faculty, including recent studies vis-à-vis some of my own survey data. Education as a consumer will be considered in the context of allowing student evaluations permeate throughout the evaluation of college teaching along complying with the recent Spelling report on higher education accountability. Chapter 5 will explore what a new policy on faculty evaluation should include based on previous work and my own pedagogical views. The focus of the chapter is to address what might work and might not work in terms of making peer evaluation the preferred system to evaluate college teaching. Student evaluations have evolved from being an informal tool to share information on courses and professors and have become a centerpiece in faculty evaluation. The policymaking process of evaluating college teaching needs to be

revisited to allow only those with adequate knowledge and experience to be the ones who effectively evaluate college faculty, a major step towards improving the quality of teaching in our colleges and universities.

CHAPTER II: LITERATURE REVIEW

COLLEGE CLIMATE IN THE 1960s

Higher education in the United States has been a place of much turmoil since the Berkeley revolt of 1964-65 (Lipset, 1971, p. xiii). Ever since protest erupted in Berkeley, numerous interpreters of what was bothering the students have suggested that the large bureaucratic educational institutions turned them off (p. xiv). Namely, the large university was seen as inherently impersonal, relying on the large lecture course, the graduate teaching assistant, and a very big student-faculty ratio. Instead of being treated as a total individual, the student in the large university found herself/himself a number on an IBM card (p. iv). It was from this perspective that idealistic students were catalyzed by two sets of issues: "the efforts to secure civil rights and economic equality following the Supreme Court's decision [to send young men to war], and later the immorality displayed by the United States in fighting a little nation-enemy in Southeast Asia"(p. xv).

Another argument presented by Lipset deals with the sociological aspects of youth while in college as the university has become more meritocratic. Young people find themselves facing a highly competitive situation. He states: (p. 19)

> "There is a variety of evidence which suggests that these tensions [emphasis on meritocratic competition in the context of sharply increased numbers of students] affect the emotional stability of many teenagers and college youth… Such tensions may find varying outlets, of which a rejection of the competitive social system which forces them into a rat race for grades is one."

This statement points to the dependence of students on grades for success. However, this situation changed during the Berkeley student protest of 1964-65 and the demonstrations against the Cambodian incursion in 1970, when large segments of the professorate stood closer to the students (p. 3).

Moreover, Lipset points out that more prestigious universities and colleges competed with each other to attract distinguished scholars, and often rewarded very publicly those faculty members most valuable in their quest for institutional prestige, thus creating animosity among some faculty members who felt neglected by their institution. Consequently, "[s]uch sentiments reinforce faculty propensities to oppose the administration of these schools, as well as the dominant values and institutions of the large society. Hence, many distinguished professors find solace in student militancy directed against the forces they hold responsible for their felt sense of status insecurity."(p. 33). Thus, this statement suggests that after the 1960s faculty and students were able to relate to each other by being neglected by the institution and saw it appropriate to join forces against the mega-university. Just how close these two entities became, it is somewhat unclear. What is clear form Lipset's work is that although faculty and students were often allied, arguments concerning grades emerged – namely, the notion of grades based on rigorous work. The power to grade was seen as an instrument of coercion, which prevented the free interchange of ideas and opinions between faculty and students (p. 226). This issue of grades in terms of grade inflation will be discussed in more detail below.

The phenomenon of grade inflation is hard to pinpoint since the studies contradict one another and there is no sufficient causality to link student evaluations of teaching to grades inflation. Initially grade inflation was blamed on the elevated grades during the Vietnam era. As Halper (1993, p. 53) states:

> "A bad grade – that is, the kind we used to issue routinely – was a one-way ticket to Nam. Draft boards did not distinguish an Ivy League GPA from a community college GPA so to protect his students and assuage his conscience, a professor would overtly or subconsciously grade higher. As the war heated up and required more bodies, so too did grades march upward."

While faculty and students might have joined forces, the military involvement in Vietnam mounted from year to year, and draft calls increased in size (Rudy, 1996, p. 165). It is then when many students felt that their chances for an education, career plans, or their lives were being threatened. In addition, student activists began to demand reform of such things as dormitory regulations, particularly *in loco parentis* rules. They also called for a prominent role in college government for student representatives, along with voice in the appointment of faculty and the approval of courses of study (p. 166). These concerns strengthened when President Johnson's need for manpower increased, and the draft rules tightened up. On 24 March 1966 Selective Service announced that college men would be deferred only if they were *full time* students who ranked in the upper half of their class (as first-year students), even higher if upper classmen (p. 171). This event suggests that the mindset of students suddenly might have changed to focus more on grades and competition in the classroom might have been exacerbated from then

on. As Rudy states: “Those at a highly competitive school like Michigan or Harvard had to consider transferring to a college where easier standards might ensure that they would attain the required class ranking. Professors were placed in a quandary also. Could they in good conscience give male students “C” or “D” grades, in effect giving them ‘a ticket to Vietnam’, or must they hand out gift “A’s” and “B’s”? (p. 172). Cahn made a somewhat similar argument when he states that the major incentive for the call to student activism was the Vietnam War (1997, p. 4). Along with this major issue there were long-standing issues such as minority rights and student freedom. He added that: “Protestors connected the nation they perceived as engaging in all varieties of social and political oppression with the educational institutions they saw as fostering bitter competition among students imprisoned within archaic traditions.” These forces coalesced to force significant changes within colleges and universities. Protests effected significant changes in general areas, among these: provision for more student involvement in the educational political process; and reconsideration of traditional content, methods, structures, and *evaluation of college instruction* [emphasis mine] (Foster and Long, 1970, p. 445). Given the range of demands, whatever the outcome of institutional responses, higher education in the United States would never stay the same as before the massive protests. The presence of faculty in the picket lines increased the legitimacy of student demonstrations (p. 370). All in all, the campus political climate in the 1960s was one of intense student political activity, which set a precedent for arguing pro student rights.

CHANGES FOLLOWING STUDENT PROTESTS

Higher education is undoubtedly a vital component of a progressive society in its capacity to both generate and to transmit the spirit of search for truth (Korn, 1970, p. 223). If higher education is committed to the search for truth, then it must be open to allow its students to question its offerings. With this in mind, student activists were committed to eliminate the prevailing notion of self-centered spirit of *rugged individualism* -- the competitive winning out over others (which makes collaborative behavior look "weak"). There was an imminent need for liberal education that would help students develop their intellect through an open-minded and critical evaluation of the fruits of civilization without the irony that such philosophies, belief systems, and human experiences were alien to them (p. 233). College teaching was seen as a process whereby an expert imparts knowledge. Up to the 1960s professors were seen as individuals more interested in their subject matter than in the intellectual and personal development of their students (p. 234). The Berkeley crisis of 1964-5 and subsequent events were expected to make reform efforts more realistic. In 1960 there were about 3.8 million students in higher education. By 1969 there were over 7.8 million students – more students than farmers, construction workers, miners, or transport workers (Steigerwald, 1995, p. 134). This was an era when the technological demands of the economy and the Great Society's emphasis on education increased the size and importance of the universities (White, 1998, p. 117). This environment encouraged young people to come together to become an important force in American politics.

Students first tried to reform the university indirectly, by creating ‘free universities’, which were intended to serve as a model for what the university should become (Reuben, 1998, p. 154). A second phase of student-led reforms included demands for new programs that would link curricula to political movements, such as black studies and women studies, which were part of a ‘relevant’ curriculum. Black and other minority students made up about 3 percent of the nation’s college students outside the Southern black colleges. Denying these minority students their demands was considered denying power to their people (Powell, Jr., 1970, p. 55). Students also demanded more participation over program development. These political processes were felt across the spectrum of higher education. The ‘free university’ was established to deal directly with contemporary and socio-political issues that encouraged activism (Reuben, 1998, p. 155). These ‘free universities’ were meant to provide a model for university reform. Some of the characteristic features of the ‘free university’ included: decreased authority of the instructor and students’ ability to choose the courses they wanted to take. Obviously, ‘free universities’ were open to everyone, and rejected grades and other mechanisms of evaluation considered irrelevant to learning. This concept of ‘free university’ was met with moderate success as it coexisted with traditional universities. Even enthusiastic supporters became doubtful of the power of free universities to transform higher education as the “unfree” universities continued business as usual without being confronted by activists (p. 156). Student activists channeled their energies into getting more participation in the traditional university. Only then could change ensue, and their ideal of what a liberal education should be could be implemented.

Meanwhile, “numerous faculties sacrificed their autonomy in a misguided attempt to appease rampaging protestors. Students were given membership on faculty committees, encouraged to vote on matters of curriculum, and asked to make decisions regarding appointment, promotion, and tenure” (Cahn, 1994, p. 86). Cahn disagrees with this practice as he argues that faculties have yielded their appropriate prerogatives to those who lack the knowledge and experience necessary to exercise legitimate academic authority. He considers the notion of a freshman’s opinion strong enough to cancel out the word of a distinguished scholar something in the proportion of an Ionesco farce. Furthermore, Cahn sees the phenomenon of giving students voice in university committees to go against Socratic ideals, since Socrates emphasized that a good teacher leads students to an awareness of their own ignorance. “Socrates, not his students knew best what they needed”, he says (p. 86). Adopting this idea of giving too much participation to college students perpetuates the notion that students have consumer rights, according to Cahn. He further argues that the consumer analogy is fundamentally flawed since academic degrees are not purchased but earned as they represent a certification of an individual’s academic achievement (p. 87). The notion of giving too much importance to student evaluations stresses the idea that students are consumers of education ‘goods’, entitled to make critical judgments of the teaching process that goes on in the college classroom.

EMERGENCE OF STUDENT EVALUATIONS OF FACULTY

Along with the inclusion of students in various university committees and the establishment of offices for student affairs, faculty evaluation booklets put together by

students became popular. Initially these were intended to help registering students by providing tips about the more or less interesting teachers and courses available at a particular school. But by the early 1970s these booklets had been transformed into complex statistics, obtained by formal procedures and used heavily by university administrators to make decisions such as an instructor's reappointment, promotion, and even tenure (Cahn, 1994, p. 38). However, research on student ratings goes back more than 80 years. Guthrie (1954) cites a study on student ratings dating back to 1924. In fact, the study of 1924 is cited as evidence to disprove the myth that students cannot make consistent judgments about the instructor and instruction because of their immaturity (Aleamoni, 1999, 153).

By 1965 the cry for student participation was coming from every direction. In the fall of 1965, during a convention of the American Council of Education, the subject was "The Student and Higher Education" and the consensus was that there should be a way for responsible student participation in educational policy-making. Course and teacher evaluations were seen as a respectable way for accomplishing such outcome (Werdell, 1967, p. 12). It was at this point that student evaluations of teaching gained prominence. There were earlier instances of survey for course and faculty dating as back as 1924, but these were more like localized attempts at some universities, particularly at Harvard. Initially these evaluations were seen as benign attempts to voice the opinions of students. A statement defending student evaluations of college teaching was given by Eckert (1950, p. 67) in which she argues that published studies had shown that students' judgments are very similar to those reached by more experienced individuals. In

addition, she cites a study published by Remmers (1928) in which he shows no relationship between the opinions students express and their own class ranks.

However, despite the overall positive implications of student evaluations described by Eckert, she admits that: “Use of this scheme [student evaluations of teachers] may influence occasional teachers to seek popularity at the expense of their function in stimulating student learning. Moreover, student ratings may be used as a kind of club by certain department heads, deans or president (p. 68). Even earlier publications on student evaluations warned about this factor. According to Brandenburg and Remmers (1927, p. 400), the validity of an instructor’s students as a true measure of the instructor’s ability is something to hesitate about. Nevertheless, they ascertain the validity of student evaluations for a specific instructor to make a careful study of her/his own teaching to improve classroom procedures (p. 402). In a survey of the relevant literature, Wittrock and Lumsdaine (1977, p. 449) conclude:

> “Most research on student ratings of teachers does not indicate that the ratings measure the effectiveness of teaching, good teaching, intellectual achievement, nor understanding of basic concepts. The ratings appear to be measuring student satisfaction, the attitudes of students toward their teachers and classes, the psychosocial needs of the student, and the personality characteristics, popularity, and speaking quality of the teacher.”

Student evaluations may yield quantifiable results that give the appearance of exactitude. But Cahn considers the use of this information for the purposes of reappointment, promotion, and tenure demeaning to faculty (1994, p. 42). Another problem that Cahn

sees with the use of student evaluations is the possibility of grade inflation as a result of not antagonizing those who evaluate them: the students. He wrote (1977, p. 4):

> "Since the early 70's, after faculties had been constantly bombarded by the claim that all members of the academic community should be treated as equals, students have been granted a major role in determining course content and the criteria by which competence is to be assessed. Clearly, a student can gain a good grade with far less difficulty when he himself is the arbiter of what is to be taught and how much needs to be learned."

The issue of grade inflation as a cause of student evaluations has not been clearly established and various studies have produced polarized conclusions about grade inflation caused (at least partly) by the prominence of student evaluations of teaching.

The function of a grade seems to be based on the student performance based upon a course's standards. Letter grades were introduced at Harvard in 1886 (Lewis, 2006, p. 115). By 1894 there were already complaints that there was soft grading taking place, as indicated by the Committee on Raising the Standard at Harvard. Grades are meant to assess the quality of a student's work. For Cahn (1994, p. 26), a grade represents the judgment of an expert (the teacher) about the quality of the student's work in a specific area. It should be added that a grade represents an expert's judgment at a specific time period, thus it represents past academic effort and an assessment of academic abilities at a given time. A grade has multiple functions: assessment of academic work, admission to graduate and professional studies, qualification for honors or advanced programs, and even to help the student formulate future plans. Although grades are typically items for

internal consumption within academia, prospective employers may use grades to assess future performance in a challenging work environment. Whereas a grade is supposed to reflect the objective assessment of a professional (the instructor), anyone looking for a grades for the purpose of making an assessment of an individual should consider the possibility that grades may have been given subjectively a particular instructor.

For the most part, grades represent an objective and convenient method to assess the quality of a student's work. By implementing a system of letter grades (A, A^-, B^+, B, B^-, C^+, C, C^-, D^+, D, and F), an instructor can provide a realistic way to compare students in a class, and to indicate a specific level of achievement in a readable, and easy to interpret fashion. Cahn addresses the argument about grades fostering competition (pp. 27-28). Although some scholars may argue that the competition for grades undermines the purpose of enriching one's educational experience, competition is necessary to stimulate students to achieve to their maximum potential. Such competition in turn helps the instructor differentiate among various levels of performance in a given course section (p. 28). Even in the event that a student fails a course, such experience may help her/him assess how to become a more effective student and whether or not a particular educational experience is right for her/him. Needless to say, the process of grading is not exempt of unfairness but perhaps that is the exception rather than the rule. Unfairness in grading is most likely due to inconsistencies on the instructor's part. A good instructor establishes *a priori* a grading system: what constitutes excellent, fair, and poor work. Also, effective instructors clarify to their students at the beginning of the term how the grade is calculated and the grading scheme for the course.

HAVE STUDENT EVALUATIONS LED TO GRADE INFLATION?

Those protesting students who demanded participation in university committees also demanded certain privileges such as fewer or no tests, which were seen as instruments that require “enormous memory, ability to take rapid notes, and an intensely competitive spirit”, also imply that where students sit on the normal curve is generally more important than what they have learned (Powell, 1970, p. 82). Powell, a student at the time, wrote in 1970 what is considered to be the representative voice at the time asking for restraining of faculty power to build a more democratic university (p. 58). Student advocates use student evaluations to make their case. He wrote: “One only has to read through the published course and teacher evaluations being put out by student governments around the country to get the clear impression that students aren’t terribly engaged by the majority of courses and professors with whom they come in contact” (p. 57). Faculty members who felt sympathetic for student advocates modified their standards. Cahn wrote: “To make their individual courses even more attractive, many teachers have abandoned examinations and based grades solely upon papers” (1977, p. 5). This approach is seen by Cahn as an attempt to obliterate academic standards. Another position taken by students was the elimination of required courses on the premise that these isolate and protect bad teaching and bad classes (Powell, 1970, p. 82). Their wish may have seemed to be granted, at least temporarily, during the early 1970s. According to Cahn (1977, p. 4):

> "…Therefore, required courses in history, literature, foreign languages, the sciences, and philosophy are now almost extinct, and as a result students have been able to avoid work in their areas of weaknesses.
>
> Due to the abolition of required courses, departments have been forced to compete for students in order to maintain the size of their faculties. One sure way is to lower standards. How many students will seek to concentrate in a department reputed to award low grades?"

Perhaps this transition into relaxing course requirements paved the way for grade inflation. However, studies supporting this assumption have not been thoroughly replicated.

By eliminating some required courses or by increasing the number of course which could be taken as a "pass" or "fail" option, student pressure might have increased further to relax academic requirements. Consequently, grades are synonymous to something completely irrelevant to the academic experience of a student. While grades can provide an effective means to determine which students deserve academic honors, and to motivate them to study, there are those who advocate the elimination of grades. Cahn presents some of the criticisms to grades that prevailed in the 1970s (1973, pp. 26-28).

- Grades are inherently inaccurate as a student's grade indicates the particular bias of her/his instructor.
- Grades traumatize students since they foster competition, arousing a bitterness and hostility in an otherwise tranquil academic environment.

- In attempting to measure people, grades succeed only in dehumanizing and categorizing them, depriving them of their uniqueness, and reducing the students to a letter of the alphabet.

The position to eliminate grades was written about by Mannello (1969, 305-308). He conducted research at his own institution (Hofstra University) to prove his case. Some of his arguments are: “The elimination of marks also removes the tension and dissension provoked by an ego-damaging inferior grade (p. 306). He argues that graded tests not only cause poor student-teacher relations, but also cause disorganizing crisis in the student’s life. A solution proposed by Mannello is to mark papers as “Acceptable” or “Unacceptable”, and a student receiving an “Unacceptable” mark could then be allowed to re-do the work thus reducing the crisis atmosphere. He also argues that a no-grade system would promote better teaching. His argument is that students would know that they will receive course credit and the instructor does not need to enforce learning leading to focus on course content and effective teaching rather than competing for a high grade (p. 307). These two arguments seem rather weak on the basis that would perpetuate a sense of entitlement among students who are not willing to work tenaciously. Cahn argues that competition brings the best in people and should be an inherent part of academic life since it is an inherent part of every aspect of life (1973, p. 27). Perhaps competition should be channeled effectively so as to minimize the negative effects of students competing against each other. Competition can be maximized so that students strive to fulfill their own potential and to focus on getting a good education rather than good grades.

Boretz points out several reasons that might have led to grade inflation: faculty conscience in the practice of building and reinforcing students' s self-esteem, increased academic popularity of mastery approaches to learning such as the practice of allowing retests and revisions, extension in the deadline for course withdrawals – which allows students to avoid low grades, and an increase in the number of students receiving federal financial aid leading to make those students work harder to maintain their eligibility for such programs (2004, p. 43). She also addresses the connection between grade inflation partly blamed on the use of student evaluations, yet studies have been controversial due to the lack of appropriate controls.

Another author who addresses possible reasons for grade inflation is Chase (1978). He discusses findings from the Carnegie Foundation in which general education at the college level was considered a "disaster" due to a combination of internal and externals forces such as: faculty interests, student concerns with the job market, 'relevant' curriculum, and social fads (p. 34). He also lists some factors that he considers responsible for the decline in academic standards and subsequently have led to grade inflation (p. 34):

- Most colleges dropped mandatory courses in writing and foreign languages and eliminated some of the distribution requirements thus increasing the number of electives students can take.
- Formation of curriculum committees including students. Those committees added consistently added new courses regardless of academic value.

- Allowing more “extensions” and “incompletes” at a high rate. For instance, the percent of students at Yale requesting incompletes in 1977 was as high as 31 percent.
- Students were permitted to drop courses as far as a few weeks before the end of the semester. In some instances the rate of drop in courses was as high as 50 percent.
- Many colleges deleted form a student’s transcript any failing grade under the notion of success, not failure, is what should be recorded.
- The establishment of committees or “courts”, which included students, and had the capacity to overrule professors on academic issues such as changing grades, granting extensions, and giving incompletes.
- As the average SAT score dropped, primarily by implementing minority programs, there was a disparity in achievement in the classroom. Consequently, some professors felt pressured to give passing grades to students even if they were failing.

These examples given above suggest that many colleges failed to maintain academic standards of excellence. It may be argued that as college demographics changed in response to changing times, academic policies changed to establish more flexible standards that would allow students more flexibility in their academic pursuits such as being less competent in certain academic subjects or avoiding foreign languages and difficult subject that could prevent them from graduating.

Chase attributes the changes stated above to a “profound change in governance, whereby considerable control of the decision-making process devolved to students and faculty” (p. 35). He points out that these institutional changes seem resistant to any academic revival because this “new academic ideology” is antithetical to quality education (p. 36). But what does the “old ideology” consist of? The years between 1945 and 1960 witnessed a period of academic renewal partly due to a large increase in college enrollment as a result of the GI Bill. In addition, liberal education was reconfigured as an education that would introduce common historical and cultural background to college students. At this point higher education, which used to be considered meritocratic until that point, turned into liberal education. Liberal arts colleges became committed to offer students a basis for the pursuit of knowledge, a vision shared by many scholars who embarked in this vision.

After 1960 a period of academic revolution ensued. The revolution consisted in the politicization of the academy. The early 1960s witnessed some crucial incidents, which destroyed the apolitical image of the college and university campus. Events such as assassinations of John F. Kennedy and Martin Luther King Jr., the war in Vietnam, the invasion of Cambodia, the Kent State incident, along with antiwar demonstrations, race riots, and campus riots led to a shift in the priority given to scholarship. Consequently, students came to college with a hostile attitude. The hostility was toward any institution or individual that would not stand against the war or for civil rights. The students demanded more relevant (with more emphasis on political issues) courses. They also demanded some freedom from grades and “outmoded” course requirements and

comprehensive examinations (p. 37). These events signal the beginning of a decline in academic excellence that in turn would trigger grade inflation.

The students' demands were met by the administration, which were sympathetic to the students' concerns and feared further violence on campuses. At the same time the civil rights movement continued to accelerate. As a consequence, federal funds for colleges were increased and colleges began to recruit more minority and disadvantaged students. As more students with less-than-adequate preparation enrolled in higher education, the academic gap between them and those well prepared was narrowed by applying liberal grading standards (p. 37).

Chase argues that as competing interests became stronger, the goal of pursuing knowledge was submerge in this prevailing environment. The diversity of competing interests created a pluralistic society within the campus, each fighting for its slice of the pie (p. 38). As the pluralistic social framework grew, the notion of one academic subject being no more worthy than another took precedence in the curriculum. Courses were offered with no explicit rationale other than demand from the students. He stated: "The colleges were 'lengthening the cafeteria line' from which students could choose while at the same time diluting the quality of education." (p. 39). It seems that grade inflation came along as education turned into a commodity where students act as consumers. The new student generation seemed more concerned with personal satisfaction and the job market leading to increasing pressure for high grades. As Chase stated: "grade

inflation… further dilutes what little meaning there is left in the phrase ‘academic credit’.”(p. 39).

Cahn addresses the issue of student evaluations of teaching and grade inflation as follows (1994, p. 43):

> “It is not coincidental that the increasing use of student evaluations occurred at the same time as grade inflation. If teacher’s livelihoods depend on the degree of their popularity with students, self-interest dictates that they award their students high grades. After all, which of us is not inclined to take a more favorable view of those whom we know take a favorable view of us? Are we not likewise apt to be less than enamored of those whom we receive cool treatment? To be blunt, it is hardly likely that students who fail a course will award the teacher their highest accolades.”

For about eighty years researchers have investigated the relation between grades and student evaluations. Almost every study conducted has reported a positive correlation between these variables, but debates continue about the causality of this association (Johnson, 2003, p. 47). The first study published in 1928 (Remmers, pp. 759-760) attempted to establish a correlation between student evaluations of teaching and grades. Since then, hundreds of observational studies have been done to look further into this relationship. Most studies report a positive correlation but as statistical methods have become more sophisticated, many of these studies have been dismissed on grounds that observational studies do not control for various factors and are thus rendered ineffective establishing causality (Johnson, 2003, p. 51).

Johnson argues that the positive correlation between grades awarded by professors and student evaluations is "intuitively obvious" (p. 48). Johnson's concern with grade inflation stems from a concern about disparities in career prospects between students who benefit from grade inflation and those who do not enjoy the perks of having grades awarded on less rigorous standards. He shows a traditionalist's point of view about the inherent danger of grade inflation within our academic fabric, and he blames the postmodern perspective on the irregular grading practices among some faculty. According to Johnson, in a postmodern view traditional grading practices have been replaced by an assessment of the biases that accompany a student's membership in "dominant groups" and lead to oppression. (p. 7). Consequently, a postmodernist is less likely to assign poor grades (p. 8). Johnson quotes the postmodern perspective of Bilimoria (1995, p. 443), which states the following:

> "Teachers' increasing awareness of the biases inherent in modern science is likely to affect their evaluations of students' acquisition of subject matter. Because disciplinary content domains are increasingly open to diverse interpretations and the inclusion of alternative representations, the scope of what is legitimate and appropriate knowledge in the academic enterprise is widened. The global questioning of tenets once held to be singularly true allows a larger number of students to display with greater diversity a legitimate and appropriate grasp of a widened content. Consequently, grade distributions are higher than they were before the advent of postmodern challenges…

He also added (Ibid):

> "As postmodern perspectives gain greater legitimacy, teachers' openness to different conclusions and more diverse methods at arriving at them favor higher grade distributions because evaluation criteria are broadened and there are many, rather than a few, acceptable discourses in which students can engage. Students are empowered to challenge not just the insights but also the methods presented by dominant orthodoxies. Failure to display reason, analysis, objective consideration of evidence, and distance is much less used as an explanation for poor grades, as these keystones of modern science are themselves shown to be biased in favor of certain, but not other, views, and are hence no more valid than any other method of arriving at conclusions."

Johnson considers this view of grading as "bizarre" (p. 8), and points out that many professors are probably inclined to this postmodern view of grading. As a consequence, there is a range in grading practices that lead to disparities in grading, leading to "inequitable assessment of students" (p. 9). Johnson concludes:

> "Clearly, students who take classes from faculty who grade leniently have a better chance of finishing college with higher GPAs – and thus better career prospects – than do students who take most of their classes with instructors who grade more stringently. Students respond to this reality in two ways. First, they preferentially enroll in classes with instructors who grade leniently. Second, they provide more favorable course evaluations to these instructors."

There are some consequences to the ways students tend to respond, according to Johnson. Students would choose courses taught by those professors who grade leniently, leaving those professors who grade stringently behind in terms of enrollment thus creating tow sets of instructors: winners and losers. Many of the professors left behind would tend to become more lenient graders themselves and assign "above average" grades. Thus, when traditional incentives for students to achieve at their maximum being eliminated, academic standards fall (p. 9), and part of the mission of higher education – to foster original thinking and hard work – gets dissipated.

A fundamental question about grade inflation is: why does it persist? Johnson considers that "grading inequities persist because their consequences are misunderstood" and continue to be perpetuated by the following myths: (p. 9)

- Student grades do not bias student evaluations of teaching.
- Student evaluations of teaching are reliable measures of teaching effectiveness.
- High grades imply high levels of student achievement.
- Student course selection is independent of expected grading practices.
- Grades have a consistent meaning across courses, departments, and institutions.

Consequently, grade enhancing leads to a university culture that might be hard to reverse. As long as students are pleased with academic outcomes consisting of high grades, the notion of students as consumers of an education product will be perpetuated. In addition, students who are "satisfied customers" will lean toward giving higher course evaluations to those instructors who reward them with high grades. Under this assumption, student

evaluations are not very good indicators of teaching effectiveness (and how much they have learned in a specific course) and should not be taken at face value.

More recently, Harry Lewis (2006, pp. 125-127) argues that grade inflation is not very important. His arguments include the idea that grades are educational tools that should not be used " as credentials for external consumption". He also considers the pressure for "meaningful" and stiff grading as anti-educational – a moralistic imperative. He considers the grade point average of little use, inflated or not (p. 139). For Lewis, grade inflation is related to differential grading standards in different fields, namely a problem of inconsistency (pp. 143-144). However, he supports the notion that student evaluations contribute to grade inflation. He writes: (p. 117)

> "Junior faculty members are in a vulnerable position. When a junior professor comes up for tenure at Harvard, the professor's teaching ratings from the undergraduate Course Evaluation Guide may be the only appraisal of teaching when promotion is discussed. Even though courses are evaluated before students know their grades, the pressure for good ratings tends to make junior professors reluctant to become known as hard graders. Favorable evaluations and higher grades have been shown to go hand in hand."

Nonetheless, Harris points out that grade inflation may be due to a number of factors. One of those factors is that there are more small courses. In the small courses grades tend to be higher "perhaps because students and faculty get to know each other better" (p.

118). Another factor might be better teaching. Good teachers tend to improve over time, and give their students more feedback, which in turn leads to better student performance by not repeating previous mistakes. As these teachers improve the quality of their teaching so does students' performance and grades (p. 119). When students' grade go up as a function of better teaching it can no longer be considered grade inflation. An additional factor is the present letter grade category using the plus/minus system. With more choices to evaluate the student's work, it then becomes more arbitrary to assign a B^+ to a paper that deserves a B, for instance (p. 120). Harris says: "The grading scale itself can make inflation more likely because grading requires assigning subjective judgments of quality to fixed categories" (p. 119). An interesting observation made by Harris is that, in large courses, papers are graded by teaching assistants and they do not have the same judgment as an experienced professor, and some are reluctant to grade harshly (p. 122), possibly for a variety of reasons. Finally, there is the possibility of "subjective judgment and human empathy of teacher for student may well result in a systematic upward bias in grades" (p. 122). Whether all these observations are true and result in 'innocent' grade leniency, grade inflation has the potential of corrupting academic excellence by promoting soft grading and not motivating the students to do their best work (p. 123). We may never be able to pinpoint an exact cause for grade inflation, and there may never be pressure to get rid of it since students are benefiting tremendously from it.

Harris takes a defensive position when it comes to grade inflation. First, he argues that after so many decades of using grades there is no consensus about which purpose they serve (p. 125). Given the lack of consistency when assigning grades, he considers the use of grades as irrelevant in undergraduate education. He adds that over the years there have been four the years there have been four rationales for grading students: to make them work harder, to recognize and reward excellence, to rank students so that those outside the institution can recognize which students are better than other, and to make them learn more (pp. 125-6). What makes these four rationales very difficult to consider within an objective framework is the rampant use of elective courses. Once students are no longer taking the same courses, a direct comparison among them is no longer feasible (p. 126). Different combinations of grade given for different courses by different teachers in different departments become meaningless when attempting to establish some useful comparisons. Thus, Harris concludes that grade inflation in an inherent outcome of the elective curriculum and inconsistencies in grading among teachers. In addition, he concludes that the four rationales for grading have their own upsides and downsides (p. 127).

Another aspect of academic life that Harris attributes to grade inflation is that notion of self-esteem. He quotes Harvey Mansfield by saying (p. 129):

> "Grade inflation has resulted from the emphasis in American education on the notion of self-esteem. According to that therapeutic notion, the purpose of education is to make students feel capable and empowered. So to grade the, or to

> grade them strictly, is cruel and dehumanizing. Grading creates stress. It encourages competition rather than harmony. It is judgmental."

He adds that a competition for grades is meaningless and anti-academic, which encourages injudicious selection of courses. In addition, he argues that cheating occurs as a result of the game students are playing to get a satisfactory grade (p. 130). Lewis' argument is that the reliance on grades makes students work hard but represents bad teaching (p. 130). He adds that instead of threatening students with bad grades, college faculty should try to make courses "more interesting, more fun, more exciting – or more connected to some part of human knowledge or experience beyond the narrow walls of the disciplinary slot into which the course falls" (p. 131). According to Lewis, a course is possibly about justifying a curricular requirement rather than a personal need for learning the subject. If that is so, then grade inflation does not really matter. However, grade inflation can be an indication of a few other things going on at the classroom or institutional level, and should be considered unethical for those students who work extremely hard to earn a high grade and get a high-quality education.

Lewis also disagrees with how grade inflation has been blown out of proportion. He says: "There have been many more predictions and worries about the evils of grade inflation than demonstrated consequences." (p. 142). He claims that the main reason why college students may not work as hard as they should is because the work they are being asked to do is not very interesting. He writes: "When the work is interesting, students at selective colleges will do it even without the potential reward of a high grade or the punishment of a low grade." (pp. 142-3). He sees the use the unappealing coursework as

the culprit in academic dishonesty and the divide between professors and students. More than grade inflation, the problem might one of grade inconsistency (p. 144). When professors in certain courses and/or department grade harder, students will avoid those courses and will enroll in easier one. In addition, grading on a curve leads to additional grade inconsistencies from one semester to the next. Lewis suggests that what needs to be done is to have a serous conversation about what constitutes an A, a B, etc. and perhaps the issue of grade inflation could be targeted more appropriately (p. 145). In addition, he suggests that other ways to eliminate grade inflation include better advising, smaller classes, better teaching and a more inspired curriculum (p. 146).

OBSERVATIONAL STUDIES BETWEEN GRADES AND STUDENT EVALUATIONS

A number of observational studies have been performed since 1928 correlating student evaluations with course grades. Observational studies are those in which the researchers do not modify the system or environment they are investigating. The first article in record by Remmers (1928) in *School and Society* addresses that a positive correlation does exists between received or expected grades and the course evaluations. However, it was not until the 1960s and 1970s that the idea of grade leniency was factored into those kinds of studies (Johnson, 2003, p. 58). Later on, around the 1980s, there was a decrease in the number of papers addressing or appealing to the grade leniency theory. Perhaps this trend was due to the arrival of construct-validity research (Greenwald and Gillmore, 1997, p. 1209). The individuals who advocated construct-

validity research pushed for the development of better evaluation instruments and discarded previous studies that showed a positive correlation between student grades and course evaluations.

Another group emerged in the 1980s, which favored the source attribution theory (Johnson, 2003, p. 59). These researchers claimed that the positive correlation between student grades and course evaluations represent a biasing effect. Gigliotti and Buchtel (1990, p. 341), discuss the concept of bias as one in which individuals tend to be biased when assigning cause to the events that affect them personally. In more detail, success is attributed to oneself, while failure is attributed to external sources. In a grading scenario, attribution theory predicts that students would attribute high grades and academic success to their own efforts, and would attribute low grades and academic failure to the instructors and academic institutions (p. 341). This concept of bias in the attribution theory can be used in two ways. If students give high evaluations, the correlation between grade leniency and student evaluations may be dismissed on the grounds that the high grades are due to students' effort. On the other hand, low course grades can then be attributed to a lack of teaching effectiveness, which may then be assessed from low course evaluations.

In addition, positive correlation between course grades and student evaluations may be explained by classroom mechanisms that do not indicate grading bias. Remmers and collaborators (1949, p. 24) wrote the following:

> "… the relationship between grades received in any individual class and the ratings given the instructor may not be direct, but the result of a third factor. This third factor may well be the level at which an instructor pitches his teaching. One whose teaching is "over the heads" of his weaker students may be held in high regard by his better students; in such a situation one might expect to find a positive correlation between grades and ratings, his instruction being at a level most appreciated by the students making the better grades. On the other hand, an instructor who directs his teaching to the weaker students and who is willing to spend an unlimited time in explaining relatively simple matters might be more "popular" with those students whose marks are relatively low; in such a situation one would probably find a negative correlation between grades and ratings."

This statement seems valid and might apply in which an instructor caters to a specific group of students. There are additional explanations for the correlation between grades and student evaluations. For instance, Marsh (1983, p. 150) proposed that prior student interest plays a major role in establishing a positive correlation between grades and course evaluations. According to Marsh, students with prior interest in the course typically work harder in the course, get a better grade, and will tend to appreciate the instructor more. In this particular case, prior interest can be a considered a biasing effect on the student evaluations. By the same token, general student motivation can be considered a factor that leads to greater student learning and higher grades (Marsh, 1984, p. 707). Just in like prior interest in the course, student motivation is not considered a

grading bias if more highly motivated students tend to learn more, receive higher grades, and appreciate the instructor more.

One last intervening factor proposed to explain the positive correlation between high grades and student evaluations is good teaching – the factor student evaluations are supposed to measure. This is perhaps the most desirable factor to account for the positive correlation between grades and student evaluations. Marsh (1984, p. 7300 argues that the search for biasing factors in student evaluations constitutes a "witch hunt". Marsh's rationale is that if there is a lack of correlation or no significant correlation, the absence of a grading bias corresponds to lack of bias in student evaluations. However, his reasoning is faulty because any intervening factor that causes a correlation between grades and student evaluations is indeed biased. Factors such as student motivation lead to favorable student evaluations but if the instructor tends to grade stringently, she or he will receive less than favorable evaluations. Thus, looking at observational studies does lead to a conclusive statement between grades and student evaluations, and tend to be flawed in some aspects of the study. Generally speaking, observational studies do not account for several environmental variables that may alter the measurement of intended relationships. There are numerous factors that prevent establishing an adequate causality between grades and student evaluations. For instance: (Johnson, 2003, pp. 72-73)

- **Grades received concurrently by students in other classes**. The studies carried out do not address the grades received by the students in other courses than the course observed. If the student receives lower grades in her/his other courses, then the evaluation for the course observed will tend to be favorable.

- **Differences in grade distributions between classes.** The same course might exhibit dramatic grade differences between different sections. A correlation established for a particular section will be downplayed if other sections exhibit a much more different grade distribution.
- **Differences in grade distribution by field of study.** Students might regard grades differently depending on the subject matter. For instance, and "A" in Calculus might be interpreted differently from an "A" in English because they think that Calculus is much harder.

The effect of the confounding factors listed above provide further evidence that observational studies provide very little insight into causality between grades and student evaluations. For that reason, a different kind of study needs to be performed.

EXPERIMENTAL STUDIES CORRELATING GRADES AND STUDENT EVALUATIONS

In order to understand the mechanism by which grades and student evaluations are correlated, more precise studies are needed to establish a positive correlation. A large number of studies have been done to resolve the causality between grades and student evaluations. However, many of these studies have been conducted in a somewhat artificial way such as in laboratory settings, and other studies involve direct manipulation of grades in college environments. A study performed by Holmes in 1972 (p. 130) studied the relationship between grade expectancy, course grades, and instructor

evaluations by using a design that manipulated student grade expectations. Each student was told that her/his final grade would be based on scores received on four exams. After each of the first three tests, the students were given their actual score and a suggested grade distribution for those tests. On the day of the last exam, students were asked what grade they expected in the course after being told that their responses would have no effect on their course grade. After the last test was scored, Holmes selected those students whose actual and expected grades matched each other, and in all cases the grades were either an "A" or a "B". Students in each were then randomly grouped into either the control or experimental group. The score of the experimental group was lowered in order to reduce their grade by one mark, but subsequently were informed of their actual grade by the end of the experiment. At the end of the last class, the students were asked to complete a teaching evaluation form. Based on the data analysis, an analysis was done to determine the impact of grade on the evaluation. The results suggested that students whose grades were manipulated downward gave lower ratings than those students whose grades were not altered (p. 131).

Given the way grades were manipulated in the experiment described above, it is difficult to argue that teaching effectiveness was solely responsible for the differences in ratings between control and experimental groups. However, the results indicate that disconfirmed grade expectancy causes students to belittle the teaching performance of their instructors (p. 132). This study was extended by Worthington and Wong (1979) by diving students into three groups based on grades received on a practice test. Within each of the three groups, the students were randomly given a grade labeled as either "good",

"satisfactory", or "poor". After receiving their practice test scores, then the students were told that those scores would predict their final course grade, and were asked to fill out a teacher evaluation. In the analysis of results, those students who received the "satisfactory" grade rated the instructor much higher (17 out of 18 items) than those who assigned a grade labeled as "poor" (p. 769).

In another study carried out by Blunt (1991), he manipulated grades in an introductory course at a community college. He also introduced anonymity by requiring half of the students to sign their evaluation forms. In the experimental design the students were given their current course grade, and then were asked to fill out the evaluation for instruction. The students were then randomly assigned into four grade treatment categories as follows: inflated grade, deflated grade, actual grade, and no assigned grade. Analysis of the results showed that two out of ten items were significantly affected by the grade treatment. The inflated and actual grade groups rated favorably the instructor. Also, the inflated grade group rated the instructor higher than the deflated grade group (p. 49).

Vasta and Sarmiento (1979) performed another grade manipulation study in which they used two sections of the same course taught by the same instructors. They manipulated the grade curve for an exam so that one section received a much higher grade than another. The author did a statistical comparison and found no significant differences in study habits, grades or demographic composition that would eliminate differences in the study attributable to student characteristics. After being told about their

grades, students were then asked to complete an anonymous instructor evaluation. Of 16 instructor items in the evaluation instrument, seven showed significant statistical differences between the groups, those graded more leniently gave more favorable ratings to the instructors. Although the magnitude of the grade manipulation was relatively small, Vasta and Sarmiento pointed out that the effect of causality effects of grades on student evaluations is potentially strong (p. 210). However, a potential flaw in the experimental design involves the possibility that students from the two sections communicated with each other and might have discovered that there was a grade discrepancy between the two groups.

It is important to point out that some of the variation observed in the observational studies might be attributed to environmental factors such as prior student interest, level of course, class size, and differences between student grade expectations. Nonetheless, the correlation observed between grade and student evaluations signals a significant causal effect between grades and student evaluations of teaching. In the experimental studies, differences in evaluation might be attributed to the instructor's grading practices. These grade-manipulation experiments imply that grading practices change student attitudes toward the instructor and are reflected in differences in the student evaluations of teaching. The differences observed cannot be explained by differences in teacher effectiveness or any other intervening factors.

A more recent (and controversial due to the large amount of opposition it created) experimental study that correlates grades with student evaluations is the DUET (Duke

Undergraduates Evaluate Teaching) web-based survey instrument during the 1998-99 academic year. The survey linked student evaluations with their course grades, it also collected information before students received their grades and after the final course grades (Johnson, 2003, p. 11). In the survey it was possible to evaluate the characteristics of the participating students, their grades in all their courses, prior student interest, instructor grading practices, and teaching behaviors that determined how students responded to the survey. By performing complex data manipulation, such as various types of regression analysis, the researchers were able to isolate the influence that grades had on course evaluations from external variables such as student interest and student aptitude, given that those variables have confounded previous studies addressing grade effects (p. 85).

The survey questions in the DUET study looked at various item categories, such as instructor interaction, student satisfaction, course difficulty, and student commitment. It was found that the items grouped in the "course difficulty" category are consistent predictors of student responses (p. 94) as increase in course difficulty tends to lead to lower student ratings. The results of the DUET study indicated that the teacher-effectiveness theory, which predicts that good teaching leads to more learning, higher grades and higher teacher evaluations is not supported by the data. The data tend to support the notion that student evaluations should not be used for administrative purposes since teaching effectiveness may not be consistently judged by the students (p. 95). In contrast, the DUET results tend to support the grade-leniency theory at the individual student level. In this theory, students reward teachers who award high grades. However,

the grade-leniency theory is not consistent with many of the coefficients at the class level. Johnson interprets these results as a mechanism by which students do not reward those teachers who award low grades to their classmates (p. 96).

The DUET study provides a high correlation for the grade-attribution theory, in which students attribute academic success to them, but attribute failure to external sources. Students who receive low grades are very likely to blame their grades on inadequate teaching. Again, results from the DUET study are analyzed both before and after students receive their course grades. However, at most institutions students complete the teaching evaluations before the final grades are known. A comparison of coefficients between predicted and received grades indicates minor differences. Johnson concludes that the effect of grading on student evaluations is unaffected by the timing of the collection of those forms (p. 97). The conclusions from the DUET study can be summarized as follows: student responses to the survey are largely affected by the grades (either expected or received), grades represent a serious bias to student evaluations of teaching, and the bias is possibly explained by the grade attribution theory. The DUET study showed causality between student grades and student evaluation of teaching. The large sample used and the items indicated an almost uniform tendency of students to rate more highly those courses and teachers for which they receive high grades (p. 108). Since the DUET experiment effectively eliminated the possibility that unobserved environmental factors could interfere, the result from the analysis can be considered conclusive evidence for prior assertions that student evaluations of teaching are consistently biased by student grades.

The issue of grade inflation continues to attract the attention of educational researchers. One thing that grade inflation has done is to damage academic ethos in general by equating learning into a commodity (Boretz, 2004, p. 46). Some colleges such as University of Indiana, Dartmouth, and more recently Cornell have taken steps to fight (or at least expose) grade inflation by reporting the letter grade along the number of students in the class, the student's numerical average, and the class average. As Nagle (1998, pp. 40-43) reports, this rationalization may be overly simplified and may not solve the problem of grade inflation. However, such action may add some veneer of transparency to the process of awarding grades. Much research remains to be done in order to make sure better inferences between SET and grade inflation are established. Perhaps these two separate phenomena have very little relation to one another. Additionally, student evaluations of teaching may need to be downplayed when measuring teaching effectiveness.

Not all course evaluations are geared toward leading students into easy course and keep them away from difficult ones. At Yale from 1939 until 1982, the campus newspaper staff published "Course Critiques" (or "Crits", for short). The writers of those critiques valued intellectual challenge and attacked "gut" (or widely easy) courses (Hampel, 2006, p. 123). The Course Critiques began as a service to students since the Yale curriculum was too broad and undergraduates had much difficulty choosing from dozens of courses beyond their major fields. By 1939 the student newspaper sponsored a forty-eight-page pamphlet called *The Undergraduate's Guide to Courses in Yale College*,

1939-1940, which evaluated 100 courses. It then became an annual publication. The Course Critiques amplified the views of Yale students unafraid of hard work (p. 124). These critiques mirrored the work ethic of the students working in the newspaper. The "Crits" reviewed primarily the lecture courses, and the descriptions ranged from two to six paragraphs, using a respectful language. After 1964, statistical information was added. In addition the staff responsible for those critiques explained in the newspaper how the information on the critiques was assembled and periodically wrote editorials pointing to the strengths and weaknesses of the undergraduate education at Yale.

The "Crits" stand out for being just more than a compilation of course ratings. They served as a source of guidance for undergraduates in search of facts and opinions rather than mere subjective opinions like those found in commercial websites of faculty ratings. Hampel (2006) examined the "Crits" to see if they represented a shortcut to learning. In other words, Hampel wanted to analyze if the course critiques offered ways for the students to try to make their education easier. What he observed was that the critiques focused primarily on the following: instructor's skills to convey the subject, quality of the readings, and distinctive features of the course. The critiques did not elaborate on whether or not the course was easy or hard (pp. 129-130). Although sometimes the critiques would discuss aspects of the workload and the grading, those "required reader interpretation because the appraisals were qualified rather than unambiguous" (p. 130). However, by 1965 the "Crits" began publishing statistics on grade distribution for the courses listed, and it is possible that this information led to a trend in which students started looking for 'easy courses'. By 1970 the use of the course

critiques “went beyond consumption” (p. 136), and led students to criticize the undergraduate curriculum and urged the students “to negotiate, not just consume”.

Hampel laments that today students can go to their computers and retrieve ratings for instructors, with more comments usually shown for larger courses, and numerical ratings based only on clarity, easiness, and helpfulness (p. 140). The comments found today in online teacher ratings do not follow the appropriateness of the comments on the “Crits”, which rarely included any cruel remarks about an instructor. Another aspect of the “Crits” was the expectation that the faculty would read the course critique for her/his course and consider some of the suggestions listed for improving the course (p. 141). Hampel argues that one big difference between the “Crits” and today’s online ratings is the lack of purpose of the latter. The Yale’s course critiques were sponsored by a campus organization, and meant for the faculty to read those critiques. All opinions were meant to be accurate and helpful rather than nasty comments. The anonymity of online ratings lacks the transparency of the “Crits”, and do not have “the responsibility of citizenship” (p. 142) that an accurate instructor evaluation should reflect.

WORLD WIDE WEB-BASED RATINGS OF INSTRUCTORS

The peer-reviewed literature on the several teacher-ratings sites developed over the past few years is almost non-existent. Those sites have emerged on the World Wide Web and their popularity is growing exponentially. A study by Kindred and Mohammed (2005) has explored various issues regarding the most popular of those sites –

Ratemyprofessors.com (http://www.ratemyprofessors.com) or RMP hereafter. RMP has become one of the most popular sites visited by college students; John Swapceinski developed it back in 1999 with the idea that students could evaluate anonymously their professors. Currently, there are about 5.8 million reviews, which include over 700,000 professors from more than 6,000 colleges (as listed on the home page of RMP).

The study by Kindred and Mohammed concludes that students' motives for accessing RMP include information seeking, convenience, and interpersonal utility. A focus group study performed by the authors indicated the accessibility of revenge or venting by students, but the most common concern among raters is instructor competence. However, the site provides a consumer-oriented and not-so-credible image by including the instructor's personality and appearance as part of the rating. Namely, students can give a chili pepper to denote "hotness" of the instructor. Most critics of RMP point to the notion that anyone can post comments, although RMP has announced that will adopt stricter controls for logging starting in August 2006.

There are number of critics of RMP. John Sutherland is among the fiercest critics of RMP and has written some newspaper pieces about the commercial nature of the site, which is about to be sold for a seven-figure dollar amount. In addition, he criticizes the lack of confidentiality and long-life of RMP's comments and how they could become available to potential employers or promotion committees (2006, p. 2). Finally, Sutherland points out how RMP reinforces the notion that education is "sold" at the market rate and could perpetuate the old cliché "the customer is always right". A study

by Felton and colleagues (2003) points to the inherent consequences of a ratings system that allows students to post their likes and dislikes. If demanding professors receive less than positive ratings, then students could set a precedent so that their fellow students do not enroll in those professors' courses. Is grade leniency the ultimate goal of RMP? It might be too soon to jump to conclusions but the possibility still exists.

PEER EVALUATION AS A SYSTEM OF FACULTY EVALUATION

The use of peer evaluation has not been given enough importance in most academic institutions. In fact, there is not a large body of literature on peer evaluation perhaps as a consequence that it is not widely exercised. Studies on peer evaluation have indicated that there is a low correlation when peers observe faculty compared to student ratings (Hutchings, 1996, p. 221). Although Guthrie in 1954 calculated a 0.54 per-student correlation, the details of these calculations were sketchy. More research needs to be done as to what has led to SET have precedence over peer-evaluation at most institutions.

According to Sullivan (1995, p. 61), it seems that ironic that in an environment where evaluation is constantly exercised, peer evaluation from institution to institution it so disproportionate. According to her, the purpose of peer evaluation is advisory and should be made in a systematic fashion, consisting of four parts. However, the process by which peer evaluation should be exercised is a matter of much dispute. For instance, the purpose of peer evaluation needs to reach consensus. It does have several purposes, and there should be a policy establishing the specific purpose peer evaluation intends to

serve. Some uses of peer evaluation are: to provide help to faculty who request it, to provide advice to help less experienced colleagues, and to evaluate a colleague for promotion and tenure (p. 62). In addition, Sullivan writes: “teaching is a multifaceted process that is best evaluated using several methods” (p. 62), and the more methods used, the more the likely- hood that strengths and weaknesses may be recognized, particularly since “the perspective of the peer is broader that that of a student” (p. 63).

Nevertheless peer evaluation is not immune to some inherent dangers. Peer evaluators might have a different (and conflicting) theoretical framework or methodological issues that could have the objectivity of the evaluation process. Also, in a department deeply fragmented peer evaluation might become extremely subjective. Another possible danger of peer evaluation is the hierarchical control it implies when senior faculty member step into the classroom of a non-tenured or less experienced faculty member. Regardless of the possible negative impact, peer evaluation should be seen as an opportunity to diversify how college faculty gets evaluated and should become a more commonly used approach in the evaluation policy of any institution.

Some possible objections to peer evaluation rest on a few issues (Chism, 1999, pp. 9-15). First, it means to open the classroom door to someone who is not a student in the class, which implies some level of anxiety in the person being evaluated. Many institutions consider peer evaluation a mere classroom observation – sometimes just a single visit to the classroom. A single visit is not enough to collect a complete repertoire of the dynamics that go on in a classroom. Another issue is how to select the appropriate

peer to conduct the evaluation, and it the peer would be as impartial as she/he should be. A major issue is time constraint, the peer evaluator might not dispose of enough time to evaluate the course materials and offer thoughtful comments. The lack of adequate standards for evaluating teaching is another factor to be considered when implementing peer evaluation. There are also concerns about internal validity (measuring what is intended to be measured) and reliability (evaluating consistently and accurately) on the part of the reviewers. Finally, there are some after-effects to be considered: concern that only one reviewer was used thus leading to a decision based on just one person's judgment, a reduction in risk-taking in teaching by not encouraging innovative techniques unfamiliar to the reviewer, and the possibility of legal challenges increasing as peer review is increased.

STUDENT EVALUATIONS AND TODAY'S MODEL OF HIGHER EDUCATION

Since the research on student ratings began, and following World War II, studies on college teaching and learning have become more common (McKeachie, 1990, p. 189). However, the focus on research on college teaching has shifted from decade to decade and consequently the literature on specific aspects of college teaching in relation to SET has been published in 'fits and starts'. In 1970, Mann published *The College Classroom* thus providing a source of insights regarding student characteristics, teacher roles, and development of a course as a working group. Mann's work addresses the nature of interpersonal and emotional events which take place in the classroom and how often student-teacher interactions are ignored (p. v). He asks: "In what capacity does the

teacher stand before the class?" (p. 2), and the answer is obviously as an expert. But there is more. The teacher is an agent not only of instruction but also of control and evaluation, and more importantly as a facilitator of learning. Therefore, in order to accomplish all these functions a college professor must combine content knowledge, pedagogical knowledge, and curricular knowledge (Shulman, 1986, p.9). Some professors may be great scholars but not necessarily experts in the craft of teaching. Cahn addresses this issue when he said: "It is a sad but indisputable fact that much of the teaching that goes on in our colleges and universities is of very poor quality" (1978, p. ix). Instructors are not the only ones to blame but also the college establishment, which has failed to recognize that scholarship and classroom competence are two different factors and anyone entering the college teaching profession should be versatile in both content and pedagogical knowledge (p. x).

It may be possible that student evaluations of teaching started as a movement to deal with poor quality of teaching. Students, feeling impotent against the authority of professors, felt the need to forewarn their fellow students about instructors who, in the students' view, did not meet those expectations of teaching adequately and/or encouraging students to learn and master a specific discipline. However, students might not be the most qualified individuals to assess the quality of an instructor. Although it may be argued by some that students spend long hours in contact with a specific instructor; thus students have the expertise to asses the instructor's teaching quality. But students are novices and their insightfulness cannot match the rigorous academic

preparation of any instructor. There may some exceptions but by no means are students the most qualified individuals to evaluate a college instructor.

Despite the possible good intentions behind student evaluations of teaching, the necessary questions are: What is the rubric behind student evaluations? Are these evaluations intended to push the system toward producing better learning or happy customers? If one interprets student evaluations of teaching with reference to educational paradigms, the evaluation instrument seems to address the transmission model of teaching but not the engaged-critical model. However, analysis of evaluation instruments reveals that these are not representative of all forms of teaching and militate against alternative modes of teaching (Kolitch and Dean, 1999, p. 27). An item-by-item analysis done by Kolitch and Dean demonstrates that most items in a standard student evaluation form are used extensively by universities in the United States are supposed to represent common features of teaching that rate the instructor's overall effectiveness (p. 31). The authors have identified assumptions such as the following: knowledge can be divided into measurable objectives; the purpose of each course is the student acquisition of a clearly defined body of knowledge; and the instructor is at the center of the classroom, whereas the student is invisible or on the periphery of the educational process (p. 38). They conclude that the student evaluation of teaching is "a cultural and educational artifact that provides students with a specific lens or conceptual framework through which to interpret their academic lives" (p. 39). However, in many academic settings the student evaluation represents a catalog of exemplary teaching characteristics designed to capture the essence of effective teaching.

An important question then is: “What constitutes exemplary college teaching?” If this question can be answered effectively, then a model of exemplary teaching can be proposed, articulated, and perhaps implemented at many academic institutions. In his book *Mastering the Techniques of Teaching* Lowman (1995) presents analyses and ideas to improve college teaching. For him, exemplary teaching is the ability to explain abstract concepts clearly and simply along with the capacity to apply innovative pedagogical approaches (pp. 289-290). Moreover, he recognizes the need for institutional support to provide a supportive teaching environment (p. 295). However, he values student evaluations as a powerful tool to improve the instructor’s future performance (p. 299). Lowman also holds the notion that students are asked to evaluate instructors because it improves student morale by acknowledging student participation in the school’s instructional programs and in their own education (p. 299). This particular ethos is similar to some of the reasons articulated for the establishment of standard student evaluation of teaching procedures in the 1960s.

CHAPTER III: INSTRUCTORS' PERCEPTIONS OF FACULTY EVALUATION QUESTIONS FROM PREVIOUS RESEARCH

Historically, faculty used to design their own course evaluations and used them primarily to improve course content. However, in more recent years, some faculty evaluation committees and academic administrators rely primarily on student evaluations to assess an instructor's competence and merit for retention, tenure, promotion, and merit pay decisions (Moore and Trahan, 1998, p. 775). In their study, Moore and Trahan studied the relationship between tenure status and grading practices. The authors expected non-tenured faculty to be engaged in behaviors designed to improve their evaluation scores. Such behavior would include giving higher grades to students with the expectation that students will reciprocate with higher evaluations of the faculty classroom performance. This type of behavior is part of the basic exchange principle that if the instructor engages in any activity leading to grade inflation, she/he will gain institutional rewards. In higher education, rewards for most faculty members could be continuing employment, tenure and promotion (p. 775). In the case of those instructors working with non-tenure implications, high course evaluations are a necessity for continuous re-appointment given their less secure academic positions.

Previous research has raised questions about the validity of student evaluations because faculty members may inflate grades to receive better evaluations from their students (Greenwald and Gillmore, 1997; Wilson, 1998). Greenwald and Gillmore found that giving grades one standard deviation above the mean leads to a full standard deviation in the instructor's rank in student evaluations (p. 1215). On the other hand,

there are studies that find little or no relationship between student evaluations and course grades (Marlin and Gaynor, 1989; Marsh, 1984; Palmer, Carliner, and Romer, 1978). However, instructors' perceptions of the relationship between student evaluations and course grades have not been considered thus giving us a reason to engage in more qualitative studies. There is also the possibility that instructors might be giving higher grades without considering that they are engaging in grade inflation. In such cases, grade inflation may be the end result of modified teaching and grading practices resulting from an over consideration of student evaluations or an evaluation system that relies too much on course evaluations.

RESULTS

OVERALL ATTITUDE ABOUT STUDENT EVALUATIONS

Table 1 presents the overall perception of faculty members at the two institutions regarding student evaluations. Numbers are given as percentages based on a sample consisting of forty instructors.

Table 1. Faculty attitude about student evaluations at two research universities.

	In favor of student evaluations (%)	**Opposed to student evaluations (%)**	**Indifferent to student evaluations (%)**
University A	**35**	**50**	**15**
University B	**40**	**45**	**15**
	N= 14	N= 19	N= 7

These results indicate that about half of faculty members are roughly divided between those who are opposed to them, and approximately one third is in favor of having students evaluate their courses. A small but nonetheless significant fraction (7 out of 40) neither endorses nor condones the use of course evaluations.

Among the instructors who favor the use student evaluations in their classroom, they consider those evaluations to provide a realistic picture of what goes on in the classroom for a semester. As one instructor of foreign languages says:

> "Student evaluations play a crucial role in the department's evaluation of faculty but senior colleagues often visit our classroom and make adequate recommendations to improve our teaching strategies. Most student evaluations for our courses are positive and indicate that students appreciate our courses and feel that they are being treated fairly. My sense is that student evaluations are in line with what our courses offer."

The instructor quoted above was the only who related student evaluations with classroom observations. Other faculty members consider that it is only fair to have students evaluate their courses given that students spend a considerable amount of time in the course and deserve the opportunity to voice their opinions and make suggestions for improvement. A biology instructor says:

> "Student evaluations are fascinating because they represent democracy at work. Students tend to be savvy about what goes on in the classroom and can distinguish good teaching from not-so-good teaching. If the instructor is fair and generous,

> student evaluations will reflect that. It's true that there is some degree of popularity embedded, but for the most part the student evaluations seem to be a good indicator of classroom performance."

It is interesting to note that this instructor acknowledges the degree of popularity that could be factored in the student evaluations, and he mentions the term "generous", which will be discussed later in this chapter in the context of grade inflation. Another instructor in favor of student evaluations refers to those evaluations as "democratic" by saying:

> "Student evaluations are part of a democratic system where students can express concerns, frustrations, and voice opinions. A faculty member should be objective enough to face students' opinions when they are not satisfied with her teaching."

This instructor uses the term satisfaction as one of the outcomes expressed in student evaluations. The idea of "customer satisfaction" will be addressed in the next chapter as one of the possible reasons for grade inflation. There are also those instructors who are in favor of student evaluations but consider these to be a genuine method to secure feedback from students. For instance, as one instructor says:

> "Student evaluations are a great system for administrators to know more about what goes on in a professor's classroom, and for the instructor to receive feedback from students about what works and what needs improvement."

The quote shown above is important because it reflects what student evaluations were meant to do when first implemented in the 1960s: a critique of a particular professor's

course in which feedback was provided. However, as other instructors interviewed pointed out, in many or most cases the courses evaluations completed by students only serve as a "course complaint form", as indicated by one of the instructors opposed to student evaluations.

The instructors who indicated being opposed to student evaluations listed various reasons for their dislike of those. Some of the reasons have to do with the consistent tone of negativity expressed by students. For instance, a mathematics instructor says:

> "For me the student evaluations tell the same story over and over. I get lots of complaints about my accent. Very few of those evaluations comment on my teaching, about how much time I spend solving problems in class, and reviewing for quizzes and exams."

A biology instructor who says the following echoes the same sense about lack of objectivity:

> "Students usually provide such immature comments, and base their evaluation primarily on expected grade and course difficulty. How can a student evaluation be reliable when it refers to me as "crazy and annoying"? For the most part student evaluations are a waste of time and administrative resources."

The quote shown above attests to the lack of reliability that student evaluations may carry. Other instructors opposed to the student evaluations do so based on the grounds that students lack the expertise to evaluate the work of a professional, and cannot offer

valuable advice on how to improve the course and/or course materials. A history instructor makes the following comment:

> "… How can undergraduates (mostly freshmen) know that my lectures are challenging? That I present recent findings and ideas? My guess is that those evaluations are a necessary evil to make students feel empowered."

The instructor quoted above makes an interesting point, which is that student evaluations might serve as a token of empowerment for the students even in the event that administrators might not consider them too seriously. In addition, it shows that evaluation forms in many instances are too generic and fail to ask the right questions. Students are limited in their knowledge and expertise of pedagogical issues, and consequently do not have the capacity to discern when a specific instructor has brought to the class recent ideas and findings. After all those students are enrolled in the course to learn about recent ideas and findings in the subject matter. Concerning the notion of judging lectures as being challenging, more specificity about whatever challenging represents is needed. For instance, is the challenging component related to the course material or a challenging lecturing style? Maybe the readings assigned are the challenging component of the course. An instructor of anthropology commented on the ambiguity of course evaluation forms when she stated:

> "Student evaluations don't really capture the classroom dynamics of my teaching. They ask vague questions that don't relate to student-centered activities. A new evaluation form needs to be developed."

An interesting consideration emerges from the statement above. The general nature of student evaluation forms is somewhat limiting for those instructors who adopt progressive teaching styles or lead mostly student discussion-based classes. The instructors who follow a different teaching style from the traditional lecturing style stand a chance of getting lower student ratings by dealing with an evaluation form constructed inappropriately. When items in the evaluation form only address some general Likert-like descriptors of a faculty member, and the scores from those items are used to compare teaching effectiveness to all other teachers, then there is no incentive for non-tenured instructors to adopt new teaching strategies that could reduce their ratings against those who use a teaching style that easily matches the student course evaluation form.

EFFECT OF STUDENT EVALUATIONS ON TEACHING AND GRADING PRACTICES

The survey asked instructors how student evaluations had affected their teaching practices, grading practices, or both. If there was any change in teaching and/or grading practices, then the instructors were asked to explain the nature of those changes. Table 2 summarizes the results. Numbers are indicated as percentages out of forty instructors surveyed, and are broken down into the two different institutions where the instructors teach.

Table 2. Changes in teaching and/or grading practices as a result of student evaluations.

	Changes in teaching practices (%)	Changes in grading practices (%)	No changes in teaching or grading (%)
University A	**35**	**55**	**45**
University B	**25**	**50**	**45**
Number	N= 12	N= 21	N= 18

One of the most striking aspects of these data is the predominant trend in changing grading practices as a result of student evaluations. More instructors were willing to make changes in how they grade rather than how they teach. A possible reason might be that altering grading practices is easier to do than changing one's teaching style. Among those respondents who indicated any changes in their teaching methods, some responses were as follows:

> "Student evaluations have totally influenced my teaching style. In my first year of teaching, students indicated a preference for discussion over monotonous lecturing and I was able to adapt to what they seemed to prefer. Also, student evaluations indicated that students disliked frequent quizzes and I've reduced the number of those. Overall, I think I've become a better teacher by taking into consideration what students like and dislike about my course."

The response shown above illustrates two important points. One is that an instructor is willing to change teaching methods to keep students satisfied, which reinforces the notion that education has become a business-like enterprise in which the goal is to keep "happy customers". Another point is that for this particular instructor, the student evaluations might have helped his teaching style by adopting more discussion rather than lecturing. This change certainly seems a positive one, by allowing students to actively engage in class and perhaps attaining better learning outcomes. It is important to point out that this instructor (who teaches biology) does not do any grading since all his exams are multiple-choice questions.

Another instructor who changed her teaching methods as a result of student evaluations did so out of concern of having too many students doing poorly in her course. She states the following:

> "Before coming to this department, I had planned to teach calculus in a certain way. Now it's totally different. I have to worry that most students pass the course. I solve almost every problem from the assigned chapters. I give quizzes that are almost identical to the problems discussed in class, and the exam problems are very similar to the quiz problems. It couldn't be any easier than it already is."

This instructor is the same one who dislikes student evaluations because the comments tend to focus on her accent. It seems that she has changed her entire teaching

methodology not only to have most of her students pass the course but perhaps to get better student evaluations by teaching a much easier course for a typically difficult subject such as calculus.

Then, there is the case of the instructor who has not only changed her teaching but has also allowed her students to have a voice in the reading assignments. She says:

> "Teaching is a craft that should be revisited and modified as often as needed. Each semester my courses are different, taking into consideration the student population, previous evaluation comments, and what students vote to read during that particular semester."

The significance of this particular change in teaching method lies in how extreme it is to the point of having students vote about what readings they would like to have assigned. What if students voted not to read anything? This instructor must be getting very positive evaluation comments based on her novel approach. But the question is: Is she teaching the curriculum she was hired for? I was assured during the interview process that there is a list of readings for various units and that most of the readings up for student selection are equivalent to one another.

Some changes in teaching methodology are only modest and seem to be geared at enhancing the classroom experience. An instructor of English comments:

> "Students always provide helpful suggestions to improve how I teach. They are very engaged in learning and are most qualified to comment on what works and

> what doesn't when it come to effective teaching strategies. For example, based on student comments I speak slower and use more classroom technology these days."

It may be true that some suggestions made by students could benefit classroom dynamics, particularly when it comes to an inexperienced instructor who may lack a wide range of resources in her/his teaching repertoire. However, considering students to be the *de facto* experts can only go so far. Students will be dealing with faculty members who might not be so willing to change teaching styles and perhaps it is in their best interest to adapt themselves to a particular teaching style.

An extreme case of changes in teaching methods comes from a writing instructor who has basically given her students almost full participation on how the course runs. She indicates the following:

> "I value every single student evaluation because each of them represents a student's perception of what took place in my classroom for the entire semester. If they say that I talk too much during class, I'm sure I do. In my class students enjoy full participation, we decide the readings for the course together on our first meeting, students choose their own deadlines for the midterm paper and when to submit the final portfolio. I've found that students develop more interest, responsibility, and enjoy the writing process."

Outside the interview process she indicated that if students don't have their portfolios ready by the end of the semester, she gives them an incomplete and they turn the writing portfolio during the following semester. When I asked about the kinds of ratings she gets from students, she quickly indicated that her evaluations are some of the best in the department. That is not surprising given the amount of flexibility that students enjoy in her course. I wonder how much students improve their writing after taking her course. I later found out that the instructor is not on a tenure-track, she works on three-year term contracts and outside teaching is an accomplished freelance photographer for various magazines; this could explain her placidity when it comes to her teaching approach.

The changes in teaching methods illustrated above not only lead to satisfied students but also seem to indirectly lead to grade inflation. By having flexible deadlines, give quizzes almost identical to problems discussed in class, and letting students choose what they want to read, these instructors are altering academic standards and consequently making their courses easier. Making courses easier has a twofold effect: students can get better grade in those instructors' courses and the student evaluations are better than for those instructors who maintain high standards.

Those instructors who admitted changing their grading practices mostly indicated that they tend to grade easily. The primary reason for that is to make sure that more students pass that course, although some instructors who admitted to grade easier also indicated that it is the only the only way some students could get an A in their courses.

One instructor of English who was denied tenure at his previous institution admitted to the following:

> "I've learned from previous experience not to be judged as a horrible teacher. I'm allowing one rewrite per student and sometimes re-grade papers when students make a reasonable request. At the end of the day you figure that as long as students are learning, you learn not to be so rigid about grading."

What transpires from the above statement is that grade inflation comes in many varieties. It is not only grading easier but also allowing resubmissions of papers and allowing re-writes. Those two 'second chances' will likely lead to a higher grade and thus increase the likelihood that students will end up with a better grade in the course than if the instructor would not engage in those types of consideration.

An instructor of Latin American Studies makes a nice case for changing his grading style; he assumes that students will pick up important concepts and analytical skills later after getting out from his course. He says:

> "To some extent I am grading higher than I used to as a teaching assistant. Upon arriving at this institution my colleagues warned me about the overall quality of students and the mixed levels of ability. I'm not so concerned about grades as much as I am about making sure that they [the students] get out of my courses what is important for them to learn and apply."

In this particular scenario the instructor is doing by default what his colleagues are probably doing as well. This instructor is one of those who strongly favor student evaluations. I assume that he must be getting lots of good student evaluations given his laid back attitude about student outcomes and grades.

A mathematics instructor who tends so be an easy grader does it on the assumption that eventually all instructors will end up inflating grades. His comment is:

> "I give lots of good grades but sometimes feel that when my colleagues end up teaching those students later on will end up lowering standards and giving high grades to a lot of students."

What is interesting about the above remark is that grade inflation is seen as something contagious. When an instructor alleges that if he inflates grades, then his colleagues will end up doing the same, is showing very little responsibility and no remorse for doing something that it is unethical. This instructor is opposed to student evaluations but has reached a point that he has ended up giving in to the pressure that those evaluations might establish.

Other faculty members do not admit to changing grading standards but implement other ways to make sure that students get high grades. A history instructor who teaches a large class explains the following:

> "My teaching assistants do all the grading based on a provided rubric. Sometimes I honor a request for re-grading a paper but for the most part students do well mainly because they get study questions, which are quite similar to the exam questions. Having discussed the study questions in class gives the students a significant advantage when taking the exam."

In this case, the strategy is to provide exam questions beforehand, discuss them, and then have those questions reappear for students to regurgitate a previously discussed answer. This instructor is opposed to student evaluations but is satisfying the student population that will evaluate him. In his department, student evaluations are taken very seriously and he is consciously or unconsciously making sure that his course evaluations are satisfactory given the easy nature of his exams.

A chemistry instructor has not changed his grading methods but is doing something that favors grade inflation, as shown below:

> "Graduate teaching assistants do all the grading in my course. Course grades are based on class average and in my course a class average automatically gets a B^-. Overall grades are high but students work hard during the entire semester. Most students are pre-med or science major and grades definitely play a major role in keeping them motivated."

I am not sure how students stay motivated in this course knowing that they can at least get a B minus by just being "average". I asked the instructor if it would be better to keep his students motivated by making it more challenging but his response was that those students are science majors who work hard regardless of the nature of the course, and who need a decent grade point average to get into graduate school and medical or other professional schools. In addition, this instructor favors student evaluations and prides himself of how those evaluations indicate overall satisfaction with his course. I would be satisfied too if my chemistry courses were as easy as this one seems to be. Chemistry is one of the subjects covered on admission exams to medical school and I would be curious to see how the students coming out of this instructor's course score on the chemistry section of the standard exam.

Some instructors are more upfront about admitting that they tend to give high grades, and that they do so to fight a grading system that they consider unfair. An instructor who is in favor of student evaluations and who has changed her teaching methods as a response to those evaluations replies as follows:

> "My idea of grades is that they are an institutional requirement to separate students into discrete groups. As long as my students do the required readings, show up in class and participate, then they deserve a high grade. Why giving students low grades if they do all the required work? Low grades, when they are underserved, prevent students from reaching their highest potential and pursue graduate studies at the institution of their choice or even compete for scholarships. I might be perceived as an easy grader but as long as my students are learning the subject, that's all that matters to me."

Instructors like this one bring a few problems to their institutions. First, they relax academic standards too much to the point that everyone in their class has a guaranteed good grade. Second, they give false hope or pretensions to their students by letting students think they are qualified for advanced study and scholarships, endeavors typically reserved for highly qualified students. Third, this type of instructor makes the job of colleagues teaching these students later on far more difficult by making students feel entitled to high grades with a minimum amount of effort. For someone who considers student evaluations to be part of a democratic system for student participation, her grading system is undemocratic, it is a 'free As for all" system.

One instructor does not admit to inflate grades but does admit that her grading has changed due to administrative pressure. She says:

> "I tend to grade in a generous way. Otherwise very few students could get a good grade in my course. Our administration is obsessed with reporting grade distributions, and I'm not planning on being singled out for having too many low grades in my course."

This statement is interesting because it shows how some faculty members end up compromising their standards due to administrative pressure. Her comments demonstrate that grade inflation may not always be driven by an instructor looking to get better student evaluations. In her case, she senses that there is pressure from the administration to maintain high grades. This instructor is not on a tenure-track position but instead her

position is a three-year contract. She must feel that if she does not comply with administrative expectations of high course grades, the possibility of reappointment might be diminished. Her situation shows that grade inflation is not entirely related to course evaluations; it could be due to administrative review of faculty without permanent status.

There are instructors who contribute to grade inflation but do so indirectly. Grade inflation in their cases is not the result of grading easier but of modifying how students get assessed. An instructor of social sciences s makes the following statement:

> "My exams are take-home and students have a weekend to work on those. I also assign a research paper, and students get to choose which questions they can answer. Getting a good grade in my course is relatively easy. When I started teaching here I realized that students could not write a critical essay during a class period. But if the essay is assigned as a take-home, everyone does a great job and no one complains."

The above statement confirms some of the earlier assumptions about how grade inflation began. By abolishing in-class examinations and lowering expectations, grade inflation then ensues. This strategy is consistent with the argument made by Cahn (1994, p. 32) when he addressed some of the origins of grade inflation as a consequence of the "misguided version of egalitarianism" of the 1960s. In addition, this instructor is aware that he is lowering teaching and grading standards so that his students can do well in his

course. Moreover, his attitude is that as long as students do well, none of them complains, which in turn simplifies his academic life possibly by getting positive student evaluations.

Finally, there are instructors who openly admit to award good grades as a response to student evaluations or to prevent negative evaluations by keeping students "happy". For instance, an instructor makes the following comment:

> "A grade is a subjective measure of student achievement. I typically curve based on the highest grade in the course and it gives me a nice distribution of mostly A's, some B's, and occasionally a C. No one fails my course unless they stop coming to class or don't submit any work. Students seem quite happy with my grading approach, and the course evaluations indicate so."

It is interesting to notice how this instructor uses the term "curve" in an introductory humanities course. What he is doing is inflating grades by stating that his grade distribution is "mostly A's". In addition, he is aware that by doing so, his students are happy. And obviously he is getting praise on the student evaluations.

Another instructor, who seems proud of how student evaluations have influenced her grading, makes the following statement:

> "My grading process has been influenced by what students expect from the course. They want to take responsibility of their own learning. Every student in my course gets an A or A^{-}, and no one from the administration has said that I

> should change my grade distribution. Students leave my course writing beautiful pieces and enjoying every aspect of the writing process. How many other instructors can say that? Student evaluations have influenced my grading because student comments continue to encourage me to continue approaching the course the way I do."

This professor is not only guilty of grade inflation, but also of 'grade compression', a term that will be discussed in detail in chapter IV. Grade compression occurs when variations in course grades are limited to the point that grades can no longer differentiate student performance. In the case of this particular instructor, college grades "yield no information whatsoever" (*The Economist*, 2002, p. 74). This instructor is the same one who allows her students to decide when to submit the midterm paper and can turn in their writing portfolios whenever the students feel like. It is no surprising that every student gets an A. She claims that her students leave her course "writing beautiful pieces" but I wonder how many times they have to rewrite a particular piece or if her students commit any plagiarism and she is unaware of it.

IMPLICATIONS ABOUT GRADING PRACTICES

Some of the survey results shown in the previous section illustrate that lack of tenure is significantly related to grading practices and grades awarded. By having more than half of the sample surveyed admitting that they have changed grading methods as a

result of student evaluations, there is some indication that there is some correlation between student evaluations and grading leniency. Almost all (thirty nine out of forty) of the instructors surveyed teach either introductory courses or courses which are part of the core course requirements for graduation. A large student population takes their courses and the consequences of lowering standards for those courses could have long-term consequences for the students and the overall quality of the undergraduate program at those two institutions.

Moore and Trahan provide three possible explanations for the relationship between tenure status and grading practices (pp. 779-780). The first explanation considers differential amounts of teaching experience. Having less experience in the classroom may not allow an individual instructor to differentiate between various levels of students' academic achievement thus leading to giving higher grades to less than exceptional work. Secondly, it is argued that students learn more in courses taught by newer or less experienced faculty. It is argued that less experienced instructors may be more enthusiastic about teaching than their senior colleagues, who may be "burned out". Consequently, students tend to learn more in courses taught by junior faculty and end up getting higher grades. So far there is not a single study supporting this assumption. A third explanation is the exchange theory. According to the exchange theory, instructors who need high ratings from students obtain those by awarding a large number of high grades. This suggestion is supported by the findings of Moore and Trahan, in which instructors with the less secure positions give the highest grades on average (p. 780). Their data imply that there is a possible relationship between tenure status and grading

practices. Further substantiation of these findings could have tremendous implications to establish a causal relationship between student evaluations and grade inflation. But more research, particularly large-scales surveys, is needed to confirm this assumption.

IN NEED OF A BETTER FACULTY EVALUATION SYSTEM

Almost fifty percent of the instructors surveyed indicated that they had neither changed their teaching nor grading practices. Interestingly, the numbers were similar in both institutions. In their responses, most of them indicated their compromise to maintain high standards, and their lack of fear about what student evaluation results show. The instructors surveyed made extensive comments on what an ideal evaluation system should be like and whether they preferred student evaluations to peer evaluation and why. Finally, they were asked to comment on the peer evaluation system at their institution and how they would change it. Table 3 shows the results of this set of questions.

Table 3. Instructor's preferences for faculty evaluation

Institution	Prefer Student Evaluations (%)	Prefer Peer Evaluation (%)	Would change current peer evaluation system (%)
University A	**20**	**75**	**55**
University B	**40**	**50**	**50**
Number	N= 12	N= 25	N= 21

Most of the instructors surveyed indicated that an ideal evaluation system should consist of various pieces of evaluation such as student evaluation, peer observations, evaluation of research and publications, and consideration of additional services to the department or the institution. Many of the instructors are not opposed to student evaluations; they just want to see other pieces of evaluation being considered along student evaluations. An instructor of foreign languages describes the evaluation system in her department as follows:

> "Our evaluation system is perfect as it is. Student evaluations indicate what students like and dislike about the course. In addition, classroom observations are meant to offer constructive criticism and each faculty member observed has the chance to review the classroom observation report and discuss it with our colleagues, and even suggest changes if needed, before it becomes part of the instructor's file."

This instructor's comments indicate that there is a clearly articulated evaluation system in her department, and that the peer evaluation system seems to be working well. What is unclear from her statement is how seriously her department considers comments from student evaluations indicating aspects of the course that they dislike. An important point made about peer evaluation is the idea of constructive criticism along with the opportunity to meet with the colleagues who made the classroom observation and discuss the observation report. Instructors could never have such opportunity with the results of student evaluations. If student evaluations are less than favorable, the only alternative for the instructor is to make changes in teaching and/or grading methods that would satisfy the students; hoping to get better evaluations the next semester.

One of the instructors comments on the complexity of an ideal evaluation system, particularly because there is not a system that is entirely objective. She says:

> "It is hard to evaluate objectively what goes on in someone's classroom. If evaluation is indeed necessary, then it should include publications, committee work, professional development activities and anything else that reflects the academic life of a professor. Student evaluations should not be considered due to the fact that they all lean toward grading practices, amount of required work, and even personality of the instructor. Some evaluations mention the country I'm from, when it should not be reflective of my teaching quality."

She makes an interesting point, which is common of many instructors on a tenure-track at research universities. Student evaluations do not seem important for those instructors. They tend to be focused on doing research and publishing, and teaching is somewhat of a burden. On the other hand, there are some aspects in her teaching that she might want to consider since student evaluations will continue to be part of the faculty evaluation system unless a major policy change goes into effect.

Other instructors, particularly at one of the institutions surveyed, indicate the lack of a systematic peer evaluation method. As one instructor notes:

> "Peer evaluation is definitely better and more reliable than student evaluations but it's not implemented in a systematic fashion. One classroom observation every

> six semesters is not enough to provide the instructor with adequate criticism, and to observe any necessary improvement over time."

The above statement is common among many of the instructors I surveyed. They are not opposed to peer observation as long as there is a clear purpose and it done periodically, not once every three years before their term contracts expire. In addition, this instructor makes a good point by indicating that an infrequent peer evaluation cannot account for improvements made upon an initial classroom observation.

Other instructors wonder about the objective nature of classroom observations. For some instructors, peer evaluation is somewhat questionable. For instance, the following instructor expresses the following concern:

> "First, there has to be a clear sense of what the [peer] evaluation system is supposed to accomplish. Then, we – the faculty members—should be told what the criteria are and how evaluations are incorporated in the overall evaluation process. So far I've been observed once in four years. Maybe peer evaluation is not meaningful to our department."

The above comments indicate that peer evaluation in many departments is rarely or never implemented. In addition, non-tenured instructors are dubious about the exact purpose of the classroom observation and how it plays within the overall evaluation system. What is puzzling about this comment is that in both universities' website there is a clear statement

indicating that faculty members are evaluated by periodic classroom observations along other pieces of information such as scholastic activity record. Both universities seem to be violating their own policy regarding evaluation of faculty by applying the peer evaluation inconsistently.

Among those instructors who prefer student evaluations to peer evaluations, there were those who were not familiar with a peer evaluation system in their department and consequently could not establish a comparison. Not surprisingly, some of the instructors who prefer student evaluations as a preferred method for faculty evaluation are those who have made changes in their teaching and/or grading practices as a result of student evaluations. One of those instructors who prefer student evaluations makes the following point:

> "Definitely student evaluations should be a major component of the faculty evaluation system. Peer evaluation is too subjective ... Two senior colleagues sitting in the same lecture can come up with such different observations. However, student evaluations are fairly consistent and more critical after students have spent so much time in the course."

SUGGESTED CHANGES FOR PEER EVALUATION CURRENTLY USED

If peer evaluation is going to be applied consistently, then there has to be a uniform mechanism for its implementation. The most common critique at one of the universities is the sporadic nature of classroom observations. Those faculty members working on three-year contracts are only observed right before their contracts are about to

expire. One senior colleague typically observes them, and the only piece of information included in the observation report has to do with classroom dynamics. There are no comments on the course syllabus, the nature of the assigned material or any other curriculum issue. At another university some tenure-track instructors have never been observed. Some of the instructors are not even aware that peer evaluation is part of the overall evaluation system.

Another main concern about the peer evaluation currently used at their institutions, is that it is done once and there is no follow-up to the classroom observation report. As one instructor comments:

> "I would change the peer evaluation currently in place by having colleagues to come my class a few times, at the start of the semester, halfway through, and near the end of the term, and then compare their observations."

This is an important suggestion because the peer evaluators can establish a baseline for the instructor, and then comment on any progress made in areas of her/his teaching that were initially found to be in need of improvement. However, it is time consuming and requires the collaboration of many tenured colleagues willing to spend time and energy evaluating and commenting on a specific instructor's teaching ability.

Another aspect of peer evaluation that some non-tenured faculty are concerned with is the objectivity of the process. One instructor makes the following point:

> "Peer evaluation is a better system than student evaluations but then you have to be concerned about departmental politics, and how accurate the observation is

> when your colleagues are not familiar with your classroom environment. Student demographics change over what time and what used to work ten or twenty years ago may no longer be an effective teaching strategy."

The above comment addresses the subjective nature of peer evaluation. The instructor told me that students call him by his first name and older colleagues might not view this classroom practice favorably. However, it is important to emphasize that any type of evaluation is subjected to a degree of subjectivity. No evaluation method is foolproof, and there is a certain measure of the evaluator's own bias. It is necessary then to have more than one colleague present during a classroom observation. A peer observation policy should be in place at the institutional level or departmental level that considers the most up-to-date methods for classroom observations and the rights of the instructor being observed.

Another concern with peer evaluation is how some instructor may feel 'invaded' when a colleague walks into her/his classroom. None of the reasons heard over the years about faculty reluctance to have other faculty members sitting in their classrooms seem convincing. One of the concerns is that no one enjoys being evaluated because there is an underlying assumption that there is a need for improvement when evaluation takes place (Arreola, 2000, p. xxi). Other concerns include: suspicion that unqualified people will evaluate them; and possible anxiety that they will be accountable for performance in an area of little training or no interest. Perhaps the most immediate reaction to peer evaluation is the lack of objectivity. The so-called sanctity of the classroom as well as the fine line distinguishing a peer from a colleague adds up to the objections to peer

evaluation. An instructor who prefers student evaluations as the primary method of evaluating faculty made the following point:

> "Peer evaluation is too artificial because it includes an hour or two of observations, and is subject to personal or political bias. Students evaluate faculty anonymously, and without knowing the final grade offer suggestions for improving the course and foster a nurturing classroom environment."

The above statement misses the point because the instructor does not acknowledge that students may be subject to personal or political bias. It is true that students have been in the classroom for a significant number of weeks, and have seen diverse aspects of the instructor's teaching style, but to consider student evaluations less subjective than peer evaluation can only go so far. Faculty members who observe their peers possibly volunteer to do so for altruistic reasons and with the best interest that the quality of teaching in their department can be improved. Since peer review takes times and the process is time-consuming, the notion that peer evaluators' own bias might be dangerous and less accurate than student evaluations goes off a tangent by disregarding the sense of contribution and commitment that engaging in peer observation of colleagues represents. It may be true that peer reviewers may be tempted to rate rather than narrate, and to judge rather than describe but by comparing reviewers' comments, any competent administrator should be able to discern subjective comments from objective ones. In addition, the instructor being evaluated should be given the opportunity to discuss the classroom observation report and dispute any statements considered to be inaccurate or inappropriate. Centra (1986, p. 1) made the following argument: "unless faculty

members are willing to leave evaluation of teaching to students, who possess only a limited view, or to administrators, who don't have the time or necessary background, then they must be willing to invest their time in efforts at peer evaluation of teaching". As part of a community of professionals, university instructors should have the disposition and responsibility to monitor one another given the complexity involved in evaluating teaching.

Evaluation of faculty continues and will continue to be a controversial issue because faculty evaluation is extremely valuable for instructors to retain their positions or to be granted tenure. It is expected that faculty would engage in practices that would improve their evaluation scores. As academic institutions increase accountability of their faculty, discussions will continue as to what piece of faculty evaluation is more valuable. More surveys need to be carried out to investigate what faculty members from across a wide range of institutions and disciplines consider the best evaluation method and how to improve peer observation. One of the instructors surveyed offers what I consider an effective response to what an ideal evaluation system should be like:

> "Proper evaluation should consist of various pieces: student evaluations (using a relevant form), peer evaluation, and portfolio evaluation. Unfortunately, peer evaluation is rarely done properly – it should not be done once every three years as it is done in our program. The most important piece to look at should be the portfolio. By looking at teaching materials the administration can get a better sense of what the faculty member has done for specific courses, along with

> publications and anything else relevant to one's academic accomplishments. Teaching is not the only academic activity we do."

As accountability continues to be part of the faculty evaluation in higher education, addressing faculty performance will become increasingly important for reappointment, tenure decisions, and merit pay. It is essential to find and establish a balance among the various methods by which faculty is evaluated so that it can be holistic, transparent and fair to the instructors.

CHAPTER IV: EXPANDING THE DISCOURSE ON STUDENT EVALUATIONS

ISSUES WITH GRADING

What is a grade?

Historically, college grades have played significant educational and social roles (Walvoord and Anderson, 1998, p. 2). Those roles can be grouped under the umbrella terms of evaluation, communication, motivation, and organization. Cahn (1994, pp. 26-33) discusses many of the issues pertaining grades. For him, grades serve multiple purposes; one of which is guiding the student toward assessing her/his own efforts and abilities to formulate future plans (p. 26). Grades also provide with a simplified system to provide information about a student's level of achievement when used fairly and consistently (p. 28). By doing so, grades provide an indicator of the educational process within academic institutions when compiled as the grade point average. The grade point average serves as the prime indicator of whether or not a student has met the institution's standards of academic rigor (Kolevzon, 1981, pp. 195-196). Parents, potential employers, students, and other organizations should be able to rely on the grade point average as the institution's system of accountability.

Economists view grades as a reward for academic performance, which in turn is an indicator to the student's academic ability along effort put into academic endeavors (Hanushek, 1979, p. 352). There is no doubt that college grading is a complex process because it is subject to both external and internal factors (both students' and instructors' perceptions), institutional policies and practices (such as grading policies), and other factors such as: student demographics, changes in curriculum, and changes in grading

practices (Birnbaum, 1977, pp.526-527). Whatever grades students get in a course, it is an indicator of that student's performance in the eye of the instructor. Therefore, it is the responsibility of students, faculty members, department, and institution to have an adequate perception of the differences among grades. As concerns about differences in grading continue to build up, discussions have emerged about the validity of using grades as indicators of student performance. In the conventional model of education, grades are used to evaluate the student's ability to comply with the course mandates (Bain, 2004, p. 152). Yet there is no other direct measure of student performance more practical to use than letter grades.

Given the level of trust placed on grades, then the quality of educational programs at institutions of higher education is measured by looking at grade point averages, among other aspects of educational programs. However, since the 1960s, confidence in the validity of grades has been affected by research exposing a phenomenon in higher education known as 'grade inflation.'

What is grade inflation?

Walvoord and Anderson argue that the meaning of grading is changing all the time (p. 12). However, in a nationwide study done by Juola in 1974, 134 colleges and universities provided data showing that grade point averaged had been raised an average of 0.404 from 1960 to 1973. In another nationwide study done by Trow (1977), it was found that the percentage of students receiving B^+ grade point averages had doubled from 18 to 36 percent in the 15-year period from 1960 to 1975.

Grade inflation is defined as when "similar quality of academic performance in a given course is awarded higher grades at the present than before" (Hu, 2005, p. 19). It indicates that there has been a "mean shift" upward in student grades in a given coursework. There have been numerous studies supporting and disputing the notion of grade inflation; and many studies done to prove grade inflation have shown contradictory data. The most obvious problem with grade inflation is that it undermines the motivational function of grades in the learning process and academic achievement discussed above. Besides grade inflation, there are problems associated with grading. "Grade compression" refers to variations in grades that so limited that it becomes really difficult to differentiate student performance. Hu (2005, p. 19) refers to grade compression as a "decreased correlation in course grades and underlying meritorious measures such as student academic performance. Thus grade compression lowers the capacity of grades to differentiate students' academic performance and reduces equity in the grading process. Grade compression is considered to be worse than grade inflation because it is a distortion in relative grades and more confusing than a uniform upward drift (*The Economist*, 2002, p. 74). For instance, at the limit of grade compression when all Columbia students get A's all the time, then the university's grade yields no information whatsoever. Some experts prefer the term 'grade compression' over 'grade inflation (Cizek, 1996; Hancher, 1994). Cizek argues that because there is no higher grade than A, A's remain as A's, B's become A's, and so on. One last phenomenon in terms of grading problems is "grading disparity" in which similar academic performances are rewarded differently in different courses or in different academic units (Hu, 2005, p. 19). A consequence of grading disparity is that courses choices are affected and could

lead to grade inflation or GPA inflation. Grade inflation is associated with grading disparity when students tend to choose courses that award higher grades.

General Explanations for Grade Inflation

Birnbaum (1977) performed one of the first studies that attempted to address the cause of grade inflation. His study collected and analyzed trends in grade point average at the University of Wisconsin-Oshkosh from 1968 to 1974. He began his work hypothesizing that grade point average increases were the result of a combination of the following factors (pp. 528-529):

- Institutions were attracting more qualified high school graduates.
- Increased student achievement and those students were learning more.
- The sex mix of students changed and more women were enrolled.
- More students majored in areas that traditionally gave higher grades.
- Students no longer took the same courses.
- Students increased their use of pass-fail or non-penalizing withdrawal options.

The results from his study did not support five of his hypotheses. Only the hypothesis dealing with the use of pass-fail option, and having students withdrawing from courses very late in the semester without penalty seemed to have a correlation with the increase in grades. Under those circumstances, low grades such as C, D, and F become a rare species. Consequently, the number of A's and B's tend to become prevalent.

At the time that Birnbaum's data were released, grade inflation was not seen as such a bad thing. Hoyt (1966) argued that he found no evidence indicating a relationship

between college grade point average and professional success. On the other hand, Davidson (1975), considered a college degree a marketable quantity and predicted that grade inflation will disadvantaged able students, resulting in a loss of "customer" to higher education. Birnbaum somehow dismissed Davidson's assumption and concluded that outside agencies can always rely on outside indicators other than grades to judge an individual's potential and ability (p. 537).

Changes in grading policy

Departmental differences have been seen as direct causes of grade inflation. Studies have found that there has been a student migration form low to high grading departments (Bearden, Wolf, and Grosch, 1992). Students then develop a shrewd sense of grades, and decide to take courses in those departments where the averages grades are higher. A consequence of this 'departmental flight' is that those departments with lower student enrollments would then succumb to pressure and could end up inflating grades to recruit and retain students (Sabot and Wakeman-Linn, 1991).

A study done by Prather, Smith, and Kodras (1979) in which they analyzed 144 courses found that the students' course-taking patterns are responsible for changes in student grade point averages. They found that challenging courses were replaced by field-centered or experience-oriented courses. The main reason for the observed course pattern has to do with the higher grades awarded in those courses. By having a large student body enrolled in those courses with higher grades, then there is a tendency for overall grade point average to increase as well.

Grading practices have been accounted for grade inflation. Those changes in grading policies could have been the result of changes in faculty characteristics. The change in faculty population over time may alter student grades (Hu, 2005, p. 52). In Goldman (1985) it was suggested that: “It seems likely that as older professors have retired from teaching, younger professors with more lenient grading practices have replaced them. Grade inflation could thereby occur even if the standards of all individual professors remained constant” (p. 102). This interpretation is supported by my own study in which non-tenured faculty have been influenced by a number of factors, and tend to award higher grades. Younger faculty may also bring their own experiences in grade inflation from high school and undergraduate work. It may be possible that they have been conditioned to grade inflation in their own experiences, and end up seeing the assignment of higher grades as something ‘normal’.

There have also been changes in administrative practices leading to grade inflation. Colleges have adopted more liberal grading practices such as allowing late withdrawal from courses, removing first-attempt grades from transcripts if students later repeat the courses, and the notorious pass-fail option. When a student has the option of pass-fail from courses considered difficult, the grade point average of the student would not change dramatically. This practice of enrolling in a course using the pass-fail option is often seen in science and mathematics courses for non-majors, and subjects such as economics. When Johnson (2003) did a study on grade inflation at Duke University, he found that students were not motivated to take courses in the lower-grading fields if not

required. If faced with the need to take a course in those low-grading fields, then the pass-fail option becomes tempting for the student.

Cole (1994) attributes grade inflation to faculty laziness, claiming that it is easier to record a good grade than a bad one; high grades normally do not need to be justified while low grades have to be defended. As Zirkel (1999) noted, "When the rationalizations are stripped away from the rationales, the basic problem is that high grades are simply easier" (p. 255). As presented in my survey results, some instructors thought that by awarding high grades everybody wins and both students and administrators would be 'happy'.

One must consider the growth of faculty by employment status since 1970 when major college expansion took place. In the period starting in 1970 and ending in 2003, Schuster and Finkelstein (2006) have documented a change from 474,000 full-time faculty members to 1.173 million. That is not such a significant increase (it is about 42 percent) considering how student enrollment has continued to grow. On the other hand, the number of part-time faculty has increased from about 104,000 in 1970 to 543,000 in 2003. The increase in part-time faculty represents a growth of about 87 percent (pp. 40-41). Such changes inevitably could lead to changes in the academic landscape by altering values, aspirations, and commitment to the workplace. As this transformation in the make-up of the faculty continues, it may be possible to see more changes in the way students' work is graded.

Another possible cause of grade inflation could be due to grading disparities among subject areas. Faculty members in disciplines where there is a lot of subjective grading involved would have to spend more time defending their grading if they were going to adopt stringent grading practices (Hu, 2005, p. 54). Having faculty members spending considerable time evaluating students' writing might result in the likelihood of adopting more lenient grading practices. Similarly, faculty members in departments who traditionally award low grades, such as science and mathematics, are faced with the dilemma to offer courses with lenient grading practices in order to boost student enrollment (Shea, 1994).

A less documented factor possibly involved in grade inflation is the blurring of the faculty-student relationship. As more departments and instructors adopt a more constructivist approach to learning, the apprenticeship model has replaced the traditional hierarchy of faculty-student relationship. In the apprentice model, an instructor may find it extremely difficult to separate the teaching activity from the process of evaluating the student. It is possible then that the sense of equality between faculty and students leads to a reduced objectivity, which then translates in the reluctance to give a low grade or fail a students thus leading to grade inflation.

Changes in curriculum

It has been argued that inflated grades are also the result of professors spoon-feeding their students, and making lectures more like an entertainment show. While in the past students had to put effort to determine how to study the course material and what was significant in a required reading, many students today are told what is important to know for an exam. Consequently, students are receiving better grades while making very little effort to master the material.

Student affect has been listed as a possible cause for cause inflation. As American education began to emphasize student self-esteem, instructors somehow developed an attitude to become part-time therapists for their students. According to Mansfield (2001, p. 1), the "Purpose of education is to make students feel capable and empowered. So to grade them, or to grade them strictly, is cruel and dehumanizing. Grading creates stress. It encourages competition rather than harmony. It is judgmental". Cahn (1994, p. 26) addressed this issue when he mentioned that some critics of grades considered them "inherently inaccurate devices that … only traumatize and dehumanize them [the students]". This paternalistic view of students could have led some instructors to award higher grades attempting not to harm students' egos and instead praising them for an academic work below expectations. Doing so only led to minimizing the significance of achievement for the sake of allowing students to get the false impression that they were more academically strong than they really were.

Student Evaluations of Teaching

Of all the suggested reasons for grade inflation, student evaluations of teaching occupy center stage. I have argued during my study that the use of student evaluations has pushed instructors to grade leniently hoping to get favorable ratings on those evaluations. Johnson (2003) provides the most compelling argument correlating student evaluations of teaching with grade inflation. After examining the relationship between grades and student evaluations of teaching along with contrasting interpretations, he concluded, "nearly all studies conducted have resulted in reports of a positive correlation between these variables, but debate continues over the cause of this association" (p. 47). In order to counteract those who propose the teacher-effectiveness theory arguing that students learn more in courses offered my more effective teachers (Hu, 2005, p. 56), Johnson argued, "If correlations between grades and student evaluations of teaching result from differences in teacher effectiveness, then correlations to teacher-course evaluations for differences in assigned grades are both unnecessary and inappropriate. If, on the other hand, positive correlations between student grades and student evaluations of teaching are explained by student attribution or grade-leniency effects, corrections to teacher-course evaluations are needed to avoid the repercussion"(pp. 73-75). He conducted the **D**uke **U**ndergraduate **E**valuate **T**eaching experiment (DUET) using responses collected from the same students both before and after they received their final course grades. Analysis of DUET data indicated that the effects of grades on student evaluations was statistically significant and concluded, "instructors can often double their odds of receiving high evaluations from students by awarding A's rather than B's or C's" (p. 83).

According to the results of the survey I conducted, about 50 percent of the faculty interviewed admitted to having made changes to their grading practices as a result of student evaluations. These results are in agreement with a previous study (Nelson and Lynch, 1984) in which a national survey of deans of colleges of education and colleges of arts and sciences indicated that about 70 percent of the deans who responded agreed that the use of student evaluations of teaching as a consideration for promotion and tenure resulted in grade inflation. Nonetheless, student evaluations continue to be used widespread and there is no sign that their use is going to be stopped anytime soon.

Disengagement Compact

Another possible factor implicated in grade inflation is the "disengagement compact" theory (Kuh, 2003). According to this theory, college students and their instructors follow the idea: "I'll leave you alone if you leave me alone." In general, by subscribing to this theory, instructors tend to award high grade without significant student effort and by doing so there are fewer grading questions to deal with. In this relationship there is mutualism and both students and instructors benefit by moving on to their respective niches. The "disengagement compact" may have emerged from the changes in priorities of the research university. Faculty are hired to teach but are primarily expected to do research and publish. As the "publish or perish" mentality continues to be a major determining factor in faculty promotion and tenure, awarding higher grades has the dual benefit of securing better student evaluations and having more time to do research. As the workload of college instructors increases, the time spent of grading objectively may be compromised, and shortcuts seem tempting.

It is evident that there are multiple factors implicated in grade inflation, grade compression, and other grading issues. As college enrollment figures continue to increase and more non-tenured faculty and adjunct instructors are hired grade inflation may continue to spiral upward. Colleges and universities need to develop more consistent policies regarding the evaluation of faculty if they want to control grading issues and provide new faculty members with a whole new set of incentives to grade fairly and concentrate of quality teaching.

Contingent Faculty

Contingent work is defined as "any job in which an individual does not have explicit or implicit contract for long-term employment or in which the minimum hours worked can vary in a nonsystematic manner" (Povlik and Nardone, 1989, p. 11). The growth in contingent faculty in higher education since the mid-1970s mirrors the increase in the use of contingent workers in the American labor market. In 1975 there were approximately 13% of full-time non-tenure track faculty and adjuncts composed about 30% of the entire faculty. By 2003, non-tenure faculty increased to 19% and adjunct faculty increased to about 46% (Umbach, 2007, p. 93). Researchers and faculty are becoming increasingly concerned about this change in the mix of tenure-track and non-tenured faculty mix (American Association of University Professors, 2003) on the basis that academic freedom is at risk. Others point that the costly nature of tenure along with decreases in faculty productivity, and the decreased ability of colleges and universities to adapt to a rapidly changing society are some of the reasons for this increase in the use of contingent faculty (Baldwin and Chronister, 2002, p. 126). Contingent faculty represent

a more flexible and less expensive source of instruction as decreases in government funding and sharp increases in student enrollment have permeated higher education institutions (Umbach, p. 92).

The major concern n with the use of contingent faculty has to do with its unintended consequences. Researchers argue that as the number of tenured faculty decreases, academic freedom will be compromised and there will be permanent damage to the academic profession (p. 92). Umbach (2007) set out to investigate the effects of contingent faculty on undergraduate teaching and learning. Specifically, he focused on faculty behaviors that engage students in good practices leading to increased learning and academic performance. His results suggest that part-time faculty are underperforming in their delivery of instruction, interact students less frequently, spend less time preparing for class, and have lower expectations that tenured and tenure-track colleagues (p. 110). In the case of full-time non-tenured faculty the results were mixed but they still were found to interact with students less often. These results are important to inform the policy process regarding how faculty are evaluated, and suggest that their reduced level of academic expectations could contribute to grade inflation. What is unclear is whether those contingent faculty members receive positive evaluations from students. When it comes to non-tenure track and adjunct faculty a proper faculty evaluation system becomes highly significant in order to provide them with adequate professional support.

PROBLEMS WITH STUDENT EVALUATIONS

One of the main concerns that critics of student evaluations of teaching have is the notion that teaching is so hard to evaluate that it cannot be done like in other professions, in which colleagues with the same or superior expertise measure the quality of the professional being evaluated. By allowing students to evaluate teaching, we are saying that teaching has no standards to be evaluated against and thus students, who spend a significant number of hours in the instructor's class has the capacity to assess teaching. But the question we must ask ourselves is: what are we trying to evaluate when we a teaching evaluation? The answer has to do with one's particular framework.

In a utilitarian framework, evaluating teaching would consist in measuring outcomes. Questions asked by an evaluator would be of the following nature, for example:

- Is the required material covered?
- Is the instructor's performance adequate?
- Does the instructor lecture clearly?
- Is the number of examinations adequate?
- Is the graded work returned promptly?
- Does the class begin on time?
- Are the students satisfied with the instructor?

The questions listed above reflect the so-called good practices but reveal very little about the kind of learning taking place in the classroom and focus on what the professor does but not about what students are learning. An effective evaluation form (not necessarily a student evaluation form) should address questions such as:

- Is the instructor helping the students learn the subject effectively?
- Is the instructor encouraging the students to learn the material?
- Is the material appropriate to the curriculum?
- Are the teaching methods used adequate for a diverse student population?

To answer these questions, the evaluator needs to be proficient with the course material, knowledgeable about pedagogy, and familiar with what constitutes effective learning. Students are definitely not qualified to answer those questions. For example, a good student might give high ratings to an instructor who fosters independent learning and constantly challenges her/his students whereas a student who expects an easy course will evaluate the same instructor mercilessly. The opposite situation may hold true if the instructor expects just plain memorization of facts and the workload is relatively light.

According to Bain (pp. 165-166), situations such as the ones described above discredit student ratings. Are students capable of evaluating their professors? Yes, but only if the right questions are asked. Students should evaluate their instructors from a learner perspective that indicates that some satisfactory amount of learning has taken place, using adequate questions that correlate to a specific course, and the specific

activities that have taken place in the course. One of the instructors I surveyed made the following comment:

> "Student evaluations can be useful if they ask the right questions, but most evaluation forms seem irrelevant for a particular course or a particular instructor's teaching style. For the most part they don't seem to contribute much information except that they tend to be either highly positive or highly negative, rarely in between."

This instructor's remarks reflect that students may tend to see teaching as a black or white issue, and may not acknowledge gray areas of someone's teaching style. Moreover, a typical evaluation form is more like a judgment form of some kind of popularity pageant because students are rarely explained what the purpose of the evaluation is. Most forms just include instructions for completing them but rarely provide a statement of the kind of information expected from the form, and how such information will be used by the university administration.

Perhaps an 'accidental' function of student evaluations is illustrated by an instructor who made the following comment:

> "The function of student evaluations is genuinely commendable…however, in practice it just doesn't work well. Students use such platform to vent their anger and frustration at having to work harder in a particular course. Whatever students write on those forms is not useful to the instructor. Unfortunately, administrators

take student comments too seriously. Just like the slogan, “the customer is always right”.”

This statement brings another aspect of student evaluations of teaching that is highly contested among their critics: the notion that those evaluations allow students to function as consumers of a product. The rise in ‘student consumerism’—the idea that universities operate like businesses for student clients has been linked to the use of student evaluation and the rise in grade inflation (American Academy of Arts and Sciences, 2002, p. 9). The idea of the “student as consumer” has gained momentum more recently. Eizler (2002, p. 499) argues that student evaluations of teaching are a form of customer satisfaction survey because many institutions use them as the primary, if not the only, basis of judgments about teaching.

ISSUES WITH PEER EVALUATION

Although peer observation of faculty teaching is often considered synonymous with peer evaluation, it is not. A peer observation of classroom teaching is just one piece of evidence within the conglomerate of what peer evaluation constitutes. In fact, it is considered that peer evaluation that consists mostly of classroom observation to be a mistake (Centra, 1999, p. 117). One of the main issues with peer evaluation is how valid it could be since it could be easily biased depending upon whether it is used for formative (improvement) versus summative (judgmental) purposes (Arreola, 2000, p.

69). The formative peer observation is essential for junior faculty as part of the teaching improvement process before promotion and tenure review. This process can prepare junior faculty for the teaching demands. Summative peer observation involves the evaluation of peer classroom behavior, and includes the review of all teaching materials. Although the outcomes of both formative and summative observation are different, formative peer observation serves as a transition for summative evaluation at the departmental level. It is thus essential to establish *a priori* a specification of the instructor's activities to be evaluated by her/his peers. By doing so, the reliability of the peer evaluation is somehow ensured. Another prominent issue in peer evaluation is who the peer evaluators are. Arreola (p. 75) suggests that a three-member committee be appointed. The department head should select one member of the committee, the faculty at large selects a second member, and the faculty member being evaluated selects a third member. It is considered that under this triad approach the faculty member being evaluated will have one person in the committee preventing possible negative from the two other members.

Another contentious issue dealing with peer evaluation is the number of classroom visits. Although there is no consensus in the peer review literature about how many visits a peer evaluation committee should consider, there are some basic considerations that should be taken. First, not all committee members should visit a classroom at the same time. The reason for this is to observe different aspects of the courses. Arreola (p. 76) recommends that each visitation should be explained to the class, and if possible, to visit the class prior to the official class observation. In addition, there

should a pre-visit conference to discuss the evaluation instrument being used, teaching and/or classroom issues that should be considered during the observation process, and any classroom environment (such as student participation or lack of) that is not truly representative of normal classroom activity (p. 76).

During my study a recurring issue was the lack or the inadequacy of peer evaluation that takes place at the two institutions I observed. As one faculty member interviewed commented:

> "Peer evaluation … consists of one senior faculty sitting in class for an hour and writing a summary of the observation."

This comment seems valid across many departments of the particular institution. This situation raises some concerns, particularly since most faculty members have a preference for peer evaluation and could benefit tremendously from the feedback associated with an evaluation system done by colleagues.

At another institution surveyed, peer evaluation is only done once over a three-year contract cycle. Most of the faculty members interviewed at that institution embraced peer evaluation. As one instructor pointed out:

> "Peer evaluation should be the only way to evaluate faculty. Students do not evaluate objectively. Colleagues are critical and have their best interests in

> improving someone else's chances of becoming better teachers and enhancing the department's image."

However, although embraced by many, classroom observations are often done just once within a contract cycle, and the amount and/or quality of appropriate feedback may be inadequate for an instructor genuinely interested in improving her/his teaching. As one faculty member interviewed noted:

> "One classroom observation every six semesters is not enough to provide the instructor constructive criticism, and to observe any necessary improvement over time."

It seems obvious that there is a large untapped potential for peers to assist in evaluating teaching but such potential is not being exploited. Peers' general impressions should serve faculty members in need to ideas and criticism. The touchy aspects of peer evaluation should be overlooked and its positive aspects should be highlighted. Even in the event that peer evaluation is weighted very low in the faculty evaluation process, those being evaluated should not be deprived of the potential benefits for self-improvement that peer evaluation can provide.

Contrary to student evaluations, little has been reported on the usefulness of peer observation of teaching. While peer observation is a common practice in the British higher education system as a mechanism for enhancing the quality of teaching and

learning, peer observation is not as prominent in the United States (Fullerton, 1999). Perhaps such lack of prominence is due to the idea that measurement and evaluation of effective teaching are ambiguous and subjective. As faculty members aim to advance their careers, they should welcome the feedback of their colleagues in order to enhance their teaching. Hence, peer evaluation needs to be an essential element in any faculty evaluation system.

A recent study by Kohut, Burnap, and Yon (2007) argues for peer observation as evidence of effective teaching that requires "a very high degree of professional ethics and objectivity, and training in observational and analytical skills" (p. 20). In their study, they compared the perceptions of observers and those observed regarding peer observation and the usefulness of peer observation as an evaluation tool that improves teaching effectiveness. They found that observers had a broad range of feelings about being trained to engage in classroom observations (p. 21), which point to the need to develop a systematic way of training properly those senior faculty members who would become peer evaluators. On the other hand, observers felt comfortable making constructive comments but were less comfortable when making critical comments (p. 22). This observation is important because anyone who agrees to be a peer evaluator must have sufficient objectivity to point out deficiencies in the teaching process if peer evaluation is meant to impact the quality of teaching and learning.

Although the studies done on peer observation are useful to unpack more knowledge about the process, more studies need to be developed to gain more insights of

the relative value of peer observation when making re-appointment and promotion decisions. As demands for greater accountability of higher education continue to increase, more accurate information about evaluating teaching performance is needed. In addition, the significance of peer evaluation as a crucial component of the faculty evaluation process needs to be reaffirmed by performing more studies that could stimulate a greater investment in institutional support, faculty time, and more interest in documenting teaching performance by other means besides student evaluations.

EDUCATION AS A CONSUMER ITEM

Higher education has been confronted with a series of new demands: changes in society, new students, changing expectations towards learning and teaching, and a range of complex economic and cultural attitudes. Changes brought about by these demands emanate from various stakeholders as our society keeps evolving. Chevaillier (2002) has deftly expressed how these changes have emerged:

> "These demands were expressions of deep changes in society and in roles assigned by it to universities. In a world where knowledge was becoming the essential resource to foster and sustain economic development, they were increasingly expected not only to produce new knowledge but also to apply it to production by taking part into the innovative process… These aspirations were accompanied, in most developed countries, by a substantial financial effort from public budgets and private sources. By bringing the attention to bear on higher education institutions, they also made them accountable to society at large… Some of these new demands came from students. New types of students entered

> universities or became a significant part of the student population… They have a choice and therefore bring competition to bear on institutions. They act as clients on various markets or on various segments of an education market. In order to attract them and to keep them, universities may resort to marketing techniques developed by business. Mission statements are turned into business plans." (303-4)

As things change in society so do the institutions of higher education. The struggle to define the nature of knowledge production becomes contested by the various actors in the policymaking process: students, politicians, institutional leaders, researchers, educators, and particularly business leaders (p. 307). As social institutions continue to change, higher education will continue to do so as well but this trend is toward an output system.

Part of the problem stated above has to do with the educational institution itself. Faculty members are expected to display all kinds of maneuvers in terms of creativity and innovation semester after semester without much opportunity to become reinvigorated (Shepperd, 1997, p. 335). The perspective by which faculty and students are defined has certainly changed. By "meeting the needs of students", we have tended to endorse different pedagogical styles, which in turn has promoted student consumerism – namely the sense of entitlement emerging when they exploit the idea of "student-as-consumer" (Delucchi and Smith, 1997, p. 336). Genuine learning has been substituted with something else. Global and commercial pressures have altered the university's teaching and research missions. As we adopt the term *globalization* more and more into our daily

lingo, the university will no longer be a social institution but an industry – subjected to market forces just like any other business.

Riesman (1980) addressed a so-called "new supremacy of the student market: shift from academic merit to student consumerism as a reversal of direction in the history of American higher education" (p. xiv) in which students are being courted as customers and view themselves as passive consumers of education. The notion being described here begins with the college marketing strategies, in which virtually every college catalog looks like a piece of real estate being advertised for rental or sale. In a market-driven society perhaps it is hard to put the blame entirely on private colleges struggling to survive on tuition money. As Riesman said:

> "… colleges are seeking in various ways to position themselves, in the best cases to try to influence student choices for the sake of the students themselves, and in many cases simply to respond to market forces by the various strategies colleges have found that will retain some members of their old clientele while inviting new ones to fill up the vacancies" (p. 105).

However, the notion of students as consumers is not new. Clark (2006) addresses this concept of students as consumers dating back to 1782 in Germany. Back then, it was thought that the university must be managed with a firm hand since the university was being compared to a factory [*wissenschaftlicher Fabrik*] that exists to serve customers (p. 379).

How are colleges and universities fulfilling this "customer-based model"? By changing the philosophy of what a higher education experience represents. As noted by Riesman: "Desperate for students who will bring in tuition, in part via federal grants and loans, and in a few cases through state support of scholarships at independent institutions, most unselective private colleges have completely abandoned any requirements either at entry point or at graduation, that would keep away students who could possibly be attracted to matriculate and remain" (p. 108). It may be possible that the larger number of students now enrolled in college is the result of effective marketing strategies where education is sold as a commodity. Leef (2006) argues that based on the premise of selling education as a commercial necessity, then education is highly oversold given that the enormous expansion of academic options has led to the deterioration of academic standards, credential inflation, and a soaring cost of college while providing little benefit on many of the students – to the point of "going past the point of diminishing returns" (p. 18).

If education is purchased as goods and services, then it seems that buyers are not getting accurate information about the product they are buying. Why is that? When a college or university is not selective among its applicant pool, they "are admitting a substantial number of students who are fundamentally disinterested in the *raison d'être* of higher education – the life of the mind" (p. 23). For some students, attending college is as easy as going shopping to a nearby mall. Distance learning, online degrees, and other methods of obtaining higher education credentials have led to the notion that a college or university education is part of the acquisition process taken for granted in a capitalist

system. This idea was set forth by Milton and Rose Friedman (1980) when they stated that college "attract(s) many young men and women who come because the fees are low, residential housing and food are subsidized, and above all, many other young people are there. For them, college is a pleasant interlude between high school and going to work." (pp. 175-6).

The idea presented above is more common in private colleges and universities, particularly those without a large endowment. These institutions need all the revenues they can get. As Leef points out:

> "Schools like Harvard and Princeton don't have to worry about their finances, but most schools are ravenous for every dollar of tuition, grant and loan money, room and board money, and student fees. They know that they accept large numbers of disengaged students, who will need remedial (or "developmental", to use the preferred euphemism) courses, and that even with those courses many will struggle in school or drop out. But many colleges and universities would face a tremendous financial problem if they only accepted serious, well-prepared students." (p. 23).

The idea of student consumerism is also enhanced by instances of intervention for consumer protection. Since the 1960s, when student movements became widespread, "there has been a mobilization of students locally and nationally to be heard on certain issues that plainly affect them in their roles as purchasers of educational services…"

(Riesman, p. 356). We now have students sitting in various university committees, including negotiation tables thus perpetuating the concept of consumers managing their rights in a market-driven economy (p. 357). To some extent this phenomenon has led to grade inflation when "faculty members, desperate for students, compete for them by an automatic grade of A and by demanding minimal amounts of work, this unprofessional (though understandable) strategy can scarcely be called fraudulent, since the students are eager rather than deceived consumers" (p. 358). What is interesting about this paradoxical situation is that if you are a consumer of education, why should you be graded as you acquire your purchase?

The idea of the university run like a business comes from the way it is organized and how it operates. Kerr (1982) further developed this idea of university organization. He suggested that the top-down management structure of higher education lends itself to the corporate approach taken by many, if not all, institutions of higher learning. He wrote that:

> "Students do have considerable strictly academic influence, however, quite beyond that with which they are usually credited. The system of electives gives them a chance to help determine in which areas and disciplines a university will grow. Their choices as consumers guide university expansion and contraction, and this process is far superior to a more rigid guild system of producer determination as in medicine where quotas are traditional. Also students, by their patronage, designate their university teachers. The faculty may, in fact, appoint the faculty, but within this faculty group the students choose the real teachers. In

> a large university a quarter of the faculty may be selected by the students to do half or more of the actual teaching; the students also "select" ten percent or more to do almost none at all." (pp. 21-22)

Once again, the above statement exemplifies how students as customers shape the way a university operates, and in turn control how some other aspects such as teaching are highly influenced by the students' choices.

Gould (2003) gives a more recent take on the corporate nature of higher education. Besides the usual nature of students as consumers who determine the output of the educational enterprise, he establishes that:

> "Universities are always in competition with one another for students and funding. State and federal governments step in to help and to make some new academic demands, but there are no thoroughgoing national controls, no common set of standards for intellectual achievement, degree quality, or assessment of learning. There are many first-rate institutions of higher learning, and the system has a remarkable way of sustaining itself with some measure of self-regulation. But higher education is always under negotiation." (pp. xi-xii)

In other words, the nature in which the higher education system has been built makes it a default system to treat its intrinsic structure as a business entity. Whether the institution is oblivious to it or not, the culture of the institution is built upon an enterprise ideal. If we consider the following argument set by Gould:

> "… the general goals of higher education focus mainly on a comprehensive social mandate for higher education, one that seems to reflect that we want such an education to offset the consumerist culture that drives our economy, even while we replicate that very same culture in the university…. Higher education, after all, is the child of its context; it is driven by social forces of change – technological, social, economic, cultural, and political – that challenge any single integrative *idea* for the university. It is also the child of its market context, which relentlessly organizes both the management and development of knowledge." (p. 6)

His argument reflects a pessimistic idea that there is no way to reverse this trend – that higher education is doomed to continue embracing a consumerist philosophy of operating, and that, if anything, this trend could only intensify due to market pressures.

This market-oriented institutional idea is nothing new. During the period of university protests and reconfiguration in the late 1960s, institutions began a sharp trend of giving in to external pressures. As Gould wrote:

> "After all, the market-driven flexibility demanded of the university informed its runaway growth in the 1960s and 1970s, its ongoing financial concerns and growing costs, its difficulty in diversifying in order to keep up with demographic changes (or its overdiversification into multi-cultural correctness, depending on whom one is speaking to), the lack of national standard and a general "dumbing down" of academic expectations, a declining interest in liberal arts degrees, the increasing culture of entitlement among students, the comprehensive spread of corporate culture in managing and generating knowledge, … universities are of central importance to society because they must reproduce the American way of

> life for succeeding generations, sustain liberal capitalism with appropriate energy, support research, and do all this within a market-driven culture." (pp. 9-10)

Gottfied (2002) considers the notion of multi-cultural correctness mentioned above. He uses the argument in a consumerist model of education, professors are supposed to make "student-customers feel comfortable about who they are" (p. 54). He argues that the current idea of education is that everyone is entitled to the "college experience" even if not everyone is college material. Consequently, many of today's college students are less academically inclined and "are kept in school through profit-accruing tricks, e.g. multiplying "hands-on" vocational curricula, making faculty salary depend on favorable student evaluations, or by having administrators browbeat anxious "learning facilitators" into passing classroom mannequins" (p. 55). He blames this trend on the multicultural ideology by the federal government to feature diversity policies, such as recruiting minorities and creating for them a non-hostile environment. Further, he argues that the endowment of major American universities includes a considerable amount earmarked for minority students. Although Gottfried somehow shies away for expanding his argument, he openly blames the multiculturalism of universities as a "particular form of commerce" (p. 57) that behaves like other profit-driven enterprises, and perpetuates the concept of diploma mills that attempt to leave student-customers unstressed.

Professor of philosophy Michael Potts (2005) gives a mode detailed account of some of the perils of a consumer model of education. First, he blames student

evaluations of faculty as a consequence of the notion that “students are our customers” (p. 54). In his view, professors are “purveyors of customer service” and students always considered to be right in disputes with a faculty member. He argues three points by which the business model of education is damaging the mission of higher education (pp. 62-63). First, when students are viewed as customer, then the student-teacher relationship no longer follows a mentoring model and instead becomes a customer-salesperson relationship. Second, a customer model could entice a student to be dishonest in the name of “doing what it takes to succeed”, such as cheating. Third, faculty members focus their classes in providing “customer satisfaction” rather than imparting knowledge and intellectual virtue. When these three notions take over higher education, the system then becomes a so-called “diploma mill”. The mission of educating and cultivating intellectual excellence is then deeply affected, and the utilitarian idea of a university takes over and the institution might no longer be an intellectual community where academic exchanges can take place; at least not necessarily.

However, one must wonder whether the institutions of higher educations are just victims of the trend on increasing marketability. Liberal education has been historically under a contest terrain. The knowledge derived from it “has always seemed to embody a tension between utilitarian outcomes and learning for its own sake” (p. 149). How far will this trend of compromising higher education go? No one knows for certain but if liberal education ever disappears from our institutions, it will be a great disservice to democracy. Compromising the mission of higher education – the pursuit of knowledge

and the analysis of values – is a dangerous idea; it assumes that the market can define who we are and where we are heading to as a society.

For-profit universities illustrate an example of response to market forces. This trend in institutions, best illustrated by the University of Phoenix, represents a response to specific market needs, and the adjustment of enrollment and hiring of part-time faculty as needed. Although these institutions do not offer tenure, it is worth considering the four areas for change that these institutions operate from: (Ruch, 2001, p. 149)

- Response to market forces
- Adaptation of the organizational structure
- Redefinition of shared governance
- Development of a strong customer orientation

Of those four areas of change that characterize the structure of for-profit universities, the fourth point: development of a strong customer orientation is important for my discussion of student consumerism. Ruch (p. 153) agrees that such strong customer-service orientation is one of the reasons why the number of students at for-profit universities continues to grow. However, he defends the position of strong-customer orientation given that he works at a for-profit institution himself. He indicates that:

> "…treating students like customers does not mean that they cease to be students as well, or that the institution must give in to all their preferences, or that faculty must give away good grades for the sake of happy customers. It simply means that the institution becomes more responsive to its students and makes serving

> them effectively the highest priority. Failing to do so, I believe, will result in students' taking their business elsewhere."

Although Ruch never explains why customer satisfaction is important, he makes two interesting points in the quote shown above. Being "responsive" to students could mean a lot of things. Being responsive to a lot of student demands could end reconfiguring the landscape of the institution such as it happened by the end of the 1960s when student demands were met by allowing them to evaluate their instructors, suggest changes to the curriculum, and having alternate options for grades in a course. What if the students' demands are not met? Then, they can take "their business elsewhere". Ruch makes this point without going into details but it implies that the institution might not be willing to let those students take "their business elsewhere", particularly when the institution's endowment is tuition-based. Students treated like customers can make unreasonable demands. What if those demands imply softening academic standards or modifying grading policies to make them more relaxed? Students could also demand the removal of a particular instructor who holds high standards thus leading to grade inflation. The possibilities for meetings students' demands are numerous and so are the opportunities for lowering the quality of education provided.

Market can also define how instructors view their positions, and how they approach teaching. During my survey, some of the instructors complained about the pressure for doing research, for publishing, and for some of them the classroom was the least important of all the activities they were responsible for. Many of the instructors

were aware that their research record and publications list would have a better chance of getting them tenured than a good teaching record. This attitude was particularly true among science and mathematics instructors, as one physics instructor stated:

> "In my department I've never heard of peer evaluation… Student evaluations don't seem so important. Basically you get tenured based on research, publications, and the ability to secure grants."

Quality of instructor may be responsive to market forces, and in a race to get tenure the stronger force is the ability to attract extramural funding and maintain a vigorous research program that can be known outside the institution.

The most recent articulation that higher education is a consumer item has been the recent report by the current Secretary of Education. In September 2006 Margaret Spellings issued a commissioned report charting the future of higher education. The language in the Preface is as follows:

> "In this consumer-driven environment, students increasingly care little about the distinctions that sometimes preoccupy the academic establishment, from whether a college has for-profit status to whether its classes are offered online or in brick-and-mortar buildings. Instead, they care – as we do – about results. (p. viii)

What else can be said against such a strong argument? When the language is crystal-clear about what our Secretary of Education attempts to do: access, affordability, quality, and accountability (p. x), change is on its way. It is not a valid argument, even in economic

terms. Greater accountability does not necessarily mean better results instantly. The report invokes for more "transparency and accountability" about how American colleges and universities operate:

> "Colleges and universities must become more transparent about cost, price, and student success outcomes, and must willingly share this information with students and families. Student achievement which is inextricably connected to institutional success must be measured by institutions on a "value-added" basis that takes into account students' academic baseline when assessing their results." (p. 4)

The implementation suggested above will put a lot more pressure on academic departments to produce favorable results, and faculty members to produce a 'measurable product'. Such scenario will lead for more faculty accountability, and more evaluation – of any kind. Student evaluations of teaching could then become more significant in the faculty evaluation process than they already are. The report also asks for innovation under the assumption that "Too many of our colleges and universities have not embraced opportunities to be entrepreneurial…" (p. 4). But the idea that shows the most business-like approach is the following:

> "The Department of Education should collect data and provide information in a common format so that interested parties can create a searchable, consumer-friendly database that provides access to institutional performance and aggregate student outcomes in a secure and flexible format. The strategy for the collection and use of data should be designed to recognize the complexity of higher

> education, have the capacity to accommodate diverse consumer preferences through standard and customizable searches, and make it easy to obtain comparative information including cost, price, admissions data, college completion rates and, eventually, learning outcomes." (pp. 20-21)

To me this report and its proposals seem like a higher education version of No Child Left Behind. By adhering to the proposal stated above, families can go on and have their "shopping for college" facilitated by looking at percentages of this and that. When our federal education agency starts suggesting value assessment at the college level, we will continue fostering a trend toward utilitarian measures that will continue undermining the entire education process. The education experience will then reach Orwellian proportions when every step of the process is constantly being overseen and manipulated.

CHAPTER V: IMPLICATIONS FOR RESEARCH, POLICY AND PRACTICE

The idea of evaluating college faculty has existed since the 1700s when teaching quality was considered when faculty members were up for promotion (Clark, 2006, p. 244). In the German university system dating back to 1756, the "applause" was an indicator of teaching effectiveness as well reputation. As Clark writes:

> "When the ministry rejected a candidate in 1756, it made an important remark. It said that seniority played no part in promotions, which must be made solely and alone for publication of useful and reasonable writings and disputations, as well as teaching. That was the new policy." (p. 280).

In addition, the principle of visitation as a precursor of peer review goes as far back as 1597, when a commission consisting of the chancellor, provost, and other administrative figures would instruct professors about the guidelines for the visitation (p. 345). The professors were given a questionnaire of 124 questions, which served as some type of self-evaluation. Then, a formal protocol would follow a due process of visitation (p. 351). Clark discusses in more detail a sample from a peer observation report done in 1789 (pp. 360-369). In the report, there were four criteria being articulated: an overall description of the professor's academic background and title, a description of how the instructor held classes and other academic responsibilities, whether he deserved a salary increase and/or any other merit, and a narrative indicating an overall assessment of the instructor's scholarship. Clark indicates that some of the comments were based on plain gossip and was subject to much subjectivity; he writes: "Peer review and private gossip – there is no way to tell which was which – issue from unnamed ones and unanimous

witnesses" (p. 369). It is interesting to note that even in the 18th century the notion of peer observation was as controversial as it still continues to be.

In the United States teaching evaluation of college faculty did not begin formally begin until the 1920s when Guthrie published the first report of student evaluations in 1924. Then in the 1960s the issue gained momentum due to a series of factors detailed in chapter II. Back when the movement advocating student evaluations took place, students were highly critical of what went on in the classroom and the organized movement that went along with it meant to provide useful information to other students about courses and grading criteria. Over time, student evaluations have become a ritual performed art the end of the course, and student evaluations have been affected by various issues such as grade expected in the course and grading leniency of the instructor. All in all, student evaluations continue to be used in virtually every classroom but the validity of those evaluations continue to lose credibility although a majority of instructors and administrators prefer to rely on those for evaluating teaching quality.

As seen in my survey results, institutions tend to rely too much on student evaluations. A study of 40,000 department chairs found that ninety-seven percent use student evaluations to assess teaching performance (US Department of education, 1991, p. 2) as opposed to fifty-three percent using peer evaluations to assess teaching. Perhaps the simplicity of distributing student evaluation forms and analyzing them seems more appealing than appointing a peer observation committee and going through the due process for establishing a reliable peer evaluation system. Cashin (1999, pp. 26-27)

argues that additional data in addition to student evaluations is needed to assess the following seven areas:

1) Subject matter mastery
2) Curriculum development
3) Course design
4) Delivery of instruction
5) Assessment of learning
6) Availability to students
7) Administrative requirements

He states that most students know next to noting about the first three, and the last one. In the survey I conduced it became clear that faculty members are divided either in favor or against student evaluations. Those in favor of student evaluations over classroom observations may do so because student evaluations, in general, tend to be more generous. Students perform a quick assessment of teaching performance in just a few minutes whereas peer observation relies on a careful assessment of classroom dynamics and teaching materials done by a senior colleague who knows the subject matter and who may be an expert observing various colleagues teaching. Given that the literature on peer evaluation is not as abundant as the one on student evaluations, then what should be done?

Student evaluations should be just one set of evidence of a faculty member's teaching effectiveness. Student evaluations could provide useful information about how

someone teaches for both formative and summative purposes. However, those evaluations should be considered along with other pieces of evidence when making personnel decisions. Student evaluations tend to focus on superficial teaching skills and do not address more substantive aspects of the teaching craft (Keig and Waggoner, 1994, p.1). The following seven considerations for using student evaluations should be kept in mind when using those evaluations as part of a college faculty member personnel file. These considerations emerge from the conversations I had during the interview process carried out in my survey:

1) *Use multiple sources of data to provide a broad range of viewpoints regarding teaching effectiveness.* A combination of data should enhance reliability and minimize bias that could emerge from student ratings.
2) *Faculty members should be aware of how student evaluations will be used by the department.* Each department should have clear guidelines of how much of the overall evaluation the student evaluations count for.
3) *Student evaluations should be analyzed based upon the type of course, enrollment, and number of courses taught by a particular instructor.* If the faculty member teachers various courses, then those courses should be equally represented in the student evaluation data. In addition, student evaluation from a course with a large enrollment should be weighed differently from those with a small enrollment. If an instructor has taught a variety of courses over a period of years, then such variety should be taken into consideration as well.
4) *The student evaluation form used should allow for a broad distinction of categories.* An appropriate evaluation form should include categories that cover

a broad spectrum of teaching strategies such as: teaching methods, organization, availability during office hours, clarity of course expectations, adequate preparation for assessing learning, and others.

5) *Take into consideration students' motivation level.* When analyzing data, differentiating between required courses versus electives should be taken into consideration. Also, the student evaluators should be differentiated between majors and non-majors in the subject area.
6) *Use an analysis of student evaluation that can be easily interpreted by the instructor*. There should be an item-by-item analysis that be easily understood by anyone who looks at the data. In addition, evaluators of the student ratings data should be properly trained to make sound interpretations of such data. A group of items should not be averaged into a single score. An analysis of student evaluation data should looked at from different angles in order to assess multiple dimensions of the instructor's teaching ability.
7) *Include space for open-ended comments in the evaluation form.* Students should be given the opportunity to indicate additional aspects of the instructor's teaching qualities that may not be addresses in the evaluation form. For example, students should have the chance to express recommendations to improve the course, to commend the instructor for strategies that were found to be effective, and/or to make additional suggestions to improve the teaching style of the instructor.

If we are going to continue using student evaluations for assessing college faculty performance, then policies related to the use of such evaluations should be developed and

implemented keeping in mind that student evaluations represent just one population group at the institution.

Developing a policy for peer evaluation

Given that there is no consensus about what the peer evaluation process should include, and the poor implementation record of a peer evaluation system, there is a much needed urgency to develop a better peer observation and evaluation policy that could assess teaching effectiveness. Although the idea of the teaching portfolio has been gaining prominence over the past twenty years, it is seldom used by institutions. In addition, there are some drawbacks in the use of the teaching portfolio. A teaching portfolio is meant to include multiple pieces of evidence of a faculty member's performance along with analysis and reflection upon peer collaboration. Some critics of the teaching portfolio argue that it is too time consuming and somewhat subjective since the individual instructor chooses which information is to be included (Zubizarreta, 1999, pp. 168-169). He argues that "It would be rare, if nonexistent, for instance, for a research scholar to submit for evaluative review a list of all rejected articles and failed projects" (p. 169). Therefore, building a teaching portfolio gives each faculty member the opportunity to reflect on their teaching and link it to support their particular teaching philosophy.

An important aspect of the teaching portfolio is that it encourages peer collaboration, and better yet, peer evaluation is a crucial piece of the portfolio assessment. It also promotes continuous revision of one's teaching strategies, and all

corresponding steps to enhance teaching performance. Another positive aspect of the teaching portfolio is that although it includes results from student evaluations, an evaluator should be less inclines to focus on those since other pieces of documentation of teaching performance are includes vis-à-vis student evaluations. In addition, the teaching portfolio covers a wide range of the faculty member's responsibilities at the institution. However, evaluating a teaching portfolio is exhausting and the evaluation process encounters some of the same issues associated with peer observation.

Peer evaluation continues to be the bone of contention in the evaluation of faculty teaching effectiveness. And the question remains: what makes the best peer evaluation system? As demands on faculty continue to increase in today's higher education system, peer evaluation might become the cornerstone of an adequate evaluation process aimed at improving teaching and learning at a particular institution. Given the inherent complexities in evaluating teaching from multiple lenses, then it is crucial that college instructors take the spotlight in the process of evaluating teaching. The challenge is how to bring the results of research in teaching evaluation into academic practice given the distinctiveness in institutional culture associated with higher education. Another challenge is how a particular institution defines effective teaching. A third challenging issue has to do with the ignorance of educational research present in many educational administrators who are directly responsible for developing and implementing institutional and/or departmental policies for teaching evaluation. Although most higher education administrators start out as instructors, by the time they become administrators there has been a lag keeping up with the recent literature on best practices in teaching, or those

administrators have never considered the issues of instructional delivery and assessment seriously.

A proposed policy on peer evaluation

Given the difficulty in evaluating teaching effectiveness in higher education, it is crucial to consider establishing a policy that looks at peer evaluation differently from existing policies. The major issue to consider is classroom teaching and developing a more concrete idea of what effective teaching is. However, teaching effectiveness could be a reflection of the institutional culture and should be defined by the institution. More importantly, peer evaluation should be implemented for both formative and summative purposes. If peer evaluation is going to be used for summative evaluation alone, then the instructors being evaluated are only being rewarded or punished but will not have a fair chance to improve their teaching before it is too late. Evaluation of teaching should not be used for personnel decisions alone but should be an ongoing effort of the department and/or institution to foster effective teaching. Once the teaching of a faculty member has been strengthened by formative evaluation, then summative evaluation may follow. Scholars have suggested that formative evaluation should be used a stand alone practice and then be followed by summative evaluation (Weimer, Kerns, and Parrett, 1988, p. 285); by “providing data, diagnostic, and descriptive feedback, with which to improve instruction”. Effective uses of both formative and summative peer evaluation systems should maintain the academic quality of a department and/or institution while providing faculty the encouragement needed to improve their teaching.

What makes an effective peer observation system?

Although it could difficult to come up with a precise answer to this question, there should be some commonalities about what is appropriate when considering peer evaluation. Although a teaching portfolio could be considered peer evaluation, my focus would be on peer observation. There are ten criteria described by Cohen and McKeachie (1980, p. 147.) that constitute effective teaching that one's colleagues can easily assess:

- Mastery of course content
- Selection of course content
- Course organization
- Appropriateness of course objectives
- Instructional materials
- Evaluative devices and methods used to teach specific content areas
- Commitment to teaching
- Concern for student learning
- Student achievement, and
- Support for departmental instructional efforts.

Peer classroom observations are useful in offering insights that cannot be obtained from relying on student evaluations or self-assessment (DeZure, 1999, p. 71). However, peer classroom observation as it is practiced at many (or most) institutions is not very reliable given that is mostly consists of a classroom observation by one or two colleagues without adequate pre-observation preparation and an appropriate follow-up to the observation. The list of criteria given above seems too mechanical to truly address

effective teaching. The criteria are too prescriptive and anyone could easily follow those and not necessarily be an effective instructor. Given that scholarly work of McKeachie, no one has disputed those criteria. All ten criteria can be easily mastered by anyone with some common sense, and mastering them does not warrant effective teaching. In fact, an instructor with little classroom experience could exhibit all ten criteria and still be ineffective transmitting knowledge and motivating her/his students to learn the subject. The ten criteria do not ensure that the instructor could exhibit the academic charisma that is essential to motivate students to go to class and learn the subject.

Peer observation is often found biased given that there could be disagreement about what constitutes good teaching, interpersonal relationships between observer and the instructor being observed, the reputation of the instructor being observed, disagreement about what to observe, lack of training of observers, and a lack of consistency in the process (p. 72). Nonetheless, there are ways by which reliability can be increased by establishing a systematic process that involves key decisions.

Peer observation should look for the following criteria, which most scholars have not documented thoroughly. The criteria are:

- Motivation to teach
- Clarity of instruction
- A clear perspective of what students should accomplish
- Understandability of students' learning behaviors
- Fairness in grading

- Returns graded material on time
- Accessibility to students
- Personality that motivates students to come to class
- Ability to motivate students to learn the subject
- An overall sense of wit when presenting the subject

An instructor who possesses these characteristics is well aware of what is the best way to teach the subject, who is to be taught (the student category), and how the material is to be presented. These criteria are in agreement with what Gurland (1978, p. 77) addresses. Ignoring any one these three categories can easily lead to some type of frustrating educational experience. An effective classroom observation should look for a simultaneous display of necessary teaching responsibilities. The notion of academic charisma entails a multiple set of strategies; some which are innate to the instructor's personality and others which are acquired with experience. Being an effective teacher in higher education requires the instructor to be a great communicator, and being someone who anticipates any possible difficulties that students may have understanding the subject matter. When students are let to be the judges in the faculty evaluation process, the necessary criteria to judge effective teaching are not being properly judged since students are not experts in discerning what effective teaching constitutes. Students may focus on the level of difficulty of the instructor or her/his personality thus relegating faculty evaluation to a popularity contest.

SOME BARRIERS TO PEER OBSERVATION

Surpassing objections to open the classroom door

Given that peer evaluation is often misunderstood as classroom observation, such confusion creates a problem of trust. Although peer observation is a crucial component of the peer evaluation process, there are other aspect of the faculty member's teaching repertoire that are considered as well. Many faculty members feel apprehensive about having a colleague present in their classroom while teaching. Such colleagues' visits often inspire a sense of distrust and a possible threat to academic freedom. While there are faculty members who may enjoy displaying their academic charisma, there are others who feel their classroom privacy invaded. Such sense of distrust creates a major problem in welcoming peer observation as an essential component in the overall teaching evaluation process. A possible solution might to establish *a priori* that peer observation is a crucial component of the evaluation process during the hiring process. A faculty member who accepts a position at a particular institution is then aware that frequent classroom observations are part of the rules of engagement in academia. If classroom observations are done frequently and consistently, for both formative and summative purposes, then university faculty might become accustomed to the process and will be less anxious about having their colleagues entering their classroom domain.

Who qualifies as a peer and who selects the peer observers?

Choosing a well-qualified peer observer is one of the most crucial aspects of peer evaluation. But first, it must be clearly established who should select the observers. Such task is usually performed by a personnel committee or the department chair. In an

ideal situation, the committee in charge of faculty evaluation should provide a list of potential observers to the faculty member being observed. As DeZure states (p. 77):

> "The advantage of this model is that it tends to reduce bias associated with inter-personal relationships with the instructor. The instructor can eliminate those she or he most distrusts, and the committee can eliminate the best friends and closest colleagues of the instructor. All parties feel that they have some control in the process and there is real choice for all."

The problem with the approach stated above is that at the very least one of the observers will be one the instructor's choices and bias could still be introduced in the observation process.

There should be at least two peers doing the classroom observation at the same time. By doing so, it may be possible that one of them picks up some details overlooked by the other observer. Inevitably, each observer is going to bring her/his own preferential approach to judge the teaching quality of the faculty member being observed. As Strenski (1995, p. 34) notes: "Ask any instructor about quality of instruction and you will hear an answer inevitably colored by that instructor's own training and model teachers". But having at least two observers reduces the possibility of having just one single observation report that could rely heavily on personal ideologies about teaching practices.

When at least two peer observers are involved, some departments opt for bringing an evaluator from a field outside the instructor's area of expertise. In one of the institutions I surveyed (University B), this is the method that is articulated on the peer observation guidelines posted. This particular approach might encounter resistance from the faculty member being evaluated on the premise that the outside evaluator would not be familiar with some the course content and thus might not serve as the best evaluator in categories dealing with content knowledge and selection of appropriate materials to teach the specific content of the course. A possible solution to this objection would be to bring a peer observer from outside the institution, just like the peer reviewer who reviews a manuscript submitted for publication. Again, there could be significant opposition to an outside peer observer: first, on the grounds that it is not economically feasible to afford observers from outside the institution; and second, the idea could face enormous resistance from faculty members claiming that it is a signal of distrust. The financial consideration could be ameliorated by having two institutions do reciprocal peer observations at various times. However, the issue of having a stranger in the classroom evaluating teaching performance could face enormous opposition and might not be approved by a faculty steering or faculty affairs committee. A policy addressing this issue of who qualifies as a peer observer and who makes the selection of peer evaluators should be well articulated based on the most recent evidence from educational research, and with a reasonable dosage of what should work in a specific department.

In terms of having an evaluator from outside the instructor's area brings a whole set of problems. First, any peer evaluator who is not familiar with the subject matter ends

up being the equivalent of a student. Since the person is not entirely familiar with how the instructor communicates the subject, a sound judgment is unlikely to result. How can a professor of medieval literature evaluate the effectiveness of someone teaching quantum mechanics? Such person ends up being the equivalent of a student who may have some interest in the subject but no expertise in measuring the effectiveness of how that particular instructor presents and addresses difficulties when teaching such an abstract subject as quantum mechanics. The same argument could be used for other subjects. Effective teaching should be evaluated by those who are experts in the same field. The effectiveness of a dentist does not get evaluated by accountants, and the effectiveness of a chef is not evaluated by a group of engineers, for example. Teaching will occupy more respect in academia when it is finally evaluated only by those who have the necessary expertise to assess an instructor's effectiveness in presenting a particular subject.

Another issue to be considered when choosing a peer evaluator is the vulnerability of the reviewer in matters of confidentiality. Chism (1999, p. 12) argues that "peer reviewers will be more candid… when they are sure their remarks are anonymous". Chism also points out that despite being in a professional setting; it is not the case that evaluators feel personally uncomfortable by having to be open about judgments involving colleagues. Records of peer review of teaching are not protected if discrimination is being suspected, based on a Supreme Court ruling in 1991(*University of Pennsylvania v. EEOC*), and also in places where an "open records" law exists. A possible solution suggested by Chism is to have the evaluation records authored by the entire review

committee or to present verbal reports to a summarizer (p. 12). However, in an academic environment there should be openness and the opportunity to express dissent. Perhaps either the department head or the academic dean should be the only person held accountable for instance in which discrimination is being claimed; thus relieving peer evaluators of certain degree of responsibility and eliminating any possible hesitation against being impartial when evaluating colleagues.

Number of classroom observations

Some institutions only require a single classroom observation per evaluation cycle, such as University B in my study, where faculty members are observed once every three years. Currently, more departments and institutions seem to be increasing the number of classroom observations required (at least that is the indication posted on departmental and/or institution websites). The issue dealing with the required number of observations is crucial for a variety of reasons. First, a single classroom observation is just a snapshot of the instructor's classroom atmosphere, and by no means is a reliable indicator of what goes on in that particular classroom most of the time. Second, the particular lesson being observed on the evaluation day might not be indicative of the instructor's teaching strengths and/or weaknesses. For instance, if the observation takes place on day following an extended holiday such as Spring Break, chances are that the students could not be fully engaged in the specific lesson. Also, if a quiz is being planned on that day or a paper is due, the students might not be fully engaged in classroom discussion. Third, the specific topic being covered on the day of the observation might not be one of the instructor's favorite topics. Consequently, multiple observations (about three) should be

carried out to get a representative sampling of the instructor's teaching ability and increase the reliability of the classroom observation report.

It may be argued that several classroom observations for a single faculty member are too time-consuming. In addition, there are pre-observation and post-observation meetings suggested to coordinate the observation process before the classroom visits and to discuss the formal evaluation letter following the observations. Each of these steps adds up to a considerable amount of time, and there may be resistance from faculty committees about undertaking such thorough classroom observation procedure. Chism suggests that people on the evaluation committee can make periodic decisions about the teaching quality of individual instructors rather than delaying decisions for a substantial period of time (p. 13). On the other hand, having regular formative classroom observations and frequent meetings with the instructor being evaluated allows making corrections to ineffective teaching thus leading to a smoother formal evaluation process and more productive peer evaluation and mentoring.

Maintaining the consistency of peer observers

Ideally, more than one observer should be used for faculty classroom observations. However, maintaining the consistency of observers becomes an issue due to inconsistencies and possible lack of reliability due to those inconsistencies. It is then necessary to make sure that when using more than one observer that each one of them brings a different type of expertise to the observation process. If one of the observers specializes in observing lecture courses and another may be an expert in observing

laboratories, then the combination of them should lead to a productive fusion of ideas and impressions about the classroom observation process leading to a fair evaluation of the faculty member and a combination of useful recommendation for improving teaching.

As with other issues described above, time could be a major obstacle in securing at least two classroom observers with different background who might be willing to make various observations together. In addition, prospective observers might complain about the labor-intensive nature of having to coordinate a series of simultaneous observations with another colleague. This situation could be ameliorated by having a colleague do a series of formative classroom observations, and then make a series of recommendations. Another observer could then start a series of classroom observations to evaluate whether or not the faculty member being evaluated has corrected has made improvements in her/his teaching based upon recommendations made by the previous observer. A possible problem emerging from this approach is that the faculty member being evaluated will end up having at least two different observers visiting the classroom numerous times leading to a sense of invasion of privacy and very little opportunity to teach without having an observer around. Whichever approach is adopted by a department, it may need to be based upon the number of faculty members in need of being evaluated and the availability of a sufficient number of qualified peer observers willing to engage in the peer observation process.

Type of courses to be observed

When a decision is to be made about evaluating a junior faculty member, the type of course(s) to evaluate is critical. A faculty member should be observed while teaching a

course that she/he likes to teach. A large lecture introductory course might be a difficult course in which to assess an instructor's teaching capabilities. Such courses tend to fulfill distribution requirements and, generally speaking, are not the favorite ones to teach by many instructors. An upper level smaller class might seem more appropriate since it may include a more diverse set of activities, more in-depth discussions, and more specialized content coverage. However, when it comes to course selection for classroom observation purposes there might not any right or wrong choice. Faculty members should be evaluated in a course in which they feel comfortable teaching. In addition, there should be a uniform policy about the type of courses in which classroom observations are to be carried out. There should not be any sense of disparity. If a faculty member is observed while teaching a large introductory lecture while another faculty member is observed teaching an advanced seminar, then the information extracted from each classroom observation might provide completely different information. Chances are that the classroom observation done in a small class setting with upper level (presumably majors) will result in a more favorable observation report. An important distinction made by DeZure (p. 79) is that classroom observations should not be made in courses that are difficult to teach. However, DeZure does not define what constitutes a difficult course to teach. Perhaps the instructor being evaluated should be allowed to decide the course in which she/he is going to be observed while teaching.

So far existing policies for classroom observation are very loose in terms of which type of course(s) junior faculty should be observed doing their teaching. It is then necessary for the institution to develop a consistent approach to decide the type of

courses that can be considered for classroom observation purposes. Regardless of whatever system is implemented, the bottom line is to establish criteria that are fair and consistent to those being evaluated. The process of establishing acceptable courses for classroom observation purposes does not need to be a fuzzy criterion; it just needs to be equitable so that any possible bias can be prevented.

Adequate training of classroom observers

An essential part of an adequate classroom observation and the overall peer evaluation process is to have adequate peer observers. The institution or the department should institute a requirement that all peer evaluators undergo a mandatory training. The purpose of the training is to add a layer of consistency to the evaluation process, which may in turn reduce bias. The nature of such training should be to include:

- **Discussions of the overall aspects of effective teaching**. Although there may not an overall agreement as to what constitute 'effective teaching', there should be at least some common rules as to what could be considered effective teaching strategies for summative evaluation purposes.
- **Adequate record-keeping procedures**. Even if it sounds a trivial issue, matters of constructive an effective narrative model, keeping track of time on task, and measuring time spent on various classroom activities are essential for an effective and fair classroom observation.
- **Attention should be given as to how the review of materials such as syllabi, exams, assignments, and other classroom materials should be evaluated**. Given the potential number of differences in the construction of these materials

among faculty members, there should be a minimum set of guidelines. For instance: Are those materials consistent with what goes on in the classroom? Do the teaching materials reflect clarity of expectations? Do they include a variety of strategies to accomplish learning of the subject matter? In general, when teaching materials are evaluated, there has to be a consensus as to how to determine that those materials reflect the best teaching practices established by the institution or the department.

- **How to provide appropriate feedback to the instructor**. Each department or institution may decide how the evaluator should provide feedback to the faculty member being evaluated. In most places, a post-observation meeting is mandatory in which both the observed and the observed meet to discuss informally the classroom observation. This meeting may also serve to clarify things for the observer in the event that she/he may have missed something while keeping track of the observation protocol. The post-observation meeting can also allow the faculty member that was observed to indicate how typical or atypical the particular session was, and to indicate her/his own assessment of how the class went on during that particular day. Such feedback may help the observer situate her/his observation within the proper perspective of what was just observed in that particular classroom on that particular day.
- **Building consensus among the reviewers**. The notion of trying to build consensus among reviewers is not only to prevent personal bias but also to establish a set of guidelines for the classroom observation process. Observers for a particular instructor should meet and discuss what each of them would like to

look for and how to combine ideas to make the observation process as efficient as possible. It is unlikely that observers may agree one hundred percent about the teaching aspects that they would like to concentrate on, but at least there would be a clear set of expectations on which to base the evaluation report. More importantly, there will be consistency in the various aspects of the observations that the observers will be focused on without having one observer being more critical than the other.

- **How to reconcile observation notes**. It may be inevitable that there will be discrepancies between the notes from the individual observers. The team of observers should meet immediately or soon after the classroom observation and address individual notes. Doing so may help reinforcing the validity if the formal evaluation and prevent future discrepancies in the event that each observer submits a report that is quite different from the one submitted by another observer. The strategy of reconciling observation notes may be minimized by addressing properly the issue of building consensus described above.
- **Writing a final classroom observation report**. An effective classroom observation is the one that collects impressions form all the observers rather than a combination of individual observations reports. First, a collective report takes away individual responsibility in the event that the faculty member being evaluated has access to the information in her/his file. It is also more efficient and less time-consuming for a tenure or promotion committee to read a single evaluation report rather than multiple reports, some of which may be utterly redundant. Also, the format of the final report should include a combination of

feedback/comments and suggestions for improvement. It should not only focus on criticism. A more through description of the classroom observation instrument will be discussed below.

There may be faculty resistance about the time-consuming aspects of getting involved in a committee devoted to peer evaluation. Some institutions can handle the time issue differently. The department can produce a peer evaluation guide that can be distributed, followed by a brief discussion of key aspects of the observation process for those who may have pressing questions and/or concerns. Some institutions may produce a training video that presents every aspect of the peer evaluation process with relevant examples. Regardless of the method used, there should be some clear guidelines about the classroom observation process and peer evaluation, in general, that those involved in the process can easily refer to at any time. Many large institutions currently have an office dedicated to address aspects of teaching effectiveness. Those centers produce written and/or audiovisual materials available to all departments thus minimizing the investment of time and resources from individual departments. In addition, those centers can offer workshops periodically for new faculty, which over a period time can have a lasting effect of improving the teaching quality of the faculty members who take advantage of such resources.

The Classroom Observation Process

The ideal purpose of the classroom observation process is to evaluate faculty by her/his peers while keeping bias non-existent or at least to a minimum. In the faculty

member's file, the report of classroom observations should be an important piece of documentation in the evaluation of subject matter competency, teaching strategies current to the instructor's field and how the instructor motivates the students to engage in the subject. None of these teaching characteristics can be effectively judged by student evaluations. However, in order to engage in the classroom observation process in an effective and equitable manner, some guidelines must be exercised. By doing an appropriate classroom observation, then the evidence collected can then be used effectively for a holistic evaluation of the teaching effectiveness along with other pieces of evidence.

Chism (pp. 76-77) lists seven general guidelines which I have modified to include some of my own ideas about what should take place during a classroom visit. The following guidelines presume that classroom observers have been carefully chose and properly guided on how to make efficient use of the classroom observation time in order to make a fair and balanced report of the classroom visit.

1. **There should be a combination of announced and unannounced visits**. This may be a very controversial issue in the classroom observation process. When a classroom observation is announced, the instructor is given time to prepare and prevent some teaching flaws from becoming visible; the instructor is also given time to teach at her/his best. If the classroom visit is not announced, the instructor might feel her/his classroom domain invaded and insecure as to the actual purpose of the unannounced visit. The observer may end up wasting time if there an exam or quiz on that particular day

and/or any lesson taking place in a passive (teacher-centered) mode. The advantage of having an unannounced visit is that the instructor is being observed without any conscious preparation for the visit and those teaching aspects that need to be corrected will be more obvious during such visit. However, an impromptu visit may allow the observer to point out aspects of teaching that need to be corrected. An unannounced visit should be made just for formative purposes, to provide informal feedback before the announced observation. Although there may be strong opposition to make unannounced classroom visits, a policy meant to incorporate those in the observation process may indicate that those visits are for the sole purpose of catching flaws that may not be visible during an announced visit, and more importantly, that such visit will not be part of the official evaluation process.

2. **Use of a teaching consultant**. As indicated previously, use of outside reviewers for evaluating teaching goes along with using peer reviewers to evaluate research grant proposals. Having an outside teaching consultant to be part of the classroom observation should elevate the teaching aspect of the college instructor to a more distinguished level. Immediate opposition may ensue based on financial grounds but, as stated previously, two institutions may engage in reciprocal peer evaluation of teaching. In addition, the department can request additional budget for hiring the teaching consultant(s). By doing so, two departments at two different institutions can potentially benefit from this arrangement thus making the teaching evaluation process to be taken more seriously. The use of a teaching consultant could

remove possible bias from the peer observation, which is done by the instructor's colleagues. The consultant may bring a new set of ideas to the evaluation process and can provide non-biased (or less biased) judgment to the peer evaluation system. Moreover, the same teaching consultant used for classroom observation may be the one used for professional development sessions of junior faculty. A potential problem of using the same teaching consultant repeatedly is that this particular individual may become biased over time once she/he gets to know members of the department that hired the consultant. In addition, the use of a teaching consultant may provide custom-made approach to evaluating teaching that some teaching development centers at many institutions fail to accomplish by not being responsive enough to the diverse in teaching approaches that could be found at any institution. After all, teaching is not a one-size-fits-all activity. All in all, the use of a teaching consultant has the potential of enhancing the teaching capabilities of less experienced faculty members in a department and can be a great asset in the faculty evaluation process.

3. **Length of the observation**. There are various ways to spend time observing a class: the entire class period; a portion of the class, including a few minutes before or after class to observe how the instructor manages transition; or just a particular teaching unit of the class. The length of the observation should depend on the length of the class itself. If a class meets once a week for three hours, then it is too physically demanding for the observer to stay for the entire class period while following guidelines for a proper classroom

observation. DeZure (p. 81) recommends that the minimum time spent in a class is 50 minutes, including a few minutes before class to observe how the instructor manages transition into or out of class and/or between topics. Most undergraduate courses at many institutions meet for either fifty or ninety minutes, which may prove convenient for the classroom observation. However, if the observer has some issues with staying in a class for ninety minutes, she/he should plan to be in the particular classroom at the beginning of the class. Walking into a classroom late is considered rude and may make the instructor being observed feel uncomfortable; it could also lead to unnecessary distraction to the students. On the other hand, having an observer leave the class early may be seen as an indication of either lack of interest or may send a wrong signal to the instructor that the teaching was evaluated poorly. In my own experience, I have had observers leaving the classroom when I have stopped teaching and moved to hands-on activities. The observers have missed the interaction of students working in groups and the rapport between students and the instructor, which is an important part of the teaching process. In other instances, I rely on instructional technology and having a classroom observer leave when I have switched to computer-based activities or laboratory exercises is an indication that those types of classroom activities are not considered pedagogically important. Doing so is an indication of an old-fashioned mentality that considers relevant only those activities in which the instructor is in charge. To prevent situations like that form happening it is essential that the instructor provides the observer with an

outline of the lesson and, whenever possible, allow the observer to become part of the various activities being performed during the classroom visit.

4. **The observer should be totally unobtrusive**. Although some scholars suggest having the classroom observer being part of the hands-on activities might add another layer to the observation report , the observer should *not* become actively involved in classroom discussions and/or make explanatory comments to what either the instructor or a student has said in class. Doing so might create a sense of animosity between the instructor and the observer, could prevent other students from participating in class or alter any other type of teaching behavior for the remainder of the observation period. An observed should be completely removed from the classroom activities and by no means should interrupt the natural flow of the class. The purpose of the observer is to assess what goes on in that particular classroom when no observer is present. Although it may be necessary to clarify to the students the purpose of having an observer in class, the role of the observer should be limited to being just that – an observer. Perhaps the best approach is to alert the students in class prior to the class being observed that there will be observers coming to class, and indicate the purpose of the classroom observation to the students. Students should not be left in the dark regarding the classroom observation visit and left to second guess the purpose of having an outsider sitting in class. Also, observers should sit in the back of the classroom and be as inconspicuous as possible. By sitting in the back of the classroom, they can assess how effectively the instructor projects her/his

voice, and if audiovisual materials can be easily seen or heard to those students sitting in the back of the classroom.

5. **The evaluation process.** When documenting a classroom observation, there are multiples approaches that can be taken by the observer. The observer may transition from low impact activities such as how promptly the class began, whether or not the instructor took attendance, any clarifications made to the previous class, and a proper introduction to the topic of the day. Then, the observer may move into a more analytical observation of the class such as:
    a) specific elements of the lecture and/or discussion;
    b) the instructor's command of the topic;
    c) socio-political dimensions, including the use of authority during discussions, invitations to engage in participation and/or maintaining rapport throughout the class; classroom structure and procedures, whether instructional materials are effective and serve their intended purpose, diversity of approaches;
    d) the curricular context, whether ideas presented in class meet the curricular requirements and follow a logical place within the course;
    e) teaching effects, how well students seem to be learning the material as indicated by questions posed during class, the overall level of attention, and how easily the instructor can change the pace and pause when needed; and

f) physical-temporal effects, the observer should take into consideration the time and day of the week, size and configuration of the room, comfort of the room, acoustics, and instructor's awareness of these issues.

6. **The classroom observation form.** Each department and/or instructor may implement the use of different formats for the evaluation forms used during the classroom observation. The most commonly used method is the open narrative, which describes the observer's perceptions of what goes on in the classroom throughout the observation. This form is considered the least reliable methods by experts such as DeZure (p. 85), who argues that it "offers no sense of priorities for the observation and no consistency among observations". However, a narrative form allows the observer to go with the flow of what goes on in the classroom although it poses the challenge of keeping up with exact details of the observation process and it takes a significant amount of time to complete it. The narrative form also lends itself to observer bias by allowing the observer to focus on certain things while neglecting others. An alternative to the narrative is the use of a checklist with scaled responses but it presents some challenge in keeping up with the observation while addressing the checklist items and responding to those. Some institutions use open-ended questions this giving the observer more flexibility in terms of how to respond to some of the questions. However, a problem with the open-ended questions form is that it may not include aspects of the teaching process that could be evident during a specific

classroom observation. Some institutions rely on videotaping the class. A major problem with videotaping is that the process does not collect the overall classroom atmosphere and could make students and/or the instructor more uncomfortable than having an observer taking notes. It is also more expensive to do the recording and it is more time-consuming for the observers involved since it may require to playback the recording to look at specific details of a particular class. Perhaps the best method for reporting the classroom observation is a combination of a checklist of items along with a short narrative. The combination of methods gives the observer some ground rules to base the evaluation on while providing some flexibility as to elaborating on specific issues not widely covered on the checklist items. Regardless of the observation form used, it should be as comprehensive as possible to allow the observer include a plethora of aspects that can possibly go on in the classroom. More importantly, the observer should be familiar with the form and comfortable using it so that completing it does not distract her/him from making global evaluation of the observation process.

What to do following a classroom observation?

Immediately following the classroom observation there should be communication between the observer and the instructor, more so when the observation is for formative purposes. The follow-up meeting is the time to provide feedback and to communicate any concerns regarding the observed teaching. When the classroom observation is done for summative purposes, the follow-up meeting is the opportunity to discuss any

clarification needed before submitting a formal evaluation report; for instance, clarification about any materials distributed in class or to just acknowledge to the instructor that what went on in the classroom was clear to the observer(s). It is best to have the follow-up meeting in person rather than by email. Not having this meeting may provide a sense of concern to the instructor, wondering if the observation did not meet the expectations of the observer(s). DeZure comments that: "[e]ven the briefest of comments at the end of the session provides a form of closure in a situation that is often fraught with high anxiety" (p. 88).

The post-observation meeting should take place within a few days after the classroom observation or as described in the institutional or departmental guidelines. During such meeting all observers should be present and there should be a preliminary copy of the observation report ready for the instructor to look at. In some place, the instructor receives a copy of the report prior to the meeting. The tone of the meeting should focus on the specific aspects of the narrative and/or checklist form, and should be as descriptive as possible so that the instructor gets a real sense of what others observed. In the vent that the classroom observation report is not a very favorable one, care must be exercised so that the entire focus of the meeting is not based on criticism, instead, there should be a balance between praise and criticism along with suggestions for improvement. In cases where criticism is exercised, there should be clear examples to support it. The advantage of videotaping is that when criticism is given, there is recorded evidence of specific instances that need improvement in they way they were

handled. Before the meeting concludes there must be some contingency as to what needs to be done; for instance, whether another observation is to be scheduled.

The final classroom observation report

The written report discussed during the post-observation meeting may be considered the final report at some institutions. However, if the peer evaluation committee is truly concerned about a particular instructor or about making the evaluation process as formative (even in the event that it is meant to be summative), then such report is considered to be a draft and amendments can be made (DeZure, p. 89). It all depends on the nature of the report and how quickly a formal evaluation report is needed.

If the written evaluation report is considered to be final, then the instructor has no chance to request modifications or to make clarifications that could have been useful to the observer(s). Such approach may be considered punitive to the instructor. Even if the written report is considered a draft, there should be some established guidelines that the instructor does not get to have the entire draft changed. Doing so would compromise the integrity of the report, and may indicate to the observer(s) that the classroom observation was not entirely accurate. Whether the observation was done for either formative or summative purposes, there should be room for the observer(s) to make judgments that the instructor does not necessarily agree with. It is then necessary to frame the observation report using a language that is not confrontational and instead focus on the “to do” aspects that need improvement. In the event that the instructor

continues to disagree with the observation report, there should an opportunity to submit a written response and, whenever possible, include it along with the written evaluation.

CONCLUSIONS

Evaluation of college faculty continues and will continue to be a thorny issue in higher education as long as there is no consensus as to how college instructors should be evaluated and the purposes of the evaluation process. Over the years student evaluations have become the preferred method to evaluate faculty and there seems some willingness from many college instructors to leave the evaluation process to the students, who often display a very narrow view of what effective in higher education constitutes. It can also be argued that many college instructors prefer the lack of rigor of student evaluations rather than engaging in the rigor and systematic approach that peer evaluation constitutes. Students, by not being experts in pedagogy, have such a limited view of teaching strategies and are often biased depending on the expected grade in the course and the 'academic charisma' of the instructor. Abrami (1985, p. 217) noted the following:

> "Students as a group may be inaccurate observers for a variety of reasons: they may be naïve and insensitive to qualitative differences in instruction; they may be collectively biased by their own expectations which distort their perceptions; [and] they may be unfairly lenient in judging teaching effectiveness."

Thus, having such a limited view of what constitutes teaching effectiveness leads to more generous teaching evaluations than those coming from a colleague who might be

genuinely interested in the teaching quality of the academic department and/or the institution.

Based on the results of my survey, it is evident that college instructors are divided in terms of how they should be evaluated. Consequently, there is a need for a faculty evaluation policy that is applied consistently at the institutional level. A faculty evaluation policy should include looking at various sources of information that can then provide a holistic view of the instructor's teaching effectiveness. An evaluation policy should also establish that there are two broad categories for faculty evaluation: formative and summative. Before any faculty member is evaluation for promotion, tenure, or reappointment decisions there should be a process by which any instructor can receive informal feedback and be given a sufficient amount of time to make changes and/or corrections to her/his teaching performance.

When it comes to evaluating college faculty, colleagues are more qualified than any other constituency within the academic community. Therefore, peer evaluation of classroom materials and peer observation should become the most important element of the faculty evaluation process. As indicated by Cohen and McKeachie (1980), faculty should assess their colleagues' course design (such as goals, content, and organization); methods and materials used for instruction; how students' academic work is evaluated; and then add additional pieces of information such as student evaluations, and any other type of assessment from peers and administrators.

A model of faculty evaluation should include as many pieces of information as possible and should be tailored to the specific needs of the institution. Various evaluation models have been proposed over the years and virtually none has become thoroughly embraced. However, in my mind, the model suggested by Soderberg (1985, p. 19) provides one of the most inclusive initiatives for a holistic evaluation of college faculty. The model is three-dimensional and all the elements intersect in such a way that it is impossible to miss any source of information or process when conducting an evaluation. The three dimensions consist of the following:

1. Interdependent processes: goals/objectives, methods and materials, and feedback. These can be assessed by asking specific questions pertaining each. The way by which each of these aspects is approached depends on the institution's mission.
2. Time phases in which the instructional process occurs: pre-interaction, interaction, and revision. Instructors can reflect on any of these processes at any time to reflect and make decisions about goals, materials, and other components of the course. Some scholars conclude that effective instructors go back and back among the three phases, and modify these events based on personal reflection and/or feedback from colleagues (Chenoweth, 1991; Yarbrough, 1989).
3. The constituencies that are in a position to provide information about faculty strengths and weaknesses: students, peers, administrators, and other sources (such as self-assessment). In Soderberg's model peers are

> the best equipped group to assess process related to pre-interaction and revision (Keig and Waggoner, 1994, p. 24).

What is significant about Soderberg's three-dimensional model is the notion that faculty evaluation is not a simple process, and that it requires the intersection of multiple factors and various constituencies to provide a comprehensive evaluation.

So far peer evaluation has been prescribed as the best method to evaluate college faculty, and the limitations of this type of evaluation have been downplayed. However, are mentioned throughout this chapter, there are a number of reasons against peer involvement in faculty evaluation that are worth considering. Issues of academic freedom have been mentioned over the years as a primary concern against peer evaluation. The main argument is that direct classroom observation goes against the faculty member's right to have complete control over what she/he teaches. However, teaching should be not considered a secluded activity that cannot be interfered with. The presence of a colleague in one's classroom should be valued as an opportunity to demonstrate one's competency and to receive feedback which may be used for improving the individual's teaching craft.

There is an issue concerning the accuracy of the peer evaluation. The main concern has to do with the inadequacy of the classroom visitation to assess the various teaching competencies of a faculty member. However, when a classroom observation is initially carried out for formative purposes it may lead to instructional improvement

before the classroom observation is done for summative purposes. In addition, the number of classroom visits should be as frequent as possible to get a more accurate picture of what goes on in that particular instructor's class.

There is also a concern regarding the methods used for evaluating classroom performance failing to capture that entire range of activities that goes on in the classroom. Given the complex nature of teaching, it is unrealistic to attempt to judge all aspects of teaching during classroom observations. Doing so may prove extremely difficult and too time-consuming. However, the evaluation system developed by the institution should focus on documenting those activities considered essential to the institution. Also, an occasional visitor cannot attempt to develop a comprehensive understanding of a particular instructor's classroom performance.

The concern about subjectivity of peer evaluation is a main objection to classroom visitations. Given the range of notions and expectations about what constitutes 'good teaching', it is unrealistic to expect a lack of disagreement about how an instructor is observed in the classroom. Whatever the objections might be, there is a special advantage to having colleagues observe someone's teaching. It brings various and different perspectives and the classroom visit provides to opportunity to reconsider the teaching approach and make some adequate changes as needed.

It still seems amusing that faculty members do not object to have their research grant proposals and manuscripts evaluated by peers but when it comes to teaching, there

is much disagreement about how it is done. The claims about morale and collegiality being affected seem weak assumptions. Perhaps if a department or institution establishes that formative peer evaluation precedes summative evaluation at various stages and for all faculty members, then more college instructors would get more involved in the process of peer evaluation and the process ca then become another ritual in the life of a college faculty member just as student evaluations are in the life of those faculty members.

More research on faculty evaluation needs to be done. However, there is discrepancy concerning the nature of research approaches. While some scholars argue for more qualitative studies that look at the perceptions of the faculty members being evaluated and the observers who perform the classroom observation and peer evaluation, others claim that such studies could be more useful and meaningful using rating scales and quantitative analyses (Centra, 1986, pp. 3-4). The problem with having the research on college faculty evaluation become highly quantitative is that it would reduce teaching assessment to a number. In addition, those numbers emerging from a quantitative study may not easily transfer to some useful considerations at individual institutions. Qualitative studies can be easily adopted as models for possibilities at some institutions, and such studies could be easily modified to examine instruction at a given institution. Teaching is and will continue to be a highly complex activity, and its interaction with learning makes it an extremely difficult to examine on numerical terms.

Student evaluations of teaching were initially developed in the U.S. as a mechanism for students to provide informal about an instructor's teaching strategies. By the 1970's those evaluations became a mechanism to exercise resistance against institutional and political practices. The student evaluations in turn were the object of faculty resistance primarily during the 1980s as a result of not knowing the long-term consequences of those evaluations could have on the college faculty evaluation process. In addition, there was a legitimate concern about how far the effect of student evaluations could have on academic freedom and how the higher education community could come to evaluate the characteristics of 'good teaching'.

Student evaluations of teaching have evolved to become a routine aspect of any classroom at any institution of higher education. Mote importantly, those student evaluations no longer seem to have their original purpose: providing an informal feedback about a particular instructor to other students, and to some extent prevent unfair grading practices from some instructors. As more institutions of higher educations have transformed into research universities, priorities have shifted to have faculty members to engage in more research and less teaching. Evaluation of teaching at some institutions has become less of a priority and more of a lip service activity. As more faculty members are evaluated on their research and scholarship, many institutions are relying more on student evaluations to evaluate teaching credentials. Having students being the predominant voice in teaching evaluation places less intrinsic value on the teaching activity and creates other problems such as grading inflation and possibly a lower quality of teaching. Students tend to be more lenient as evaluators and their judgment does not provide a complete and accurate picture of the teaching performance of a college faculty member.

Consequently, those evaluations which seemed threatening at some point in history have become the preferred method of assessing teaching quality.

The implication of continuing to use student evaluations is that the teaching is not going to be properly evaluated and efforts to improve teaching will be thwarted as long as higher education faculty seem content with what student evaluations communicate. As seen in my survey results, the faculty sample interviewed at tow research universities seems roughly divided into those who favor being evaluated by students, and those who would prefer being evaluated primarily by their peers. Those who prefer student evaluations over any other method tend to do so based on the popularity they enjoy from students and refuse to let their classroom doors to colleagues who can provide a sound judgment of what goes on in their classrooms. College teaching evaluation is and will continue to be a complex issue as long as there is a conceptual difficulty understanding what constitutes good teaching. More research studies need to be designed and performed, and results need to be examined rigorously so that policies that can lead to instructional improvement can be developed and implemented. A successful program for evaluating college faculty is only a means to an end, which is the improvement of teaching to enhance student learning.

REFERENCES

Abrami, P.C. 1985. Dimensions of Effective College Instruction. *Review of Higher Education* 8: 211-228.

Aleamoni, L.M. 1999. Student Rating Myths Versus Research Facts from 1924 to 1998. *Journal of Personnel Evaluation in Education* 13(2): 153-166.

Aleamoni, L.M., and Yimer, M. 1973. An investigation of the relationship between colleague rating, student rating, research productivity, and academic rank in rating instructional effectiveness. *Journal of Educational Psychology* 64: 274-277.

Alper, P. 1993. Grades and grade inflation and flyoverland fairy tale. *Higher Education Review* 25(2): 57-61.

American Academy of Arts and Sciences. 2002. Evaluation and the Academy: Are We Doing the Right Thing? www.amacad.org. Retrieved on 25 July 2007.

American Association of University Professors. 2003. *Policy statement on contingent appointments and the academic profession.* Washington, DC: Author.

Arreola, R.A. 2000. *Developing a Comprehensive Faculty Evaluation System: A Handbook for College Faculty and Administrators on Designing and Operating a Comprehensive Faculty Evaluation System* (2nd ed.). Bolton, MA: Anker Publishing Company.

Astin, A.W. 1997. The changing American college student: Thirty year trends, 1966-1996. *Review of Higher Education* 21: 115-135.

Bain, K. 2004. *What the Best College Teachers Do*. Cambridge, MA: Harvard University Press.

Baldwin, R., and Chronister, J. 2002. What happened to the tenure track? In R. Chair (Ed.), *The questions of tenure* (pp. 125-159). Cambridge, MA: Harvard University Press.

Bearden, J., Wolf, R., and Grosch, J. 1992. Correcting for grade inflation in studies of academic performance. *Perceptual and Motor Skills* 74: 745-746.

Bilimoria, D. 1995. Modernism, postmodernism, and contemporary grading practices. *Journal of Management Education* 19: 440-458.

Birnbaum, R. 1977. Factors Related to University Grade Inflation. *Journal of Higher Education* 48(5): 519-539.

Blunt, A. 1991. The effects of anonymity and manipulated grades on student ratings of instruction. *Community College Review* 18: 48-54.

Boesel, D., and Fredland, E. 1999. *College for all? Is there too much emphasis on getting a 4-year college degree?* Washington, DC: U.S. Department of Education.

Boretz, E. 2004. Grade Inflation and the Myth of Student Consumerism. *College Teaching* 52(2): 42-46.

Brandenburg, G.C. and Remmers, H.H. 1927. Rating Scales for Instructor. *Educational Administration and Supervision* 13: 399-406.

Braskamp, L.A., and Ory, J.C. 1999. *Assessing faculty work.* San Francisco: Jossey-Bass.

Bruner, J. 1996. *The Culture of Education*. Cambridge, MA: Harvard University Press.

Burns, C.W. 1998. *Peer Evaluation of Teaching: Claims vs. Research*. Washington, DC: ERIC Clearinghouse on Assessment and Evaluation. (ERIC Document Reproduction Service No. ED421470).

Cahn, S. M. 1973. *The Eclipse of Excellence: A Critique of American higher Education*. Washington, DC: Public Affairs Press.

Cahn, S.M. 1977. *Why Johnny Can't Fail: Grade Inflation in Our Schools*. Paper presented at the 92nd New England Association of Schools and Colleges New Hampshire/Vermont Regional Meeting, October 19, 1977.

Cahn, S.M. (ed.). 1978. *Scholars Who Teach: The Art of College Teaching.* Chicago: Nelson Hall.

Cahn, S.M. 1994. *Saints and Scamps: Ethics in Academia* (revised ed.). Lanham, MD: Rowman & Littlefield.

Carr, J.W., and Padgett, T.C. 1992. Legal liability when designing or adopting student rating forms. In R.A. Arreola (Ed.), *Proceedings of the first annual CEDA conference on practical issues in faculty evaluation* (pp. 68-76). Memphis, TN: Center for Educational Development and Assessment.

Cashin, W.E. 1999. Student Ratings of Teaching: Uses and Misuses. In Seldin, P. (ed.), *Changing Practices in Evaluating Teaching: a Practical guide to Improved Faculty Performance and Promotion/Tenure Decisions*. Bolton, MA: Anker Publishing Company.

Centra, J.A. 1975. Colleagues as raters of classroom instruction. *Journal of Higher Education* 46: 327-337.

Centra, J.A. 1986. *Colleague evaluation: The critical link.* Paper presented the annual meeting of the American Educational Research Association, San Francisco, CA. (ERIC Documentation Reproduction Service No. 275 722).

Centra, J.A. 1999. *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness.* San Francisco: Jossey-Bass.

Chase, A. 1978. Skipping Through College: Reflections on the Decline of Liberal Arts Education. *Atlantic Monthly* 242(3): 33-40.

Chenoweth, T. 1991. Evaluating Exemplary Teaching. *Journal of Personnel Evaluation in Education* 4: 359-366.

Chevaillier, T. 2002. Higher education and its clients: Institutional responses to changes in demand and in environment. *Higher Education*33: 303-308.

Chism, N.V.N. 1999. *Peer Review of Teaching: A Sourcebook.* Bolton, MA: Anker Publishing Company.

Cizek, C.J. 1996. There's no such thing as grade inflation. *Education Week*, April 17, p. 32.

Clark, W. 2006. *Academic Charisma and the Origins of the Research University.* Chicago: The University of Chicago Press.

Cohen, P.A., and McKeachie, W.J. 1980. The role of colleagues in the evaluation of teaching. *Improving College and University Teaching* 28: 147-154.

Cole, W. 1994. By rewarding mediocrity we discourage excellence. *Chronicle of Higher Education* (January 6), p. B1.

Delucchi, M., and Smith, W.L. 1997. Satisfied Customers Versus Pedagogic Responsibility: Further Thoughts on Student Consumerism. *Teaching Sociology* 25: 336-337.

DeZure, D. 1999. Evaluating Teaching Through Peer Classroom Observation. In Seldin, P. (ed.), *Changing Practices in Evaluating Teaching: a Practical guide to Improved Faculty Performance and Promotion/Tenure Decisions*. Bolton, MA: Anker Publishing Company.

Eckert, R. E. 1950. Ways of Evaluating Teaching. *School and Society* 71(1833): 65-69.

Feldman, K.A. 1989. Instructional effectiveness of college teachers as judged by teachers, themselves, current and former students, colleagues, administrators, and external (neutral) observers. *Research in Higher Education* 30: 137-194.

Felton, J., Mitchell, J., and Stinson, M. 2003. Web-based student evaluations of professors: The relations between perceived quality, easiness and "hotness". *Social Science Research Network Electronic Library*. Accessed 9 August 2006 from hhtp://ssrn.com/abstract=426763.

Foster, J., and Long, D. 1970. The Dynamics of Institutional Response. In Foster, J. and D. Long (Eds.), *Protest! Student Activism in America*. NY: William Morrow & Company.

Friedman, M., and Friedman, R. 1980. *Free to Choose*. New York: Harcourt.

Fullerton, H. 1999. Observation of teaching. In *A handbook for teaching and learning in higher education*, H. Fry, S. Ketteridge, and S. Marshall (eds.), 226-241. London: Kogan Page.

Gigliotti, R.J., and Buchtel, F.S. 1990. Attributional bias and course evaluations. *Journal of Educational Psychology* 82: 341-351.

Goldman, L. 1985. The betrayal of the gatekeepers: Grade inflation. *Journal of General Education* 37(2): 97-121.

Gottfried, P. 2002. About Consumerist Education. *Academic Questions* 16(2): 53-57.

Gould, E. 2003. *The university in a corporate culture.* New Haven: Yale University Press.

Greenwald, A.G., and Gillmore, G.M. 1997. Grading leniency is a removable constraint. *American Psychologist* 52: 1209-1217.

Gurland, R.H. 1978. Teaching Mathematics. In Cahn, S.M. (ed.). *Scholars Who Teach: the art of college teaching.* Chicago: Nelson-Hall. Pp. 76-97.

Guthrie, E.R. 1954. *The Evaluation of Teaching: A Progress Report*. Seattle, WA: University of Washington.

Hampel, R.L. 2006. The Academic Work Ethic at Yale, 1939-1982. In Geiger, R.L. (ed.) *Perpectives on the History of Higher Education* 25: 123-144.

Hancher, M. 1994. The consequences of grade inflation [Letter to the editor]. *Chronicle of Higher Education*, February 2, p. A53.

Hanushek, E. 1979. Conceptual and empirical issues in the estimation of educational production function. *Journal of Human Resources* 14: 351-388.

Holmes, D.S. 1972. Effects of grades and disconfirmed grade expectancies on students' evaluation of their instructor. *Journal of Educational Psychology* 63: 130-133.

Hoyt, D.P. 1966. College Grades and Adult Accomplishments: A review of Research. *Educational Record* (Winter 1966), 70-75.

Hu, S. (Ed.) 2005. Beyond grade Inflation: Grading Problems in Higher Education. *ASHE Higher Education Report* 30(6). San Francisco: Jossey-Bass.

Hutchings, P. 1996. The peer review of teaching: Progress, issues, and prospects. *Innovative Higher Education* 20: 221-234.

Johnson, V.E. 2003. *Grade Inflation: A Crisis in College Education*. New York: Springer.

Juola, A.E. 1974. Grade inflation: The facts, the reasons, the consequences, the future. Paper presented at the Annual Meeting of the American College Personnel Association, Chicago.

Kaplan, W.A. 1978. *The law of higher education: A comprehensive guide to legal implications of administrative decision making.* San Francisco: Jossey-Bass.

Kerr, C. 1982. *The Uses of the University*, 3rd ed. Cambridge, MA: Harvard University Press.

Keig, L., and Waggoner, M.D. 1994. *Collaborative Peer Review: The Role of Faculty in Improving College Teaching.* ASHE-ERIC Higher Education Report No. 2. Washington, D.C.: The George Washington University, School of Education and Human Development.

Kindred, R., and Mohammed, S. 2005. "He will crush you like and academic ninja!": Exploring teacher ratings on ratemyprofessors.com. *Journal of Computer-Mediated Communication* 10(3), article 9. http://jcmc.indiana.edu/vol/issue3/kindred.html.

Kohn, A. 1999. The costs of overemphasizing achievement. *The School Administrator* 56: 40-46.

Kohut, G.F., Burnap, C., and Yon, M.G. Peer observation of teaching: Perceptions of the Observer and the Observed. *College Teaching* 55(1): 19-25.

Kolevzon, M.S. 1981. Grade Inflation Higher Education: A Comparative Study. *Research in Higher Education* 15(3): 195-212.

Kolitch, E., and Dean, A.V. 1999. Student Ratings of Instruction in the USA: hidden assumptions about 'good' teaching. *Studies in Higher Education* 24(1): 27-42.

Korn, H. A. 1970. Individualism and Higher Education. In Sampson, Edward E., and Harold A. Korn (Eds.), *Student Activism and Protest: Alternatives for Social Change*. San Francisco: Jossey-Bass, Inc.

Kuh, G.D. 2003. What we're learning about student engagement from NSSE. *Change* 35(2): 24-32.

Kuh, G.D., and Hu, S. 1999. Unraveling the complexity of the increase in college grades from the mid-1980s to the mid-1990s. *Educational Evaluation and Policy Analysis* 21: 297-320.

Kuh, G.D., Hu, S., and Vesper, N. 2000. They shall be known by what they do: An activities-based typology of college students. *Journal of College Student Development* 41: 228-244.

Kwon, L., Kendig, N., and Bae, M. 1997. Grade inflation form a career counselor perspective. *Journal of Employment Counseling* 34: 50-54.

Korn, H. A. 1970. Individualism and Higher Education. In Sampson, Edward E., and Harold A. Korn (Eds.), *Student Activism and Protest: Alternatives for Social Change*. San Francisco: Jossey-Bass, Inc.

Lavin, D.E. 1965. *The prediction of academic performance: A theoretical analysis and review of research.* NY: Russell Sage Foundation.

Leef, G.C. 2006. The Overselling of American Education. *Academic Questions* 19(2): 17-34.

Levine, A., and Cureton, J.S. 1998. Collegiate life: An obituary. *Change* 30: 12-17.

Lundrum, R.E. 1999. Student expectations of grade inflation. *Journal of Research and Development in Education* 32: 124-128.

Mann, R.D. *et al*. 1970. *The College Classroom*. NY: John Wiley & Sons.

Mansfield, H.C. 2001. Grade inflation: It's time to face the facts. *Chronicle of Higher Education.* (April 6), pp. 1-3.

Marlin, J.W. Jr., and Gaynor, P.E. 1989. Do Anticipated Grades Affect Student Evaluations? *College Student Journal* 23: 184-192.

Marsh, H.W. 1983. Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/course/instructor characteristics. *Journal of Educational Psychology* 75: 150-166.

Marsh, H.W. 1984. Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology* 76: 707-754.

Mathis, B.C. 1974. *Persuading the Institution to Experiment: Strategies for Seduction.* Occasional Paper No. 9, Center for the Teaching Professions. Evanston, IL: Northwestern University.

McKeachie, W.J. 1990. Research on College Teaching: The Historical Background. *Journal of Educational Psychology* 82(2): 189-200.

Moore, M., and Trahan, R. 1998. Tenure Status and Grading Practices. *Sociological Perspectives*, 41(4): pp. 775-781.

Nagle, B. 1998. A proposal for dealing with grade inflation: The Relative Performance Index. *Journal of Education for Business* (September/October): 40-43.

Nelson, J., and Lynch, K. 1984. Grade inflation, real income, simultaneity, and teaching evaluations. *Journal of Economic Education* 15: 21-37.

Palmer, J., Carliner, G., and Romer, T. 1978. Leniency, Learning, and Evaluations. *Journal of Educational Psychology* 71: 293-302.

Pascarella, E.T., and Terenzini, P.T. 1991. *How college affects students: Findings and insights from twenty years of research.* San Francisco: Jossey-Bass.

Potts, M. 2005. The Consumerist Subversion of Education. *Academic Questions* 18(2): 54-64.

Povlik, A.E., and Nardone, T. 1989. On the definition of "contingent work." *Monthly Labor Review* 11(2): 9-16.

Powell, Jr. R. S. 1970. Participation is Learning. *Saturday Review* 53: 55.

Prather, J.E., Smith, G., and Kodras, J.E. 1979. A longitudinal study of grades in 144 undergraduate courses. *Research in Higher Education* 10: 11-24.

Remmers, H.H. 1928. The relationship between students' marks and student attitude toward instructors. *School and Society* 28: 759-760.

Remmers, H.H., Martin, F.D., and Elliot, D.N. 1949. Are student ratings of instructors related to their grades? *Purdue University Studies in Higher Education* 66: 17-26.

Reuben, J. 1998. Reforming the University: Students Protests and the Demand for a 'Relevant' Curriculum. In Degroot, G.J. (Ed.). *Student Protest: The Sixties and After*. London: Addison Wesley Longman Limited.

Riesman, D. 1980. *On Higher Education: The Academic Enterprise in an Era of Rising Student Consumerism.* San Francisco: Jossey-Bass Publishers.

Rubin, A.M. 1993. Audience activity and media use. *Communications Monographs* 60(1): 98-105.

Ruch, R.S. 2001. *Higher Ed, Inc.* Baltimore: The Johns Hopkins University Press.

Rudy, W. 1996. *The Campus and a Nation in Crisis: from the American Revolution to Vietnam*. Madison: Fairleigh Dickinson University Press.

Sabot, R., and Wakeman-Linn, J. 1991. Grade inflation and course choice. *Economic Perspectives* 5: 159-170.

Seldin, P. 1993. How Colleges Evaluate Professors: 1983 Versus 1993. *AAHE Bulletin*, Oct. 1993a, pp. 6-8.

Schuster, J.H., and Finkelstein, M.J. 2006. *The American Faculty: The Restructuring of Academic Work and Careers*. Baltimore, MD: The Johns Hopkins University Press.

Shea, C. 1994. The subtleties of grade inflation. *Chronicle of Higher Education* (Jan. 5). A45-A46.

Shepperd, J.W. 1997. Relevance and Responsibility: A Postmodern Response. Response to "A Postmodern Explanation of Student Consumerism in Higher Education". *Teaching Sociology* 25: 333-335.

Shulman, L.S. 1986. Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher* 15(2): 4-14.

Soderberg, L.O. 1986. A Credible Model: Evaluating Classroom Teaching in Higher Education. *Instructional Evaluation* 8: 13-27.

Sonner, B.S. 2000. A is for "adjunct": Examining grade inflation in higher education. *Journal of education for Business* 76(1): 5-9.

Steigerwald, D. 1995. *The Sixties and the End of Modern America*. New York: St. Martin's Press.

Strenski, E. 1995. Two cheers for peer review: Problems of definition, interpretation, and appropriate function. *Journal of Excellence in College Teaching* 6(3): 31-49.

Sullivan, T.A. 1995. Teaching Evaluation by Peers. *Teaching Sociology* 23(1): 61-63.

Sunderland, J. 2006. Rank ratings. *Education Guardian*. April 12. Accessed on 14 August, 2006. http://education.guardian.co.uk/print/0,329455329-108228,00.html.

The Economist. 2002. An eye for an A. March 9, p. 74.

Trout, P.A. 1997. What the numbers mean: Providing a context for numerical student evaluation of courses. *Change* 29: 24-30.

Trow, M. American Higher Education, 1969-1975. *Chronicle of Higher Education*, May 1977.

Umbach, P.D. 2007. How Effective Are They? Exploring the Impact of Contingent Faculty on Undergraduate Education. *The Review of Higher Education* 30(2): 91-123.

US Department of Education. 1991. Assessing teaching performance. *The Department Chair: A Newsletter for Academic Administrators* 2(3).

U.S. Department of Education. 2006. A Test of Leadership: Charting the Future of U.S. Higher Education. Washington, D.C.

Vasta, R., and Sarmiento, R.F. 1979. Liberal grading improves evaluations but not performance. *Journal of Educational Psychology* 71: 207-211.

Walvoord, B.E., and Anderson, V.J. 1998. *Effective Grading: A Tool for Learning and Assessment*. Sa n Francisco, CA: Jossey-Bass Publishers.

Weimer, M., Kerns, M.M., and Parrett, J.L. 1988. Instructional Observation: Caveats, Concerns, and Ways to Compensate. *Studies in Higher Education* 13: 285-293.

Werdell, P.R. 1967. *Course and teacher evaluation.* Washington, DC: United States National Student Association.

White, C. 1998. Two Responses to Student Protest: Ronald Reagan and Robert Kennedy. In Degroot, G.J. (Ed.). *Student Protest: The Sixties and After*. London: Addison Wesley Longman Limited.

Willingham, W.W., Lewis, C., Morgan, R., and Ramist, L. 1990. *Predicting college grades: An analysis of institutional trends over two decades.* NY: College Board and Educational Testing Service.

Wilson, R. 1998. New Research Casts Doubt on Value of Student Evaluations of Professors. *Chronicle of Higher Education* 44: 12-14.

Wittrock, M.C. and Lumsdaine, A.A. 1977. Instructional Psychology. *Annual Review of Psychology* 28: 417-459.

Worthington, A.G., and Wong, P.T. 1979. Effects of earned and assigned grades on student evaluations of an instructor. *Journal of Educational Psychology* 79: 764-775.

Yarbrough, D.B. 1989. A Cognitive Psychological Perspective on Teacher Evaluation. *Journal of Personnel Evaluation in Education* 2: 215-228.

Young, J.R. 2002. Homework? What homework? *Chronicle of Higher Education* (Dec. 6), A35-A37.

Zirkel, P.A. 1999. Grade inflation: A leadership opportunity for school of education? *Teachers College Record* 101: 247-260.

Zubizarreta, J. 1999. Evaluating Teaching Through Portfolios. In Seldin, P. (ed.), *Changing Practices in Evaluating Teaching: a Practical guide to Improved Faculty Performance and Promotion/Tenure Decisions*. Bolton, MA: Anker Publishing Company.