

[Education Policy Analysis Archives](#)

Volume 5 Number 6

February 12, 1997

ISSN 1068-2341

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass Glass@ASU.EDU.

College of Education

Arizona State University, Tempe AZ 85287-2411

Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES. Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Academic Freedom, Tenure, and Student Evaluation of Faculty: Galloping Polls In The 21st Century

[Robert E. Haskell](#)

University of New England¹

This is the first of four articles by Haskell on this subject. The other articles can be found at

- [Volume 5 Number 17](#)
- [Volume 5 Number 18](#)
- [Volume 5 Number 21](#)

Abstract: Despite a history of conflicting research on the reliability and validity of student evaluation of faculty (SEF) it has typically not been viewed as an infringement on academic freedom. When it is suggested that SEF may impinge on academic freedom, it is often considered an attack on either student rights, or on the process of evaluating faculty performance in general. Faculty and educational administrator views and surveys are reviewed as SEF is used in salary, promotion and tenure decisions. It is suggested that the literature shows that SEF infringe on instructional responsibilities of faculty by providing a control mechanism over curricular, course content, grading, and teaching methodology. It is further suggested that SEF play a significant role in current attacks on tenure, and that its role in a demographically diverse 21st century educational system has changed from its benign historical origins. It is concluded that contrary to current views, SEF is a serious unrecognized

infringement on academic freedom.

..... Despite the long line of research on the validity of SEF it has not typically been viewed as impinging upon academic freedom. At first glance, the very notion seems counterintuitive, if not downright mean spirited. Indeed, when it is suggested, it is seen as not only a novel idea, but as an attack on either students, or a general attack on evaluating faculty. In fact, one recent handbook for college administrators, (Weeks, 1996) refers to such an idea as "a rather novel attack on the use of student evaluations in assessing a faculty member's performance." Indeed, it has generally been taken for granted that SEF is appropriate and necessary. Unlike the body of research on the methodology of SEF construction and the validity of the ratings, there is a paucity of data on the issue of its impact on academic freedom.

..... The likely reasons for this paucity of data are that (1) there has been little professional mention of SEF as an infringement on academic freedom, (2) some faculty are embarrassed to admit that student evaluations may influence their professional behavior in the classroom, (3), to question the right of students to evaluate faculty may be considered unprofessional if not undemocratic, (4) to question the right of students to evaluate faculty may be seen as self serving, and (5)SEF tends not to be high status research. Finally when SEF is recognized to have an impact, unlike traditional threats to tenure and academic freedom, the deceptive appearance of SEF does not seem to warrant serious concern.

..... In a recent and otherwise carefully reasoned book on academic freedom (Menand, 1996), the issue of SEF is not mentioned.² At best, the subject of SEF has been the orphan in discussions on academic freedom with no direct logical lineage. At worse, it has been considered illegitimate. There are signs, however, indicating wider faculty recognition of this issue. As administrative policy, the use of SEF has largely evolved in a de facto manner, this paper will review the issue in search of a more tutored policy.

..... It is important to note at the outset, that it is not SEF per se that is the issue, but the impact of its use on salary, promotion, tenure decisions, and equally important, its impact on the delivery of quality education. In the evolution of any policy the accumulation of data, judgements, and arguments around an issue need to be coalesced. As the history of legal rights demonstrates, issues not considered to have legitimate standing only come to have standing after a long process of advocacy. It is, therefore, the purpose of this paper to coalesce arguments, and data, and explore implications of SEF scattered throughout the literature and to thus facilitate this developmental process.

..... In explicating SEF, many closely related issues must be substantially bracketed. These related issues include (1) its validity (Cahn, 1987; Damron, 1996; Greenwald, 1996; Greenwald and Gillmore, 1996; Scriven, 1993; Seldin, 1984; Tagomori, Bishop, and Laurence. 1995),³ (2) the problem of defining teaching effectiveness,⁴ (3) general variables affecting SEF scores,⁵ (4) alternatives to SEF's (5) classroom politically correct or popular standards and perceptions, (6) low student academic preparation, (7) age and gender discrimination issues (Feldman, 1983, 1993) (8) strategies for change, and (9) other integrally related issues such as their being largely responsible for lowered course standards, and grade inflation.⁶ Though these are important and related issues, they can only be addressed here in so far as they directly impact the focus on SEF and academic freedom.

..... While SEF was apparently first used in the early 1920's at the University of Washington, beginning in about the 1960's SEF has been increasingly used by universities in decisions on tenure and promotion. One of the reasons that SEF was instituted---and rightly so---was for informational feedback so that faculty might be more aware of student needs. The instrument has not, however, been used just for informational feedback to professors. If this were the case, then SEF would presumably not be a problem.⁷ As Cashin (1996), Director of the Kansas State

University, Center For Faculty Evaluation and Development, notes, "The higher education rhetoric is almost universal in stating that the primary purpose of faculty evaluation is to help faculty improve their performance. However, an examination of the systems--as used--indicates that the primary purpose is almost always to make personnel decisions. That is, to make decisions for retention, promotion, tenure, and salary increases." Herein lies the problem.⁸

In comparing two studies of the same 600 liberal-arts colleges, the author found that the number of institutions using student ratings to evaluate instructors had escalated from 29 per cent to 68 per cent to 86 per cent. The author noted that no other method of evaluation has approached that level of usage (Seldin, 1993). Another survey found that most business schools now use SEF for decision making, with 95% of the deans at 220 accredited undergraduate schools always making use of them as a source of information (Crumbley, 1995). Two nationwide studies of accounting department Chairpersons, indicated that reliance upon SEF was second only to research publications in professional journals (Yunker and Sterner, 1988). Department chairs and Deans often weigh student ratings heavily in the faculty evaluation process. Perhaps no other method of evaluation has become so sacrosanct. SEF are used not only in the U.S. but in Australian, Canada, Europe and Great Britain. Unlike in the U.S., however, in Great Britain SEF by formal questionnaire, despite apparently no formal mandate, are increasingly used, though not weighed as heavily as is information gathered by other means. (Husbands and Fosh, 1993).

While considerable research---both past and present--- has been conducted to assess the validity of SEF, formal reports or studies suggesting that they might impinge on academic freedom are virtually non existent. There is likewise a paucity of direct hard data to support widespread but informal evidence addressing the issue of SEF as impinging on academic freedom.⁹ Informal and reasoned analyses of the issue indicate that because SEF is used for faculty salary, promotion, and tenure decisions, there is pressure to comply with student classroom demands regarding teaching style, grading and a host of others demands(see below). It is suggested that it is this pressure to comply with student demands that directly leads to an infringement upon academic freedom.

SEF is not simply a salary, promotion, and tenure issue--- as important as these are for individual faculty; nor is the issue simply that students evaluate faculty. As the findings of this paper suggest, SEF are not the benign instrument they may appear to be or may once have been. Their primary impact goes to the core of academic freedom and to quality of instruction.

A Brief Look At Academic Freedom

..... Academic freedom and tenure are two sides of the same coin. Following an initial statement of principles in 1915, the current view of tenure was established in 1940 when the American Association of University Professors (AAUP) and the Association of American Colleges (AAC) officially sanctioned it for purposes of preserving faculty's right to academic freedom. Legally, it assures faculty the right to pursue any line of inquiry in the course of their teaching or research without being censored, penalized or fired by university administrators.¹⁰ In 1973, the Commission on Academic Tenure in Higher Education (jointly sponsored by the AAUP and the AAC) made the following recommendation:

The commission believes that "adequate cause" in faculty dismissal proceedings should be restricted to (a) demonstrated incompetence and dishonesty in teaching and research, (b) substantial and manifest neglect of duty, and (c) personal conduct which substantially impairs the individual's fulfillment of institutional responsibilities. The burden of proof in establishing cause for dismissal rests upon the institution (University of Michigan, 1994).

..... The AAUP's Statement on Teaching Evaluation suggests that

Casual procedures, a paucity of data, and unilateral judgments by department chairs and deans too often characterize the evaluation of teaching in American colleges and universities.... A judicious evaluation of a college professor as teacher should include: (1) an accurate factual description of what an individual does as teacher (AAUP Committee C , 1975).¹¹

..... How SEF establishes incompetence, or neglect of duty is problematic, having largely to do with issues of validity. To further complicate matters the concept of academic freedom, like most abstract terms is logically fuzzy around its edges. Moreover, unlike the legal categories of academic freedom and tenure, there is no equivalent legal category of SEF. Consequently, published legal rulings on this issue are scarce.

..... While academic freedom has not been recognized universally by the courts as equivalent to a constitutional right, it has nevertheless been viewed as a right which the courts have deemed must not be violated in the performance evaluation process. In addition, academic freedom has been associated with the First Amendment right of free speech. Some courts have considered it to be a First Amendment-right in and of itself. Academic freedom is a "special concern of the First Amendment, which does not tolerate laws that cast a pall of orthodoxy over the classroom." While the two rights are not necessarily the same, they frequently and sufficiently overlap to trigger judicial scrutiny when faculty performance evaluation process threatens to impinge on the First Amendment (Copeland and Murry, 1996).

Faculty Assessment of How SEF Infringe Upon Academic Freedom

..... Mention of this issue within the text of a number of articles throughout the literature provides a kind of uncontrolled data base indicating that an increasing number of faculty consider SEF an impingement on academic freedom. There are, however, a few formal surveys. While the judgement of a profession's membership on a subject is not necessarily a valid indicator, especially when that the judgement renders them the primary beneficiaries, in the absence of significant evidence to the contrary it would seem reasonable that a profession's judgement should be accorded considerable weight in any analysis or process leading to policy decisions.

..... Before reviewing what the literature seems to show is the predominant faculty view, a statement of at least one faculty who does not think SEF impinges on academic freedom will illustrate the opposite perspective. This faculty was quoted as saying,

Do I sense more than a little self-disdain in the comments of my colleagues who portray faculty members as cowed by the prospect of a less than enthusiastic course evaluation by their students? That is not my impression of my colleagues, and I can assure you it is not an accurate representation of me. It is my impression that faculty grade students with quality of performance as the primary, if not sole, criterion (Kresl, 1995).

At least some faculty, then, believe that SEF does not influence their behavior and is therefore not an abridgement of academic freedom.

..... While formal surveys of faculty views on this issue are relatively rare, one study found that at least one third of faculty respondents reported lowering their grading standards and course level in response to their student evaluations (Ryan, Anderson, and Birchler, 1980). Another study found that 64.8% of respondents believed that "student evaluation forms are responsible for lenient grading." According to another study, 39% of accounting administrator respondents

admitted being aware of faculty who altered their instructional behavior in order to improve evaluation scores (Crumbley and Fliedner, 1995). Faculty were also in nearly universal agreement that SEF is important in promotion (86.6%) and tenure (88.2%) reviews (Kolevzon, 1981, see also Avi-Itzhak and Lya, 1986).¹² At the very least, such reviews are strong prima facie evidence of SEF as an infringement on academic freedom.

..... In two nationwide studies of accounting departments it was shown that 37% of faculty respondents were dissatisfied with their present evaluation system. The response of 561 accounting faculty to the statement that the present system of SEF is well-designed and properly implemented, reported that 15.3% indicated "strong disagreement," 31.7% "disagree," and 25.5% indicated a "neutral" response (Yunker and Sterner, 1988; Bures, DeRidder and Tong, 1990).¹³

..... Unlike the paucity of formal surveys, there are numerous statements by faculty in the research literature clearly arguing that SEF is an infringement on academic freedom. These statements by faculty contend that SEF (1) is prime facie evidence of administrative intrusion into the classroom, (2) are often used as an instrument of intimidation forcing conformity to politically correct standards (Young, 1993), (3) create pressure for a self-policed lowered teaching standard (Bonetti, 1994), (4) are responsible for a considerable amount of grade inflation (Greenwald, 1996, Greenwald and Gillmore, 1966),¹⁴ (5) function as prescriptions for classroom demeanor (Damron, 1996), (6) when used for promotions, salary raises or continued employment, SEF becomes a potent means of manipulating the behavior of faculty (Stone, 1995), (7) when salary and promotion are possible consequences of SEF there is pressure for faculty to teach in a manner that results in higher student evaluation (Damron, 1996), (8) contrary to their original intent of improving instruction, do not eliminate poor or below-average teachers but instead increases poor teaching practices (Carey, 1993), (9) illustrate a mercantile philosophy of "consumerism" (Benson, and Lewis, 1994), which erodes academic standards (Goldman, 1993; Renner, 1981), (10) have thus lowered the quality of U.S. education (Carey, 1993; Crumbley, and Fliedner, 1995; Young, 1993), (11) lead to the inappropriate dismissal of faculty (Parini, 1995), and (12) constitute a threat to academic freedom (Dershowitz, 1994; Stone, 1995). Finally, it would seem that SEF creates an educational conflict of interest between faculty and student impacting on the quality of instruction.¹⁵

..... At least one Faculty Senate tenure and promotion committee at a large state university along with the ACLU (Heller, 1986), found that given the crude state of SEF and its subsequent use, such evaluations are a "de facto violation of academic freedom." (p.14) Despite these concerns, however, there are some faculty who, given the alternative of peer evaluation, consider SEF as less harmful (Carey, 1993).

..... Academic freedom involves more than just external political control of course content (Schrecker, 1986), and a Scopes Trial type gag order; it entails pressure to censor unpopular subjects within the popular cultural belief systems of students. In this regard, Alan Dershowitz (1994), the well known Harvard Law School professor, and appellate attorney believes that the administrative use of SEF is a direct threat to academic freedom and to quality education. After teaching on a controversial subject, and receiving negative feedback from students, he says he "realized how dangerous it would be for an untenured professor" to teach about such subjects. He goes on to point out, however, that "Most of the students appreciated the diversity of viewpoints," and evaluated him fairly, noting that he was "very good at presenting alternative views," and "helped me get a less dogmatic view of the law," that he was "open to criticism," that the class was "the most engaging class on campus," that he was "the most intellectually honest professor I've had," and was "fair in presenting sides that usually aren't raised."

..... There was, however, a sizable group of students who, says Dershowitz, "used the power of their evaluations in an attempt to exact their political revenge for my politically incorrect teaching." He was lecturing on the (sensitive) legal analysis of rape.¹⁶ He notes, that

One student said that I do "not deserve to teach at Harvard" because of my "convoluted rape examples." Another argued that women be allowed an "option" not to take my class because I "spent two days talking about false reports of rape," another demanded that my "teaching privileges" be suspended. One woman purported to speak for others: "Every woman I know in the class including myself found his treatment of rape offensive and disturbing." Another woman felt "oppressed throughout the course." Although I always try to learn from my evaluations, I will not be bullied into abandoning a teaching style that I believe is best designed to stimulate thinking. (italics added, p.118-119)

..... Dershowitz says that it takes no courage for him to exercise his academic freedom, since he holds the rank of full professor and has tenure, but he seriously wonders if he were an untenured assistant professor, if he would I have the same courage. He concludes:

Are other less established teachers being coerced into changing their teaching by the fear of negative evaluations, which can be fatal to tenure? You bet they are, and it poses a real danger to academic freedom and good education. (italics added, p.118-119)

Dershowitz perhaps overstates his case on the security provided by tenure for faculty on many campuses. Being at Harvard, he is not the typical faculty member. Unlike at the Harvards, the Seven Sisters, and the Ancient Ones, in the trenches of most campuses SEF is accorded inordinate weight (at least by administrations) over research. He is also "politically" protected by being so well known, not to mention that his famous cases and books have probably made him financially independent.

Classroom and Curricula Infringements Based on SEF

..... The following three recent examples will illustrate the extreme to which SEF can lead in terms of impinging on academic freedom. Each exemplifies a different aspect of academic freedom being infringed upon by SEF as used at most institutions. There are likely a number of cases similar to a recent one reported in the Chronicle of Higher Education of faculty being dismissed because of negative student evaluations (Magner, 1995).

..... With four weeks remaining in the semester, a non-tenured faculty was removed as the instructor of an introductory chemistry course at the University of Montana.¹⁷ His dismissal was prompted by a student petition signed by 100 of the 200 students in one of his classes. These students maintained that he demonstrated an "inability to teach new information" and a "willful academic demoralization of students." Three different views of this situation have emerged. Some contend that he was a tough teacher who became the victim of "whiny" students looking for an easy instructor to spoon-feed them the course material. Supporters maintained that many students in the course planning professional or graduate programs, feared that a low grade in a difficult subject would diminish their chances of admission. Others, maintained he ignored student questions, and made tests more difficult than they needed to be.

..... The faculty member maintained he received good student evaluations in the past with no problems in his other courses. The faculty believe the reasons for the complaint is that he is a demanding teacher who requires students to go beyond simple memorization and regurgitation of facts on his exams. He says he has "never knowingly belittled a student." He sees an anti-intellectualism at work, maintaining that "There is an attitude on...campus that any display of learning or erudition is arrogance." He also maintains that "it's not the students' prerogative to decide who teaches a course, and by extension, how that course should be taught."

Administration is quoted as saying, "We're an open-admission university. A very large fraction of the class was completely unable to compete." As it turns out, the previous year the administration warded off a similar student petition against a tenured full professor of chemistry. In that case, the administration says they "clearly felt we were dealing with some people [students] who wished to get better grades for less work." Granted, all the facts are not in on this case. The faculty was replaced with a professor who is popular, with students. Nevertheless, it is clear that student evaluations were the primary mechanism leading to a faculty's dismissal when, whatever the facts are, they are at least open to serious question.

..... As well as academic freedom in the classroom, other cases affect tenure, promotion, and more importantly, curricula. A faculty (Goldman, 1993) at Wichita State University notes that in a thirty-faculty department which is responsible for certification of teachers, in the past 25 years, six faculty have been hired as assistant professors to teach Foundations of Education. All apparently well qualified, receiving their doctorates from excellent universities, only one of these faculty was awarded tenure; none was promoted. According to Goldman (1993), the reason for this is student evaluations.

..... In general, as the data show, required courses hold less interest and receive lower evaluations than elective courses; in general, students who are drawn to become teachers are concrete-sequential, and are less interested in the abstract and theoretical content of the foundations of education course. This leads the faculty who teach the foundational course to receive lower student evaluations than other education faculty. Because student evaluations are often the major, if not the only, gauge of teaching quality, and since teaching quality usually out ranks research and scholarly productivity on most campuses, when tenure, promotion, and salary increases are awarded, these rewards will not be evenly distributed to the foundational, educational psychology faculty. Rewards will accrue to the concrete-oriented methods faculty whose courses will further intensify the concrete orientation of teacher preparation. And the downward spiral will continue.¹⁸

..... The third example of SEF as infringement on academic freedom reportedly occurred in a business department (in Crumbley, 1995). A rigorous instructor teaching a basic business class gave D's and F's and received SEF scores in the one range (on a 5-point scale). As a consequence, she was removed from teaching that class and assigned to a non required graduate course where she proceeded to give a 50-50 split of As and B's each semester. She received SEF scores of up to 4.9 after making this adjustment. She informs the students at the start of the semester that only As and B's will be given. This administrative strategy of assigning "tough" grading faculty to non required courses allows students to force easy grading by self-selecting away from the more stringent graders, and thus "censors" certain instructors courses and by implication the content of what they teach.

..... These examples suggest how SEF can be used to shape faculty behavior, curricular content, and the kind of faculty that are retained in programs. They also suggest how SEF is selectively used as a primary mechanism of dismissal of faculty who do not conform to student and administrative demands. Though a faculty may (a) receive good evaluation in other courses, and (b) receive good evaluations from the majority of students, a few negative evaluations may be determinative of disciplinary action toward faculty. The examples also raise the issue of who and what determines academic standards. On campuses with "open admissions," meaning a lower level of student in courses than has historically been the case, that an otherwise competent instructor is required to adapt his or her course to whatever level of student is enrolled, or be subject to the above consequences based on student dissatisfaction. Regardless of the specifics, the case illustrates the direct aspects of student evaluations affecting educational standards and raises the question of academic freedom of faculty in the classroom.

Legal Considerations of SEF

..... There are multiple latent legal issues engendered by SEF. Unlike academic freedom and tenure, as a legal search category SEF does not exist.¹⁹ In a subsequent paper I will coalesce and deal more fully with legal rulings involving SEF in the denial of reappointment and tenure.

While the following ruling is Canadian and has no standing in the U.S., it is nevertheless initially relevant, especially given that Canadian protection of free speech is more narrow than in the U.S.

..... Under Canadian law, a situation recently developed at the University of Regina where a faculty was denied tenure partly on the basis his student evaluation scores (Education

Employment Law News, 1994).²⁰ A board of arbitration ruled that

tenure decisions could not be based solely on assessments which were completed by students who had never been made aware of the ramifications of their statements....To base serious career decisions narrowly on student evaluations is not to be encouraged..... This is particularly so when...the students are not advised of the potential use of the evaluation tool... [I]f evaluations are to be used for serious career development purposes those completing them should be aware of the potential consequences of their participation.

At least one similar ruling exists under U.S. law.

..... At the University of Guam, a ruling to remove anonymous student evaluations from professors' tenure files was handed down by an arbitration board as the result of a rare challenge to the use of such evaluations in tenure and promotion decisions (Blum, 1990). The action was in response to a grievance filed by the university's faculty union, the Guam Federation of Teachers, which is affiliated with the American Federation of Teachers (AFL-CIO). The union said the use of SEF violated the union's contract with the university, which provides that anonymous documents or those "based on hearsay" should not be included in a faculty member's file. The union also argued that the university improperly interpreted the data from the evaluations.

..... Some of the issues here are (1) students not being made aware of the purpose and ramifications of their evaluations, (2) the anonymous nature of student evaluations, (3) the invalid analysis of SEF, and therefore, (4) SEF in effect being anecdotal and hearsay data. Since most SEF results are prepared anonymously, an instructor has no recourse to confront his/her evaluators. As will be addressed below, the anonymous nature of SEF is beginning to also be questioned by arbitration boards.

..... After reviewing the research, a Director of the Office of Educational Assessment at one large state university (Gillmore, 1984) concluded that, "If student rating are to qualify as evidence in support of faculty employment decisions, questions concerning their reliability and validity must be addressed." (p.561) Perhaps the most comprehensive review and statistical analysis of SEF validity is that of Greenwald (1996).

..... The philosopher of science, Michael Scriven (1995, 1993, 1991), who has conducted extensive analyses on faculty evaluation methodology suggests of faculty evaluation in general that,

All are face-invalid and certainly provide a worse basis for adverse personnel action than the polygraph in criminal cases. Based on examination of some hundreds of forms that are or have been used for personnel decisions (as well as professional development), the previous considerations entail that not more than one or two could stand up in a serious hearing.²¹

The cost of lawsuits notwithstanding, given the above, it is surprising that there have not been more lawsuits by faculty.²² It appears, however, that in regard to faculty evaluation in general the courts have not been concerned with validity issues, though this may be changing as well (Rebell,

1990).

Releasing SEF to Students and the Public

..... In exploring possible legal implications of SEF, it should be made clear that I am not an attorney and approach this section on the basis of the "reasonable man" legal standard. To begin, some faculty believe that due process and defamation issues are involved in SEF (Crumbley, 1996). It has been suggested that faculty are entitled to at least the same rights as students. The Fourteenth Amendment requires, for example, due process before a public institution may deprive one of life, liberty, or property. Given the problematic nature of SEF, due process is in question. In a university, a faculty's reputation is considered a liberty right, and for tenured faculty the courts have pronounced the possession of tenure a property right. Presumably, any inappropriate action depriving faculty of these rights would be open to legal action.

..... Though it is illegal to post a student's grades using a social security number or date of birth,²³ on the majority of campuses scientifically questionable SEF and other anecdotal student remarks about faculty teaching are not only used in determining faculty salary increases, promotion and tenure decisions, they are openly published on some university campuses and sanctioned by some administrators and state government officials. In what many faculty see as an outrageous attempt to control the academic classroom, some state governments have sanctioned the release of SEF to the campus community and in some cases to the general public by publishing faculty student evaluations on the university's world wide web pages, thus making them not only available on campus but globally.

..... Case in point: At the University of Wisconsin, the Chancellor refused to release the SEF, citing a statute allowing personnel evaluations to be withheld from public view. The students took the chancellor to court. However, after being advised to do so by the state's Attorney General, citing Wisconsin's open-records law, the University of Wisconsin's campus will open students' evaluations of professors for public view. To the credit of the student and faculty senates, they passed resolutions in support of the Chancellor's refusal, and the university's lawyer concurred. Despite these resolutions, the Attorney General disagreed, writing that "the requested records are public records and the university's stated reasons for withholding access do not outweigh the public interest in the records" (*Chronicle of Higher Education*, 1994a, 1994b).

..... Other schools also published SEF. One recent survey of accounting departments found that 11.4% of the respondents indicated that SEF scores are made available to students (Crumbley and Fliedner, 1995). Indeed, a search using "faculty evaluation" on the world wide web will return numerous examples of published SEF. All this while faculty are restricted from divulging information on students (see Pennsylvania State University, 1996). Articles are, however, beginning to appear that question the legality of publically releasing SEF (Robinson and Fink, 1996).

..... It has been suggested that if a university damages a faculty's reputation by publishing false and anecdotal data from SEF, faculty should be able to sue for libel or defamation. The concept of defamation typically refers to communication that causes a person to be shamed, ridiculed, held in contempt by others, or their status lowered in the eyes of the community, or to lose employment status or earnings or otherwise suffer a damaged reputation. Legally, while defamation is governed by state law, it is limited by the first amendment (Black, 1990).²⁴ According to one source, however, the courts have generally protected administrators from defamation charges resulting from performance evaluations (Zirkel, 1996). It would seem, however, that these older precedents applied when administrative evaluations were conducted in private and not publically distributed.

..... University administrators are often allowed to release SEF to students when the release of

personnel information is apparently allowed in no other phase of personnel or other key management functions. An Idaho ruling upheld the release of SEF to students by reasoning that students were not the general public and therefore faculty evaluations were not protected under the privacy rights of the Idaho Code (Evaluating Teacher Evaluations, 1996). Given such apparent breaches of confidentiality and privacy, it will be instructive to see how the courts will continue to rule. It would seem that a university should be held responsible for insuring that data made public are valid.

..... Finally, in typical personnel evaluations, professional validation studies are not permissible unless shown by professionally acceptable methods to be "predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are being evaluated." In Title VII of the Civil Rights Act of 1964 the employer must meet "the burden of showing that any given requirement (or test) has a manifest relationship to the employment in question" (in *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971)). "In view of the possibility inherent in subjective evaluations, supervisory rating techniques should be carefully developed, and the ratings should be closely examined for evidence of bias" (EEOC Guidelines, 99 CFR 1607.5 (b) (4). (in *Crumbley*, 1996).

SEF As Social Judgement and Diagnosis

..... Given courts uncritical assumptions regarding the validity of faculty evaluations in general, and the untrained judgement of those making decisions, it would seem appropriate to address the issue of judgement and decision making in SEF. Frequently, in interpreting SEF the method is simply "eyeballing" them. The "actual" consistency of findings of SEF validity at this point in time, notwithstanding, one judge (Rebell, 1990) clearly points out, "because the state of the art concerning teacher-evaluation practices is at a sensitive developmental stage, extensive court intervention at this point can substantially influence ---for better or worse---the future direction of basic practice in the field." (p.344) He goes on to say that seen in this light, it is clear that whether increased judicial intervention will have a positive or a negative impact on professional evaluation practice lies more with educators and psychometricians than with the judges, meaning that faculty control over future rulings depends upon providing clear evidence to the courts.

..... Thus, part of influencing future court rulings is demonstrating relevant research. With the typically unsystematic method by which SEF are interpreted, and the considerable weight accorded selective negative comments by a few students in making tenure and promotion decisions, the meaning of SEF (i.e., diagnosis) as they are applied to the analysis of teaching effectiveness is, in fact, all too frequently an intuitive one.

..... As to the issue of the validity of findings in SEF, as indicated at the opening of this paper, more recent and rigorously analyzed statistical findings cast serious doubt on their validity. The research literature on judgement and decision making would seem applicable to both students making such judgements and to those interpreting (i.e., diagnosing) the results, especially given the intuitive and non-data based foundations of such judgements and decisions. In the course of reviewing the research over the years on social judgement and clinical diagnosis, it seems clear that the manner in which nearly all SEF data are analyzed is a subset of the research on social judgement and clinical diagnosis, with both student and interpreter being involved in the same logical and cognitive biases and distortions that result in the pervasive low accuracy level of social judgement in general and clinical diagnosis in specific.

..... Psychological research has recognized the severe cognitive problems and limitations of "intuitive," and "experience-informed" everyday judgements for over thirty years, (Dawes, Faust, and Meehl, 1989; Faust, Guilmette, Hart, Arkes, Fishburne and Davey, 1988; Garb, H. N. 1989; Hayes, 1991; Larkin, McDermott, Simon, and Simon, 1980; Rabinowitz, 1993) yet the mistakes

continue in everyday practice situations. Interpretation of SEF are no different. As Franklin & Theall (1990) independently observe relative to SEF

Even given the inherently less than perfect nature of ratings data and the analytical inclinations of academics, the problem of unskilled users, making decisions based on invalid interpretations of ambiguous or frankly bad data, deserves attention. According to Thompson (1988, p. 217) 'Bayes Theorem shows that anything close to an accurate interpretation of the results of imperfect predictors is very elusive at the intuitive level. Indeed, empirical studies have shown that persons unfamiliar with conditional probability are quite poor at doing so (that is, interpreting ratings results) unless the situation is quite simple.' It seems likely that the combination of less than perfect data with less than perfect users could quickly yield completely unacceptable practices, unless safeguards were in place to insure that users knew how to recognize problems of validity and reliability, understood the inherent limitations of rating data and knew valid procedures for using ratings data in the contexts of summative and formative evaluation. (79-80).

The authors conclude by noting "It is hard to ignore the mounting anecdotal evidence of abuse. Our findings, and the evidence that ratings use is on the increase, taken together, suggest that ratings malpractice, causing harm to individual careers and undermining institutional goals, deserves our attention." (pp. 79-80). Recognizing such problems is not methodological nit-picking; they are pragmatic and fundamental.

..... It would therefore be a reasonable recommendation to suggest those researching the validity and especially interpreting SEF familiarize themselves with the psychological social judgement and clinical diagnosis literature. Few states or courts require evaluators to be trained in evaluation methods (Rebell, 1990).²⁵ It is these cognitive problems that need to be addressed in the SEF literature prior to otherwise statistically sophisticated of the data being performed analyses---when rigorous methods are used.

SEF and Administrative Control of Academic Freedom

..... In addition to legal aspects, there are the pragmatics of the SEF which maintain its use. The literature clearly suggests that administrators tend to strongly oppose the elimination of SEF being used for faculty salary, promotion and tenure decisions. There are three reasons for opposing the elimination of SEF. The first seems to be a lack of practical alternatives to SEF (Greenwald, 1996),²⁶ the second is administrative control, and the third is that student input facilitates student retention in numerous ways.

..... SEF provides a mechanism of control in a system otherwise lacking direct control over faculty, and are a powerful tool in assuring classroom changes that lead to the retention of student tuition dollars by assenting to student consumer demands and of parents who foot the tuition bill. Academic issues such as teaching, grading, curricular requirements, and other academic standards have by tradition and expertise been the exclusive province of faculty. Although academic freedom and the protection of tenure would appear to insure faculty classroom independence, the extent to which faculty thought and behavior are administratively shaped is the extent to which both are infringed upon (see below).²⁷ Such administrative control mechanisms, of which SEF is one, are therefore seen as infringements on academic freedom.

..... Control mechanisms are more widespread and intricately embedded in the everyday operations of the university than is generally acknowledged. After a review of the research, Stone (1995), observes that SEF opens the "door to the direct application of bureaucratic control to

academic decisions. It is the very kind of policy that, for example, has enabled educational administrators to mandate the "politically correct" at the expense of the "academically credible." He further observes,

In great part, the reason for this outcome is that the faculty comprising colleges, departments, and programs are subjected to the same incentives that impact the institution as a whole. Chief among these incentives is internal flow of institutional resources. The allocation of institutional resources is administratively controlled in such a way that the academic units with the greatest enrollment growth are generally afforded the greatest resources. The incentives are such that no individual faculty member can comfortably question activities that build his or her department's or college's budget. After all, the job they save may be their own. The same can be said for peer evaluations of faculty for promotion, tenure, and merit pay. Faculty typically find it difficult to negatively evaluate colleagues who are high credit hour producers and who have high student ratings even if their academic standards are suspect. Credit hours and student ratings are routinely and carefully monitored by the administrative bureaucracy. Learning outcomes are given much less attention. Units that fail to grow not only remain under supported, they are sometimes cannibalized, and their share of the budget is channeled to the areas demonstrating greater potential for growth....The internal competition for resources shapes individual faculty behavior as well.... They are, therefore, in the interest of self-preservation, especially inclined to accommodate academic standards to the student market.

Stone concludes that administrative influence, including the use of SEF, has become so effective in shaping faculty that their collective voice has been seriously compromised.²⁸ It should be noted that some administrators, and university presidents do recognize the importance of tenure and support its continuance. (e.g., Cotter, 1996).

..... Contrary to some published reports, conditions such as weakened standards, fragmented curriculum, and inflated grades do not simply arise from a spontaneous deterioration of faculty into so-called "deadwood." Neither are most faculty incompetent at teaching. As Stone (1995) points out, "Rather, these problems seem likely to have developed as a result of the continuing insidious pressure placed on teaching and grading practices by the imperative to keep students happy and enrollments up." While most students are positively disposed to SEF, not all students are so oriented.²⁹

..... Many students understand the above described ensuing consequences. A glance at articles from online student newspapers reveals strong sentiments against what many students consider the erosion of standards created by SEF. One student writer went so far as to say "We therefore suggest a boycott of the 1995 student/teacher evaluations. This boycott will provide a more effective means of communication than anything written on the evaluation itself. Something must be done about the trend of grade inflation. We as students refuse to contribute to the downfall of academia " (Stern and Flynn, 1995). Some students are thus quite aware of the effects of SEF on their education. Thus the foregoing are something to seriously consider in the light of the principle of academic freedom.

Faculty Complicity and Adaptation to SEF

..... There is the argument that academic freedom is not abridged if faculty voluntarily participate in the SEF process. This argument is problematic for three reasons. First on most campuses SEF is required by administrative policy, and is therefore neither individually nor

collectively voluntary.³⁰ Second, as noted by Stone (1995) above, given their mandated use in salary, promotion and tenure decisions, SEF creates significant pressure to conform to administrative and student demands. The third is that the second condition does not seem to meet the requirements of what is legally considered "informed consent" in a contractual agreement. Like most people, given that faculty behaviors are subject to reward structures, it is not surprising that faculty would tend to teach to the evaluations and self-limit their academic freedom prerogative. From a purely psychological/behavioral learning theory perspective, when a faculty's livelihood is at stake, it is understandable why many will adjust their grading and course content level in order to receive a good evaluation. It is also predictable: behaviors that are rewarded tend to increase the probability of their reoccurrence.³¹ Similarly, as two research studies note (Crumbley, 1995; Nelson and Lynch, 1984), from an economic business accounting point of view, If a faculty can choose teaching styles, grading levels, and course content, s/he will naturally prefer choices that are expected to result in higher SEF scores; if faculty know the variables affecting their careers, they will meet these criteria. There are laws that regulate ones financial statements so as to reduce the manipulation and opportunistic behavior regarding income, yet there is no regulation of SEF. Faculty "have a high incentive to manage SEF, even more so than managers have the incentive to enhance earnings"(Crumbley, 1995). Unfortunately student ratings tend to discourage instructors from espousing views that might offend popular student prejudices. "Expecting that even highly professional individuals will disinterestedly adhere to academic and intellectual ideals in the face of pervasive incentives to do otherwise is not realistic." (Stone, 1995) Even one of the strongest advocates of the validity of SEF (Seldin, 1984) warns: "The confidentiality of the data must at all cost remain inviolate. If data are shared, it must be with the consent and at the discretion of the appraised professor. For data surreptitiously to be used for personnel decisions, it will have an immediate chilling--even fatal--effect on the credibility of the entire evaluation program." (p.129) Certainly when SEF are published not only confidentiality has been violated but so is trust. Again, given the above pressures, it would seem that "voluntary" faculty participation in the SEF process does not meet the requirements of what is legally defined as informed consent. As two legal scholars (Copeland and Murry, 1996) noted in a different context is directly applicable here. They note, "Of course the right to academic freedom means very little if by exercising it the educator suffers, or runs the risk of suffering, financial consequences." (p.249)

SEF As Trojan Horse In The Movement to Eliminate Tenure

..... There is a final and even more serious threat that SEF can be said to pose to academic freedom than the simple act of student pressure and administrative control of rewards based on them. This last threat is not generally recognized. This threat is the use of SEF in a continuing national movement against tenure and to reduce costs. For about the last twenty years there has been considerable criticism about the concept of tenure. The public has always viewed the purpose of academic tenure as job security, just like seniority in a trade union. And as with unions, the perception is that job security leads to lowered productivity. Accordingly, there have been continuing periodic attempts to eliminate tenure. More recently, however, these attacks have increased. The attacks come from all levels of society and from inside academia itself, including boards of trustee and State Legislatures attempting to either eliminate or revamp academic tenure, the latest being the widely publicized bid by the University of Minnesota Board of Regents (Guernsey, 1996; Healy, 1996; Magner, 1996; *The Chronicle of Higher Education*, 1995). Typically, the arguments against tenure have been economic. With shrinking budgets,

these old economic arguments have taken on an even more bottom line approach to education, arguing, at the least, for the revamping of tenure for the purpose of enabling schools to become programmatically and fiscally more flexible in a changing world. While views of tenure are changing, including those of the AAUP, traditionally AAUP has primarily held that "Institutions of higher education are conducted for the common good and not to further the interest of either the individual teacher....[and]..The common good depends upon the free search for truth and its free exposition. Academic freedom in its teaching aspect is fundamental for the protection of the rights of the teacher in teaching and of the student to freedom in learning" (AAUP, 1970).

..... At least one court has ruled, "The purpose of tenure is to protect academic freedom---the freedom to teach and write without fear of retribution for expressing heterodox ideas"(in Copeland and Murry, 1996, p.250).³² Herein lies the point: academic freedom can only be guarded within a system of tenure.³³ Certainly tenure does not guarantee academic freedom. Nevertheless, though tenure is not a sufficient condition, it is a necessary condition for academic freedom.

..... Some tenure system makes it procedurally difficult to fire faculty who may disagree with popular beliefs. There is a long history, continuing to the present, of professors being fired because of their unpopular critiques. The classic examples, of course, include the 1925 trial of John T. Scopes, the Tennessee high school teacher who was fired for teaching the Darwinian theory of evolution. In the 1950s, we saw Senator Joseph McCarthy's communist witch hunts that resulted in professors being fired and/or blacklisted. During the 60's and 70's entire departments were wiped out or put in special trusteeship because of faculty critiques of the Viet Nam war. More recently, "political correctness" about gender, race, and other issues, both on the left and the right, has led to the firing of professors, both tenured and non tenured. So even tenure is not an absolute guarantee for academic freedom.

..... Part of maintaining safeguards to academic freedom is that the university be insulated (not isolated) from the popular culture. To understand tenure, then, it is necessary to understand that the university is a unique organization and in its most important aspects can not be compared or likened to other organizations in society. First, unlike the business corporation, the university should not be consumer oriented (though there are, of course, certain economic realities). The role of the university is leadership, not as a servant of consumer (i.e., student) demands.

..... The university tenure function insures one of the only places in society where open dialogue on any issue no matter how unpopular or unorthodox can be critically examined without consideration of the political cost, without fear of reprisal, without the pressures of social taboos, social norms, faddish movements, personal notions of etiquette, and other immediate pragmatic pressures that exist in the culture in which it is embedded. Above all other roles, this is the defining feature of the university in a democratic society. Without this role, there is no place where truth--- with either a capital "T" or lowercase "t"---can be sought. Once the university becomes indistinguishable from the popular culture, or becomes political, it ceases to be a credible place for the creation and dissemination of knowledge.³⁴

..... If tenure is compromised, we will no longer have a university as we know it. It may look like a university, but it will not in fact be one.³⁵ It will be something else. If we want to play chess without using the King, we can certainly elect to do so, but whatever game we are then playing, let us be clear that it is not chess.³⁶

..... Undergirding this push to eliminate or revamp tenure lies a metaphor that has been transformed into a literal concept: The university as a business entity. The university considered as a business carries with it the attendant and associated ideas of students as consumers in an educational market place, fiscal efficiency, and being a servant of the consumer community in which it is physically embedded. (See Sommer, 1995). The fact is, a university is not like a business. Some courts have recognized inappropriateness of this metaphor. One has noted

(EEOC V. Franklin & Marshall College, 1985)

I do not agree with the majority's assumption that academic institutions are the same as any other employer. At least insofar as their administrative and governance structures are concerned, colleges and universities differ significantly from garden variety private employers. In the context of application of the provisions of the National Labor Relations Act the Supreme Court has counseled that "principles developed for use in the industrial setting cannot be imposed blindly on the academic world." NLRB v. Yeshiva University, 444 U.S. 672, 681, 100 S.Ct. 856, 861, 63 L.Ed.2d 115 (1980) p.120

..... Academic tenure, then, is not like seniority in business, civil service, or a union where the purpose of "tenure" is for the protection of the worker. Unlike these organizations, academic tenure is for the protection of the education function, not individual faculty jobs; it is for the protection of the role of the university in a democratic society. In fact, the metaphor of student as consumer is more appropriately replaced by the metaphor of student as worker or apprentice. Enter SEF.

..... SEF has become the newest weapon in the attack on tenure. But the assault is not a frontal one. SEF is increasingly used in evaluation of faculty for tenure and low evaluation is taken as prima facie evidence of teaching ineffectiveness, thus justifying not awarding tenure to a faculty. A recent legal handbook for college administrators (Weeks, 1996)says, "Student ratings are an effective tool to respond to litigation...[and]...They can become important evidence if a college should decide to terminate a tenured professor for reasons of competence." Of course findings from SEF may indeed validly describe incompetence, but given (a) the conflicting data on their validity, (b) the way many institutions have constructed SEF instruments, (c) the often unsystematic statistical method by which SEF are interpreted, and especially (d) given the considerable weight accorded negative comments by only a few students in making tenure and promotion decisions, it would seem SEF can all too easily be used as a covert instrument for the elimination of tenure candidates and other faculty who may threaten student tuition dollars and perhaps ideological and popular culture agendas.

..... As noted above SEF can be a powerful administrative control mechanism on granting tenure. Abolishing tenure also lowers operating costs. Older faculty who may be unpopular, and who lost more, can be fired and younger, cheaper, faculty can be hired. It is often older faculty who consistently receive lower student evaluations than younger faculty (Feldman, 1983). McMurtry (1991) has noted that, education has always been subject to external pressures whose purpose is to subordinate it to vested interests of various kinds, whether it is slave-holding oligarchies, theocratic states, political parties or the prevailing dogmas of collective beliefs. The difference today, however, is that threats to academic freedom come from within: the consumer student.³⁷ As the well known scholar David Reisman (1981) noted years ago in this regard "This shift from academic merit to student consumerism is one of the two greatest reversals of direction in all the history of American higher education; the other being the replacement of the classical college by the modern university a century ago." (p.xi) Within a consumer model of education, to deny consumer demands is difficult. While it is difficult enough to deal with political, ideological and economic pressures, dealing with consumer pressures has become nearly impossible. If denying fiscal efficiency is viewed as unreasonable, irresponsible, and even irrational, to deny "consumer's" their demands is viewed as undemocratic and downright mean spirited.

..... It might be asked that if tenure is not joined at the hip with academic freedom, why are Supreme Court Justices granted lifetime tenure? It might be asked, too, why it is that virtually no reasonable person suggests revamping academic freedom, yet many have little qualm--if

any---about eliminating or at least about revamping tenure? The short answer is that for many tenure has come not to be seen as crucial to academic freedom. The longer answer is that at least in principle, revamping academic freedom would not be acceptable in a democracy. At worse, it would be viewed as the equivalent of Gestapo busting in the front door of the university (or as a kind of Tiananmen square of higher education). What is not widely understood is that SEF is often a kind of Trojan Horse in the battle against tenure and academic freedom.³⁸ It often becomes a stealth mechanism by which to covertly abrogate both tenure and academic freedom.

SEF and Academic Freedom in the 21st Century

..... As noted above, arguments against tenure have typically been economic ones. As higher education enters the 21st century, and its associated demographic changes, however, arguments against tenure are changing. A paradigm shift is taking place in arguments against academic freedom and tenure, a paradigm that is based in the changing demographics of the student population. It is said that academic freedom tends to be viewed from the perspective of a bygone era when the university faculty and student population were relatively homogeneous. Accordingly, eliminating tenure or at least radically revamping it is increasingly being justified not on matters of principle but by political and other expedient considerations.

..... It is currently suggested that eliminating or revamping tenure would be fairer to minorities, to unemployed, and to part-time faculty, (Murray, 1996; Parini, 1995; Wilson, 1996) presumably by opening up faculty positions for young professors and minorities. Thus critics of tenure maintain that a university going into the 21st century can no longer afford the luxury of the traditional view of academic tenure. While these are valid problems changing the parameters of academic freedom and tenure is a dangerous way to solve them. To the contrary, in a 21st century global society, academic freedom and tenure become more important than it has been in the past, not less. This is why:

..... With increasing diversity in both the general culture and the university campus, including not only diversity among the student body, but diversity among the faculty, comes increasing conflict of ideas, values, and perspectives. To manage these changes, increasing attempts to reduce the conflict by limiting freedom of speech and action in the name of protecting minorities are being implemented. For example, it has been charged (Schrank, 1993) that,

As now defined, academic freedom...ignores the intersubjectivity of all persons in the setting... such an approach conceals the vulnerability of women and other historically excluded groups who are still marginal in the academy, and does not take account of the historic advantages enjoyed by white, heterosexual, able-bodied males...it does not acknowledge power imbalances in relations based on gender, race, sexuality, class, and other dimensions of difference....An exclusive focus on words and ideas can obscure the fact that we are talking about whole people. For many people, ideas and words may well be referents for highly significant experiences which have powerful meanings, evoke strong emotions, and are not simply ideas and words. The meanings and emotions attached to ideas and words render people extremely vulnerable in what, for others, might be a benign intellectual exercise....The argument for including civility recognizes that ideas are not separate, public, and objective but real, emotional, and personal.

..... Certainly, insensitivities in the classroom are to be discouraged. What this quote reflects is a fundamental paradigm shift in the parameters of academic freedom as historically conceptualized. SEF can and do reflect these and other political and cultural conflicts, creating what the courts in other contexts have called a "chilling" effect on academic freedom (Recall the

above emotional evaluation by students described by Allan Dershowitz).

..... This paradigm shift in the parameters of academic freedom, however, is but a subset of a more overarching social shift in the first amendment right to free speech outside of academia. In an extensive article in a noted law review journal (Matsuda, 1989), the author opens with the statement, "This article rejects an absolutist first amendment position." (p.2321) Such views argue for restricting academic freedom on the basis of the feelings of various and changing constituencies. The problem, as Dershowitz (1994) points out, "is that everyone has a different...exception, and in a nation of equal protection, it is difficult to pick and choose among the proffered exceptions. If we were to accept them all, there would be little left of the First Amendment." (p.41) In academia, this means there would be little left of academic freedom and quality education.³⁹

..... Contrary to popular views, tenure and the traditional concept of academic freedom has not simply protected the majority, it has and will protect the newly entering minorities. Indeed, women and other minority studies programs in the university largely owe their very existence to the traditional principle of academic freedom. As Nadine Strossen, (1995) head of the ACLU points out in her well-documented book, restrictions on free speech originally meant to protect minorities often end up being used against them.

..... In reviewing legal cases involving SEF in denial of tenure, it is clear that SEF can be used as an instrument of covert discrimination against minorities. As a legal handbook (Weeks, 1996) for college administrators points out, "Student ratings are an effective tool to respond to litigation involving discrimination, since they provide both qualitative and quantitative data on teaching." The handbook was suggesting that SEF can be used to show that the institution was not discriminating on the basis of gender or race but instead on the basis of SEF. In fact courts have taken the position that even if discrimination can be shown, if it can be shown that the faculty member is "really" being fired because of poor student evaluation of their teaching, discrimination is not relevant. While prime facie this is a reasonable judgement, disentangling gender and race bias from student evaluations is extremely problematic given the way evaluation forms are typically constructed, analyzed, and applied.

..... In conclusion, what this article suggests is that SEF is far from the benign instrument it may once have been in a more homogeneous political, gender, racial, and academically prepared student environment. Unfortunately, on many campuses the traditional model of student and teacher belongs to a past age.⁴⁰ Faculty now teach in a litigious context. The new role and impact of SEF need to be reassessed accordingly.

Summary

..... Validity: As currently used (and perhaps under more stringent conditions as well) the validity of SEF as used in salary, promotion, and tenure decisions, are in question relative to methodology: (1) untrained interpreters using (2) intuitive, and (3) eyeballing methods of analysis to analyze SEF, (4) not controlling for contaminating variables such as (a) level of course, (b) instructors' standards (c) grading practices, (d) subject matter or discipline, (e) personality and (f) interest of student, (g) academic level of student, (h) required course v.s elective course, (i) class size, (j) age, and (k) gender of instructor, and (l) a host of other variables; (5) Validity is also in question regarding student ability to validly render judgements regarding instruction and curricular.

..... Administrative Pressures To Relinquish Faculty Control: (6) SEF is typically required by administrative policy, (7) are therefore involuntary, (8) are used for salary, promotion, and tenure decisions, which allow (8) administrative intrusion into the classroom, (9) create economic incentives for shaping faculty behavior to assure (10) assent to general consumer demands for the

type and level of education, (11) assuring classroom changes in accordance with educational fads, (12) political ideology, (13) to teach in a manner that results in student satisfaction, (14) leading to the retention of students, (15) their tuition dollars, resulting in (16) institutional growth.

..... Control Over Academic Standards: Given the above, pressures are created (17) for faculty to assent and adjust their teaching to whatever level of student the (open enrollment) institution elects to admit, (18) for conforming to student demands in the classroom by lowering standards which include (19) to lower curricular changes, (20) make tests less rigorous, (21) inflate grading, (22) fewer classroom requirements and prerequisites, (23) for easier and less course content, (24) adjusting courses to popular culture belief and notions, (25) sets up a conflict of interest between the instructor and quality of education, all of which result (26) in the opposite of the original intent of SEF which was the improvement of instruction.

..... General Legal Implications: Depending on their use, SEF (27) is often in conflict with usual personnel practices and procedures, (28) when published can lead to defamation of faculty reputation, (29) be discriminatory with regard to age, gender, race, and other variables, (30) involuntarily imposed, (31) not meeting what is considered contractual informed consent of faculty for their use,

..... Academic Freedom and Tenure: SEF can be used to (32) inappropriately dismiss competent faculty, (33) abrogate tenure, and (34) abridge academic freedom.

References

Abrami, P. C., Dickens, W. J., Perry, R. P., & Leventhal, L. (1980). Do teacher standards for assigning grades affect student evaluations of instruction? *Journal of Educational Psychology*, 72, 107-118.

Avi-Itzhak, T., and Lya K. (1986). An investigation into the relationship between university faculty attitudes toward student rating and organizational and background factors. *Educational Research Quarterly*, 10 . 31-38

Benson, D. E. Lewis, J. M. (1994). Students' evaluation of teaching and accountability: implications from the Boyer and the ASA reports. *Teaching Sociology*, 22, 195-99

Blum, D. E. (1990, October 3). U. of Guam removes evaluations from files. *The Chronicle of Higher Education* Section: Personal & Professional, p. A21

Bonetti, S. (1994). On the use of student questionnaires. *Higher Education Review*, 26, 57-64

Bures, A. L., DeRidder J. and Hsin-Min Tong. (1990). An empirical study of accounting faculty evaluation systems. *The Accounting Educators' Journal*. (Summer). 68-76.

Cahn, S. (1987, October 14). Faculty members should be evaluated by their peers, not by their students. *Chronicle of Higher Education*, p. B2

Carey, G. W. (1993). Thoughts on the lesser evil: student evaluations. *Perspectives on Political Science*, 22, 17-20.

Cashin, W. E. (1996). Developing an effective faculty evaluation system. Idea Paper No. 33, Manhattan: Kansas State University, Center For Faculty Evaluation and Development (January)

Cashin, W. E. (1995). Student ratings of teaching: The research revisited. IDEA Paper No. 32.

Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.

Chacko, T. I. (1983). Student ratings of instruction: A function of grading standards. *Educational Research Quarterly*, 8(2), 19-25.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, 51, 281-309.

Copeland, J.D. and Murry, J. W., Jr. (1996). Getting tossed from the ivory tower. *Missouri Law Review*, 61, 233-327

Cohen, P.A. (1983). Comment on "a selective review of the validity of student ratings of teaching." *Journal of Higher Education*, 54, 448-458.

Cotter, William. (1996). Why tenure works. *Academe*. (Jan/Feb) 26-29

Crumbley, L.. (1996). Due process/defamation issues [Not to be construed as legal advice] Society for A Return to Academic Standards. [Online] Available: <http://tamu.edu:8000/~crumble/sfertas.html>

Crumbley, L. (1995). On the dysfunctional atmosphere of higher education: games professors play. *Accounting Perspectives*, 1, (Spring). [Online] Available: <http://tamu.edu:8000/~crumble/sfertas.html>

Crumbley, L. D. and Fliedner, Eugene (1995). Accounting administrators' perceptions of student evaluation of teaching (set) information. Manuscript, Department of Accounting, Texas A&M University.

Damron, J. C. (1996). Instructor personality and the politics of the classroom. Manuscript, Douglas College, New Westminster, British Columbia, Canada V3L 5B2. (An Earlier Versions in the June 1994 Issue of Faculty Matters (No. 5, Pages 9-12) and the September, 1994 Issue of Update (The Newsletter of the Okanagan University College Faculty Association). Douglas College)

Dawes, R. M., Faust, D., and Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243, 1668-1674

Dershowitz, A. (1994). *Contrary to popular opinion*. New York: Berkley Books

Dilts, D. A., Samavati, H., Moghadam, M.R., and Haber, L.J. (1994). Student evaluation of instruction: Objective evidence and decision making. *Journal of individual employment rights*, 2,73- 86.

Divoky, J. J. and Rothermel, A. (1988). Student perception of the relative importance of dimensions of teaching performance across type of class. *Educational Research Quarterly*, 12. 40-45

Dowell, D. A., & Neal, J. A. (1982). A selective view of the validity of student ratings of teaching. *Journal of Higher Education*, 53, 51-62.

DuCette, J. and Kenney, J. (1982). Do grading standards affect student evaluations of teaching? Some new evidence on an old question. *Journal of Educational Psychology*, 74. 308-314.

Education Employment Law News (1994). How big a role should student evaluations play in the assessment of a professor for tenure? (January) pp. 3-4.

EEOC V. Franklin & Marshall College, No. 84-1739, United States Court of Appeals for the Third Circuit, 775 F.2d 110; 3 Fed. R. Serv. 3d (Callaghan) 282; 85 A.L.R. Fed. 669; 39 Fair Empl. Prac. Cas. (Bna) 211; 38 Empl. Prac. Dec. (Cch) P35,644, August 5, 1985, Argued, October 21, 1985, Decided

Evaluating Teacher Evaluations (1996). Article in the Topical Interest Group: Assessment in Higher Education submitted by glen9579@novell@uidaho.edu on 11/12/96. Available: <http://marsquadra.tamu.edu/TIG/GeneralArticles/evaluatingteacherevaluations.html>

Faust, D., Guilmette, T. J., Hart, K., Arkes, H.R., Fishburne, F. J. and Davey, L. (1988). Neuropsychologists' training, experience, and judgement accuracy. *Archives of Clinical Neuropsychology*, 3, 145-163.

Feldman, K. A. (in press). Identifying exemplary teachers and teaching: Evidence from student ratings. In R. P. Perry & J. C. Smart (Eds.), *Effective teaching in higher education: Research and practice*. New York: Agathon Press.

Feldman, K.A. (1993). College students' view of male and female college teachers: Part II--Evidence from students' evaluations of their classroom teachers. *Research In Higher Education*, 34, No. 2. 151-191.

Feldman, K.A. (1983). Seniority and experience of college teachers as related to evaluations they receive from students. *Research in Higher Education*, 18, 3-124.

Franklin, J., & Theall, M. (1990). Communicating student ratings to decision makers: Design for good practice. In Theall, M. & Franklin J. (Eds.), *Student Ratings of Instruction: Issues For Improving Practice*. San Francisco: Jossey-Bass.

Garb, H. N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin*, 105, 387-392.

Gilbaugh, John W. (1982). Renner substantiated. *Phi Delta Kappan*, 63 (6 February) 418.

Gillmore, G. (1984) Student ratings as a factor in faculty employment decisions and periodic review. *Journal of College and University Law*, 10, 557-576 .

Goldberg, G. and J. Callahan. (1991). Objectivity of students' evaluations of instructors. *Journal of Education for Business*, 66, (July/August) 377-378.

Goldman, L. (1993). On the erosion of education and the eroding foundations of teacher education. *Teacher Education Quarterly*, 20, 57-64.

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

Greenwald, A. G. (1996). Applying social psychology to reveal a major (but correctable) flaw in student evaluations of teaching. University of Washington, Draft Manuscript, March 1, 1996.

Greenwald, A. & G. Gillmore. (1996). No pain, no gain? The importance of measuring course workload in student ratings of instructions. University of Washington. Manuscript submitted for

publication.

Guernsey, L. (1996, September 20) Metropolitan State College reviews its tenure policy. *The Chronicle of Higher Education*, p. A13.

Hayes, S. C. (1991). The emperor's clothes: Examining the "delusions" of professional psychology: The healthy skepticism of David Faust. *Scientist Practitioner*, 1, 22-25.

Healy, P. (1996, January 26). Newly appointed trustees stir up public colleges. *The Chronicle of Higher Education*, p. A26.

Heller, S. (1986, December 17). Proper use of student evaluations at issue in professor's legal appeal to regain job. *The Chronicle of Higher Education*, p. 14.

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

Howard, G. S., & Maxwell, S. E. (1982). Do grades contaminate student evaluations of instruction? *Research in Higher Education*, 16, 175-188.

Howard, G. S., & Maxwell, S. E. (1980). Correlation between student satisfaction and grades: A case of mistaken causation? *Journal of Educational Psychology*, 72, 810-820.

Howard, G. S., Conway, C. G., & Maxwell, S. E. (1985). Construct validity of measures of college teaching effectiveness. *Journal of Educational Psychology*, 77, 187-196.

Husbands, C. and Fosh, P. (1993). Student's evaluation of teaching in higher education: Experiences from four European countries and some implication of the practice. *Assessment & Evaluation on Higher Education*, 18, 95-114.

Kemp, B. W. and G. S. Kuman, (1990). Student evaluations: are we using them correctly? *Journal of Education for Business*, 65, (November/December). 106-111.

Kolevzon, M. S. (1981). Grade inflation in higher education: a comparative study. *Research in Higher Education*, 15, 195-212.

Kresl, P. (1995). Grade inflation must be examined before discussed. [Online] Available: <http://www.bucknell.edu/bucknellian/sp95/03-09-95/ops/4002.html> (March 7).

Larkin, J., McDermott, J., Simon, and Simon, H. (1980). Expert and novice performance in solving physics problems. *Science*, 208, 1335-1342.

Leatherman, C. (1996a, September 13). Commission recommends strengthening college presidents' power. *The Chronicle of Higher Education*, p. A43.

Leatherman, Courtney (1996b, October 25). More faculty members question the value of tenure. *The Chronicle of Higher Education*, p. A12.

Magner, D. K. (1996, September 20) Minnesota regents' proposals stir controversy with faculty. *The Chronicle of Higher Education*, p. A11.

Magner, D. K. (1995, May 19). Mid-semester removal of professor roils university of montana. *The Chronicle of Higher Education*, p. A25.

- Marsh, H. W., & Dunkin, M. J. (1992). Students' evaluations of university teaching: A multidimensional perspective. *Higher Education: Handbook of Theory and Research*, 8, 143-233.
- Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76, 707-754.
- Marsh, H. W. (1982). Validity of students' evaluations of college teaching: A multitrait-multimethod analysis. *Journal of Educational Psychology*, 74, 264-279.
- Marsh, H. W. (1980). The influence of student, course, and instructor characteristics on evaluations of university teaching. *American Educational Research Journal*, 17, 219-237.
- Matsuda, M. J. (1989). Public response to racist speech: considering the victim's story. *Michigan Law Review*, 87, 2320-2378.
- McCallum, L.W. (1984). A meta-analysis of course evaluation data and its use in the tenure decision. *Research in Higher Education*, 21, 150-158.
- McKeachie, W. J. (1979). Student ratings of faculty: A reprise. *Academe*, 65, 384-397.
- McMurtry, J. (1991). Education and the market model. *Journal of Philosophy of Education*, 25, 209-217.
- Menand, L. (1996). (Ed.), *The Future of Academic Freedom*. Chicago: University of Chicago Press.
- Menand, L. (1993). The future of academic freedom. *Academe*. (May/June) 11-17.
- Murray, B. (1996). Attacks on tenure lead to new standards: To contain spiraling higher-education costs and be fairer to minority faculty, the tenure system should be reformed, educators say. *American Psychological Association Monitor* (May).
- Nelson, J. P. Lynch, K. A. (1984). Grade inflation, real income, simultaneity, and teaching evaluations. *Journal of Economic Education*, 15, 21-37.
- Parini, J. (1995, July 14). Tenure and the loss of faculty talent. *The Chronicle of Higher Education*, Section: Opinion, POV.
- Pennsylvania State University (1996). Policy On Confidentiality Of Student Records.[Online] Available: [://psuinfo/Policies-Rules for Students/Policies and Rules/Records](http://psuinfo/Policies-Rules%20for%20Students/Policies%20and%20Rules/Records). Also available via gopher from info.psu.edu.
- Pinsker, S. (1989). Teaching in a litigious age. *Change*, 21, (July/August), 50-54.
- Powell, R. W. (1977). Grades, learning, and student evaluation of instruction. *Research in Higher Education*, 7, 193-205.
- Rabinowitz, J. (1993). Diagnostic reasoning and reliability: a review of the literature and a model of decision-making. *The Journal of Mind and Behavior*, 14, 297-316.
- Rebell, M, A. (1990). Legal issues concerning teacher evaluation. In Millman, Jason & Darling-Hammond, Linda, (Eds.), (pp. 337-355) *The new handbook of teacher evaluation*.

Beverly Hills: CA. SAGE.

Reisman, D. (1981). *On higher education: The academic enterprise in an era of rising student consumerism*. San Francisco: Jossey Bass.

Renner, Richard, R. (1981). Comparing professors: how student ratings contribute to the decline in quality of higher education. *Phi Delta Kappan*, 63 2 Oct. 128-30.

Robinson, R. K., and Fink, Ross, L. (1996). Public disclosure of teaching evaluations: privacy and liability considerations. *Journal of Education for Business*. May/June 284-287.

Ryan, J. J., J.A. Anderson, and A.B. Birchler. (1980). Student evaluations: The faculty responds. *Research in Higher Education*, 12 (December) 317-333.

Sacks, P. (1996). *Generation X Goes To College. An Eye Opening Account of Teaching in Postmodern America*. LaSalle, Il: Open Court.

Schrank, B. (1993). Academic freedom and the inclusive university. Paper presented at the Learned Societies Meeting, Carleton University, Ottawa, Ontario, 5 June 1993. [Online] Available: <http://www.caut.ca/swc/afiu.html>

Schrecker, E. W. (1986). *No ivory tower: Mccarthyism & the universities*. Cambridge: Oxford University Press.

Scriven, M. (1995). A unified theory approach to teacher evaluation. *Studies in Educational Evaluation*, 21, 111-129.

Scriven, M. (1993). The validity of student ratings: In teacher evaluation. Evaluation & Development Group.

Scriven, M. (1991). *Evaluation thesaurus*. (4th ed.) Newbury Park, Calif. : SAGE Publications.

Seldin, P. (1993, July 21). When students rate professors. *The Chronicle of Higher Education*. Opinion p. POV.

Seldin, P. (1984) *Changing practices in faculty evaluation*. San Francisco: Jossey-Bass.

Snyder, C. R., & Clair, M. (1976). Effects of expected and obtained grades on teacher evaluation and attribution of performance. *Journal of Educational Psychology*, 68, 75-82.

Sommer, J. W. (Ed.) (1995). *The academy in crisis: The political economy of higher education*. (pp.95-126) London: Transactions Publishers.

Stern, J., and Flynn, P. D. (1995). Students propose a course of action for grade inflation. *The Bucknellian*. (Feb, 20).[Online] Available: <http://www.bucknell.edu/bucknellian/sp95/03-02-95/ops/4165.html>

Stone, J. E. (1995). Inflated grades, inflated enrollment, and inflated budgets: an analysis and call for review at the state level. *Education Policy Analysis Archives*, 3, Number 11 (2075) [Online] Available: <http://olam.ed.asu.edu/epaa/v3n11.html>

Strossen, N. (1995). *Defending pornography: Free speech, sex and the fight for womens's rights*. New York: Anchor/Doubleday.

Tagomori, H. T. Bishop, L. A. 1995. Student evaluation of teaching: Flaws in the instruments. *Thought & Action*, 11, 63-78.

The Chronicle of Higher Education. (1995, March 31). South Carolina lawmaker would abolish tenure, p. A24.

The Chronicle of Higher Education (1994a, April 27). State calls for release of faculty evaluations, p. A4.

The Chronicle of Higher Education. (1994b, December 14). Faculty evaluations to go on line at U. of Texas at Austin, p. A21.

University of Michigan (1994). Toward a definition of tenure. Senate Advisory Committee on University Affairs Standing Subcommittee on Tenure Endorsed Unanimously by the University of Michigan Senate Assembly, December 12.[Online] Available:
<http://www.umich.edu/~aaupum/tenure05.htm>

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

Weeks, K. M. (1996). *Managing departments: Chairpersons and the law*. Nashville, Tennessee: College Legal Information, Inc.

Wilson, R. (1996, June 7) Scholars off the tenure track wonder if they'll ever get on. *The Chronicle of Higher Education*, p. A12.

Worthington, A. G., & Wong, P. T. P. (1979). Effects of earned and assigned grades on student evaluations of an instructor. *Journal of Educational Psychology*, 71, 764-775.

Young, R. D. (1993). Student evaluation of faculty: a faculty perspective. *Perspectives on Political Science*, 22, 12-16.

Yunker, B.J. and Sterner, J (1988). A survey of faculty performance evaluation in accounting. *Accounting Educators' Journal*, (Fall). 63-74.

Zirkel, P. A. (1996) *The law of teacher evaluation*. Bloomington, IN: Phi Delta Kappa Educational Foundation.

Notes:

1. Address correspondence and notification of commentary to: Robert E. Haskell, Ph.D. Professor of Psychology, Department of Social and Behavioral Sciences, University of New England, Biddeford, Me. 04005. Email haskellr@maine.rr.com. I would like to thank Larry Crumbly of Louisiana State University for providing me with some of the sources cited in this article. I would especially like to thank professors John Damron, of Douglas College, and John Stone, of the Andrew Jackson Institute, for sources and for reading an early draft of this manuscript.

2. The closest and only allusion to SEF is found in Chapter Four by Cass R. Sunstein. In discussing the formal and informal regulation of speech (i.e., certain viewpoints) the author notes, "the evaluation of students and colleagues cannot occur without resort to content, and it

would be most surprising if viewpoint discrimination did not affect many evaluations." (p.106)

3. Since the issue of SEF validity, in terms of learning, is so central a few observations of the literature is necessary. Greenwald and Gillmore (1996) have categorized some of the significant reviews and empirical research that find in favor of validity of SEF as measures of quality of instruction, for example, Cashin (1995), Cohen (1981), Feldman (in press), Howard, Conway, and Maxwell (1985), Howard and Maxwell (1980, 1982), Marsh (1980, 1982, 1984), Marsh and Dunkin (1992), and McKeachie (1979). Reviews and empirical critiques that are critical of the validity of SEF include, Chacko (1983), Dowell and Neal (1982), Holmes (1972), Powell (1977), Snyder and Clair (1976), Vasta and Sarmiento (1979), and Worthington and Wong (1979). Positions, suggesting cautious support for validity of SEF while at the same time expressing concerns about the adequacy of their support, include, Abrami, Dickens, Perry, & Leventhal (1980). The recent methodologically sophisticated research of Greenwald (1996), and Greenwald and Gillmore (1996) find strong evidence inconsistent with the common dismissive interpretation of the relationship between SEF and high student grades as reflecting a relationship between amount learned and student ratings.

4. Cahn (1987) suggests that student ratings do not measure the instructional effectiveness or the intellectual achievement of students. SEF measure student satisfaction, attitudes toward instructors course, student personality, and the psychosocial needs of the student. Cahn suggests, students know if instructors are likeable, not if they are knowledgeable; they know if lectures are enjoyable, not if they are reliable. In a meta-analysis Cohen (1983) concludes from his study, "While the magnitude of the average rating/achievement correlation for the thirty-three multisection courses is not overwhelming [14.4% of shared variance between ratings and the criterion], the relationship is certainly stronger and more consistent than we were led to believe..." (p. 455). Dowell & Neal (1982) conclude that "The research literature can be seen as yielding unimpressive estimates of the validity of student ratings. The literature does not support claims that the validity of student ratings is a consistent quantity across situations. Rather, the evidence suggests that the validity of student ratings is modest at best and quite variable (p. 59). "The variability in obtained validity coefficients even in studies with reasonable methodological requirements. . . lead us to suspect that the validity of student ratings is influenced by situational factors to such an extent that a meaningful, generalizable estimate of their validity does not exist. In general . . .no meaningful estimate of the validity of student ratings can be provided with confidence that is generalizable enough to be useful..." (60-61)

5. For example, studies demonstrate the following variables: Age, gender, class size, year of student, level of student, instructor style, subject matter, major or elective course, student interest in subject matter, instructor grading difficulty, anonymous v.s signed ratings, whether students are informed of their use, instructor present v.s instructor absent while completing the evaluation (see for example, Divoky and Rothermel, 1988).

6. At least one study found that 70% of students "indicated that the grade they thought they would get influenced the level at which they rated their professors" (in Goldman, 1985). The grade inflation literature shows that the typical percentage of A's and B's on many campuses is anywhere from 70-90%, and that the percentage of graduates who are graduating with cum laude and above is somewhere in the range of 50-70%.

7. If used correctly (see Copeland and Murry, 1996; Kemp and Kuman, 1990; Scriven, 1995, 1993, 1991; Seldin, 1984), SEF can be very useful instructionally, and when used in conjunction with other methodologically sound evaluation procedures and criteria, it can assist in informing an institution when a faculty does not pass muster as an effective teacher. The question, of course

is: does student feedback to faculty result in improved teaching and student learning. In a review of studies, Marsh (1984) suggests that there is a small positive correlation for improved student learning of SEF feedback to faculty if used in a carefully constructed collegial consultation process. They are seldom used in this manner.

8. Bona Fides: I am a full professor, so I'm not personally concerned with SEF effect on promotion, tenure, and performance evaluations. I typically receive a 3.4 or 5 on a 4. scale on my evaluations, and in the classroom I do a fairly good dog and pony show as well.

9. Other countries using SEF questionnaires seem not as concerned with research on their validity as is the U.S., there is apparently almost no published technical literature on validity and bias in SEF in German universities. There is however, a debate in German higher education as to whether SEF is an invasion of academic autonomy (Husbands and Fosh, 1993).

10. While faculty are entitled to freedom of discussion and inquiry in their classroom, it is a generally recognized limitation that they should not introduce controversial matter which has no relation to their subject. What constitutes "controversial" and "no relation," however, often remains an open question.

11. AAUP Committee C on College and University Teaching, Research, and Publication. It was adopted by the Council of the American Association of University Professors in June 1975 and endorsed by the Sixty-first Annual Meeting as Association policy. They also state: "An important and often overlooked element of evaluating teaching is an accurate description of a professor's teaching. Such a description should include the number and level and kinds of classes taught, the numbers of students, and out-of-class activities related to teaching. Such data should be very carefully considered both to guard against drawing unwarranted conclusions and to increase the possibilities of fairly comparing workloads and kinds of teaching, of clarifying expectations, and of identifying particulars of minimum and maximum performance. Other useful information might include evidence of the ability of a teacher to shape new courses, to reach different levels and kinds of students, to develop effective teaching strategies, and to contribute to the effectiveness of the individual's and the institution's instruction in other ways than in the classroom.

12. Kolevzon found an interesting breakdown of these responses focused upon the finding that faculty with fewer years of teaching experience at the university were more impressed with the importance of the role played by student evaluations when compared with senior faculty. p.208 Faculty with at least 6 years of teaching experience were more strongly supportive of the position that student evaluation forms were responsible for lenient grading. Avi-Itzhak and Lya found---perhaps counterintuitively---that senior faculty were the most opposed to SEF being used for salary, promotion and tenure.

13. It should go without saying, that not all students are the same. SEF vary by maturity, intellectual level, i.e., graduate student evaluations v. undergraduate (See, Divoky and Rothermel, 1988; Dilts, Samavati, Moghadam, and Haber, 1993) and therefore probably by campus and program. This may account in part for some of the wide variation in faculty attitudes toward SEF.

14. Other studies also show that inflating grades leads to higher SEF (See, DuCette and Kenney, 1982; Goldberg and Callahan, 1991; Kemp and Kuman, 1990).

15. Given that SEF serve a dual function of both promotion termination, another area of possible

counterproductiveness is the inherent conflict in the necessary cooperation for faculty development purposes; SEF used for such purposes is thus in conflict with faculty self-protection. Unlike in the past, now that SEF are an increasingly important instrument at many institutions, its underlying problems become increasingly evident.

16. Last semester, says Dershowitz, "a small group of students complained about my teaching...[subject]... from a civil liberties perspective." I responded that it was important for the students to hear a variety of perspectives about...[subject]..., just as they hear, without objection, about other crimes. I also reminded them that the majority of students who speak in class present the "politically correct" views. I told them that the answer to an offensive argument is not to censor but rather to come up with a better argument. A few days later, one of the students told me that I should expect to be "savaged" in the evaluations that each teacher receives from the students. (*italics and brackets added, p.118-119*)

17. I use these examples as merely illustrations of the role of SEF, making no attempt to assess validity or non validity of the examples.

18. I use this case because I have a certain degree of prima facie confidence in the instructor's view. I had a similar experience teaching a required abnormal psychology course to a group of vocational majors (Occupational Therapy). I have been teaching for twenty years, and received good student evaluations and have earned tenure twice at institutions that rely heavily on SEF. I recently experienced my first student "revolt." The students were under pressure not to get two D's. My course was the last course before they were formally admitted into the major. I received no support from administration, including the department chair. Moreover some of the vocational faculty (who incidently are not doctoral level faculty), making it known to the students, complained to administration about my course. For interesting similar personal accounts, See Peter Sacks (1996). *Generation X Goes To College. An Eye Opening Account of Teaching in Postmodern America. LaSalle, Il: Open Court*

19. I have contacted a number of legal scholars who have written compendia of legal cases regarding the denial of tenure and promotion. Few were able to refer me to cases involving SEF. I would like to express my appreciation to Professor William A. Kaplin, School of Law, Catholic University of America, and to Michael Rebell, of Rebell and Katzive, New York city.

20. I wish to than Dr. John Damron of Douglas College, for this reference.

21. Scriven has done considerable rigorous work on evaluation procedures, particularly on the justification inferring from ratings to conclusions about the merit of teaching on the basis of statistical correlations between ratings and student learning gains. He suggests that such inferences are invalid, unless a number of stringent conditions are met on the design, administration, and use of such rating forms.

22. Litigation resulting from invalid data in Great Britain has occurred. As Husbands and Frosh (1993) note, "As far as we know, there have been no comparable cases in European courts but, if European universities follow the American example of using student evaluations largely or exclusively for summative purposes, it is only a matter of time before there is external examination of the techniques being used, and of their suitability for the purposes for which they are intended." p.103. They conclude, "it is a sad commentary on the gullibility of some people in the face of numerical data that it required the intervention of the courts to force the discontinuation of the more gross forms of this type of interpretation" p. 103.

23. Many faculty are not aware of the extent of the confidentiality of student information. For example, (1) student scores or grades can not be posted publicly by name, social security numbers, or any other identifier that can be known by anyone but the instructor and student, (2) student papers or lab reports that have student names and grades can not be left in places that are accessible to others, (3) other students may not have revealed to others in a class, (4) faculty are not to request student information without a legitimate educational reason, (5) student grades or other educational information may not be shared with other faculty unless the faculty has a specific legitimate reason to know, (6) libraries are apparently prohibited from revealing to instructors class reading material s/he has specifically put on reserve in the library for students to read to see if they have read the material, (7) student grades or other educational information can not even be revealed to the parent of the student (who may be paying for the student's education) without written permission of the student. There are many other restrictions as well.

24. According to Black's Law Dictionary (With Pronunciations Sixth Edition (1990). St. Paul, Minn.: West Publishing Co.) "Libel. A method of defamation expressed by print, writing, pictures, or signs. In its most general sense, any publication that is injurious to the reputation of another. A false and unprivileged publication in writing of defamatory material. *Bright v. Los Angeles Unified School Dist.*, 51 Cal.App.3d 852, 124 Cal.Rptr. 598,694. "A maliciously written or printed publication which tends to blacken a person's reputation or to expose him to public hatred, contempt, or ridicule, or to injure him in his business or profession. *Corabi v. Curtis Pub. Co.*, 441 Pa. 432, 273 A.2d 899,904. "Accusation in writing or printing against the character of a person which affects his reputation, in that it tends to hold him up to ridicule, contempt, shame, disgrace, or obloquy, to degrade him in the estimation of the community, to induce evil opinion of him in the minds of right thinking persons, to make him an object of reproach, to diminish his respectability or abridge his comforts, to change his position in society for the worse, to dishonor or discredit him in the estimation of the public, or his friends and acquaintances, or to deprive him of friendly intercourse in society, or cause him to be shunned or avoided, or where it is charged that one has violated his public duty as a public officer. Almost any language which upon its face has a natural tendency to injure a man's reputation, either generally or with respect to his occupation. *Washer v. Bank of America Nat. Trust & Savings Ass'n*, 21 Cal.2d 822, 136 P.2d 297, 300."

25. According to Rebell (1990), Florida specifically mandates the school board to provide professional training programs to "ensure that all individuals with evaluation responsibilities understand the proper use of the assessment criteria and procedures" (Fla. Educ. Code, /sec 231.29(2)). p.345. Apparently, however, such mandates are not applied to higher educational institutions.

26. Dowell & Neal (1982) suggest that "The attraction of student ratings to higher education officials lies in the seductive way they seem to reduce a complex human activity, teaching, to simple numbers. Administrative over reliance upon ratings may encourage faculty to "perform to criterion," to teach in such a way as to accrue good ratings, even though this may not result in optimal educational practice. Ratings cannot substitute adequately for more informed judgment about teaching effectiveness, which might be obtainable from reviewing course materials, visiting classes, and discussing teaching practices with faculty. Unfortunately, these detailed qualitative methods are currently regarded as too expensive and difficult to use in most institutions. Evidence presented in [Cohen's 1983] review reinforces our earlier conclusion that student ratings are inaccurate indicators of student learning and therefore are best regarded as indices of "consumer satisfaction" rather than teaching effectiveness" (p.60-61)

27. Recently, a group of college and university presidents have proposed that faculty input into

decisions be reduced. The commission's 50-page report, "Renewing the Academic Presidency: Stronger Leadership for Tougher Times," makes recommendations to presidents, professors, trustees, and public officials. Copies of the report, available for \$12.95 for members of A.G.B. and \$19.95 for non-members, may be obtained from the Association of Governing Boards of Universities and Colleges, One Dupont Circle, Suite 400, Washington 20036; (202) 296-8400. From: Leatherman, (1996a)

28. It is perhaps ironic in a democracy that there should be so much scrutiny of faculty behavior in the classroom. One faculty who has taught and lived in totalitarian states observed (in Sacks, 1996), " I grew up under a totalitarian government and worked as a teacher there, and never did anybody come to visit my classroom to make sure I was toeing the line in that sense...there was more "oppression" of teachers the United States, a country that boasts the ideals of freedom speech and thought, and that there was more pressure to conform to "acceptable" ideas in U.S. classrooms." (p.33)

29. See endnote # 12

30. For purposes of this paper I consider the mandated use or non use of SEF for salary, promotion and tenure by union contract agreements a special case of the "voluntary" nature of SEF. Some union contracts prohibit their use in salary, promotion and tenure decisions. Other require that students sign their evaluations.

31. There is all too frequently negative administrative support for faculty when students complain. For years I've seen faculty struggle with this issue. In the end, they usually conform to it--at least to some degree. When I informed one administrator that I was going to change my grading system from my already drastically modified Bell curve to to the campus norm of giving mostly A's and B's, because I just could no longer "compete" for students in my courses, he soberly looked me straight in the eye and firmly said, "That's a good idea."

32. See, *Grimes v. Eastern Ill. Univ.*, 710 F.2d 386, 388 (7th Cir. 1983). Cited in, Copeland and Murry, (1996).

33. A recent article on the nation-wide attacks on tenure, repeatedly referred to it as "job security" for professors. It was noted that American Association for Higher Education (AAHE) researchers found that tenure serves an important purpose by "rewarding dedicated, productive professors." Other articles maintain that a tenure system helps to recruit and retain quality faculty. And a representative of the American Association of Psychology Professors, argued that "tenure is only awarded to faculty who conduct quality research and teaching" (Murray, 1996) For similar justifications see Leatherman (1996b). After talking about job security in Business and in the military, both of which have nearly vanished, one faculty said "I can't say there's some obvious reason why higher education should be immune to it"(in Leatherman,1996b). As a consequence, there seems to be a merging of the popular culture and the academic views of tenure, with the direction of flow from the popular culture to the academic. Arguments in defense of tenure increasing exhibit a tone of, "...and, oh! by the way, it's to help insure academic freedom as well." faculty could be given five and ten-year contracts. While this may address the irrelevant issue of job security in some measure, it neither solves the problem, nor addresses the issue of academic freedom. This is why it doesn't: First--and this seems to be not widely recognized--it merely postpones being "dehired." Second, it exacerbates the job security issue. By the time a faculty has been at an institution for ten years, s/he is either an associate professor, or is at a salary range that reduces the chances of finding another position. Increasingly, with institutions cutting costs, most positions are entry level ones. So variable length contracts just

provide an additional cost cutting mechanism for eliminating senior ---or soon to be senior---faculty. And third, as a consequence, such contracts tend to put a damper on faculty feeling free to examine and critique issues that are unpopular. With such contracts, the tenure function does not in fact exist.

34. I am not an uncompromising academic isolationist advocating the development of a kind of Monroe Doctrine applied to higher education. I see the value of some changes that have taken place in the university which could probably only have occurred as the result of outside pressures. But such changes should be made with extreme caution, reason, logic, and evidence. Most importantly, any changes should be made within the boundaries of what a university needs to be in order to perform its role in society. To tamper with tenure is to clearly cross one of those boundaries.

35. This is not intended as a blanket apologia for academia. There are many problems within the academy. In many other areas, I am a severe critic of my colleague's collective behavior.

36. Increasingly, and perhaps understandably, junior faculty are questioning the system of tenure. After all, not only do they not have it, but it is becoming increasingly difficult to attain, politically and in many other ways. A recent survey by the Higher Education Research Institute at the University of California at Los Angeles shows that more than a third of some 34,000 faculty either strongly or somewhat agreed that tenure is an outmoded concept (an increase from 1989), seeing it as "elitist," a "holdover," and a tiered system of "haves and have nots" (in Leatherman, 1996b). But it seems that even older faculty are beginning to question tenure. The foundations of such reasoning, it seems to me, is very egocentric, shortsighted, and dangerous to the very concept of a university in a democratic society.

37. McMurtry delineates the inappropriate use of the consumer metaphor in education. "The best product on the market, as we know, is the one which is the most 'problem-free' for its purchaser--delivered ready made for instant easy use', 'guaranteed replacement' if it does not work, and 'repaired cost-free' whenever it needs maintenance attention. The best education, on the other hand, is the opposite on all standards of excellence. It cannot be produced or delivered by another at all, is never ready-made nor instant, and cannot be guaranteed replacement or service cost-free if it is not working. The higher the standards it has, the less it can be immediate in yield, the more work it demands of its owner, and the more its failures must be overcome by its possessor's own work. An education can never be 'problem free', and poses ever deeper and wider problems the higher the level of excellence it achieves. Freedom in the market is the enjoyment of whatever one is able to buy from others with no questions asked, and profit from whatever one is able to sell to others with no requirement to answer to anyone else. Freedom in the place of education, on the other hand, is precisely the freedom to question, and to seek answers, whether it offends people's self-gratification or not....What is the best policy for buying a product--to assert the customer's claim 'as always right'--is the worst possible policy for a learner. What is the best policy for selling a product--to offend no-one and no vested interest--ma): be the worst possible policy for an educator. The principles of freedom here are contradictory, and become the more so the more each is realized." p.213-214

38. Extending the ancient Greek account, in computer science a Trojan Horse refers to a set of instructions hidden within a legitimate program, causing a computer to perform illegitimate functions.

39. A recent edited book on *The Future of Academic* by Louis Menand (1996) provides an extensive series of closely reasoned chapters addressing the multicultural-based attacks on the

First Amendment. See especially Henry Louis Gates, Jr.'s "Critical Race Theory and Freedom of Speech."

40. See, for example Pinsker, (1989).

About the Author

Robert E. Haskell

Brief Bio ~ Robert E. Haskell has been teaching college and university level courses for over twenty years. He earned his Ph.D. from the Pennsylvania State University in Psychology and Social Relations, his M.A., and B.A. from San Francisco State University. His areas of research and teaching include: transfer of learning, analogical reasoning, small group dynamics. Major publications include: four books, the latest of which is, *The Future of Education and Transfer of Learning: A Cognitive Theory of Learning and Instruction For The 21st Century* (forthcoming), and numerous presentations, chapters, and research articles in national and international journals. He also serves on several editorial review boards, and is Associate Editor of The Journal of Mind and Behavior. He is former Chair, and currently Professor of Psychology, Department of Social and Behavioral Sciences, University of New England.

Professor of Psychology
University of New England
Biddeford, Maine 04005

UNE Home Page: <http://home.maine.rr.com/une/>

E-mail: haskellr@maine.rr.com

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is
<http://olam.ed.asu.edu/epaa>

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

EPAA Editorial Board

[Michael W. Apple](#)

University of Wisconsin

[John Covaleskie](#)

Northern Michigan University

[Alan Davis](#)

University of Colorado, Denver

[Greg Camilli](#)

Rutgers University

[Andrew Coulson](#)

a_coulson@msn.com

[Sherman Dorn](#)

University of South Florida

[Mark E. Fetler](#)
California Commission on Teacher Credentialing

[Thomas F. Green](#)
Syracuse University

[Arlen Gullickson](#)
Western Michigan University

[Aimee Howley](#)
Marshall University

[William Hunter](#)
University of Calgary

[Daniel Kallós](#)
Umeå University

[Thomas Mauhs-Pugh](#)
Rocky Mountain College

[William McInerney](#)
Purdue University

[Les McLean](#)
University of Toronto

[Anne L. Pemberton](#)
apembert@pen.k12.va.us

[Richard C. Richardson](#)
Arizona State University

[Dennis Sayers](#)
University of California at Davis

[Michael Scriven](#)
scriven@aol.com

[Robert Stonehill](#)
U.S. Department of Education

[Richard Garlikov](#)
hmwkhlp@scott.net

[Alison I. Griffith](#)
York University

[Ernest R. House](#)
University of Colorado

[Craig B. Howley](#)
Appalachia Educational Laboratory

[Richard M. Jaeger](#)
University of North
Carolina--Greensboro

[Benjamin Levin](#)
University of Manitoba

[Dewayne Matthews](#)
Western Interstate Commission for Higher
Education

[Mary P. McKeown](#)
Arizona Board of Regents

[Susan Bobbitt Nolen](#)
University of Washington

[Hugh G. Petrie](#)
SUNY Buffalo

[Anthony G. Rud Jr.](#)
Purdue University

[Jay D. Scribner](#)
University of Texas at Austin

[Robert E. Stake](#)
University of Illinois--UC

[Robert T. Stout](#)
Arizona State University

Contributed Commentary on
Volume 5 Number 6: Haskell *Academic Freedom, Tenure, and Student Evaluation of Faculty*
and
Volume 5 Number 8: Stake *Response to Haskell's "Academic Freedom ... & Student Evaluation"*

**On Michael Theall's (and implied et al.)
"A Reply to Haskell and to Stake"**

Robert E. Haskell

In "On Drawing Reasonable Conclusions about Student Ratings of Instruction: a Reply to Haskell and to Stake," Michael Theall (1997), Director of the Center for Teaching and Learning at the University of Illinois at Springfield, opens his critique of my article (Haskell, 1997a) by lamentably describing my piece as rhetorical. He then goes on to characterize my article: as an example of (1) faculty who "fulminate" against SEF, as (2) "simplistic," (3) "loaded with misinterpretations of the literature," as (4) "mythology," (5) exhibiting an "ignorance of evaluation/measurement literature," as containing (6) "sweeping generalizations," (7) "misinformation," as (8) "simply ridiculous!" (9) "ripe" with "hysterical rhetoric," as (10) assuming a "mythical group of better students" of some bygone era, asserting that (11) SEF "are the cause of grade inflation," as suggesting (12) "we do away with ratings." Continuing, Theall wrote that (13) "Perhaps the weakest part of his article is what isn't there: constructive suggestions for improvement," that I (14) "suggested that ratings are a violation of academic freedom," and finally---but not exhaustively---(15) that "Academic freedom has been defined in many ways, but never before in a way that suggests the construct (tradition? principle? tenet?) is vulnerable to the influence of student ratings." The last two items, at least, are correct. With this said, let me say I am perplexed by Theall's response.

I am perplexed for three reasons. First, because of what my article actually clearly said and did not say, second, because of Theall's own findings, and third because of the tone. Let me begin with what my article actually clearly said and did not say, and Theall's reaction to it. This will lead naturally into his own findings which, ironically, seem to support much of my thesis.

Theall's Inaccuracies and Misinterpretations

In anticipation of responses like Theall's, I clearly laid out the boundaries of my article quite carefully. Evidently, he either did not notice these boundaries or for some reason elected to ignore them. I will start with the more general accusations and proceed to the more specific, both of which, as I will document, are filled with inaccuracies. He says of my article, among other things that "Perhaps the weakest part of his article is what isn't there: constructive suggestions for improvement." I had indicated that I would not be addressing this issue, nor dealing in detail with findings on validity of SEF, or with grade inflation. I clearly indicated the boundaries: I said

In explicating SEF, many closely related issues must be substantially bracketed. These related issues include (1) its validity (Cahn, 1987; Damron, 1996; Greenwald, 1996; Greenwald and Gillmore, 1996; Scriven, 1993; Seldin, 1984; Tagomori, Bishop, and Laurence, 1995), (2) the problem of defining teaching effectiveness, (3) general variables affecting SEF scores, (4) alternatives to SEF's (5) classroom politically correct or popular standards and perceptions, (6) low student academic preparation, (7) age and gender discrimination issues (Feldman, 1983, 1993) (8) strategies for change, and (9) other integrally related issues such as their being largely responsible for lowered course standards, and grade inflation. Though these are important and related issues, they can only be addressed here in so far as they directly impact the focus on SEF and academic freedom.

Despite these clearly stated boundaries, and despite the fact of providing example citations to many of the areas like validity that I could not cover---I refer the reader to endnotes #3, 4, 5, 6 and 14 in my paper. Disregarding these limits, Theall proceeded to critique my article by concentrating on (a) validity and (b) grade inflation, exactly those areas that I said would have to be bracketed. I refused to be pulled into these two areas in my paper, and I refuse to be hauled into them now.

Both validity and grade inflation are too important to be briefly dealt with within the confines of this response by simply offering, as did Theall, a citation here and a citation there. Such an ad hoc approach to these important issues is little more than a statistical analogue of using verbal anecdotes to resolve the issue. In a more appropriate forum, if Theall would like to come up with something a little more systematic than the statistical anecdotes he offered, it might be worth responding. (Note 1) In any event, Theall refused to substantially address the main issue of my article: academic freedom

and SEF .

I will now address some of the undocumented specific inaccuracies of Theall's critique. From numerous places in his response, he says,

Robert Haskell suggests that we do away with ratings....The thrust is simply that we should do away with student ratings.... I am targeting the use (perhaps, misuse) of these issues in efforts to discredit or do away with faculty evaluation practices in general and student ratings in particular....building his case that ratings were an essentially unreliable form of data that should be done away with.

I did not say that SEF should be done away with. In fact in (a) my original article and (b) my response to Stake, (Haskell, 1997b) both of which Theall evidently read, I clearly stated:

It is important to note at the outset, that it is not SEF per se that is the issue, but the impact of its use on salary, promotion, tenure decisions, and its impact on the delivery of quality education.

In addition in endnote #6 I said, again:

If used correctly (see Copeland and Murry, 1996; Kemp and Kuman, 1990; Scriven, 1995, 1993, 1991; Seldin, 1984), SEF can be very useful instructionally, and when used in conjunction with other methodologically sound evaluation procedures and criteria, it can assist in informing an institution when a faculty does not pass muster as an effective teacher.

Let me make it clear once again: I am not against SEF , only their use for administrative purposes. As I said in my paper, while it is true that

SEF may indeed validly describe incompetence...given (a) the conflicting data on their validity, (b) the way many institutions have constructed SEF instruments, (c) the often unsystematic statistical method by which SEF are interpreted, and especially (d) given the considerable weight accorded negative comments by only a few students in making tenure and promotion decisions, it would seem SEF can all too easily be used as a covert instrument for the elimination of tenure candidates and other faculty who may threaten student tuition dollars and perhaps ideological and popular culture agendas.

Note that I said "conflicting data on their validity," not that the literature demonstrates conclusively that they are completely invalid.

Given the number of instances on which Theall used the word "practical," I would think, given the time and resource intensive requirements for rendering SEF use valid, he would appreciate the fact that the likelihood of its appropriate use is not high. (In fact he does. See below). Theall , further claims that I make causal attributions from SEF to grade inflation. He wrote,

Lately, there has been a lot of discussion attempting to causally link student ratings to problems such as grade inflation'.... Ratings" become the cause of the downfall of higher education.... Robert Haskell (1997) suggested that ratings are.... the cause of grade inflation)....it is very risky to point to one reason for these changes.

Nowhere in my article was it either suggested that SEF is "the cause" or was the term "cause" used or such an attribution made about SEF. I did suggest that SEF is

largely responsible for lowered course standards, and grade inflation....are responsible for a considerable amount of grade inflation.

As scholars and rigorous researchers will likely agree there is a considerable difference between attributing a single causal connection and suggesting they contribute to grade inflation. I guess the statistical concept of variance was conveniently forgotten.

Theall claims that I "blame" students. He wrote, "Don't blame students." In anticipating such a reading of my article, I stated in the opening that to question SEF is often

seen as not only a novel idea, but as an attack on either students, or a general attack on evaluating faculty.

I evidently anticipated this claim correctly. In further anticipation of such a reading of my article, in endnote #12 I said,

It should go without saying, that not all students are the same. SEF vary by maturity, intellectual level, i.e., graduate student evaluations v. undergraduate.

Then later in the text I stated,

Many students understand the above described ensuing consequences. A glance at articles from online student newspapers reveals strong sentiments against what many students consider the erosion of standards created by SEF . One student writer went so far as to say "We therefore suggest a boycott of the 1995 student/teacher evaluations. This boycott will provide a more effective means of communication than anything written on the evaluation itself. Something must be done about the trend of grade inflation. We as students refuse to contribute to the downfall of academia (Stern and Flynn, 1995) Some students are thus quite aware of the effects of SEF on their education.

Theall then goes on to claim that if there is grade inflation that one of the parties we should blame is faculty; that we should "Blame the perpetrators." There is a section in my paper entitled, "Faculty Complicity and Adaptation to SEF," in which I analyze this problem, but he does not mention any of the arguments I advance. I come to quite a different conclusion, however, than does Theall .

To my amazement, Theall simply notes, almost in passing, that I suggest that SEF "somehow violate the great traditions of academic life." What does he mean, somehow? The entire article was a fairly detailed "let me count the ways"--a listing, explaining, and accounting of just how I suggest SEF do violate academic freedom, both in the abstract and with concrete examples. So when he simply says of my article that "Academic freedom has been defined in many ways, but never before in a way that suggests the construct (tradition? principle? tenet?) is vulnerable to the influence of student ratings," I would suggest that Theall, tell us more about SEF and academic freedom, since this was the main thesis of the article.

I would now like to note very briefly and comment on some other claims made by Theall . My comments will be brief because, again, I dealt with them in my paper. First, he says that students "are the appropriate providers of data," mainly because they are paying "tuition dollars." I addressed these issues in my paper in discussing the misplaced (a) political metaphor of democracy applied to instruction, and (b) corporate metaphor of student as consumer.

Theall also claims that

ratings are collected before students get their final grades and thus, their opinions must be based on expectations which are, in turn, based on performance to date. Ratings thus can not be said to reflect a disconfirmed expectancy about the overall course outcome (i.e., the course grade).This would only be the case if ratings were gathered after final grades were distributed and the final grade disconfirmed what was expected as a result of the experiences during the semester. So the ratings relationship is limited to experiences and results during the term rather than to the final grade.

In many if not most cases Theall's claim is incorrect. This is why: (a) In many, and maybe most, courses students receive test grades and the class curve/distribution (or at least should) throughout the semester, (b) the student grapevine is quite efficient in informing students what instructors grading practices are. So ratings thus can be said to reflect a disconfirmed expectancy about the overall course outcome.

Theall at least did make one technically valid set of points. Assuming SEF are on their face valid and reliable because there is "little variance across responses", says Theall ,

The question now becomes one which deals with: 1) the appropriateness of course content; 2) the standards used for grading; and 3) the question of whether lenient coverage and grading were deliberately chosen in order to influence ratings. The first two items can be and should be dealt with via curricular mechanisms such as departmental review of courses and content, and faculty agreement on standards for student work.

Granting for the moment SEF validity and reliability (I probably would even really accept their reliability), I would like to address 1) and 2) in the above quote (I addressed the third above). On a collective level, I agree with Theall that faculty have not been responsible in developing and enforcing standards. In fact in endnote #34 I wrote:

This is not intended as a blanket apologia for academia. There are many problems within the academy. In many other areas, I am a severe critic of my colleague's collective behavior.

Theall's Work Supports Much of My Article

I noted at the beginning of this response that I was perplexed at much of Theall's reaction. I am particularly perplexed at his response to a passage I cited from his work. He says,

Haskell included a long quote from a chapter Jennifer Franklin and I (1990) did in our *New Directions for Teaching and Learning* issue #43 (Theall & Franklin, 1990b) While our point was that ratings practice must be improved, Haskell used the quote to supplement other citations as evidence in building his case that ratings were an essentially unreliable form of data that should be done away with ... a gross misinterpretation of our intent.

I am perplexed for two reasons. First, it is really neither here nor there that the quote apparently does not reflect Theall's (et

al.) intent. Whatever an author's intent was in using a quote does not necessarily reflect on other uses of the material, especially when the quote is not used to suggest an author's position.

Second, I am perplexed because much of Theall's chapter revolves around other material that he suggests demonstrate some of the severe problems in using SEF as they are currently used. The full quote to which Theall refers is as follows:

Even given the inherently less than perfect nature of ratings data and the analytical inclinations of academics, the problem of unskilled users, making decisions based on invalid interpretations of ambiguous or frankly bad data, deserves attention. According to Thompson (1988, p. 217) "Bayes Theorem shows that anything close to an accurate interpretation of the results of imperfect predictors is very elusive at the intuitive level. Indeed, empirical studies have shown that persons unfamiliar with conditional probability are quite poor at doing so (that is, interpreting ratings results) unless the situation is quite simple." It seems likely that the combination of less than perfect data with less than perfect users could quickly yield completely unacceptable practices, unless safeguards were in place to insure that users knew how to recognize problems of validity and reliability, understood the inherent limitations of rating data and knew valid procedures for using ratings data in the contexts of summative and formative evaluation. (79-80).

The authors conclude by noting "It is hard to ignore the mounting anecdotal evidence of abuse. Our findings, and the evidence that ratings use is on the increase, taken together, suggest that ratings malpractice, causing harm to individual careers and undermining institutional goals, deserves our attention." (pp. 79-80).

Then in endnote #26 of my paper, Theall objects to my quoting Dowell and Neal noting that: "...Cohen's (1983) review reinforces our earlier conclusion that student ratings are inaccurate indicators of student learning..." (Note 2) Theall omits the rest of the Cohen's quote which said, "and therefore are best regarded as indices of consumer satisfaction' rather than teaching effectiveness." Even given that Theall (et al.) is talking about the logic in use of SEF, not its inherent validity, he goes on to say,

Conversations with faculty and administrators...led increasingly to concerns about what users [e.g., chairmen; deans] were doing with the information we were providing. We saw that some departmental administrators, who routinely use ratings to make decisions about personnel, evaluation policy, and resource allocation, were not familiar enough with important ratings issues to make well informed decisions...Clearly stated disclaimers regarding the limitations of ratings data in particular circumstances appeared to have little effect on the inclination of some clients to use invalid or inadequate data...There are some fundamental concepts for using numbers in decision making. To the degree that these concepts are ignored, interpretations of data become, at best, projective tests reflecting what the user (e.g., a chairperson or dean) already knows, believes, or perceives in the data. Treating tables of numbers like inkblots ('ratings by Rorschach') will cause decisions to be subjective and liable to error or even litigation...

Theall (et al.) continue in that chapter to lay out carefully the methodological precondition for the reliable use of SEF. For example, in that chapter Theall's (et al.) wrote:

Three types of errors come to mind immediately. The first involves interpretation of severely flawed data, with no recognition of the limitations imposed by problems in data collection, sampling, or analysis. This error can be compared to a Type I error in research -- wrongly rejecting the null hypothesis -- because it involves incorrectly interpreting the data and coming to an unwarranted conclusion. In this case, misinterpretation of statistics could lead to a decision favoring one instructor over another, when in fact the two instructors are not significantly different...(p.87-88)

The second type of error occurs when, given adequate data, there is a failure to distinguish significant differences from insignificant differences. This error can be compared to a Type II error. -- failure to reject the null hypothesis - because the user does not realize that there is enough evidence to warrant a decision. In this case, failure to use data from available reports (assuming the reports to be complete, valid, reliable, and appropriate) may be prejudicial to an instructor whose performance has been outstanding but who, as a result of the error, is not appropriately rewarded or worse, is penalized. (p.87-88)

The third type of error occurs when, given significant differences, there is a failure to account for or correctly identify the sources of differences. This error combines the other two types and is caused by misunderstanding of the influences of relevant and irrelevant variables. In this case, a personal predisposition toward teaching style..., may lead a user to attribute negative meanings to good ratings, or to misinterpret the results of an item as negative evidence when the item is actually irrelevant and there is no quantitative justification for such a decision. Any of these errors can render an interpretation entirely invalid.(p.87-88)

Summarizing, Theall (et al.) wrote,

Let us...state our goal in the following way: "The user will make decisions that are based on valid, reliable

hypotheses about the meaning of data." In this case, the user should receive or construct working hypotheses that do the following things:

- Take into account problems in measurement, sampling, or data collection and include any appropriate warnings or disclaimers regarding the suitability of the data for interpretation and use.
- Do not attempt to account for differences between any results when they are statistically not significant (probably $<.05$).
- Disregard any significant differences that are merely artifacts (for example, small differences observed in huge samples), which can technically be significant but are unimportant).
- Account for any practically important, significant differences between results in terms of known, likely sources of systematic bias in ratings or reliably observed correlations, as well as in terms of relevant praxio logical constructs about teaching or instruction.
- The user should also refrain from constructing or acting on hypotheses that do not meet these conditions...(pp. 87-89)...

Again, I said in my paper,

SEF may indeed validly describe incompetence...given (a) the conflicting data on their validity, (b) the way many institutions have constructed SEF instruments, (c) the often unsystematic statistical method by which SEF are interpreted, and especially (d) given the considerable weight accorded negative comments by only a few students in making tenure and promotion decisions, it would seem SEF can all too easily be used as a covert instrument for the elimination of tenure candidates and other faculty who may threaten student tuition dollars and perhaps ideological and popular culture agendas.

So what is the problem of "intent" to which Theall refers?

Inappropriate Tone of Theall's Commentary

Theall's also makes brief mention of Stake's (1997) response to my article. While Stake can take care of himself, I would like to comment on one of Theall's comments as in my view it is indexical of Theall's setting up "straw men." Theall says,

Stake's reference is probably to "educational seduction", the skill of the infamous "Dr. Fox" who supposedly entertained students and received high ratings despite the fact that he delivered no content (Naftulin, Ware, & Donnelly, 1973). Many who do not care for ratings find one study that supports their position but ignore subsequent work (e.g. Perry et al., 1979; Marsh & Ware, 1982) which points out problems with the original study and proceeds to clarify the issue.

The point is that Stake does not even come close to mentioning the Dr. Fox study, a study that used an actor to deliver lectures with nearly no content. Setting up such straw men, allows Theall to then claim to have demolished them. What the purpose of this was, I can not say.

Finally, in a critique which rather cavalierly takes on not only my article, but Stake (1997) and Greenwald, and Greenwald and Gilmore (1996), (Note 3) perhaps what is most disturbing is that Theall's attempt at rebuttal is, in tone, unjustifiably condescending and, in word, *ad hominem*. In referring to a research article of his (et al.), from which I quote in my article much to his dissatisfaction, he says,

Frankly, I don't like to recommend articles like this to those not actively involved in ratings research or practice because such writings can mislead readers who aren't really familiar with the cited ratings literature. I'm sorry if that sounds elitist: it isn't intended to be but I do have a reason for noting it.

Theall should be more than "sorry;" as a scholar presumably knowledgeable in science, he should be embarrassed. He should also not flatter himself by suggesting his statement may be considered elitist; arrogant, perhaps, but not elitist. (Note 4) This is why: First, he has no legitimate reason to assume that I am not familiar with the literature (see below), just because I disagree with him and have reservations about SEF . His assumption seems to be based on the most fundamental of logical and statistical inference fallacies. Theall says,

We (Franklin & Theall, 1989) found that ignorance of evaluation/measurement literature and methods correlated significantly with negative faculty opinions about students and student ratings. I note this because discussions about ratings are so often filled with misinformation.

Now as any first year college student learns in statistics, (a) correlation is not causation, and (b) one can not infer from a statistical generalization of a group, that an individual in that group is an indicator of that generalization. In fact this "logic" is the same as that used by people who exhibit prejudice: applying some generalization or central tendency characteristic of a specific group of people to an individual. Scriven (1988) makes this point in a similar context.

Second, were his claims not so clearly uninformed, they would be an insult not only to me, but to many other scholars as well. This is why that is: To say that scholars not actively involved in doing statistical or experimental research on a subject are not qualified to read, understand and comment on this research, immediately puts most theoretical physicists, theoretical biologist, and philosophers of science out of business. Hmmm. (Note 5)

References

- Barnett, Larry, D. (1996). Are Teaching Evaluation Questionnaires Valid? Assessing the Evidence. *Journal of Collective Negotiations in the Public Sector*, 25, (4) 335-349
- Cohen, P. A. (1983). Comment on a selective review of the validity of student ratings of teaching. *Journal of Higher Education*, 54, 448-458.
- Franklin, J. L. & Theall, M. (1990). Communicating student ratings to decision makers: design for good practice. In M. Theall & J. Franklin (Eds.) *Student ratings of instruction: issues for improving practice*. New Directions for Teaching and Learning # 43. San Francisco: Jossey Bass.
- Franklin, J. L. & Theall, M. (1989). Who reads ratings. Knowledge, attitudes, and practices of users of student ratings of instruction. Paper presented at the 70th annual meeting of the American Educational Research Association. San Francisco: March 31.
- Greenwald, A. G. & Gillmore, G. M. (1997). No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology* (in press).
- Greenwald, A. G. & Gillmore, G. M. (1996). Applying social psychology to reveal a major (but correctable) flaw in student evaluations of teaching. University of Washington, Draft Manuscript, March 1.
- Haskell, R. E. (1997b). Contributed Commentary on: Stake Response to Haskell: Academic Freedom, Tenure, and Student Evaluation of Faculty. *Educational Policy Analysis Archives*, 5 Available online: <http://olam.ed.asu.edu/epaa/v5n8c1.html>.
- Haskell, R. E. (1997a). Academic freedom, tenure, and student evaluations of faculty: Galloping polls in the 21st century. *Educational Policy Analysis Archives*, 5. Available online: <http://olam.ed.asu.edu/epaa/v5n6.html>.
- Natufin, D. H. Ware, J. E., & Donnelly, F. A. (1973). The Dr. Fox lecture. A paradigm of educational seduction. *Journal of Medical Education*, 48, 630-635.
- Scriven, M. (1988). The New Crisis in Teacher Evaluation: The Improper Use of 'Research-based' Indicators. *Professional Personnel Evaluation News*. (January) 1-8.
- Stake, J. E. (1997). Response to Haskell: academic freedom, tenure, and student evaluations of faculty. *Educational Policy Analysis Archives*, 5. Available online: <http://olam.ed.asu.edu/epaa/v5n8.html>.
- Theall, M. (1997). On Drawing Reasonable Conclusions about Student Ratings of Instruction: a Reply to Haskell and to Stake. *Educational Policy Analysis Archives*, 5. Available online: <http://olam.ed.asu.edu/epaa/v5n8c2.html>.
- Thompson, G.E. (1988). Difficulties in interpreting course evaluations: Some Bayesian insights. *Research in Higher Education*, 28, 217-222.

Notes

1. Theall, suggested that I misinterpreted research on the SEF. For example, he says,

Peter Cohen's most recognized contribution to the ratings literature is the meta-analysis of multisection validity studies in which performance on a common final exam was correlated with ratings (1981). There was a .43 correlation between ratings and performance on those exams. In other words, sections with highly rated instructors had higher average scores which could not be attributed to differential grading standards or sampling errors. This is evidence for the validity of ratings, the exact opposite of what Haskell says. Calling that relationship grade inflation is simply ridiculous!

Further, in at least partial defense before readers who may not be conversant with the literature, let me offer at least one somewhat lengthy statistical anecdote from Barnett (1996) that addresses Theall's belief in .43 correlations. In analyzing some of Cohen's (1981) data, Barnett explains,

I turn now to a second statistical matter the interpretation of correlation coefficients. The numerical value of a correlation coefficient can be deceiving, even when it is 'statistically significant' (i.e., even when it is unlikely

to have occurred by chance if no relationship exists). A correlation coefficient measures the degree to which change in the amount of the explanatory variable is accompanied by change in the amount of the effect variable, but the most beneficial feature of a coefficient is not its numerical value, which has no inherent, practical meaning. Rather, the square of the numerical value is the most advantageous aspect of a correlation coefficient, for the square indicates the proportion of variation in the effect variable that can be statistically attributed to variation in the explanatory variable. The research summary by Professor Cashin reported a correlation coefficient of .44 between student ratings of an instructor 'overall' and examination grades [8]. This coefficient means that 19.4 percent of the variation in student learning (as measured by course grades) is explained by variation in instructional quality (as measured by student ratings). If accurate, a correlation of this magnitude is 'practically useful,' as Professor Cashin said, though one must keep in mind that four-fifths of the variation in course grades remains unexplained and is attributable to other factors.

But does this correlation coefficient accurately estimate the relationship between student evaluations of teaching and student achievement? The best research on the magnitude of the relationship is the 'multisection validity study.' When it is ideally designed, such a study possesses the following features: each course included in the study has numerous sections; students are randomly assigned to sections; the sections of a course have different instructors but a common textbook and the same examination(s); all examinations for a course are constructed by a person who does not teach any section; and subjective (essay) components of examinations are graded by the person who developed them. A review of multisection validity studies cites one work that, the author of the review asserts, eliminates at least in part 'many of the criticisms of the multisection validity study' and 'provide[s] strong support for the validity of students' evaluations of teaching effectiveness' [9, p. 721]. However, the cited work which subjected the results of other multisection studies to a statistical analysis did not control a number of critical variables that could have generated or enlarged the relationship between student ratings of teachers and student achievement [10]. Among the missing variables that might have explained the relationship was the rigor of the requirements of the instructor (such as checks for student preparation and amount of material assigned), a factor that may vary considerably across sections of a single course. If the variable was related both to student ratings of instructional quality and to student achievement, a control for the variable could have markedly weakened or entirely eliminated the relationship originally found between student ratings and student achievement p.339

Another variable that the work omitted was the students' level of interest in the subject matter of the course prior to exposure to the teacher they later evaluated. As will be suggested below, neither of these variables should have been excluded from the analysis and left uncontrolled.

While the work did not incorporate a number of potentially important variables into its data analysis, the work is the source of a set of correlation coefficients (including the coefficient of .44) that Cashin suggested are credible estimates of the relationship between student ratings of teachers and student achievement. A reader of the reproduced coefficients can easily be misled, however, because Cashin failed to make clear that the coefficients may have been seriously confounded by variables whose influence was not removed. The failure to clarify this point is surprising inasmuch as Cashin explicitly stated that a control is necessary for one of the variables omitted by the work, namely, the interest students initially exhibited in the subject.' [12, p. 5]. P.340

Cohen treated student judgments of 'the amount and difficulty of the work the teacher expects of students' as one component of instructional quality. Whether it is an element of teaching quality or a separate factor, the amount and difficulty of work should be controlled under the conditions mentioned because it may explain much or all of the relationship detected between, on the one hand, evaluations of an instructor overall or on specific dimensions and, on the other, performance on examinations. From the studies he reviewed, Cohen calculated a negligible mean correlation coefficient for the relationship between the amount/difficulty of work and student achievement, but he also found a substantial range for the coefficients reported by the studies. Specifically, the interval for 95% of the coefficients extended from -.42 to +.39. Id, at 293, 295. Individual studies may thus involve a nontrivial association between the perceived difficulty of teachers and the examination performance of students (p.341).

I would also suggest that he reader see Greenwald's historical review of past studies and Barnett's review of some of the most cited pieces that claim the validity of SEF.

2. I can understand how such quotes might be embarrassing (a) to an author who professes that SEF is basically if not completely valid instrument, and (b) to a practitioner trying to convince faculty and administrators to use them in faculty evaluation programs. I have never heard of a consultant who suggests that SEF have been shown to be invalid. Given Theall's critique let me make it very clear that I am not here indicting consultants. I have done and occasionally do consulting myself. I would like to say that Theall (et al.)---in word, I can not speak for deed--- does recognize most of what needs to be done to make SEF use in administrative decisions reasonably ethical.

3. Requests for prepublication copies of Greenwald's articles should be sent to <http://weber.u.washington.edu/~agg/>

4. Theall goes on to say, "Stake says that ratings undermine the faith and trust students must place in teachers." Given his opinion of ratings, I assume that he must then feel that teachers have no need to place faith or trust in their students. This

paternalism is antiquated and unrealistic. The logic by which it is unacceptable for Stake not to trust (valid?) perceptions of students, but acceptable for Theall not to trust the perceptions of scholars, escapes me, just as does his calling Stake's view of students paternalistic and not seeing that his own views of his not trusting scholars to correctly read research is paternalistic. Though he does recognize his view sounds elitist, he says it is really not.

5. I wonder what Theall would say about child psychologists who have no children of their own, or about healthy psychotherapists who treat the mentally disordered.

**Contributed Commentary on
Volume 5 Number 6: Haskell *Academic Freedom, Tenure, and Student Evaluation of Faculty***

and

Volume 5 Number 8: Stake *Response to Haskell: Academic Freedom, Tenure, and Student Evaluation of Faculty*

21 May 1997

Michael Theall

**On drawing reasonable conclusions about student ratings of instruction:
a reply to Haskell and to Stake.**

Kenneth Ebel was right when he said (1983, p. 65) "No corner of the university lacks faculty members who fulminate against student evaluations, with little or no examination of the large body of research ...that underlies the practice." Lately, there has been a lot of discussion attempting to causally link student ratings to problems such as "grade inflation," and to suggest that ratings somehow violate the great traditions of academic life. I do not mean to trivialize these issues. They are important and deserve discussion. Rather, I am targeting the use (perhaps, misuse) of these issues in efforts to discredit or do away with faculty evaluation practices in general and student ratings in particular.

"Grade-inflation" (i.e., the rise in "average" grades since 1990 or is it 1950, or 1890, or the dawn of time) is always a hot topic because it's easy to make sweeping generalizations about "today's students," especially in comparison to some more-than-likely-mythical group of "better" students from some time past (especially a time "...when I was a student"). While it is probably true that a lot of things have changed about higher education including the standards used to grade, the expectations of both faculty and students, and the overall grade profile, it is very risky to point to one reason for these changes. There just isn't evidence for such a simplistic leap.

When coupled with the sometimes hysterical rhetoric about student evaluations, the inflation complaint itself gets inflated. "Ratings" become the cause of the downfall of higher education. They are, as one anonymous and angry professor wrote to me "...a corrupt practice of the 60's....A sop to students from administrators who are unwilling or unable to do anything to really improve teaching." More recently, in an article rife with rhetoric and loaded with misinterpretations of the literature, Robert Haskell (1997) suggested that ratings are a violation of academic freedom (and of course, the cause of grade inflation). Haskell included a long quote from a chapter Jennifer Franklin and I (1990) did in our *New Directions for Teaching and Learning* issue #43 (Theall & Franklin, 1990b) While our point was that ratings practice must be improved, Haskell used the quote to supplement other citations as evidence in building his case that ratings were an essentially unreliable form of data that should be done away with ... a gross

misinterpretation of our intent.

Frankly, I don't like to recommend articles like this to those not actively involved in ratings research or practice because such writings can mislead readers who aren't really familiar with the cited ratings literature. I'm sorry if that sounds elitist: it isn't intended to be, but I do have a reason for noting it. We (Franklin & Theall, 1989) found that ignorance of evaluation/measurement literature and methods correlated significantly with negative faculty opinions about students and student ratings. I note this because discussions about ratings are so often filled with misinformation. For example, Haskell, (in reference note #26) says that "...Cohen's (1983) review reinforces our earlier conclusion that student ratings are inaccurate indicators of student learning...". Peter Cohen's most recognized contribution to the ratings literature is the meta-analysis of multisection validity studies in which performance on a common final exam was correlated with ratings (1981). There was a .43 correlation between ratings and performance on those exams. In other words, sections with highly rated instructors had higher average scores which could not be attributed to differential grading standards or sampling errors. This is evidence for the validity of ratings, the exact opposite of what Haskell says. Calling that relationship grade inflation is simply ridiculous!

Academic freedom has been defined in many ways, but never before in a way that suggests the construct (tradition? principle? tenet?) is vulnerable to the influence of student ratings. The only comparable complaint I've heard came from a faculty member who missed classes frequently because he attended numerous religious services. He told me that since the ratings instrument contained an item about the instructor being available and his ratings on that item were depressed due to his frequent absences, the student ratings process violated his constitutionally guaranteed freedom of religion. Fortunately, in a reply to Haskell in a later issue of EPAA ([V5#8](#)), Jeffrey Stake (1997) does us the favor of providing a strong counter-argument to the academic freedom issue. Unfortunately, he then proceeds to propose that "Almost anything that can be done to undermine the administrative practice of getting students to evaluate teaching ought to be done" (p. 1). Stake says that ratings undermine the faith and trust students must place in teachers. Given his opinion of ratings, I assume that he must then feel that teachers have no need to place faith or trust in their students. This paternalism is antiquated and unrealistic. Stake also proposes that administrative use of ratings creates an image that institutions are asking if they "guessed right" about the faculty they hired. Drawing this in the other direction, I suppose that assigning grades creates the image that institutions are asking if they "guessed right" about the students they admitted. Finally, Stake proposes that asking students' opinions "... focuses the attention of students on the acting and special effects, rather than the message." (p. 2) I haven't seen many evaluation forms which contain items on either of these topics. Stake's reference is probably to "educational seduction," the skill of the infamous "Dr. Fox" who supposedly entertained students and received high ratings despite the fact that he delivered no content (Naftulin, Ware, & Donnelly, 1973). Many who do not care for ratings find one study that supports their position but ignore subsequent work (e.g. Perry et. al., 1979; Marsh & Ware, 1982) which points out problems with the original study and proceeds to clarify the issue. In this case, the determination that while style is no substitute for substance, "expressiveness" in a lecture can be a powerful aide to students' motivation and attention during lectures and thus, to recall and other aspects of eventual performance.

Jennifer Franklin and I did two studies which involved looking at grades and ratings: one specifically on grade inflation (1991) and one on the attributions of students about their academic performance (1990a). In the first study we looked at thousands of courses from one institution over a five year period during which mandatory ratings were initiated. We found a statistically significant but numerically minute increase in average grades but we also found that average

ratings actually decreased in the same time period. So, even if faculty were trying to "buy" higher ratings by giving higher grades, students "weren't buyin' it."

In the attribution study we found that ratings of instructors were very consistent (i.e. reliable) across all grades. In other words, the ratings were essentially the same from 'A' students and 'F' students. This held up when class average ratings showed that the instructor was considered to be "among the best" as well as when the instructor was considered to be "average" or "among the worst." There was no revenge for low grades or reward for high grades.

Major reviews (e.g., Marsh, 1987) have reported either near zero correlations between workload, difficulty, and ratings, or positive relationships indicating that more work and/or more difficult courses often receive higher ratings (e.g., Cashin, 1988). Greenwald & Gillmore (1996a, 1996b) however, report a positive grades-ratings relationship (i.e., higher grades - higher ratings) and a negative grades-workload relationship (i.e., higher grades - less work). However, they do not then propose a direct, negative ratings-to-workload relationship (i.e., less work- higher ratings) although they suggest it when they say (p. 14) "...finding a negative path between expected grade and workload is critically diagnostic of a causal effect of grading leniency." One might extrapolate that if less/easier work is related to higher grades, and higher grades are related to higher ratings, then less/easier work should also relate to higher ratings. Whether the syllogism holds remains a question. Even if it does, there is still a question about its implications for the student ratings process. Does a relationship - even a causal one, mean that ratings are inherently invalid or unusable? Greenwald and Gillmore note that ratings are imperfect but that flaws are "correctable" (1996b) and suggest making improvements rather than dismissing ratings. For the record, the relationship between grades and ratings in other studies is relatively consistent, with correlations normally between .3 and .4 (e.g., Feldman, 1976).

The counter-argument to the grades-ratings-inflation idea is that there is (and should be) a relationship between grades and ratings because good teaching begets good learning. Good learning results in both good grades and satisfied learners. This is still a compelling notion but Greenwald and Gillmore (1996A, 1996B) have clarified the definitions and controlled some of the variables more precisely. On the surface, their work suggests that there is a relationship which does not rely on the counter-argument and that when all else is held constant, giving higher grades will result in getting higher ratings. In other words, you can't account for the relationship solely on the basis of the "good teaching begets good learning" counter-argument.

One question about their findings is whether their finding is a psychometric, an instructional, or a psychological phenomenon. Consider that ratings are undeniably a measure of the satisfaction of learners with their learning experience perhaps more than they are a direct or absolute measure of the total quality of instruction. Though this sounds like heresy from a ratings proponent, it is a position that most researchers have held for some time. Consider also, that ratings are collected before students get their final grades and thus, their opinions must be based on expectations which are, in turn, based on performance to date. Ratings thus can not be said to reflect a disconfirmed expectancy about the overall course outcome (i.e., the course grade). This would only be the case if ratings were gathered after final grades were distributed and the final grade disconfirmed what was expected as a result of the experiences during the semester. So the ratings relationship is limited to experiences and results during the term rather than to the final grade.

So what could account for the grades-ratings relationship if not "better teaching begets better learning"? I propose the following.

Given that student satisfaction is related to the students' perceptions that they have received something of value in return for their tuition dollars, and given that a series of successful classroom and related experiences has unfolded during the semester, it is reasonable to assume that students will believe that they have learned something, will be satisfied, and will provide positive ratings. Now, in many classes (especially the lower level, undergraduate courses that populate many evaluation databases and by virtue of their enrollments and numbers, may disproportionately influence results) students are not the best judges of the breadth and depth of the instructor's knowledge or the extent to which the instructor has provided a complete, current, or even adequate treatment of the subject. If students have received good grades and positive feedback from their instructors they should: 1) be satisfied; 2) feel that they have learned something; and 3) be able to honestly rate their experiences and their instructors highly [even if the experience might be marginal using some other rating criterion coming from some other group]. The data they provide is thus both valid and reliable. Valid, because they are the appropriate providers of data and (we will assume for the moment) the instruments used ask at least face valid questions. Reliable because there will be little variance across responses (Marsh, 1987) and because their responses will not change much over time (Frey, 1976).

The question now becomes one which deals with: 1) the appropriateness of course content; 2) the standards used for grading; and 3) the question of whether lenient coverage and grading were deliberately chosen in order to influence ratings. The first two items can be and should be dealt with via curricular mechanisms such as departmental review of courses and content, and faculty agreement on standards for student work. The third item represents an ethical dimension that has much less to do with ratings than it does with behavior. If it is possible to manipulate students covertly (as in "dumbing down" courses so that students are led to believe that they are learning a lot when, in fact, they are only scratching the surface), then the problem belongs to the instructor and/or the department. Don't blame students or the ratings process. If the situation involves overt manipulation (as in making a 'ratings for grades' deal), then the problem belongs to the instructor, the students, the institution, and to higher education itself. Blame the perpetrators and those who let them get away with it, not the process of collecting, reporting, and using ratings data. Ultimately, I believe this suggests that ratings are reliable and valid indicators. If what they indicate doesn't please those who get and use the data, then the users, their departments, or their institutions should correct the situation, rather than discard the data or the processes of collecting, analyzing, or reporting it.

There are three points here. The first is that no system, however well constructed is safe from manipulation. The second is that there are probably legitimate instances of situations in which the established findings from the ratings literature are violated even though these situations are infrequent enough to influence the results of large database analyses which have been the bases of most research. The third is that ratings are a symptom, not a cause. Many opinions about ratings (like those of Haskell and Stake) seek to "blame" ratings for phenomena like grade inflation. Legitimate concern for the quality of teaching and learning should prompt us to examine the contexts of teaching and learning as well as the ratings that are provided by students, and to carefully review the other available kinds of data before jumping to erroneous conclusions about what "causes" what.

Robert Haskell suggests that we do away with ratings. Perhaps the weakest part of his article is what isn't there: constructive suggestions for improvement. The thrust is simply that we should do away with student ratings. One could just as well argue that we should do away with grades. Either way, no more nasty inflation! WRONG! There are appropriate reasons and uses for both. What's needed is more substantial research, less rhetoric, and better informed faculty and administrators who can distinguish between opinion and evidence, separate mythology from

established findings, and explore the situational context in order to arrive at fair and equitable conclusions about faculty performance. Doing this won't violate anyone's rights. In fact, it should lead to better evaluation, and if data are carefully collected, reported, and used, better teaching as well.

REFERENCES

- Cashin, W. E. (1988) Student ratings of teaching: a summary of research. (IDEA paper No. 20) Kansas State university, Division of Continuing Education. ERIC # ED 302 567
- Cohen, P. A. (1983) Comment on a selective review of the validity of student ratings of teaching. *Journal of higher Education*, 54, 448-458.
- Cohen, P.A. (1981) Student ratings of instruction and student achievement: a meta-analysis of multisection validity studies. *Review of Educational research*, 51, 281-309.
- Ebel, K. (1983) *The aims of college teaching*. San Francisco: Jossey Bass.
- Feldman, K. A. (1976) Grades and college students' evaluations of their courses and teachers. *Research in higher Education*, 4, 69-111.
- Franklin, J. L. & Theall, M. (1991) Grade inflation and student ratings: a closer look. Paper presented at the 72nd annual meeting of the American Educational Research Association. Chicago: April 7. ERIC # ED 349 318
- Franklin, J. L. & Theall, M. (1990) Communicating student ratings to decision makers: design for good practice. In M. Theall & J. Franklin (Eds.) *Student ratings of instruction: issues for improving practice*. New Directions for Teaching and Learning # 43. San Francisco: Jossey Bass.
- Franklin, J. L. & Theall, M. (1989) Who reads ratings. Knowledge, attitudes, and practices of users of student ratings of instruction. Paper presented at the 70th annual meeting of the American Educational Research Association. San Francisco: March 31.
- Frey, P. (1976) Validity of student instructional ratings. Does timing matter? *Journal of Higher Education*, 47 (3), 327-336.
- Greenwald, A. G. & Gillmore, G. M. (1996a) No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology* (in press). Available HTTP: <http://weber.u.washington.edu/~agg/paingain.htm>
- Greenwald, A. G. & Gillmore, G. M. (1996b) Applying social psychology to reveal a major (but correctable) flaw in student evaluations of teaching Available HTTP: <http://weber.u.washington.edu/~agg/asptramf.htm>
- Haskell, R. E. (1997) Academic freedom, tenure, and student evaluations of faculty: galloping polls in the 21st century. *Educational Policy Analysis Archives* V5 #6. Available HTTP: <http://olam.ed.asu.edu/epaa/v5n6.html>
- Marsh, H. A. (1987) Students' evaluations of university teaching: research findings, methodological issues, and directions for future research. *International Journal of Educational*

Research, 11, 253-388.

Marsh, H. W. & Ware, J. E. (1982) Effects of expressiveness, content coverage, and incentive on multidimensional student ratings scales. New interpretations of the Dr. Fox effect. *Journal of Educational Psychology*, 74, 126-134.

Natuflin, D. H. Ware, J. E., & Donnelly, F. A. (1973) The Dr. Fox lecture. A paradigm of educational seduction. *Journal of Medical Education*, 48, 630-635.

Perry, R. R., Abrami, P. A., & Leventhal, L. (1979) Educational seduction: the effect of instructor expressiveness and lecture content on student ratings and achievement. *Journal of Educational Psychology*, 71, 107-116.

Stake, J. E. (1997) Response to Haskell: academic freedom, tenure, and student evaluations of faculty. *Educational Policy Analysis Archives* V5 # 8. Available HTTP: <http://olam.ed.asu.edu/epaa/v5n8.html>

Theall, M. & Franklin, J. L. (1990a) Attributions or retributions: student ratings and the perceived causes of performance. Paper presented at the 71st annual meeting of the American Educational Research Association. Boston: April 16. ERIC #ED 319 764

Theall, M. & Franklin, J.L. (1990b) Student ratings of instruction: issues for improving practice. *New Directions for Teaching and Learning* # 43. San Francisco: Jossey Bass.

About the Author

Michael Theall, Ph.D.

Associate Professor & Director

**Center for Teaching and Learning University of Illinois At Springfield
Springfield, IL 62794-9243**

phone: 217-786-7157

fax: 217-786-7188

email: theall@uis.edu

Education

- BA St Anslem College (1963; English)
- MS Syracuse Univ. (1976; Instructional Technology)
- PhD Syracuse Univ. (1980; Instructional Design, Development, & Evaluation)