

Education Policy Analysis Archives

Volume 5 Number 17

August 25, 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, Promotion, Reappointment, Tenure and The Administrative Use of Student Evaluation of Faculty: (Part II) Views From the Court

Robert E. Haskell
University of New England¹

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

- [Volume 5 Number 6](#)
- [Volume 5 Number 18](#)
- [Volume 5 Number 21](#)

Abstract: Though a controversial history of research on the reliability and validity of student evaluation of faculty (SEF) exists, it has not been typically viewed as an infringement on academic freedom, promotion, reappointment, and tenure rights. As a consequence, legal aspects of SEF are neither readily apparent, nor available. Unlike academic freedom, tenure, and other issues, which exist as legal categories, SEF as a category is virtually absent in legal compendia on higher education law. The question of its judicial standing is important to any suggestion of abridging faculty rights. In this second of four articles, legal rulings are

categorized and abstracted verbatim from cases where SEF is integral to the denial of academic freedom, tenure, promotion and reappointment are reviewed and provide and provide an initial resource of legal ruling on SEF. Seventy-eight findings are summarized from the abstracted textual material.

Table of Contents

- [Use of SEF](#)
 - [Legal Rulings on The Reliance and Validity of SEF](#)
 - [Reliance on SEF v. Peer Evaluation](#)
 - [Popularity and Effectiveness](#)
 - [Transcendent Value of a Professor Over Teaching Quality](#)
 - [The Courts' Approach to the General and Psychometric Accuracy/Validity of SEF](#)
 - [Numerical Ranking of Faculty](#)
 - [Rulings on The Qualitative and Subjective Assessment of SEF](#)
 - [Acceptance of Administrative Subjective Judgements of SEF Data](#)
 - [Use of Qualitative Written Student Comments](#)
 - [Mixed Student Comments](#)
 - [Variables Not Taken Into Account When Assessing SEF](#)
 - [Student Biases in Affecting SEF](#)
 - [Other Variables Not Taken Into Account](#)
 - [Courts Impact on Academic Matters](#)
 - [Impact on Academic Standards](#)
 - [Methods of Instruction and Academic Freedom](#)
 - [Release of SEF To Students and To the Public](#)
 - [Conclusion](#)
 - [References](#)
 - [Appendix A: Summary of 78 Findings](#)
 - [Appendix B: Summary of Case Outcomes](#)
 - [Appendix C: Abstracted Case Material](#)
-

.....Unlike the extensive body of conflicting research on the construction, assessment, validation, and interpretation of student evaluation of faculty (SEF), until recently there has been virtually no categorical mention of their impact on academic freedom and tenure rights (Haskell, 1997). When SEF is recognized to have an impact, unlike traditional threats to tenure and academic freedom, it does not appear to warrant serious concern. Moreover, as I noted in my first article, when it is suggested that SEF may impinge on academic freedom, tenure, promotion and reappointment, it is generally considered an attack on either student rights, the process of evaluating faculty performance in general, or both. Unlike academic freedom, tenure, and other educational issues, as a retrieval category SEF is virtually absent in legal compendia on higher education law. Sources citing legal cases on the issues of academic freedom, tenure and promotion do not typically address SEF.

.....A recent booklet on "*The Law of Teacher Evaluation*" (Zirkel, 1996) contains no mention of SEF cases. Nor does a recent comprehensive legal guide for educational administrators, (Kaplin and Lee, 1995), nor do other reports (Poch, 1993) on the legalities of academic freedom, tenure and promotion. Two other recent sources, however, do address the issue, one very briefly, (Baez, and Centra, 1995), the other somewhat more extensively (Weeks, 1996). This absence is both surprising and important given that courts have recognized for some time "that in recent years increasing weight has been placed upon teaching ability" (Johnson v. University of Pittsburgh, 1977).

.....In a rare mention of SEF as an infringement on academic freedom, a recent handbook for college administrators, refers to the concept as "a rather novel attack on the use of student

evaluations in assessing a faculty member's performance" (Weeks, 1996). It has generally been taken for granted, not only by many administrators but by faculty as well, that SEF is both appropriate and necessary.² There are, however, increasing signs indicating the examination of this issue by faculty.³ It is important to note at the outset that it is not SEF *per se* that is the issue in this paper, but rather the legal and educational implications of its use in salary, promotion and tenure decisions. Neither is the primary concern with faculty welfare. As I will delineate in a following paper (Part IV), the primary issue is the implications for academic freedom in the service of standards and quality of student instruction. Such signs would suggest a review of legal rulings involving SEF.⁴

.....In addition to not considering SEF a significant issue in the abridgement of academic freedom, tenure, promotion, and reappointment, a further reason for the paucity of articles on this issue is the difficulty in searching for legal cases. It appears that cases involving student evaluation of faculty are most always only a part of a much wider set of issues regarding denial of academic freedom, promotion, tenure, and reappointment.⁵ In addition, it seems that such cases are often settled not in courts of law but by Labor Relation Arbitration Boards rulings. Since rulings on SEF are embedded within other cases, locating them through key word searches is extremely difficult.⁶ Accordingly, there exists no systematic cataloging of how courts have specifically viewed SEF in cases involving denial of academic freedom, tenure, promotion, and reappointment.

.....Because administrative policy regulating the use of SEF has largely evolved in a de facto or default manner, this paper will examine legal reasoning and rulings on SEF. In the evolution of any legal standard, 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 legal standing only come to have standing after a long and arduous process of analysis and advocacy. It is the purpose of this paper to coalesce the legal rulings on SEF and to thus facilitate this developmental process, and to serve as an initial resource for faculty organizations and other policy makers. In doing so, I would like to make clear that I am not an attorney and approach this paper and court rulings from the legal concept of the "reasonable man" standard.

.....In the first of what will be four articles on SEF (Haskell, 1997), I suggested that SEF is deceptive regarding its negative implications for higher education. In this second article, based on the content of located cases, I have developed a set of categories and have abstracted the corresponding rulings. More importantly, it should be noted that the issues involved in SEF go beyond problems involved in general faculty evaluation procedures. Delineating the ramifications of SEF provides a concrete vehicle for further understanding the abstract concept of academic freedom.

.....To render the boundaries of this paper manageable, its focus will be delimited to how courts⁷ have addressed SEF issues within various legal challenges to the denial of academic freedom, tenure, promotion, and reappointment by institutions of higher education. It is important to note that there are multiple legal variables that determine an action or final outcome in any given case. For example particular statutory considerations, cause of action, level of prima facie evidence, standards and burden of proof, and level of judicial action, i.e., whether it is pre trial, lower level court, or appellate, etc., (see Kaplin and Lee, 1995, section 1.4.3.6). For my purposes here, I will not be concerned with these variables. Accordingly, this paper will neither be concerned with the validity or wisdom of the legal rulings, nor with the complex reasoning on which the rulings were based. My purpose is to review the general reasoning of the courts on SEF. Only when specifics of a case are contextually warranted in order to clarify the court's view of SEF will contextual specifics be noted.⁸ Not only are such cases prima facially complex, but when specific legal definitions (e.g., disparate impact

and treatment) and other special Congressional Acts (e.g., EEOC) are superimposed on them, they become logically unwieldy, not just to the non legal scholar, but apparently even to the Courts.⁹

.....I should note that the collective rulings are an extremely disparate and frequently contradicting corpus. I should note, too, that in presenting this complex legal corpus, in order to avoid any misinterpretation, I felt necessary to quote extensively from the original material rather than paraphrasing and requiring the reader to constantly refer to the Appendix, which is provided as a resource. For ease of reading and convenient referencing, I have placed the quoted legal rulings, along with my own necessary transitional phrases connecting the quotes, within an indented text format Finally, I will be making no distinctions between the various levels of judiciary proceedings.

Use of SEF

.....Since the 1960's SEF has been increasingly used in decisions on tenure and promotion. 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 (Seldin, 1993) noted that no other method of evaluation has approached that level of usage. 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. With the exception of publications, no other method of faculty evaluation has become so sacrosanct.¹⁰ The reason that SEF was initially instituted was for informational feedback to faculty so they might be more aware of student needs.

.....Not only does SEF have legal standing, but in some states it is mandated for secondary teachers. According to one author, student evaluations in some states have legal weight, and teachers have been fired on the basis of low ratings on their student evaluations. In California SEF is apparently mandatory under law "whenever practicable" (in, Allegre, and Guista, 1993). SEF is used not only in the U.S. but in Australian, Canada, Europe and Great Britain. Unlike in the U.S., however, in other countries SEF by formal questionnaire, despite apparently no formal mandate, is increasingly used, though not weighed as heavily as is information gathered by other means (Husbands and Fosh, 1993). The instrument has not been used just for informational feedback, however. If it was only used for feedback, then SEF would presumably not be a problem.¹¹ 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 a nest of problems.

Legal Rulings on The Reliance and Validity of SEF

.....This section, will examine courts' approach to SEF in the denial of academic freedom, tenure, promotion, and reappointment decisions. In so doing, I have abstracted from located

cases the following categories: (1) Reliance on SEF v. Peer Evaluation, (2) Popularity and Effectiveness, (3) Transcendent Value of a Professor Over Teaching Quality, (4) Establishing the General and Statistical Accuracy/Validity of SEF, (5) Numerical Ranking of Faculty, (6) Acceptance of Administrative Subjective Judgements Of SEF Data, (7) Use of Qualitative Written Student Comments, (8) Mixed Student Comments, (9) Student Biases in Affecting SEF, (10) Variables Not Taken Into Account When Assessing SEF, (11) Courts View of Who Sets Academic Standards, (12) Methods of Instruction and Academic Freedom (13) Release of SEF To Students and To The Public.¹²

.....Within these categories, I have included Canadian court findings.¹³ I have chosen to include Canadian cases for three reasons. First, to provide a larger data base illustrating how legal bodies reason and rule on cases involving SEF, second, to provide a wider array of an otherwise relatively small number of U.S. cases; and third, to provide a set of comparative views from a culture and legal system that has both an interesting set of similarities and differences.¹⁴ In two following papers, I will further address and analyze the impact and implications of these legal rulings in relation to (a) psychometric Validity, and (b) academic freedom and instruction.

Reliance on SEF v. Peer Evaluation

.....One important issue is how the courts view the relative weighting of SEF in administrative decisions of teaching competence. Is it acceptable, for example, to rely heavily on them, or must they be used in conjunction with other evaluative methods?

In *Johnson v. University of Pittsburgh* (1977), the court noted that it (5) ¹⁵ "has placed little reliance on students' surveys....students in a given course rating a teacher, or professor, some of them as excellent, others as terrible and in between, many who say passable, mediocre etc.... we cannot say it was unreasonable for the tenured faculty to consider this along with other matters" (p.1359). (8) "It is also obvious that the court and the administration of universities cannot permit students to exercise a veto over professors who may be world renowned scientists" (p.1366-7). A similar view was expressed in *Yu Chuen Wei v. Vermont State Colleges Faculty Federation* (1995).

In *Peters v. Middlebury College* (1977), the court gave some weight to an administrative devaluing of a set of positive student evaluations of a faculty that said (2) "The department chair sent a letter to the president of the college, saying, 'The course of action I recommend is not likely to be popular with students who, though they in part recognize her intellectual limitation, are warmly responsive to her enthusiasm, energy, openness and ready human concern' " (p.860).

In *Carley v. Arizona Board of Regents*, (1987), the court said, (23) "Carley has cited no authority that *relying primarily or solely on student evaluations would be impermissible*. We have found none" (p.1105, italics added).

In *Guam Federation of Teachers v. The University of Guam* (1990), the Guam Federation of Teachers challenged the use of SEF in tenure and promotion decisions (Blum, 1990). The Board (1) ruled to remove anonymous student evaluations from professors' tenure files, (2) The union said the use of SEF violated the union's contract with the university, (3) which provides that anonymous documents or those "based on hearsay" should not be included in a faculty member's file, (4) The court further ruled that (5) students should be made aware of the purpose and ramifications of their evaluations, and (6) anonymous student evaluations should not be used.

In *Robert Kramer v. The President of the University of British Columbia* (1992), the Board noted that (18) "The most important perceived error in the teaching evaluation, in the opinion of the Board, is the reliance solely upon the student evaluations and written comments...There was no peer review at all; no member of the Department audited any of Dr. Kramer's lectures" (p.10).

In *University of Regina Faculty Association v. University of Regina* (1993) a Canadian Arbitration Board ruled that (3) "With respect to teaching, it is our opinion that the evidence of unsatisfactory performance is very weak indeed... It is important to note that the basis of the comments, particularly the negative ones in the fall of 1992, were written student assessments... [and] Although these assessments are expressly recognized in Art. 17.19 of the collective agreement, to base important career decisions on them only does not seem justified" (p.4). The Board further ruled (4) 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. (5) [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" (p.4) (8) "To base serious career decisions narrowly on student evaluations is not to be encourage... (9) If teaching is to be seriously evaluated for career purposes, whether for positive or negative purposes, it seems incumbent upon Faculties not to rely only on classroom administered evaluations but to broaden the base of assessment" (p.4)

In *Christopher Turner v. The President of the University of British Columbia* (1993), the Board ruled, "(9) while the [Faculty Association] Agreement permits, but does not mandate either student reviews or peer reviews, and the methods of assessment 'may vary', we do conclude that the reliance placed on these very limited student reviews must have been great, since there was no other evaluation referred to. Where there is no other evidence sought, student comments will have an apparent importance and credibility that they may not deserve... (10) We would strongly recommend peer review in the reconsideration which we are requiring" (p.7). The board further noted that (8) "This board has been asked on a number of occasions to pass judgment on the relevance of student evaluations to the [Faculty Association] Agreement criteria for good teaching. Good teaching is an elusive concept. Students may not be good judges during a course; their judgment might be quite different several years later in life" (p.7).

Summary: From these cases, it can be seen that court rulings range from saying that (1) relying primarily or solely on student evaluations is acceptable, to (2) placing little exclusive reliance on SEF, (3) in rare cases SEF can not be permitted to stand in the way of promoting or retaining professors who are excellent in non teaching areas, (4) tenure decisions can not be based solely on SEF by students who have not been made aware of the ramifications of their evaluations, (5) anonymous documents or those "based on hearsay" should not be included in a faculty member's file, (6) students should be made aware of the purpose and ramifications of their evaluations of faculty, (7) anonymous student evaluations should not be used, (8) peer evaluations must also be a part of evaluating teaching.

Popularity and Effectiveness

.....A significant issue in SEF is the extent to which it measures popularity not teaching effectiveness. Accordingly, it is instructive to see how courts view this issue.[16](#)

In *Johnson v. University of Pittsburgh* (1977), the court said, "It is also obvious that the court and the administration of universities cannot permit students to exercise a veto over professors who may be world renowned scientists and yet if the students rate them

unfavorably can be terminated at any time because of unpopularity" (p.1366-7).

In *Carley v. Arizona Board of Regents*, (1987), he (8) he maintained his popularity suffered as reflected in his low student evaluations; in *Robert Kramer v. The President of the University of British Columbia* (1992), he maintained that (14) student evaluations were considered from the standpoint of his popularity, not his effectiveness; and in *Brian Maclean v. President of The University of British Columbia* (1991), (35) The Faculty Agreement specified that "Evaluation of teaching shall be based on the effectiveness rather than the popularity of the instructor." Courts have ruled in various directions on this issue.

In *Robert Kramer v. The President of the University of British Columbia* (1992), the board noted (21) "As for the 'popularity vs. effectiveness' debate, a discouraging or hostile attitude is a part of effectiveness as much as it is of popularity" (p.8).

In *Christopher Turner v. The President of the University of British Columbia* (1993), the Board ruled, (8) "while popularity is not competence nor effectiveness, to the extent that it encourages students it has some relation to both" (p.7).

Summary: From these cases, it can be seen that court rulings range from saying that (9) in cases of exceptional research faculty that popularity should not play a role in termination due to teaching, to (10) in normal cases that a measure of popularity is related to teaching effectiveness.

Transcendent Value of a Professor Over Teaching Quality

.....Despite the importance placed on teaching, there is precedent for both school policy and the courts---under certain conditions---to ignore poor teaching as indicted by SEF.

In *Johnson v. University of Pittsburgh* (1977), the court said, "It is also obvious that the court and the administration of universities cannot permit students to exercise a veto over professors who may be world renowned scientists" (p.1366-7), noting, "It is obvious that a professor may be possessed of excellent qualifications as a research scientist and not necessarily be able to prove his or her worth as a teacher, concluding that, (9) "in cases where one has an outstanding scientist of national or international reputation, one may decide to promote and give tenure notwithstanding inability to come across as a teacher, this however is not one of those cases" (p.1366-7).

In *Yu Chuen Wei v. Vermont State Colleges Faculty Federation* (1995), (31) Wei's last claim charged that the College violated the Contract by denying her a promotion, even though both her scholarly performance and professional activities were exceptional. Article 22(E) of the College provides for otherwise granting promotion if the President decides that "performance in one of three areas has been exceptional" (p.314). The Board concluded that "Although Grievant had a significant publication record, most of it was developed before coming to Castleton" (p.315). (33) In terms of exceptional scholarship, Dr. Wei maintained she had solved a significant mathematical problem (apparently published). The Board's response was, (34) "although Grievant claimed to have solved the Erdos conjecture, [the]Dean reasonably concluded that she had not established that she actually had solved the conjecture. Under these circumstances, and given our consideration of the discrimination issue previously discussed, we conclude that (35) Grievant has not established discrimination. The Colleges reasonably, and based on legitimate reasons, concluded that Grievant had met the tenure standards in this performance area but that her performance was not exceptional" (p.315).

In *Dr. Brian Maclean v. President of The University of British Columbia* (1991), the court said, (34) "while a superior research and publication record cannot overcome a poor teaching record, it might tip the scales where the teaching record was 'on the edge'" (p.10).

Summary: Courts have ruled that (11) the courts and educational administrations can not allow low SEF to stand in the way of promoting or retaining professors who may be world renowned scientists, (12) deemed nationally or internationally exceptional as a researcher, courts have variously ruled that SEF may be disregard and not disregarded, (13) at least in these two cases the courts did not find the faculty exceptional.

The Courts' Approach to the General and Psychometric Accuracy/Validity of SEF¹⁷

.....An issue directly related to the relative reliance on SEF for administrative purposes is validity. It would presumably appear from any "reasonable man" standard that the more valid the SEF data in a given case, the more justifiable is the reliance on them.

In *Johnson v. University of Pittsburgh* (1977), the court said (7) "We have repeatedly approved the use of statistical proof where it reached proportions comparable to those in this case to establish a prima facie case of racial discrimination in jury selection cases . . . Statistics are equally competent in proving employment discrimination. We caution only that statistics are not irrefutable. They come in an infinite variety and, like any other kind of evidence they may be rebutted. In short, their usefulness depends on all of the surrounding facts and circumstances" (8) The court further said in Footnote # 20: "Considerations such as small sample size may of course detract from the value of such evidence" (p.1361).

In *Peters v. Middlebury College* (1977), it was maintained that (5) "A professor's value depends upon his creativity, his rapport with students and colleagues, his teaching ability, and numerous other intangible qualities which cannot be measured by objective standards" (p.860).

In *Fields V. Clark University* (1987), the court noted that (10) Fields' "attacks" the university's use of her student evaluations because they were not gathered and evaluated according to accepted standards of scientific polling procedures. In response, the court agreed, saying, "She is probably correct. The use made of the student evaluations in her case, however, followed the practice at the defendant's university in other tenure decisions" (P.671).

In *Cynthia J. Fisher v. Vassar College* (1995), the court noted that (7) "statistical analyses may be a part of a plaintiff's effort to establish discriminatory treatment" (p.1209).

In *Yu Chuen Wei and the Vermont State Colleges Faculty Federation* (1995), the court ruled that (4) "The Court need not consider the accuracy of these administrative determinations, and that (24) tenure criteria "are not drawn with mathematical nicety." The board further ruled that (25) "the Dean and the President, both reviewed Grievant's student evaluations carefully. Their failure to take it a step further, and perform a statistical comparison of Grievant's student evaluations with those of other faculty members who have been granted tenure was not arbitrary and was reasonable; (26) Such a comparison is nowhere required by the Contract, [and] (27) we decline to hold such an involved comparison is necessary before a reasonable tenure determination can be made" (p.311).

In *Dr. Brian Maclean v. President of The University of British Columbia* (1991), the court

concluded (38) "that the instrument was not perfect, that it had flaws, and that the very limited number of samples (because of the very limited number of courses and students surveyed over the period) impaired its reliability. (p.30). (39) "However, we accept the evidence of Dr. [X] that the instrument has some value, directed toward the specified factors. The court noted that (28) "One problem with the questionnaire is that it solicits bad points as well as good points. Despite that caveat, we conclude that the inclusion of the qualitative comments was not a significant error" (p.32).

In *Robert Kramer v. The President of the University of British Columbia* (1992), the Board said, (19) Given certain Departmental procedures, "there is a danger that some negative class commentary will dominate the discussion and will not be the 'independent' opinion of all of the students. (20) This is especially true in the context of the direction to assess "effectiveness" versus "popularity" (p.10). They further noted, (18) Given that "There was no peer review at all; no member of the Department audited any of Dr. Kramer's lectures. There was, therefore, nothing to guide the Department but the student comments," and "no way to test the accuracy or fairness of the undoubtedly disturbing comments in Asian Studies" (p.10).

In *University of Regina Faculty Association v. University of Regina* (1993), The Board argued (6) that "the University was under an obligation to verify negative comments before acting on them" (p.4).

In *Christopher Turner v. The President of the University of British Columbia* (1993), the Board said, (7) "while not ignoring some student unhappiness with Dr. Turner's teaching style, we think that the comments and emphasis on the size of Dr. Turner's classes as evidence of poor teaching are open to objection and constitute errors of procedure and/or evidence" (p.6).

Summary: With regard to establishing the general and statistical accuracy of SEF in the above cases, the range of opinion is from (14) statistical analyses may be a part of a plaintiff's effort to establish discriminatory treatment if it reached proportions comparable to those in cases establishing a prima facie racial discrimination, (15) cautioning that statistics are not irrefutable, with their usefulness depending the surrounding facts and circumstances of a case, (16) the Court maintaining that it need not consider and is under no obligation to establish the accuracy of administrative interpretations of SEF, (17) that tenure criteria are not drawn with "mathematical nicety," (18) administrator's failure to perform statistical comparisons is not arbitrary and is reasonable, (19) especially if such is not required by a Faculty Association Contract, (20) nearly any use made of SEF is acceptable if it followed the standard practice of the university, (21) that creativity, rapport with students and colleagues, teaching ability, and other qualities are intangibles which cannot be measured by objective standards.¹⁸

Numerical Ranking of Faculty

.....It seems to be common practice to rank and compare faculty to each other according to average SEF scores, considering decimal differences between faculty as significant.

In *Dyson v. Lavery* (1976), a student evaluation ranked her 46th of 48 teachers.

In *Lieberman v. Grant* (1979), the court noted (4) a compilation of student ratings showed that the cumulative ratings for members of the department ranged from a low of 4.09 to a high of 8.95. She had a cumulative rating of 7.06, which ranked her 12th out of the 15 junior faculty members. The 7.06 figure included the ratings from a previous semester in

which the plaintiff received a rating of 8.18. Prior to this rating in the spring of 1972, the plaintiff's cumulative rating was 6.7.

In *Carley v. Arizona Board of Regents* (1987), it was noted that (1) of the 13 faculty in his department of art, he was ranked fifth, (2) by his chairman he was ranked 7th, (3) student evaluations, however, ranked him last: 13th of 13 (p.1105).

In *Robert Kramer v. The President of the University of British Columbia* (1992), the court noted (24) scores in the other two courses were higher---3.45 in one, 3.91 in another, against a "faculty average" of 4.22. The board further noted, "In the result, one got a 2.82 and one got a 3.07...the difference is statistically invalid in any event" (p.10).

Summary: Numerical scores often result in faculty (22) being compared relative to other faculty, (23) being ranked relative other faculty, (24) with distinctions often being made on the basis of tenths of a decimal, (25) with most courts accepting these fine decimal distinctions.

Rulings on The Qualitative and Subjective Assessment of SEF

Acceptance of Administrative Subjective Judgements of SEF Data

.....An issue directly related to both the reliance on and validity of SEF is views of the court regarding accepting or not accepting subjective administrative judgements of SEF.

In *Dyson v. Lavery* (1976), the court found that despite questionable errors it concluded that administrative judgements were acceptable because, "they were sincere and grounded on some evidentiary basis" (p.111); and (5) "In the absence of a finding that same were sexually motivated, the administration's professional judgment must be respected" (p.111).

In *William Sypher v. Vermont State Colleges Faculty Federation* (1982), (7) sufficient evidence exists from which the Dean and President could have reasonably concluded Sypher was not above average in his teaching effectiveness; (8) the Board went on to say that if they adopted the Colleges' view that Sypher was not reappointed because of his teaching effectiveness, no argument advanced by him defending his teaching was likely to persuade the President because his decision was made on the "vigor and variety of student criticisms" (p.135).

In *Carley v. Arizona Board of Regents* (1987), The court ruled (18) the University president was free to consider factual findings made by minority members of the academic freedom and tenure committee and any other evidence which he found relevant in determining whether to deny renewal of teaching contract to non tenured instructor. The president was not bound by factual findings made by majority members of committee (p.1103).

In *Yu Chuen Wei v. Vermont State Colleges Faculty Federation* (1995), it was noted that (28) The Dean and the President obviously had much experience in reviewing student evaluations, and could reasonably draw on that experience in each tenure review. (p.311); judgements "were not arbitrary or capricious and were exercised honestly upon due consideration,"that Deans and Presidents have "much experience in reviewing student evaluations, and could reasonably draw on that experience" (p.311).

In *Dr. Brian Maclean v. President of The University of British Columbia* (1991), the court

said, (40) "The relevance and quality of the scores are "a matter of weight for the various decision-makers, and we assume that they were reasonably aware of the limitations of student evaluations and gave them the weight they deserve" (p.30).

In *Robert Kramer v. The President of the University of British Columbia* (1992), the board concluded, "In the final analysis, we feel that this review of the Head's comments on teaching, which would be the sole evidence upon which the Dean and the President could rely, shows that it was incomplete and might have been misleading" (p.12-14).

In *University of Regina Faculty Association v. University of Regina* (1993), the Board said teaching was wrongfully evaluated, but upheld denial of tenure on grounds of inadequate scholarship.

In *Christopher Turner v. The President of the University of British Columbia* (1993), The board concluded that (11) "there were sufficient errors of procedure and/or evidence to return the case for reconsideration" (p.11).

Summary: Courts have tended to accept administrative subjective judgements if (26) they are deemed sincere (27) grounded on some evidentiary basis (28) if made on the "vigor and variety of student criticisms" (29) "not arbitrary or capricious and were exercised honestly upon due consideration," (30) based upon "much experience in reviewing student evaluations, (31) reasonably draw on that experience (32) and have ruled that Presidents are not bound by factual findings made by majority members of a faculty.

Use of Qualitative Written Student Comments

.....Over and above quantitative data, the use of written comments, often single instances, by students on their SEF seems wide spread by both educational administrators, faculty evaluation committees, and the courts.

In *Dyson v. Lavery* (1976), the court said (1) "A number of students apparently had voiced displeasure over the quality of her class preparation and presentation" (p. 111 (3) "These *impressions*" said the court, "were largely confirmed after the initial decision to not rehire her had been made, by a student evaluation that ranked her 46th of 48 teachers in the Business Department" (p.111, italics added).

In *Johnson v. University of Pittsburgh* (1977), the court said, (3) "we have the *instance* referred to in Finding 27 (p.1359, italics added).

In *Lieberman v. Grant* (1979), the court noted (3) based on complaints received from "*several students*," to the effect that Lieberman's interest in feminism caused her to ignore other themes in literature (p.873, italics added).

In *William Sypher v. Vermont State Colleges Faculty Federation* (1982), (1) some of the student comments noted that, "When students try to disagree he shoots you down and tries to degrade you in front of the class," (p.115), while others said, "encourages student participation as much as possible... encourages student to express their ideas freely and not worrying how 'dumb' it may sound...always wants you point of view." (P.115) (2) With regard to the numerical ratings, the Board's opinion was that (3) "regardless of a strong majority of students' rating his teaching as above average, (4) the existence of a significant minority of students feeling degraded, humiliated, and embarrassed can reasonably lead an evaluator to question a teacher's effectiveness" (p.115).

In *Yu Chuen Wei v. Vermont State Colleges Faculty Federation* (1995), the Board said,

(22) "the statistical comparison does not take account of the comments made by students on the evaluation forms. Grievant's student evaluations are striking in how often mention is made of Grievant's communication difficulties, particularly language difficulties (p.304-5). The board further noted with respect to comments that while some students had written that she was a "slant eyed bitch," and that she should "go back to China," (30) "We also are not persuaded that the racism evident in the student evaluations of Grievant made student evaluation results unreliable. The percentage of evaluations in which racism by students was evident was approximately one percent of the total evaluations" (p.306).

In *Robert Kramer v. The President of the University of British Columbia* (1992), (2). The department Head viewed Kramer's 1989-90 course evaluations "with some alarm"....(4) Even more disturbing to the department Head was that a considerable number of students in their written comments stated that Dr. Kramer was biased sarcastic, and hostile to the material and that a number of students had stated that Dr. Kramer's teaching would cause them to stay away from the Asian Studies department. (5) There were also some diametrically apposed positive comments" (p.10).

In *University of Regina Faculty Association v. University of Regina* (1993), The Board argued (6) that the University was under an obligation to verify negative comments before acting on them. Consequently, (7) the fact that Dr. Jalan had received some negative evaluations from students could not be used to undermine the otherwise generally favorable comments he had received in his annual performance reviews" (p.4).

In *Dr. Brian Maclean v. President of The University of British Columbia* (1991), the court noted that (25) "With respect to the "qualitative" scores---i.e., the "comments," there was a clear error. The qualitative comments from a number of courses were read and commented on, and conclusions were drawn from them which went into the "file." Both Reviewing faculty read and commented on them, as did the Department Chair in her letter to the Dean. Yet the Dean had clearly stated in a departmental memo that the qualitative comments were not to be used for administrative or promotion purposes. (26) While in the abstract there is no reason why such comments would not be relevant, if the Department had a rule against their use, or in other words if they were "for the professor's eyes only," then it was a significant breach of Departmental rules to use them" (p.31). (27) In the opinion of the Board, so long as the comments were fairly presented, they offered the PAT [Promotion and Tenure Committee] and others a better balanced view of the teaching qualities and problems of Dr. MacLean than the quantitative statements alone" (p.31). (28) The court noted that "One problem with the questionnaire is that it solicits bad points as well as good points. Despite that caveat, we conclude that the inclusion of the qualitative comments was not a significant error" (p.32).

Summary: The use of student comments ranges from (33) placing importance on a single comment (34) to several comments as significant information, (35) maintaining that statistical analyses of SEF need to be bolstered by individual comments, (36) maintaining that while some very negative---e.g., racist, sexist---comments may be found, the court may find that they do not rendered the SEF unreliable, (18) that such instances or "impressions" may be validated after the fact, (37) negative comments often seem to outweigh positive ones, and (38) comments may often outweigh numerical data to the contrary, (39) negative comments need not be verified before acting on them, (40), that negative comments can not be used to undermine otherwise generally favorable comments received in an annual performance review.

Mixed Student Comments

.....Just as quantitative SEF data may be bimodal, so too written student comments may

also be bimodal or mixed. How do courts view and pronounce on such data?

In *Johnson v. University of Pittsburgh* (1977), the court noted (2) they "approached this question of teaching ability with considerable doubt, in view of the fact that in prior years there does not appear to have been any criticism of her teaching and also in view of the fact that...there was evidence that the department chairman, had informed her after one of her lectures in 1971 what a great lecture it had been"; On the other hand, the court said (3) "we have the *instance* referred to in Finding 27 (p.1359, italics added).

In *Fields V. Clark University* (1987, it was observed (3) a few of which, from students in Fields' seminars, were "wildly enthusiastic" about her enthusiasm, commitment and presentations; (4) a few were ambivalent; (5) with a considerable number being extremely negative, particularly (6) with regard to her large lecture classes in basic courses in sociology.

In *Yu Chuen Wei v. Vermont State Colleges Faculty Federation* (1995), moreover, they said, (19) "The statistical comparison demonstrates that Grievant was evaluated higher by students than [her male colleague] with respect to upper level classes, but that (20) [the male colleague] was evaluated higher than Grievant in lower level classes. Given (21) this "mixed" result, the statistical comparison of evaluations does not demonstrate by a preponderance of the evidence that Grievant's students rated her the same, or better, than [male colleague]" (p.305).

In *Dr. Brian Maclean v. President of The University of British Columbia* (1991), it was noted that (20) In general, the in-class peer reports were mixed but favourable. The in-class discussions were more problematic. (p.30). (21) While the knowledge, interest and enthusiasm of Dr. MacLean were acknowledged, "the problem appeared to be one of style or personality." It was further noted that (29) "As against the low figures, they disclosed a number of good qualities in Dr. MacLean---enthusiasm for his subject, wide knowledge of the literature, much out of class assistance to students, and a commitment to seeking good work from students. (p.31). (30) The reviewing faculty report noted the comments about Dr. MacLean's "derogatory manner, biased opinion, unwillingness to listen," were matched by "clear, stimulating, very helpful after class." And, (31) "some students have told us that the comments made were not representative of the class as a whole and were unduly influenced by the process" (p.41). (32) "A number of students, both from earlier years and from his current classes, furnished letters of support, and in preparation for the appeal, some furnished affidavits with respect to particular matters such as the 'intimidation' discussion in Soc. 250 and events in Soc. 490 and 520 in the fall of 1989." (p.33)

In *Robert Kramer v. The President of the University of British Columbia* (1992). (16) While a number of negative student comments were quoted in the department Head's letter, there were a number of very positive comments, and these were not mentioned at all. (25) "We have examined all of these written comments. There was a very wide range of comments. There were not 29 comments saying sarcastic and biased comments; but there were certainly 29 comments which included either cynical, sarcastic, biased, insulting, negative, condescending, belittling, opinionated, arrogant, nihilist, and destructive.... (29) However, it would only be fair to add that there were a number of comments in favour of Dr. Kramer, stating that the student "liked the course immensely," "now interested in Asian Studies"; "helps create a relaxed atmosphere," "really enjoyed him," "very approachable and knowledgeable," "very enthusiastic," "captivates audiences with his humour," "very effective" (p.12). (30) "In the other two courses, both small, both Japanese language, there were also some negative comments" (p.12).

In *Christopher Turner v. The President of the University of British Columbia* (1993), the board noted that (6) "While there is no question of Dr. Turner's competence as a teacher at all levels, teaching evaluations for the last several years show that his effectiveness is

marred by what students perceive as excessive formality, lack of enthusiasm and dullness....In a previous promotion attempt, his teaching was briefly described as "very competent" but student evaluations indicate further improvement to be "better than adequate" (p.2).

Summary: With regard to non numerically assessed written student comments, they are often qualitatively characterized as (41) a few were ambivalent, (42) a considerable number, (43) of mixed result, and selectively recognized: (44) it would only be fair to add that there were a number of comments in favor, (45) there were also some negative comments, (45) sometimes placing the greater weight on past evaluations of teaching over current comments, (47) sometimes placing greater weight on current comments over past positive evaluation of teaching.

Variables Not Taken Into Account When Assessing SEF

Student Biases in Affecting SEF

.....A significant issue is how courts view variables in assessing the reliability and validity of SEF is student bias affecting their evaluation of faculty.

In *Johnson v. University of Pittsburgh* (1977), the court noted that (10) "It has also been pointed out that in some cases difficult courses have to be given to the students and the material is such that it is difficult for even the best teacher to get it across.

In *Carley v. Arizona Board of Regents* (1987), he (7) characterized his professional style as being a "demanding teacher contrary to some student expectations," (8) Because of this, he maintained his popularity suffered and resulted in low student evaluations, (9) examination of his student comments indicated that Carley was correct in his assessment as 61% (49 out of 80) negative student comments focused on these values. The court ignored these findings.

In *Dr. Brian Maclean v. President of The University of British Columbia* (1991), it was noted that (21) While the knowledge, interest and enthusiasm of Dr. MacLean were acknowledged, "the problem appeared to be one of style or personality."

In *Robert Kramer v. The President of the University of British Columbia* (1992), the Board noted that (26) It was obvious that almost all of the classes were upset about an examination which was considered more geography than Asian Studies, and (27) they didn't like the marking. (28) They also felt the workload was far too heavy for an "introductory" course. The Board apparently only noted this variable.

Summary: Student bias variables include (48) being a demanding teacher, (49) thus thwarting student expectations, (50) difficult examinations (51) grading, (52) heavy workload in a course. (53) While most courts ignore these student biases in SEF, (54) occasionally a court will recognize that difficult courses have to be given to the students and that such material is difficult for even the best teacher to get the material across.

Other Variables

.....Cases concerned with the validity of SEF note various factors that were not controlled in the evaluation of the faculty process. Courts sometimes weigh these factors heavily, sometimes they do not.

In *Lieberman v. Grant* (1979), Lieberman attempted to introduce approximately ten personnel files concerning the tenure proceedings of other faculty in the English department for comparison. (6) Recognizing that such evidence would have had some minimal probative value, the Court, exercised its discretion under Fed. R.Ev. 403, and excluded it on the ground that "such probative value would be substantially outweighed by the delay and waste of time, which introduction of such evidence would have necessarily entailed....The plaintiffs case without such evidence seemed almost interminable, consuming 52 trial days over a two-year period. That is long enough" (p.873).

In *Fields V. Clark University* (1987) the court notes but does not admonish the non separation of student remarks from small seminar courses and those from large lecture classes.

In *Cynthia J. Fisher v. Vassar College* (1995), the district court found (2) that the biology department distorted Fisher's teaching recommendations by (3) "selectively exclud[ing] favorable ratings," by "focus[ing] on the two courses in which Dr. Fisher had difficulties" and (4) by "applying different standards to her than were applied to other tenure candidates" (p.1209).

In *Yu Chuen Wei v. Vermont State Colleges Faculty Federation* (1995), it was noted that (19) "The statistical comparison demonstrates that Grievant was evaluated higher by students than her [male colleague] with respect to upper level classes, but that (20) [male colleague] was evaluated higher than Grievant in lower level classes." Given (21) this "mixed" result, the statistical comparison of evaluations does not demonstrate by a preponderance of the evidence that Grievant's students rated her the same, or better, than [male colleague]" (p.305). Wei maintained that (16) her students rated her the same or higher than the male colleague's students rated him. The Board disagreed, saying, (19) "We note that the comparison offered by Grievant is somewhat weak since [male colleague] was tenured in 1988, and those student evaluations of his which were compared with Grievant post-dated his tenure review by a number of years" further saying, "we decline to hold such an involved comparison is necessary before a reasonable tenure determination can be made" (p.305).

In *Dr. Brian Maclean v. President of The University of British Columbia* (1991), the Board noted that (19) the reviewing faculty held in-class discussions about his teaching.

In *Robert Kramer v. The President of the University of British Columbia* (1992), Kramer argued that the most significant mistake was the failure to consider all aspects of his teaching. For example, only his teaching in 1989-90 was considered, whereas (9) he had taught a wide range of courses over the previous three years (10) had three new courses that year, (11) plus a graduate course. Moreover, (17) The department head indicated that his teaching was not up to the departmental "standard." The standard appeared to be the performance of the tenure-track faculty, though Kramer was one of the most junior faculty members (p.8). (15) Only one of the more than thirty numerically rated questions was used: "Rate instructor bad to good." (16) While a number of negative student comments were quoted in the department Head's letter, there were a number of very positive comments, and these were not mentioned at all.

In *Christopher Turner v. The President of the University of British Columbia* (1993), the Dean said, "there were few students in undergraduate literature courses since 1986/7---(3,8, and 6 respectively," thus mistaking student 'response' figures for actual

student enrolment. The Board concluded that (5) "This misunderstanding is in our opinion sufficient in itself for a reconsideration, since teaching was the focus..." (p.3), and (7) "we think that the comments and emphasis on the size of Dr. Turner's classes as evidence of poor teaching are open to objection and constitute errors of procedure and/or evidence" (p.6).

Summary: The variables noted in these cases include, (55) not controlling for class size, i.e., those obtained in small seminars from those obtained in large lecture classes, (56) those obtained from tenured faculty from those obtained from non tenured junior faculty, (57) not performing appropriate comparisons of SEF with other faculty, (58) noting SEF in all courses, not just to problem courses, (59) not mistaking student 'response' figures for actual student enrolment figures when using them to determine student attraction to a course, (60) using all courses taught, (61) taking into consideration the faculty teaching a wide range of courses, versus those with lighter teaching loads, (62) number of new courses taught in a year, (63) whether graduate courses were taught at the same time as teaching undergraduate courses, (64) selectively mentioning only negative student comments, or (65) overly weighting negative comments, and (66) different procedures for gathering student opinion.

Courts Impact On Academic Matters

Impact on Academic Standards

.....It has historically been the case that faculty as a collective body and the institution set the expected level of student performance.

In *William Sypher v. Vermont State Colleges Faculty Federation* (1982), it was observed (9) With regard to the "political" aspect of the case, Sypher had written a letter in defense of his student rating level at the college which said, "it is certainly distressing when very good is not good enough, especially at a college with a modestly-talented student body that often discourages efforts at subtlety, wit and deeper penetration of subjects." (10) The Board responded to this letter saying, "other actions and statements by Grievant constituted legitimate reasons for not retaining him. In a May, 1980, letter to, Dean Beston, Grievant expressed his contempt for Castleton students" (p.135), (11) concluding, "Accordingly, we find credible the College's contention that Grievant was not reappointed because of his teaching effectiveness. [Italics added] (p.135).V.

In *Lovelace v. Southeastern Massachusetts University* (1986), the court said, (4) A school sets itself up to attract and serve only the best and the brightest students or whether it instead gears its standard to a broader, more average population is a policy decision which, we think universities must be allowed to set....matters such as course content, homework load, and grading policy are core university concerns (p.424).¹⁹

In *Robert Kramer v. The President of the University of British Columbia* (1992), the Board said (30) "One perceptive student noted that some of the unhappiness came from the fact that the levels of Japanese language ability were badly divided; some found it easy, others very hard" (p.12).

Summary: From the above limited cases, the courts have clearly said (67) universities must be allowed to set (68) standards, including (69) course content, (70) homework load, and (71) grading policy.

Methods of Instruction and Academic Freedom

.....It is generally accepted that academic freedom clearly pertains to speech in the classroom. In like manner, most faculty seem to believe that academic freedom also pertains also to how a course is taught. How do courts view teaching method in relation to SEF an academic freedom?

In *Lovelace v. Southeastern Massachusetts University* (1986), the court's ruled, (1) "It is important to note what plaintiffs first amendment claim is and to separate speech from action. Plaintiff has not contended that he was retaliated against simply because he advocated that the university elevate its standards.... Plaintiffs complaint instead is that he was retaliated against when he refused to change his standards" (p.425), (2) citing other cases, the court rejected his contention that university teacher has first amendment right to disregard established curriculum content, that the first amendment does not prevent a university from terminating an untenured faculty whose pedagogical style and philosophy does not conform to those of the school's administration, (3) "is a policy decision which, we think, universities must be allowed to set" (p.426). Further, the court ruled that (4) "We will assume for purposes of this opinion that plaintiffs refusal to lower his standards was a substantial motivating factor, (see *Mount Health Board of Education v. Doyle*, 429 U.S. 274, 283-284, 97 S.Ct. 568, 574-575, 50 LEd.2d 471 (1977), in the decision not to renew his contract.

In *Carley v. Arizona Board of Regents* (1987), (4) he claimed as "protected speech" his teaching methods where his goal in his commercial art course was to promote a business atmosphere by requiring attendance, promptness, and self-reliance, and required them to meet deadlines. The court ruled, (19) his teaching style is not a form of speech protected under the First Amendment. (20) Decision not to retain a non tenured instructor, even if based, in part, upon student evaluations expressing disapproval of his teaching methods, did not violate instructor's First Amendment right to academic freedom; (21) Carley was not denied a contract because of expressing unpopular opinions or otherwise presenting controversial ideas to his students. (22) Thus, we conclude that the decision not to retain Carley, even if based, in part, upon student evaluations expressing disapproval of his teaching methods, did not violate his first amendment rights. (p.1103).

Summary: Teaching method (72) because it is an action, not speech, (73) is not a form of free speech, nor (74) covered under academic freedom, (75) except if noted in specific contractual faculty agreements.

Release of SEF To Students and To the Public

.....In addition to the administrative use of SEF, in recent years, other uses of SEF have become controversial, including releasing SEF to students and to the public.

University of Idaho: A legal ruling, cited on the World Wide Web site of the *Topical Interest Group: Assessment in Higher Education* (Evaluating teacher evaluations, 1996), [20](#) notes that the University of Idaho also recently went to court over the issue of whether SEF can be published. The student newspaper initiated a lawsuit when it was refused access to SEF for publication. The legal question was whether SEF is protected under privacy rights by the Idaho Code. In a ruling that seems to strain logical credulity the court ruled that since the University did not consider students as the general public, the University was not breaking the law by allowing students access to the evaluations.

Further, the opinion of the court was that according to State law, teacher evaluations are not protected as part of personnel records.

University of Wisconsin: In another ruling, the Chancellor at the University of Wisconsin refused to release SEF, citing a statute that disallows personnel evaluations from being released to public view. Students took the chancellor to court. However, after being advised by the state's Attorney General, citing Wisconsin's open-records law, the University of Wisconsin's campus released SEF to the public. Both the student and faculty senates, 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).

Summary: (76) Unlike most personnel records, SEF can be released to students and the public, on the grounds, that (77) students are not considered the general public, and (78) that SEF are public records and withholding them from public access does not outweigh the public interest in them.

Conclusion

.....From the above cases, it is clear that "a view from the courts" on SEF does not mean a coherent set of rulings. Given the lack of legal attention to SEF, this is not surprising. A collectively incoherent set of rulings, however, is not restricted to the concept of SEF but for the time-honored concept of academic freedom as well. As Kaplin and Lee (1995) conclude,

The foregoing discussions have made clear that academic freedom is an area in which the law provides no firm guidelines for administrators. This is particularly true for private institutions, since the decided cases are almost all constitutional decisions applicable only to public schools. Even the constitutional cases are sometimes incompletely reasoned or difficult to reconcile with one another. The fact that decisions often depend heavily on a vague balancing of faculty and institutional interests in light of the peculiar facts of the case makes it difficult to generalize from one case to another (p.192).

When we speak of "the courts," then, we are speaking of a hydra with some common but many disparate voices.

.....Given the above rulings and the courts propensity to accept faculty/institutional agreements, it would seem clear as Kaplin and Lee advise, that, "it is especially crucial for institutions to develop their own guidelines on academic freedom and to have internal systems for protecting academic freedom in accordance with institutional policy" (p. 192). This would appear to be especially true for SEF policy.

.....From most of the above cases---even given that, as challengers, the burden of proof has been on faculty --- it seems clear that the courts have not been kind to faculty with regard to student evaluations.²¹ It is the burden of the two following papers to explore, analyze, and comment on the implications of the views from the court presented here.

Notes

1. Address correspondence to: Robert E. Haskell, Ph.D. Professor of Psychology, Department of Social and Behavioral Sciences, University of New England, Biddeford, ME 04005. Email: rhaskell@maine.rr.com. I would like to thank Mr. Nicholas DiGiovanni, Jr., attorney at law, of Morgan, Brown & Joy, Boston, MA, Mr. Gary Founds of College Legal Information, Inc., especially Professors William A. Kaplin, School of Law, Catholic University of America, and John Damron, of Douglas College for continually providing me with sources and support.

[\[BACK to document\]](#)

2. The issue of needing some form of faculty performance evaluation is an immediate and pragmatic one that can not be dealt with here. If used correctly (see Copeland and Murry, 1996; Kemp and Kuman, 1990; Scriven, 1995, 1993, 1991,1988; 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 may be having problems with his or her teaching and/or relations with students.

[\[BACK to document\]](#)

3. In addition to there having been no categorical mention on the professional level of SEF as an infringement on academic freedom, the reasons for SEF not being examined as an infringement on academic freedom and tenure rights are that many faculty (1) are often embarrassed to admit that student evaluations may influence their professional behavior in the classroom, (2), seem to accept that to question the right of students to evaluate faculty is considered unprofessional if not undemocratic, (3) seem also to accept that to question the right of students to evaluate faculty is seen as self serving, if not downright mean spirited, and (4) see research on such SEF as not high status research.

Based on a review of statements in the literature, in a previous paper I have, outlined faculty attitudes on SEF impinging on tenure and academic freedom (Haskell, 1997). 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 selected standards (Young, 1993), (3) create pressure for lowered teaching standard (Bonetti, 1994), (4) are responsible for a considerable amount of grade inflation (Greenwald, 1997), (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 evaluations (Damron, 1996), (8) contrary to their original intent of improving instruction, do not eliminate poor or below-average instructors but instead increases poor teaching practices (Carey, 1993; Barnett,1996), (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). Finally, SEF creates an educational conflict of interest between faculty and student that negatively impacts on the quality of instruction. [See also Furedy, 1995.] [\[BACK to document\]](#)

4. Based on a review of statements in the literature, in a previous paper I have, outlined faculty attitudes on SEF impinging on tenure and academic freedom (Haskell, 1997). 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 selected standards (Young, 1993), (3) create pressure for lowered teaching standard (Bonetti,

1994), (4) are responsible for a considerable amount of grade inflation (Greenwald, 1997), (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 evaluations (Damron, 1996), (8) contrary to their original intent of improving instruction, do not eliminate poor or below-average instructors but instead increases poor teaching practices (Carey, 1993; Barnett, 1996), (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). Finally, SEF creates an educational conflict of interest between faculty and student that negatively impacts on the quality of instruction.

[\[BACK to document\]](#)

5. Increasingly, SEF are becoming a primary and single factor in denial cases.

[\[BACK to document\]](#)

6. I had a Lexis Nexus search conducted but with only a couple of usable cases found. I have been in contact with 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.

[\[BACK to document\]](#)

7. The term "court" as used here will also include rulings by state Arbitration Boards. Since this paper is only concerned with how ruling bodies view SEF data, this paper will not distinguish between state courts, federal courts or arbitration boards.

[\[BACK to document\]](#)

8. Because the concept of context in itself is complex, it would take this paper too far afield to attempt to explicate why certain case data were included while other omitted.

[\[BACK to document\]](#)

9. I will explore the consequences of this issue for SEF my third paper.

[\[BACK to document\]](#)

10. The weight given to SEF of faculty vary according to the institution, with research oriented institutions not according them the weight that teaching oriented institutions do. The latter, of course, constitute the vast majority of campuses.

[\[BACK to document\]](#)

11. 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 if SEF feedback to faculty if used in a carefully constructed collegial consultation process. A condition which seldom occurs.

[\[BACK to document\]](#)

12. In the appendix, I have provided an abstracting of each case relative to its SEF content. What is needed are legal scholars to locate and review a larger data set and to perform a statistical content and comparative analysis of court rulings on SEF relative to different contextual variables.

[\[BACK to document\]](#)

13. While there are certainly insufficient cases cited here to draw any firm conclusions, it does appear that at least in relations to SEF that Canadian Boards seem to take a more critical look at their use than do courts in the U.S. As I understand it, arbitrations are adjudicated by a panel made up of a arbitrator who is a lawyer and is officially recognized by the government; typically, two other lawyers are selected by the university the Faculty Association. This procedure is circumscribed in provincial law and decisions rendered by the panel have legal standing. I would like to specifically thank Dr. John Damron of Douglas College for providing me with all of the Canadian cases with the assistance of the Canadian Association of University Teachers (CAUT).

[\[BACK to document\]](#)

14. While many educational issues in Canada and the U.S., are quite similar (see Emberley, P. C. & Newell, W.R. 1994; Emberley, P. C., 1995), it is well known that Canadian jurisprudence does not define free speech and academic freedom as broadly as does the U.S., system. While there are certainly insufficient cases cited here to draw any conclusions, it does appear that at least in relations to SEF that Canadian Boards seem to take a more critical look at their use than do courts in the U.S.

[\[BACK to document\]](#)

15. For reference convenience, the numbers in parentheses have been retained and correspond to the numbers in the abstracted appendices relative to the specific case.

[\[BACK to document\]](#)

16. The numbers in these brief summaries will be cumulative throughout the following summaries and a cumulative summary is presented in the appendix.

[\[BACK to document\]](#)

17. The problem, however, is assessing the validity of SEF data. At this time, state-of-the-art analysis of SEF Barnett, 1996; Greenwald, 1997; Greenwald & Gillmore (1996) for state of the art statistical analyses of previous studies suggesting their validity.

[\[BACK to document\]](#)

18. In general, courts have tended to only require precise accuracy in cases where EEOC issues are involved. I will address this issue in more depth in a following paper.

[\[BACK to document\]](#)

19. An important issue here is: who is the university? Faculty or administration. According to Kaplin (personal communication, April 1, 1997) "When courts seem to equate the university administration or a single administrator with the "university," it is generally because the university's governing body has delegated the pertinent authority to the administration or administrator." I will be exploring the implications of this issue in more detail in Part IV.

[\[BACK to document\]](#)

20. See <http://marsquadra.tamu.edu/TIG/GeneralArticles/evaluatingteacherevaluations.ht>.

[\[BACK to document\]](#)

21. As one scholar (Damron, personal communication, April, 1997) who read a draft of this paper observed: from

"the legal decisions you review in your paper it is clear that untenured and/or politically incorrect faculty are often considered to be "fair game" by administrators, with literally _any_ superficially plausible excuse serving as a rationale for dismissal. Use of such

strategies reveal that faculty are often regarded as little more than term employees who are as disposable (and replaceable) as tissues. Clearly, there is a very serious ethical issue here, and a hugely hostile attitude toward academic freedom and faculty in general...the great variety of decisions you've reviewed and their assorted implications for the coherence and ethics of the legal processes that gave rise to them...it seems to me that many judges and arbitration panelists have little sense of how to proceed in hearing[s] involving academics."

[\[BACK to document\]](#)

22. In general, courts have tended to only require precise accuracy in cases where EEOC issues are involved. I will address this issue in more depth in the fourth and final paper of this series.

[\[BACK to document\]](#)

References

- Allegre, M., and Guista, M. (1993). What every English teacher ought to know about using student evaluations. *Teaching English in the Two-year College*, 20(1), 57-63.
- Baez, B, and Centra, J. (1995). Tenure, promotion, and reappointment: Legal and administrative implications. *Ashe-eric Higher Education Report*, No. 1.
- Barnett, L.D. (1996). Are teaching evaluation questionnaires valid? Assessing the evidence. *Journal of Collective Negotiations in the Public Sector*, 25(4), 335-349.
- Benson, D.E., and 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, A21.
- Bonetti, S. (1994). On the use of student questionnaires. *Higher Education Review*, 26, 57-64.
- Brian Maclean v. President of The University of British Columbia (1991). Held at Vancouver, B.C. January 28--June 20.
- Carey, G. W. (1993). Thoughts on the lesser evil: Student evaluations. *Perspectives on Political Science*, 22, 17-20.
- Carley V. Arizona Board of Regents, 737 P.2d 1099 (Ariz. App. 1987).
- 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.
- Christopher Turner v. The President of the University of British Columbia (1993). Arbitration Appeals Board Case No.: AI-256 Vancouver, B.C. July 19,20 and 21.
- 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

Chuen Wei and the Vermont State Colleges Faculty Federation, 18 VLRB 261 (1995)

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

Crumbly, L.D., and Fliedner, E. (1995). *Accounting administrators' perceptions of student evaluation of teaching (SET) information*. Manuscript, Department of Accounting, Texas A&M University.

Cynthia J. Fisher v. Vassar College (1995). United States Court of Appeals for the Second Circuit Nos. 1179, 1303, 2275 Docket Nos. 94-7737, 94-7785, 94-9125

Damron, J.C. (1996). *Instructor personality and the politics of the classroom*. Manuscript, Douglas College, New Westminster, British Columbia, Canada V3L 5B2. (An Earlier Version 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). Available online at http://vax1.mankato.msus.edu/~pkbrandon/Damron_politics.html

Damron, (1997). Personal communication.

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

Dyson V. Lavery, 417 F.supp. 103 (E.d. Va. 1976).

Evaluating teacher evaluations. (1996). This article submitted by glen9579@novell@uidaho.edu on 11/12/96. <http://marsquadra.tamu.edu/TIG/Forum.html>. (See, Topical Interest Group: Assessment in Higher Education at <http://marsquadra.tamu.edu/TIG/GeneralArticles/evaluatingteacherevaluations.ht>)

Fields V. Clark University, Civil Action No. 80-1011-s, United States District Court for the District of Massachusetts, 40 Fair Empl. Prac. Cas. (Bna) 670, March 14, 1986; Vacated and Remanded May 8, 1987.

Furedy, John J. (1995). Academic freedom, opinions and acts: The voltaire-mill perspective applied to current canadian cases *University of New Brunswick Law Journal*, 44, 131-4.

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

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, submitted for publication.

Greenwald, A., and Gillmore, G. (1997). No pain, no gain? The importance of measuring course workload in student ratings of instructions. *Journal of Educational Psychology* (forthcoming).

Haskell, R.E. (1997). Academic freedom, tenure, and student evaluation of faculty: Galloping

polls in the 21st century. *Education Policy Analysis Archives*, 5(6), [Refereed Online Journal]
Available: <http://olam.ed.asu.edu/epaa/v5n6.html>

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

Johnson V. University of Pittsburgh, 435 F.supp. 1328 (W.d. Pa. 1977).

Kaplin, W.A. (1997). Personal communication. January 20th, School of Law, Catholic University of America.

Kaplin, W.A., and Lee, B. (1995). *The law of higher education: A comprehensive guide to legal implications of administrative decision making (3rd Ed)*. San Francisco: Jossey Bass.

Kemp, B.W., and Kuman, G.S. (1990). Student evaluations: Are we using them correctly? *Journal of Education for Business*, 65, (November/December). 106-11.

Lieberman V. Grant, 474 F.supp. 848 (D.conn. 1979).

Lovelace v. Southeastern Massachusetts University, 793 F.2d 419 (1st Cir.1986).

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

Parini, J. (1995, July 14). Tenure and the loss of faculty talent. *The Chronicle of Higher Education*, Section: Opinion, POV.

Peters V Middlebury College, 409 F.supp. 857 (D. Vt. 1977).

Poch, R.K. (1993). Academic freedom in American higher education: Rights, responsibilities, and limitations. *Ashe-eric Higher Education Report No. 4*.

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

Robert Kramer v. The President of the University of British Columbia (1992). Arbitration Appeals Board Case No. AI-245. Held at the University April 22 and 23.

Scriven, Michael. (1988). The new crisis in teacher evaluation: The improper use of 'research-based' indicators. *Professional Personnel Evaluation News*. (January) p.4.

Scriven, M. (1991). *Evaluation Thesaurus*. (4th Ed.) Newbury Park, CA : Sage Publications.

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

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

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

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>

University of Regina V. University of Regina Faculty Association and Dr. Pradeep Jalan (1993). Case No.: AI-298 Regina, Saskatchewan, July 9.

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

William Sypher v. Vermont State Colleges Faculty Federation, 5 VLRB 102 (1982).

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.

Appendix A: Summary of 78 Findings

.....From the cases reviewed, it can be seen that court rulings range from saying that (1) relying primarily or solely on student evaluations is acceptable, to (2) placing little exclusive reliance on SEF, (3) in rare cases SEF can not be permitted to stand in the way of promoting or retaining professors who are excellent in non teaching areas, (4) tenure decisions can not be based solely on SEF by students who have not been made aware of the ramifications of their evaluations, (5) anonymous documents or those "based on hearsay" should not be included in a faculty member's file, (6) students should be made aware of the purpose and ramifications of their evaluations of faculty, (7) anonymous student evaluations should not be used, (8) peer evaluations must also be a part of evaluating teaching, that (9) in cases of exceptional research faculty that popularity should not play a role in termination due to teaching, to (10) in normal cases that a measure of popularity is related to teaching effectiveness.

.....If (11) the courts and educational administrations can not allow low SEF to stand in the way of promoting or retaining professors who may be world renowned scientists, (12) deemed nationally or internationally exceptional as a researcher courts may disregard SEF, (13) but at least in these two cases they did not find the faculty exceptional. With regard to establishing the general and statistical accuracy of SEF in the above cases, the range of opinion is from (14) statistical analyses may be a part of a plaintiff's effort to establish discriminatory treatment if it reached proportions comparable to those in cases establishing a prima facie racial discrimination, (15) cautioning that statistics are not irrefutable, with their usefulness depending the surrounding facts and circumstances of a case, (16) the Court maintaining that it need not consider and is under no obligation to establish the accuracy of administrative interpretations of SEF, (17) that tenure criteria are not drawn with "mathematical nicety," (18) administrators failure to perform statistical comparisons is not arbitrary and is reasonable, (19) especially if such is not required by a Faculty Association

Contract, (20) nearly any use made of SEF is acceptable if it followed the standard practice of the university, (21) that creativity, rapport with students and colleagues, teaching ability, and other qualities are intangibles which cannot be measured by objective standards.²²

.....Numerical scores often result in faculty (22) being compared relative to other faculty, (23) being ranked relative other faculty, (24) with distinctions often being made on the basis of tenths of a decimal, (25) with most courts accepting these fine decimal distinctions. Courts have tended to accept administrative subjective judgements if (26) they are deemed sincere (27) grounded on some evidentiary basis (28) if made on the "vigor and variety of student criticisms" (29) "not arbitrary or capricious and were exercised honestly upon due consideration," (30) based upon "much experience in reviewing student evaluations, (31) reasonably draw on that experience (32) and have ruled that Presidents are not bound by factual findings made by majority members of a faculty. The use of student comments ranges from (33) placing importance on a single comment (34) to several comments as significant information, (35) maintaining that statistical analyses of SEF need to be bolstered by individual comments, (36) maintaining that while some very negative---e.g., racist, sexist---comments may be found, the court may find that they do not rendered the SEF unreliable, (18) that such instances or "impressions" may be validated after the fact, (37) negative comments often seem to outweigh positive one, and (38) may often outweigh numerical data to he contrary, (39) negative comments need not be verified before acting on them, (40), that negative comments can not be used to undermine otherwise generally favorable comments received in an annual performance review.

.....With regard to non numerically assessed written student comments, they are often qualitatively characterized as (41) a few were ambivalent, (42) a considerable number, (43) of mixed result, and selectively recognized: (44) it would only be fair to add that there were a number of comments in favor, (45) there were also some negative comments, (45) sometimes placing the greater weight on past evaluations of teaching over current comments, (47) sometimes placing greater weight on current comments over past positive evaluation of teaching. Student bias variables include (48) being a demanding teacher, (49) thus thwarting student expectations, (50) difficulty examinations (51) grading, (52) heavy workload in a course. (53) While most courts ignore these student biases in SEF, (54) occasionally a court will recognize that difficult courses have to be given to the students and that such material is difficult for even the best teacher to get the material across.

.....Some of the variables noted in cases include, (55) not controlling for class size, i.e., those obtained in small seminars from those obtained in large lecture classes, (56) those obtained from tenured faculty from those obtained from non tenured junior faculty, (57) not performing appropriate comparisons of SEF with other faculty, (58) noting SEF in all courses, not just to problem courses, (59) not mistaking student 'response' figures for actual student enrolment figures when using them to determine student attraction to a course, (60) using all courses taught, (61) taking into consideration the faculty teaching a wide range of courses, versus those with lighter teaching loads, (62) number of new courses taught in a year, (63) whether graduate courses were taught at the same time as teaching undergraduate courses, (64) selectively mentioning only negative student comments, or (65) overly weighting negative comments, and (66) different procedures for gathering student opinion. Courts sometimes weigh these factors heavily, sometimes they do not (67) universities must be allowed to set (68) standards, including (69) course content, (70) homework load, and (71) grading policy Teaching method (72) because it is an action, not speech, (73) is not a form of free speech, nor (74) covered under academic freedom, (75) except if noted in specific contractual faculty agreements, (76) since students may not be considered the general public students may have access SEF, (77) in some State law, teacher evaluations are not protected as part of personnel records. (78) SEF are public records because withholding access does not

outweigh the public interest.

Appendix B: Summary of Case Outcomes

I. Dyson v. Lavery, (1976).

.....Patricia Dyson, a lawyer, in not renewing her contract, filed suit alleging sex and salary discrimination in both her initial employment and in the decision not to renew her contract. The Court concluded that Dyson's termination was not the product of sex discrimination.

II. Johnson v. University of Pittsburgh, (1977).

Sharon Johnson, at the end of her second three-year contract as an assistant professor was denied promote to the rank of associate professor and wan not granted tenure. She contended that her termination was based on sex discrimination. The University of Pittsburgh, claimed that the denial was a result of Johnson's inadequacy as a teacher and as a researcher and writer.

.....The court found that plaintiff's teaching was not adequate, upholding the university decision.

III. Peters v. Middlebury College, (1977).

.....Joan Peters file suit for sexual discrimination after being denied reappointment on the basis of her scholarship and teaching. Court found that she did not merit tenure, but was discriminated against.

IV. Lieberman v. Grant, 474 F.Supp. 848 (D.Conn. 1979).

.....Marcia R. Lieberman, denied tenure, claimed sex discrimination. The court concluded that sexual prejudice did not play any significant role in the tenure proceedings.

V. William Sypher v. Vermont State Colleges Faculty Federation (1982).

.....Sypher contented his denial of reappointment after his third year violated his academic freedom rights as stated in the faculty collective bargaining agreement. The faculty contended he was not reappointed because of his "political" activity on campus. The college contented he was not reappointed because of his inadequate teaching evaluations. The court found credible the College's contention that Grievant was not reappointed because of his teaching effectiveness.

VI. Lovelace v. Southeastern Massachusetts University, 793 F.2d 419 (1st Cir.1986).

.....Matthew Lovelace, a non tenured faculty challenged the decision not to renew his contract based on various constitutional claims, including a claim that the use of student ratings violated his academic freedom. He contended that his contract had not been renewed because he was a demanding teaching. When he refused to change his standards, his contract was not renewed. The court upheld the university.

VII. Fields V. Clark University (1987).

.....Fields contended she was being discriminated against on the basis of her sex. The university claimed she was denied reappointment because of inadequate teaching. The court ruled in Fields' favor.

VIII. Carley v. Arizona Board of Regents, (1987).

.....Denny Carley was denied reappointment largely on the basis of student evaluations of his teaching. This appeal is from a superior court judgment affirming a decision of Northern

Arizona University President to deny the renewal of a teaching contract to Carley. The issues on appeal involved: (1) whether his right to academic freedom was violated because student evaluations were utilized as the primary tool to determine his teaching effectiveness. The court concluded that the decision not to retain Carley, even if based, in part, upon student evaluations expressing disapproval of his teaching methods, did not violate his first amendment rights.

IX. Robert Kramer v. The President of the University of British Columbia (1992).

.....Kramer, an Assistant Professor, was denied his first two-year reappointment. He maintained that there were errors of procedure, failures to obtain relevant evidence, and that the denial was based on a defective dossier. The Board upheld Kramer.

X. Christopher Turner v. The President of the University of British Columbia (1993).

.....Christopher Turner was denied promotion to full professor. The board concluded that (11) there were sufficient errors of procedure and/or evidence to return the case for reconsideration.

XI. University of Regina Faculty Association v. University of Regina (1993).

.....Pradeep Jalan, an Associate professor in his fifth year at the University. Denial of tenure was on the basis that he had not met the standard for an Associate Professor in the areas of scholarly performance and teaching. The Board ruled that the evidence relied upon by the employer to suggest that the standard was not met, in our view, was not sufficient.

XII. Cynthia J. Fisher v. Vassar College, (1995).

.....Cynthia Fisher was denied tenure. The college maintained the denial was because of her performance. Following a bench trial, the district court found that the college had discriminated against her by reason of her sex and age.

XIII. Yu Chuen Wei v. Vermont State Colleges Faculty Federation (1995).

.....Yu Chuen Wei was denied tenure. The case involved SEF as a part of a wider issue of racial discrimination. The Board concluded that Wei did not established discrimination. The Board concluded that Wei had not established that the tenure decision in her case was unreasonable, arbitrary or based on erroneous reasons.

XIV. Guam Federation of Teachers v. The University of Guam (1990).

.....

XV. Dr. Brian Maclean v. President of The University of British Columbia (1991).

.....Denied reappointment after his first two-year term. The court upheld the university decision.

Appendix C: Abstracted Case Material

.....From each case, I abstracted out the material that was relevant to teaching effectiveness and SEF.

I. Dyson v. Lavery, (1976).

.....Patricia Dyson, a lawyer, hired to teach in the Business Department at Virginia Polytechnic Institute (VPI) in February, 1970, When VPI failed to renew her contract, she filed suit against the president of the university alleging sex and salary discrimination in both

her initial employment and in the decision not to renew her contract. The court carried out a complete statistical analysis of salary and promotion by gender.

.....In terms of her teaching, the court said, "the Business Department Administrators testified that (1) the input received through normal channels for Ms. Dyson's teaching performance had been generally negative. A number of students apparently had voiced displeasure over the quality of her class preparation and presentation" (p. 111) and (2) she refused to follow normal procedures not posting student grades; (3) "These impressions" said the court, "were largely confirmed after the initial decision to not rehire her had been made, by a student evaluation that ranked her 46th of 48 teachers in the Business Department." Further, it was ruled that, (4) "The Court need not consider the accuracy of these administrative determinations, for it concludes that they were sincere and grounded on some evidentiary basis" (p.111); and (5) "In the absence of a finding that same were sexually motivated, the administration's professional judgment must be respected" (p.111 all italics added). The Court concluded that Dyson's termination was not the product of sex discrimination.

II. Johnson v. University of Pittsburgh (1977).

.....Sharon Johnson was an assistant professor at the University of Pittsburgh in the Biochemistry Department of the School of Medicine. At the end of her second three-year contract as an assistant professor, the university decided not to promote her to the rank of associate professor and not to grant her tenure. She contended that her termination was based on sex discrimination. The University of Pittsburgh, on the other hand, claimed that the denial of promotion and tenure was a result of Johnson's inadequacy as a teacher and as a researcher and writer.

.....The tenured faculty (1) described her teaching as "disastrous" and stated she demonstrated "a virtually complete inability to organize and convey a substantive body of information to students"; (2) the court noted they "approached this question of teaching ability with considerable doubt, in view of the fact that in prior years there does not appear to have been any criticism of her teaching and also in view of the fact that at...there was evidence that the department chairman, had informed her after one of her lectures in 1971 what a great lecture it had been"; On the other hand, the court said, (3) "we have the instance referred to in Finding 27 and the fact that in recent years increasing weight has been placed upon teaching ability"; (4) "We further have an evaluation by another faculty who impressed the court as completely impartial and credible who stated...that plaintiff's teaching performance was inadequate"; in addition, (5) "The court has placed little reliance on students' surveys. The value of these surveys was deprecated by the Dean and it was further testified that student surveys always result in the students in a given course rating a teacher, or professor, some of them as excellent, others as terrible and in between many who say passable, mediocre etc"; (6) "While the court is reluctant to rely upon student's surveys as demonstrating the fitness or unfitness of a faculty member nevertheless we cannot say it was unreasonable for the tenured faculty to consider this along with other matters in the meeting" (p.1359).(7) Citing a previous case (*International Brotherhood of Teamsters v. U. S. -- U.S. --, 97 S.Ct. 1843, 52 L.Ed.2d 396 (1977)*, where in a class action case the court cited *McDonnell Douglas* for the use of statistics) went on to say: "We have repeatedly approved the use of statistical proof where it reached proportions comparable to those in this case to establish a prima facie case of racial discrimination in jury selection cases . . . Statistics are equally competent in proving employment discrimination. We caution only that statistics are not irrefutable. They come in an infinite variety and, like any other kind of evidence they may be rebutted. In short, their usefulness depends on all of the surrounding facts and circumstances" (8) The court further in Footnote # 20 said: "Considerations such as small

sample size may of course detract from the value of such evidence" (p.1361); the court also noted, "It is obvious that a professor may be possessed of excellent qualifications as a research scientist and not necessarily be able to prove his or her worth as a teacher. It is also obvious that the court and the administration of universities cannot permit students to exercise a veto over professors who may be world renowned scientists and yet if the students rate them unfavorably can be terminated at any time because of unpopularity"; concluding that, (9) "in cases where one has an outstanding scientist of national or international reputation, one may decide to promote and give tenure notwithstanding inability to come across as a teacher, this however is not one of those cases. Finally, it said, (10) "It has also been pointed out that in some cases difficult courses have to be given to the students and the material is such that it is difficult for even the best teacher to get it across. (11) The court finds the weight of the evidence is that plaintiff's teaching was not adequate and while there are some allegations to the contrary and there had previously been no criticism of her performance, nevertheless the court cannot say that the tenured faculty was unreasonable in considering this factor and in arriving at the conclusion they did" (p.1366-7).

III. Peters v. Middlebury College, (1977).

.....Joan Peters file suit against Middlebury College for sexual discrimination after her being denied reappointment on the basis of her scholarship and teaching. The basic reason for the denial of reappointment was that she did not meet the standards of professional competence, especially in the upper division courses to which the English department aspired.

.....With regard to her (1) student evaluations, they were varied but with support for Peters predominating, but it was said that the students "ability to evaluate in the advanced and specialized areas of the English department was not established" (p.867). (2) The department chair sent a letter to the president of the college, saying, "The course of action I recommend is not likely to be popular with students who, though they in part recognize her intellectual limitation, are warmly responsive to her enthusiasm, energy, openness and ready human concern" (p.860). (3) The court ruled that "the reasons given by the institution for non promotion showed that by a preponderance of the evidence she was not professionally qualified for tenure and that the reasons given were not pretextual, and that there was substantial evidence to show that plaintiff's teaching was not satisfactory....(4) An evaluation of one's teaching ability is necessarily a matter of judgment. (5) A professor's value depends upon his creativity, his rapport with students and colleagues, his teaching ability, and numerous other intangible qualities which cannot be measured by objective standards" (p.860). Court found that she did not merit tenure, but was discriminated against.

IV. Lieberman v. Grant, 474 F.Supp. 848 (D.Conn. 1979).

.....Marcia R. Lieberman was hired by the University of Connecticut English department. Among her allegations was sex discrimination, but prior to her tenure review, she had been repeatedly warned by her department and its personnel committee that she had inadequacies in both teaching and scholarship.

.....(1) A five member departmental promotion and tenure committee all voted for tenure on the initial preliminary vote. After (2) being sent to a joint faculty committee on her department chair's negative recommendation (3) based on complaints received from "several students", to the effect that Lieberman's interest in feminism caused her to ignore other themes in literature, on close vote she was denied tenure; (4) a compilation of student ratings was showed that the cumulative ratings for members of the department ranged from a low of 4.09 to a high of 8.95. She had a cumulative rating of 7.06, which ranked her 12th out of the 15 junior faculty members. The 7.06 figure included the ratings from a previous semester in which the plaintiff received a rating of 8.18. Prior to this rating in the spring of 1972, the

plaintiff's cumulative rating was 6.7; (5) Lieberman also attempted to introduce approximately ten personnel files concerning the tenure proceedings of other faculty in the English department for comparison; (6) Recognizing that such evidence would have had some minimal probative value, the Court, exercised its discretion under Fed. R.Ev. 403, and excluded it on the ground that "such probative value would be substantially outweighed by the delay and waste of time, which introduction of such evidence would have necessarily entailed....The plaintiffs case without such evidence seemed almost interminable, consuming 52 trial days over a two-year period. That is long enough" (p.873). The court concluded that sexual prejudice did not play any significant role in the tenure proceedings.

V. William Sypher v. Vermont State Colleges Faculty Federation (1982).

.....A faculty member contended that he was denied reappointment by Castleton State College, after his third year violated his academic freedom rights as stated in the faculty collective bargaining agreement. The faculty contended he was not reappointed because of his "political" activity on campus. The college contended he was not reappointed because of his inadequate teaching evaluations.

.....(1) Some of the student comments noted that, "When students try to disagree he shoots you down and trys to degrade you in front of the class," (p.115), while others said, "encourages student participation as much as possible... encourages student to express their ideas freely and not worrying how 'dumb' it may sound...always wants you point of view." (P.115) (2) With regard to the numerical ratings, the Board's opinion was that (3) regardless of a strong majority of student's rating his teaching as above average, (4) the existence of a significant minority of students feeling degraded, humiliated, and embarrassed can reasonably lead an evaluator to question a teacher's effectiveness; (5) the negative student evaluations are buttressed by the Chair of the Education Department; (6) were it not for this incident, it would be difficult for the Dean and President to support their subjective opinion that Sypher was not substantially above average in teaching effectiveness, given a statistical analysis of student evaluations demonstrating a strong majority of students believed to the contrary; (7) sufficient evidence exists from which the Dean and President could have reasonably concluded Sypher was not above average in his teaching effectiveness; (8) the Board went on to say that if they adopted the Colleges' view that Sypher was not reappointed because of his teaching effectiveness, no argument advanced by him defending his teaching was likely to persuade the President because his decision was made on the "vigor and variety of student criticisms."

.....(9) With regard to the "political" aspect of the case, Sypher had written a letter in defense of his student rating level at the college which said, "it is certainly distressing when very good is not good enough, especially at a college with a modestly-talented student body that often discourages efforts at subtlety, wit and deeper penetration of subjects." (10) The Board responded to this letter saying, "other actions and statements by Grievant constituted legitimate reasons for not retaining him. In a May, 1980, letter to, Dean Beston, Grievant expressed his contempt for Castleton students" (p.135), (11) concluding, "Accordingly, we find credible the College's contention that Grievant was not reappointed because of his teaching effectiveness. [*Italics added*] (p.135).

VI. Lovelace v. Southeastern Massachusetts University, 793 F.2d 419 (1st Cir.1986).

.....Matthew Lovelace, a non tenured faculty at Southeastern Massachusetts University, challenged the decision not to renew his contract based on various constitutional claims, including a claim that the use of student ratings violated his academic freedom. He contended that his contract had not been renewed because he had refused to inflate his grades or lower his standards, and therefore his First Amendment academic freedom rights had been violated. More specifically, he maintained that in response to student complaints that his homework

assignments were too time consuming and that his courses were too hard, the university first threatened not to renew plaintiff's contract unless he appeased the students. When he refused to lower his standards, his contract was not renewed.

.....In the court's ruling it said, (1) "It is important to note what plaintiffs first amendment claim is and to separate speech from action. Plaintiff has not contended that he was retaliated against simply because he advocated that the university elevate its standards.... Plaintiffs complaint instead is that he was retaliated against when he refused to change his standards" (p.425), (2) citing other cases, the court rejected his contention that university teacher has first amendment right to disregard established curriculum content, that the first amendment does not prevent a university from terminating an untenured faculty whose pedagogical style and philosophy does not conform to those of the school's administration, (3) "is a policy decision which, we think, universities must be allowed to set" (p.426). Further, the court ruled that (4) "We will assume for purposes of this opinion that plaintiffs refusal to lower his standards was a substantial motivating factor, (see *Mount Health Board of Education v. Doyle*, 429 U.S. 274, 283-284, 97 S.Ct. 568, 574-575, 50 LEd.2d 471 (1977), in the decision not to renew his contract...Whether a school sets itself up to attract and serve only the best and the brightest students or whether it instead gears its standard to a broader, more average population is a policy decision which, we think universities must be allowed to set....matters such as course content, homework load, and grading policy are core university concerns, integral to implementation of this Policy decision. (p.424).

VII. Fields V. Clark University (1987).

.....Fields (Neither first name nor academic degree level was cited) taught at Clark University in the sociology department. While Field's scholarship and contributions to the university were considered adequate, they were deemed by various reviewing groups not to be so outstanding as to overcome deficiencies in her teaching. Fields contended she was being discriminated against on the basis of her sex.

.....In support of the collective opinion that (1) Fields was a poor teacher, the university (2) submitted a group of student evaluations, (3) a few of which, from students in Fields' seminars, were "wildly enthusiastic" about her enthusiasm, commitment and presentations; (4) a few were ambivalent; (5) with a considerable number being extremely negative, particularly (6) with regard to her large lecture classes in basic courses in sociology. The students (7) complained that the plaintiff's courses were disorganized. that there was no correlation between reading assignments and the course work and (8) that the plaintiff herself was often unprepared and (8) unresponsive to students' needs, with (9) one student saying that her performance was so bad that it was a principal reason for her leaving the university.

.....The court noted that (10) Fields' "attacks" the university's use of her student evaluations because they were not gathered and evaluated according to accepted standards of scientific polling procedures. In response, the court agreed, saying, "She is probably correct. The use made of the student evaluations in her case, however, followed the practice at the defendant university in other tenure decisions" (P.671). On the findings of sexual discrimination the court ruled in Fields' favor.

VIII. Carley v. Arizona Board of Regents, (1987).

.....Denny Carley was denied reappointment by Northern Arizona State University where he taught in the department of art as assistant professor. The denial was largely on the basis of student evaluations of his teaching. This appeal is from a superior court judgment affirming a decision of Northern Arizona University President to deny the renewal of a teaching contract to Carley. The issues on appeal involved: (1) whether his right to academic freedom was violated because student evaluations were utilized as the primary tool to determine his teaching effectiveness, (2) whether the President abused his discretion by

rejecting the findings of the majority of the Academic Freedom and Tenure Committee and (3) whether there was substantial evidence to support the President's decision.

.....In term of his teaching, (1) of the 13 faculty in his department of art, he was ranked fifth, (2) by his chairman he was ranked seventh, (3) student evaluations, however, ranked him last: 13th of 13. (4) Carley identifies as "protected speech" his teaching methods where his goal in his commercial art course was to promote a business atmosphere by requiring attendance, promptness, and self-reliance, and required them to meet deadlines. (6) He considered this approach part of the content of his class. (7) He characterized his professional style as being a "demanding teacher contrary to some student expectations." (8) Because of this, he maintained his popularity suffered as reflected in his low student evaluations. (9) An examination of student comments on the evaluation forms indicated that Carley was correct in his assessment as 61% (49 out of 80) negative student comments focused on these values.(10) Carley maintained that because student evaluations were critical of those methods, the student evaluations challenged his academic freedom rights, concluding that the student evaluations cannot be used as the primary basis for failing to renew his contract (p.1101). (11) With the ACLU, Carley filed suit, with his attorney saying, "Our primary argument is that it's a violation of academic freedom---he wasn't doing anything that wasn't in line with being a demanding teacher.

.....(12) A departmental review committee voted for non-reappointment, (13) but the department chair disagreed, recommending to the dean that the appointment be renewed. (14) The University-wide Academic Freedom and Tenure committee, with a 6 to 3 vote, also decided in the Carley's favor, finding "improper use of the student evaluation instrument, a de facto violation of academic freedom, [and] personal prejudice" which "constitutes substantive violation of due process." (15) They further found that the judgment of his teaching effectiveness had been "based wholly on student evaluations" which have "no objective validity as a measure of teaching effectiveness." (See Heller, 1986, italics added) (16) Three dissenting members of the committee filed a minority report, finding that his department colleagues had treated the faculty member fairly and that student evaluations were only one of several criteria in judging his teaching effectiveness.

.....(17) The president reviewed Carley's case and found justification for not supporting his reappointment. The court ruled (18) the University president was free to consider factual findings made by minority members of academic freedom and tenure committee and any other evidence which he found relevant in determining whether to deny renewal of teaching contract to non tenured instructor. The president was not bound by factual findings made by majority members of committee.

.....(19) It was said that Carley "paints with too broad a brush the concept of academic freedom," since his teaching style is not a form of speech protected under the First Amendment. (20) Decision not to retain a non tenured instructor, even if based, in part, upon student evaluations expressing disapproval of his teaching methods, did not violate instructor's First Amendment right to academic freedom; (21) Carley was not denied a contract because of expressing unpopular opinions or otherwise presenting controversial ideas to his students. (22) Thus, we conclude that the decision not to retain Carley, even if based, in part, upon student evaluations expressing disapproval of his teaching methods, did not violate his first amendment rights. p.1103. (23) Carley has cited no authority that relying primarily or solely on student evaluations would be impermissible. We have found none. p.1105 Denial of reappointment was upheld.

IX. Robert Kramer v. The President of the University of British Columbia (1992).

.....Kramer, an American citizen, was Assistant Professor in the Department of Asian Studies, Faculty of Arts, whose first two-year reappointment was denied. Kramer maintained

that there were errors of procedure, failures to obtain relevant evidence, and that the denial was based on a defective dossier.

.....(1) The department vote was 4:3, against reappointment. The department Head also did not recommend reappointment. (2). The department Head viewed Kramer's 1989-90 course evaluations "with some alarm." (3) The numerical scores were among the lowest of the tenure track faculty. (4) Even more disturbing to the department Head was that a considerable number of students in their written comments stated that Dr. Kramer was biased sarcastic, and hostile to the material and that a number of students had stated that Dr. Kramer's teaching would cause them to stay away from the Asian Studies department. (5) There were also some diametrically apposed positive comments. (6) Since the course was the general introduction to the subject and the Department, such negative comments were of great concern to the department Head.

.....(7) Kramer argued that the most significant mistake was the failure to consider all aspects of his teaching, as outlined in Article 4.02. (8) Only his teaching in 1989-90 was considered, whereas (9) he had taught a wide range of courses over the previous three years (10) had three new courses that year, (11) plus a graduate course. The basic course, Asian 105 had 220 students, making it numerically the largest class per faculty member in the department (Even though it was shared between two instructors). (12) his graduate teaching was entirely ignored, and (13) only the formal student evaluations were considered, with no peer evaluation. (14) Student evaluations were considered from the standpoint of his popularity, not his effectiveness. (15) Only one of the over thirty numerically rated questions was use: "Rate instructor bad to good". (16) While a number of negative student comments were quoted in the department Head's letter, there were a number of very positive comments, and these were not mentioned at all. (17) Department head indicated that his teaching was not up to the departmental "standard". The standard appeared to be the performance of the tenure-track faculty, though Kramer was one of the most junior faculty members.

.....The board of appeals responses included, (18) "The most important perceived error in the teaching evaluation, in the opinion of the Board, is the reliance solely upon the student evaluations and written comments for the 1989 course evaluations. There was no peer review at all; no member of the Department audited any of Dr. Kramer's lectures. There was, therefore, nothing to guide the Department but the student comments," and "no way to test the accuracy or fairness of the undoubtedly disturbing comments in Asian Studies" (p.10). (19) "Given that the Departmental procedure is to set aside a class hour for class discussion without faculty present, with the written evaluations and comments made during the same hour, there is a danger that some negative class commentary will dominate the discussion and will not be the 'independent' opinion of all of the students. (20) This is especially true in the context of the direction to assess "effectiveness" versus "popularity" (p.10). However, the board noted (21) "As for the 'popularity vs. effectiveness' debate, a discouraging or hostile attitude is a part of effectiveness as much as it is of popularity" (p.8). (22) "We recognize that the strong negative comments...may have [been] received...too late in the first year to have any peer teaching assessments in that particular class, but there does not appear to have been any attempt to see if the apparent problem had continued in September or October of 1990 in any courses, or (23) if the student perception was accurate and fair" (p.10).

.....The board further noted, "In the result, one got a 2.82 and one got a 3.07, and Dr. Overmyer cautiously and generously assigned the higher rating to Dr. Kramer the difference is statistically invalid in any event. But any rating at all in this course is suspect. (24) Further, it is the largest class in the Department, and large classes generally attract lower scores. Dr. Kramer's scores in the other two courses were higher--3.45 in one, 3.91 in another, against a "faculty average" of 4.22. As his counsel pointed out, that is a comparison against regular faculty, mostly senior and tenured" (p.10-11). (25) "We have examined all of these written

comments. There was a very wide range of comments. There were not 29 comments saying sarcastic and biased; but there were certainly 29 comments which included either cynical, sarcastic, biased, insulting, negative, condescending, belittling, opinionated, arrogant, nihilist, and destructive... [but] (26) It was obvious that almost all of the class were upset about an examination which was considered more geography than Asian Studies, and (27) they didn't like the marking. (28) They also felt the workload was far too heavy for an "introductory" course. (29) However, it would only be fair to add that there were a number of comments in favour of Dr. Kramer, stating that the student "liked the course immensely", "now interested in Asian Studies"; "helps create a relaxed atmosphere", "really enjoyed him", "very approachable and knowledgeable", "very enthusiastic", "captivates audience with his humour", "very effective" (p.12). (30) "In the other two courses, both small, both Japanese language, there were also some negative comments. One perceptive student noted that some of the unhappiness came from the fact that the levels of Japanese language ability were badly divided; some found it easy, others very hard" (p.12).

.....The board concluded, "In the final analysis, we feel that this review of the Head's comments on teaching, which would be the sole evidence upon which the Dean and the President could rely, shows that it was incomplete and might have been misleading, due to the total focus on student evaluations and the lack of peer opinion on teaching ability....The original Departmental vote was so close, we may find it easier to conclude that a different result might (We do not say must) have occurred, and that we should return the case: here. to the departmental level" (p.12-14).

X. Christopher Turner v. The President of the University of British Columbia (1993).

.....Christopher Turner, who taught Russian literature was denied promotion to full professor. In not recommending promotion, the Departmental review document in part stated:

.....(1) Turner's main issue was that there were a number of inaccuracies and misleading descriptions in the critiques of his teaching, with teaching weaknesses the main reason for the negative decision on his promotion, noting (2) "the small number of student evaluations handed in from his undergraduate literature course (3, 8, and 6). Turner provided the Board with the class lists in those courses, showing that there were 4, 16, and 16 students who had registered in those courses, and that 4, 14 and 16 had completed them. Thus the actual numbers were proportionately much higher than the 'responses' referred to... (3) It was argued that a low 'response' rate was proof that rather few students were actually attending, at least on the date of the evaluation, but that is a dangerous inference; some students simply do not fill in responses; some students in courses with no final exam may not attend the final evaluation lecture. (4) It would be easy for readers to infer from the [Department Head's] figures that Dr. Turner's classes were smaller than comparable classes in the department, a conclusion which Dr. Turner was able to rebut with course enrolment figures. In fact, it is clear that [the dean] misread [Department Head's] figures; in her letter of reasons to Dr. Turner...she states 'there were few students in undergraduate literature courses since 1986/7---(3,8, and 6 respectively...', and thus has taken 'response' figures for actual enrolment. Thus the general criticism in the Head's letter and the Dean's later letter of reasons, about relatively small classes compared with others in the department, was not supported in the material available to the University. (5) This misunderstanding is in our opinion sufficient in itself for a reconsideration, since teaching was the focus..." (p.3).

.....The board noted that (6) "While there is no question of Dr. Turner's competence as a teacher at all levels, teaching evaluations for the last several years show that his effectiveness is marred by what students perceive as excessive formality, lack of enthusiasm and dullness....In a previous promotion attempt, his teaching was briefly described as "very competent" but student evaluations indicate further improvement to be "better than

adequate." (p.2). (7) "while not ignoring some student unhappiness with Dr. Turner's teaching style, we think that the comments and emphasis on the size of Dr. Turner's classes as evidence of poor teaching are open to objection and constitute errors of procedure and/or evidence" (p.6).

.....The board further noted that (8) "This board has been asked on a number of occasions to pass judgment on the relevance of student evaluations to the Agreement criteria for good teaching. Good teaching is an elusive concept. Students may not be good judges during a course; their judgment might be quite different several years later in life. On the other hand, while popularity is not competence nor effectiveness, to the extent that it encourages students it has some relation to both" (p.7). (9) While the Agreement permits, but does not mandate either student reviews or peer reviews, and the methods of assessment 'may vary', we do conclude that the reliance placed on these very limited student reviews must have been great, since there was no other evaluation referred to. Where there is no other evidence sought, student comments will have an apparent importance and credibility that they may not deserve... (10) We would strongly recommend peer review in the reconsideration which we are requiring" (p.7).

.....The board concluded that (11) "there were sufficient errors of procedure and/or evidence to return the case for reconsideration" (p.11).

XI. University of Regina Faculty Association v. University of Regina (1993).

.....Dr. Pradeep Jalan was an Associate professor in the Faculty of Administration at the University of Regina. In February 1993, during Jalan's fifth year at the University, the Faculty Review Committee recommended that he be granted tenure. This recommendation was not accepted by the Dean, who recommended denial of tenure on the basis that Jalan had not met the standard for an Associate Professor in the areas of scholarly performance and teaching.(Education Employment Law News, 1994). Jalan appealed the Dean's recommendation to an appeals committee, which denied the appeal. The University of Regina Faculty Association then filed a grievance on Jalan's behalf, claiming he had been wrongfully denied tenure. The University justified its denial of tenure on two grounds. (1) that Jalan's teaching was unsatisfactory, submitting student evaluations completed over the course of five years in support of this conclusion, and (2) arguing that he had not met the standard of scholarship required for tenure, as Jalan had only two publications, neither of which had been published in a refereed journal; they appeared only in publications of conference proceedings.

.....The Board ruled, (3) "With respect to teaching, it is our opinion that the evidence of unsatisfactory performance is very weak indeed ...It is important to note that the basis of the comments, particularly the negative ones in the fall of 1992, were written student assessments...[and] Although these assessments are expressly recognized in Art. 17.19 of the collective agreement, to base important career decisions on them only does not seem justified" (p.4). The Board further ruled (4) 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. (5) [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" (p.4) The Board argued (6) that the University was under an obligation to verify negative comments before acting on them. Consequently, (7) the fact that Dr Jalan had received some negative evaluations from students could not be used to undermine the otherwise generally favorable comments he had received in his annual performance reviews. (8) "To base serious career decisions narrowly on student evaluations is not to be encourage...Indeed, in a document prepared by the University... titled, 'The Assessment of Teaching Effectiveness: Guidelines for Teaching Evaluations' it is recognized that teaching evaluations can involve a variety of techniques. The document states that: 'These may include

direct observations, informal feedback from students (both past and present), student course/instructor evaluations, review of instructional material selected by the instructor, assessment of the appropriateness of textbooks, examination of class outlines, assessment of the appropriateness of tests and examinations set by an instructor, and grading practices.' In the case before us, only one of these techniques was relied upon. If (9) teaching is to be seriously evaluated for career purposes, whether for positive or negative purposes, it seems incumbent upon Faculties not to rely only on classroom administered evaluations but to broaden the base of assessment" (p.4) The Board said teaching was wrongfully evaluated, but upheld denial of tenure on grounds of inadequate scholarship. University of Regina V. University of Regina Faculty Association And Dr. Pradeep Jalan (1993).

.....With respect to teaching, it is our opinion that the evidence of unsatisfactory performance is very weak indeed. It will be recalled that in the first two years of his appointment there were negative comments with respect to Dr. Jalan's teaching. However, in year three and four comments were quite positive. Then in the fall of 1992 again the comments were negative. It is important to note that the basis of the comments, particularly the negative ones in the fall of 1992, were written student assessments that were universally applied in the Faculty of Administration. Although these assessments are expressly recognized in art. 17.19 of the collective agreement, to base important career decisions on them only does not seem justified. Indeed, in a document prepared by the University dated February 15, 1991 titled, "The Assessment of Teaching Effectiveness: Guidelines for teaching Evaluations" it is recognized that teaching evaluations can involve a variety of techniques. The document states that:

These may include direct observations, informal feedback from students (both past and present), student course/instructor evaluations, review of instructional material selected by the instructor, assessment of the appropriateness of textbooks, examination of class outlines, assessment of the appropriateness of tests And examinations set by an instructor, and grading practices (p.19).

20. In the case before us, only one of these techniques was relied upon. If teaching is to be seriously evaluated for career purposes, whether for positive or negative purposes, it seems incumbent upon Faculties not to rely only on classroom administered evaluations but to broaden the base of assessment. The University itself has recognized this in the February 15th, 1991 document. To base a serious career decision narrowly on student evaluations is not to be encouraged. That is not to say that student evaluations properly structured and properly administered cannot be an important part of the evaluation but, as the University has recognized, it should not be the only factor. This R is particularly so when, as was strenuously argued by Ms. Rasmussen for the Association, the students are not advised of the potential use of the evaluation tool. The point which is very valid, is that if the evaluations are to be used for serious career development purposes these completing them should be aware of the potential consequences of their participation.

.....It is our conclusion that the grievor met the onus placed on him to establish that he was a competent teacher. This onus was met through the statements made in the Annual Performance Reviews and the evidence relied upon by the employer to suggest that the standard was not met, in our view, was not sufficient.

XII. Cynthia J. Fisher v. Vassar College, (1995).

.....Dr. Cynthia Fisher taught in the biology department at Vassar College. An appeal by defendant Vassar College from final judgments of the United States District Court for the Southern District of New York (Motley, J.). Following a bench trial, the district court found that, in denying plaintiff tenure, defendant had discriminated against her by reason of her sex

and age.

.....An analysis of Fisher's teaching ability included a review of her student evaluations, which (according to the report) (1) reflected "consistent problems with clarity and her ability to illuminate difficult material" but were otherwise generally positive; the district court (2) found that the biology department distorted Fisher's teaching recommendations by (3) "selectively exclud[ing] favorable ratings", by "focus[ing] on the two courses in which Dr. Fisher had difficulties" and (4) by "applying different standards to her than were applied to other tenure candidates" (p.1209). Further, the district court observed (5) that "the males tenured while Dr. Fisher was on the faculty were praised for their teaching while Dr. Fisher was criticized for her teaching, despite (6) the facts on which the Committee's determinations were based (student evaluations, Biology Majors Reports and [Student Advisory Committee] reports) revealing that Fisher's evaluations were superior to theirs" (p.1211); the court noted that (7) statistical analyses may be a part of a plaintiff's effort to establish discriminatory treatment. Finally, (8) the court . . . found that the termination of Fisher's employment resulted from a pretextual and bad faith evaluation by Vassar of her qualifications, not from any inadequacy of her performance, qualifications, or service.

XIII. Yu Chuen Wei v. Vermont State Colleges Faculty Federation (1995).

.....Dr. Yu Chuen Wei taught in the Mathematics department at Castleton State College where she was denied tenure. The case was arbitrated by the Vermont Labor Relations Board (VLRB 261, 1995), Wei's case involved SEF as a part of a wider issue of racial discrimination. An extensive review Wei's student evaluations were described by the Board.

.....With regard to her student evaluations, (1) Wei contended that the College acted unreasonably and arbitrarily by denying her tenure on grounds of teaching effectiveness without first performing a systematic statistical comparison of the student evaluations. (2) She brought in an expert who had researched and written an article on attitudes of student as a source of contamination on SEF to testify that there were serious problems with the validity of SEF under the best of statistical circumstances. (3) As an expert in the field of statistical assessment of SEF, her findings indicated that SEF relative to the quality of instruction is significantly affected by inappropriate attitudes, thereby creating suspicion about the objectivity of undergraduate student evaluators. (4) In specific, the expert found: (a) the race, sex or ethnic background of the professor; (b) the grade the student expected to receive; (c) whether the course was a required one; and (d) whether the student an undergraduate or graduate student influenced how students evaluated professors. (5) In analyzing Wei's evaluations, the expert compared the mean or average responses of three years of SEF questions for Wei to the mean or average responses to the same questions for a male colleague for the same period of time who was the most recently tenured professor in the Mathematics Department. Both an older and a newer evaluation form were analyzed. (6) A statistical comparison utilized a t-test. A t-test is standard for determining whether the difference in two means is significant, i.e., to determine whether differences are valid or due to chance or random error. (7) The results indicated: For (8) upper level classes using the "old" evaluation form, Wei's mean responses were significantly higher than the male colleague's mean responses in four out of the five questions pertaining to instructor attributes. (9) In one category, there was no significant difference between Wei's mean value of responses and the male colleague's mean value of responses. For (10) lower level classes using the "old" evaluation form, in two categories there was no significant difference between Wei's mean value of responses and the male colleague's mean value of responses.

.....(11) In three categories, the male colleague's mean value of responses was higher than Wei's. In (12) upper level classes using the "new" evaluation form, Grievant's mean value of responses was significantly higher than that of the male colleague for three of the questions.

(13) In the other two categories, there was no significant difference between Wei's mean value of responses and the male colleague's mean value of responses. For (14) lower level courses using the "new" evaluation form, the male colleague's mean value of responses was significantly higher than Wei's in two categories. (15) In the other three categories, there were no significant difference between Wei's mean value of responses and her male colleague's mean value of responses. (16) Wei maintained that her students rated her the same or higher than the male colleague's students rated him.

.....The Board disagreed (17) "that the statistical comparison of Grievant's student evaluations with those of [male colleague] contributes to a conclusion that the Colleges' articulated reasons for denial of tenure and promotion constituted a pretext for discrimination. (18) We note that the comparison offered by Grievant is somewhat weak since [male colleague] was tenured in 1988, and those student evaluations of his which were compared with Grievant post-dated his tenure review by a number of years."

.....Moreover, they said, (19) "The statistical comparison demonstrates that Grievant was evaluated higher by students than [male colleague] with respect to upper level classes, but that (20) [male colleague] was evaluated higher than Grievant in lower level classes. Given (21) this "mixed" result, the statistical comparison of evaluations does not demonstrate by a preponderance of the evidence that Grievant's students rated her the same, or better, than [male colleague]" (p.305). Further, the Board said, (22) "the statistical comparison is insufficient by itself to demonstrate how students evaluated the respective faculty members. The statistical comparison does not take account of the comments made by students on the evaluation forms. Grievant's student evaluations are striking in how often mention is made of Grievant's communication difficulties, particularly language difficulties. No similar problem, or as serious a problem, appears on [male colleague's] student evaluations with the frequency in which Grievant's communication difficulties are mentioned....In any event, (23) even assuming the statistical comparison was valid, the comparison does not support the conclusion that Grievant was treated in a discriminatory manner compared to her male colleagues." (p.304-5).The board judged that (24) the tenure criteria "are not drawn with mathematical nicety." The board further ruled that (25) the Dean and the President, both reviewed Grievant's student evaluations carefully. Their failure to take it a step further, and perform a statistical comparison of Grievant's student evaluations with those of other faculty members who have been granted tenure was not arbitrary and was reasonable; (26) Such a comparison is nowhere required by the Contract, and (27) we decline to hold such an involved comparison is necessary before a reasonable tenure determination can be made; (28) The Dean and the President obviously had much experience in reviewing student evaluations, and could reasonably draw on that experience in each tenure review. (p.311).

.....The Board went on to say, (29) In many other evaluations, in which students refer to communications problems in Grievant's teaching, references to difficulty in understanding Grievant may reasonably be interpreted as expressing a concern about her ability to communicate" (p.305). Finally the board noted that while some students had written that she was a "slant eyed bitch," and the she should "go back to China," (30) "We also are not persuaded that the racism evident in the student evaluations of Grievant made student evaluation results unreliable. The percentage of evaluations in which racism by students was evident was approximately one percent of the total evaluations." (p.306)

.....(31) Wei's last claim charged that the College violated the Contract by denying her a promotion, even though both her scholarly performance and professional activities were exceptional. Article 22(E) of the College provides for otherwise granting promotion if the President decides that "performance in one of three areas has been exceptional" (p.314).

.....With regard to this Article, the Board concluded that (32) "Grievant has established no such discrimination. Although Grievant had a significant publication record, most of it was

developed before coming to Castleton" (p.315). (33) In terms of exceptional scholarship, Dr. Wei maintained she had solved a significant mathematical problem (apparently published). The Board's response was, (34) "although Grievant claimed to have solved the Erdos conjecture, Dean Mark reasonably concluded that she had not established that she actually had solved the conjecture. Under these circumstances, and given our consideration of the discrimination issue previously discussed, we conclude that (35) Grievant has not established discrimination. The Colleges reasonably, and based on legitimate reasons, concluded that Grievant had met the tenure standards in this performance area but that her performance was not exceptional" (p.315). (36) The Board concluded that Wei had not established that the tenure decision in her case was unreasonable, arbitrary or based on erroneous reasons.

XIV. Guam Federation of Teachers v. The University of Guam (1990).

.....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.

XV. Dr. Brian Maclean v. President of The University of British Columbia (1991).

.....Held at Vancouver, B.C. January 28--June 20.

.....Dr. Brian MacLean was an Assistant Professor in the Department of Anthropology and Sociology, whose first two-year term of employment was not renewed.

.....(1) The Departmental Committee voted 16:1 against renewal, as did the Head, the Dean's Advisory Committee, the Dean and the President. (2) With three years of teaching experience in Saskatchewan, and London, his scores for Anth/Soc. 100 on one of the critical questions--- "Would you take another course from this instructor" were very low--29%(fall) and 34% (spring), compared with an average of 69% for all other instructors in the course. (3) They were the lowest for all other sections of the course, and for all first and second year courses. (4) Moreover, he had the lowest scores on another important question--- "would you recommend this course to a friend". (5) In his speciality course, Introduction to Criminology, his scores on the above questions were 61.2% for the instructor, 74.6% for the course, which, according to the Department Chair were well below other specialized courses. (6) In his senior course, in Criminology, his ratings were 60% for the instructor and 40% for the course [precise rankings were not provided] (7). When MacLean was not awarded a "merit" salary increase, the Chair told him that this was because of his low teaching ratings.

.....(8) His Department Chair testified to the Appeals Board that this problem was discussed with Dr. MacLean and it was strongly suggested he consult with "good" teachers in the department. (9) The Chair stated that MacLean's response to the bad ratings was that he did not believe them, that (10) he denied that there was a problem, arguing that "the students were not very good students" (p.6).

.....(11) Peer reviews of his teaching were conducted. (12) Before the classroom visits, the

two reviewing faculty members had been informed by the Dean that Dr. MacLean had a "teaching problem" the previous year. (13) Both faculty then attended a large lecture about 80 students. (14) Both reviewing faculty reported a favourable opinion of the lecture, noting that Dr. MacLean's speech was clear, his "delivery" was good, (15) despite the disadvantage that it took place in one of the more "difficult" and unpopular lecture rooms. (16) Neither faculty attended lectures in Dr. MacLean's other two small courses because they had only four students in each and (17) "hence were not amendable to the usual 'lecture' style and could be easily disrupted by a visitor" (p.7). (18) Instead, in these two courses, a class visit was made in the absence of Dr. MacLean.

.....(19) In addition to the student questionnaires, the reviewing faculty held in-class discussions about his teaching. (20) In general, the in-class peer reports were mixed but favourable. The in-class discussions were more problematic. (p.30). (21) While the knowledge, interest and enthusiasm of Dr. MacLean was acknowledged, "the problem appeared to be one of style or personality." (22) He was criticized for "putting them down", for belittling their responses to his questions in class, and that he was "hostile, aggressive, volatile, biased" (p.9). (23) It was noted that, "The students had obviously been talking among themselves before this class visit". (24) One of the reviewing faculty asked the Dean if the Departmental meeting could be postponed until after grades were submitted so that the students would feel protected. Inquiries to the President's office found that this was acceptable. (p.9).

.....The court noted that (25) "With respect to the "qualitative" scores---i.e. the "comments", there was a clear error. The qualitative comments from a number of courses were read and commented on, and conclusions were drawn from them which went into the "file". Both Reviewing faculty read and commented on them, as did the Department Chair in her letter to the Dean. Yet the Dean had clearly stated in a departmental memo that the qualitative comments were not to be used for administrative or promotion purposes. In evidence she admitted that they were used, and indeed as we have said they formed parts of the "file". (26) While in the abstract there is no reason why such comments would not be relevant, if the Department had a rule against their use, or in other words if they were "for the professor's eyes only", then it was a significant breach of Departmental rules to use them" (p.31). (27) In the opinion of the Board, so long as the comments were fairly presented, they offered the PAT [Promotion and Tenure Committee] and others a better balanced view of the teaching qualities and problems of Dr. MacLean than the quantitative statements alone" (p.31). (28) The court noted that "One problem with the questionnaire is that it solicits bad points as well as good points. Despite that caveat, we conclude that the inclusion of the qualitative comments was not a significant error." (p.32).

.....It was further noted that (29) "As against the low figures, they disclosed a number of good qualities in Dr. MacLean---enthusiasm for his subject, wide knowledge of the literature, much out of class assistance to students, and a commitment to seeking good work from students.(p.31). (30) The reviewing faculty report noted the comments about Dr. MacLean's "derogatory manner, biased opinion, unwillingness to listen", were matched by "clear, stimulating, very helpful after class". And, (31) "some students have told us that the comments made were not representative of the class as a whole and were unduly influenced by the process" (p.41). (32) "A number of students, both from earlier years and from his current classes, furnished letters of support, and in preparation for the appeal, some furnished affidavits with respect to particular matters such as the 'intimidation' discussion in Soc. 250 and events in Soc. 490 and 520 in the fall of 1989." (p.33). (34) "We agree with the appellant that these comments are of some importance, since we agree that while a superior research and publication record cannot overcome a poor teaching record, it might tip the scales where the teaching record was 'on the edge'" (p.10).

.....(35) The Faculty Agreement specified that "Evaluation of teaching shall be based on the effectiveness rather than the popularity of the instructor, familiarity with recent developments in the field, preparedness, presentation, accessibility to students and influence on the intellectual and scholarly development of students" (p.30), (36) MacLean sought expert opinion about the validity of the student evaluation instrument. Both MacLean and the University called expert witnesses to testify as to the reliability and relevance of the evaluation instrument. (37) The evidence, submitted by two experts, one for MacLean and one for the university, was conflicting. (p.30).

.....With regard to SEF validity, the court concluded (38) "that the instrument was not perfect, that it had flaws, and that the very limited number of samples (because of the very limited number of courses and students surveyed over the period) impaired its reliability. (p.30). (39) "However, we accept the evidence of Dr. [X] that the instrument has some value, directed toward the specified factors. (40) The relevance and quality of the scores is a matter of weight for the various decision-makers, and we assume that they were reasonably aware of the limitations of student evaluations and gave them the weight they deserve" (p.30).

.....Finally, the court said, (41) "As we have said, there were some errors of procedure and of evidence. We have set them out in some detail; and although they have some substance, they must be viewed against the much larger body of evidence which was either not questioned or was confirmed during the hearing" (p.38). "The evidence before us led us to conclude that Dr. MacLean has worked hard, has been active in research, and has some teaching talents of value. (42) It is clear that he appeals to a proportion of students who find him stimulating and helpful. (43) A number were willing to testify on his behalf or to supply written testimonials" (p.38). (44) "Dr. MacLean's faults appeared to be faults of style or approach...(45) Nonetheless, his colleagues have made a judgment based upon what we have found to be an adequate foundation of evidence" (p.38).

.....The court upheld the university.

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

[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

[Greg Camilli](#)

Rutgers University

[Andrew Coulson](#)

a_coulson@msn.com

[Sherman Dorn](#)

University of South Florida

[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

[Michael Scriven](#)

scriven@aol.com

[Robert Stonehill](#)

U.S. Department of Education

[Robert E. Stake](#)

University of Illinois--UC

[Robert T. Stout](#)

Arizona State University
