College Students' Evaluation of University Teaching Effectiveness: A Review of Literature

Assefa Berhane[*]

ABSTRACT: *At college and university levels, student ratings of teaching are considered to be valuable data sources for improving the teaching-learning process, providing feedback to instructors, and evaluating teaching for use in personnel decisions, like tenure or merit salary increases. However, student ratings are also being considered as a hindrance to tenure and an administrative interference in academic freedom. Student ratings have not been successful, mainly because less attention is given to the process and the factors that affect student ratings. This paper discusses the historical context of student ratings and its controversies and factors that influence student ratings of university teaching: the nature of the course, the instructor, and administration of the evaluation. Finally conclusions and recommendations are given.*

Introduction

At the college and university levels teaching effectiveness can be evaluated by current students, former students, peers, administrators, or trained observers. Effective teaching or the degree to which a teacher enhances student achievement, requires a willingness to interact with students, to engage them in discussion, encourage them to think critically and creatively, to listen to their comments and questions, and so on (Park, 1996). Most people interested in improving teaching see the primary purpose of college students' ratings as providing feedback to teachers that will be helpful for improvement (Mckeachine, 1997). However, due to the fact that the primary purpose of student ratings of faculty *is almost always to make personnel decisions ... for retention, promotion, tenure, and salary increase,* student ratings are not free of controversies, and unanswered questions are still prevalent (Cashin, cited in Haskell, 1977, p. 2). While a few faculty with strong pedagogical training enjoy students' evaluation of teaching, some others call it a *necessary evil* (Benson & Lewis, 1994).

[*] Principal, Adwa Teacher Training Institute

College Students' Evaluation of University Teaching Effectiveness: A Review of Literature

Assefa Berhane[*]

ABSTRACT: *At college and university levels, student ratings of teaching are considered to be valuable data sources for improving the teaching-learning process, providing feedback to instructors, and evaluating teaching for use in personnel decisions, like tenure or merit salary increases. However, student ratings are also being considered as a hindrance to tenure and an administrative interference in academic freedom. Student ratings have not been successful, mainly because less attention is given to the process and the factors that affect student ratings. This paper discusses the historical context of student ratings and its controversies and factors that influence student ratings of university teaching: the nature of the course, the instructor, and administration of the evaluation. Finally conclusions and recommendations are given.*

Introduction

At the college and university levels teaching effectiveness can be evaluated in current students, former students, peers, administrators, or trained observers of their teaching. Effective teaching or the degree to which a teacher enhances students' in achievement, requires a willingness to interact with students, to engage them in discussion, encourage them to think critically and creatively, to listen to their comments and questions, and so on (Park, 1996). Most people interested in improving teaching see the primary purpose of college students' ratings as providing feedback to teachers that will be helpful for improvement (Mckeachine, 1997). However, due to the fact that the primary purpose of student ratings of faculty *is almost always to make personnel decisions ... for retention, promotion, tenure, and salary increase,* student ratings are not free of controversies, and unanswered questions are still prevalent (Cashin, cited in Haskell, 1977, p. 2). While a few faculty with strong pedagogical training enjoy students' evaluation of teaching, some others call it a *necessary evil* (Benson & Lewis, 1994).

[*] Principal, Adwa Teacher Training Institute

One of the leading countries that intensively carry out student ratings of teaching is the United States of America (USA). Ethiopia may learn from the experience of the USA. This paper is exclusively based on US college students' evaluation of university teaching effectiveness. It analyzes the impact of college students' evaluations of faculty teaching effectiveness. The following guiding questions are asked: Are students' evaluations valid indications of effective teaching? What advantages do students' evaluations have on teaching effectiveness? And what disadvantages do students' evaluations have on teaching effectiveness? If students' evaluations have disadvantages how could they be improved?

The paper starts with the sources used to research this topic, historical context and definition of teaching. It then moves to discuss factors that influence student ratings such as the nature of the course (required or elective, course level, discipline, and class size), the instructor (academic rank, gender, year of teaching service, adjunct or full time, personality behavior), and administration of the evaluation (timing, presence or absence of an instructor during evaluation) are discussed. Finally conclusions and recommendations are given.

Sources of Data

Secondary source documents are used in this review. These include the three major encyclopedias: Encyclopedia of Educational Review, 1992; International Encyclopedia of Teaching and Teacher Education, 1987; and International encyclopedia of Higher Education, 1977. Several cited sources in books and materials were also consulted.

Besides, relevant articles from computer search were referred to. ERIC and SSCI data base indices were used for journal, magazine and newspaper articles.

Historical Context and Definition of Teaching

The evaluation of teaching has a long history. Evaluation started in 399 BC when Socrates was executed for having corrupted the youth of Athens with his teachings (Marsh, 1987b). Also, around 350AD in Antioch, any father who was not happy with the kind of instruction given to his son was able to

examine the case. If the teacher was found to have neglected his duties the father could file a formal complaint to a panel of teachers and laymen and could immediately transfer his son to another teacher (Marsh, 1987b).

There is good reason to believe that this evaluation of teaching and learning process lays the basis for recognition and reward of good teaching. Evaluation provides the knowledge and understanding that directs improved teaching and learning. One of the different techniques used in the evaluation of faculty teaching effectiveness is students' evaluations of their instructors (Divoky & Rothermel, 1989; Dressel, 1977). Information from student ratings of teaching usually are summarized and given to the teachers. Student ratings may help teachers to improve their instruction and to aid students in course or instructor selection (Cohen, 1980). Furthermore, student ratings may also help administrative evaluation of teaching effectiveness, and the development of faculty competence.

In modern perspective students' ratings of teachings were introduced in Harvard University, Washington University, Purdue University, the University of Texas, and other institutions in the mid 1920s. Remmers of Purdue University did the first systematic research in students' evaluations of faculty teaching effectiveness in 1927. For the design of instruments of students' evaluations of teaching effectiveness, Remmers proposed the following three principles:
    i)   The list of traits must be short enough to avoid halo effects and
         carelessness due to student boredom;
    ii) The traits must be those agreed upon by experts as the most
         important; and
    iii) The traits must be susceptible to student observation and
         judgement. (Marsh, 1987b, p. 257)

What is Teaching?

Although it is difficult to classify, define, or describe the work of the faculty, recently many authors (Boyer, 1990; Rice, 1991; Braskamp & Ory, 1994) have written about the work of faculty. According to Braskamp and Ory (1994) the diversity of the faculty work may be classified as teaching, research and creative activity, practice and professional service, and citizenship. Teaching

includes instructing in the classroom, mentoring interns and advanced graduate students, tutoring students individually, and advising students on such topics as appropriate educational programs and career opportunities. Active learning and cooperative learning between students and teachers, for example, problem solving skills, critical thinking, higher order reasoning, and joint student-faculty research projects, are considered teaching strategies that enhance student learning.

Good teaching demands knowledge of the subject and the ability to organize, synthesize, and communicate that knowledge in a meaningful way (Park, 1996). In evaluating good teaching, both the act of teaching and the content or subject matter taught are considered to be important. Additionally, students should be able to indicate the objectives of the course and the relationships of the various assigned tasks to the achievement of the objectives. Hence, evaluation should include an appraisal of the teaching, the assignments, the examinations, the textbooks used, enthusiasm (like expressive speech, movements while lecturing, gestures, facial expressions, uses of humor, etc.) of the instructor, interpersonal skills of the instructor, friendliness and helpfulness, organization, and so on (Murry, 1984; Kaplan, Mets, & Cook, n.d).

Effective teaching is a multi-dimensional complex process (Mckeachine, 1997). A variety of instructional methods, how students comprehend a concept, apply it, and integrate it, and creating a convenient teaching and learning environment that enhances student achievement are considered to be effective techniques. Effective teaching is the creation of a learning environment that includes, but is not limited to, the classroom. Effective teaching is considered as the creating of favorable conditions in which appropriate learning occurs and the building of those conditions that foster the process (Braskamp & Ory, 1994).

Student Evaluations

Dressel (1977) noted that a low rating of instruction by students or even by a single student raises a question that needs the instructor's attention. "No freshman or sophomore can judge the scholarship of a professor, but even a freshman can react to clarity of lectures, individual assistance provided, quality

of assignments, content of tests, and fairness of grades" (Dressel, 1977, p. 1487).   Others say teaching is such a multifaceted complex process that students could be considered important judges of the instructional process, but not for the content of what is being taught (Abrami, 1989; Benson & Lewis, 1994). According to Abrami, the evaluation of curricular materials has to be the duty of faculty.

Students' evaluations can improve teaching (McKeachie, 1987). However, the degree of improvement of teaching is significantly enhanced when the ratings are supplemented by the services of an educational consultant responsible for taking data and giving recommendations; or by the inclusion of supplementary feedback methods such as videotapes of teacher performance, discussions with students, and interaction analysis (Tiberius, et al., 1989). Teachers who received student ratings feedback with consultation maintained higher student ratings over a ten year period than instructors who received student ratings feedback without consultation (Stevens, 1987).

Although students' evaluations of teaching are an accepted method of evaluating faculty teaching in many institutions and are said to be more statistically reliable than colleague ratings, it remains controversial (Bain, 1996; Marsh & Roche, 1997; Haskell, 1997; Stake, 1997; Williams & Ceci, 1997).   Heated debates concerning the merits and demerits of students' evaluations of teaching continue, despite their use as indicators of teaching effectiveness (Marsh & Roche, 1997).

Usually, teachers question the validity of student ratings when their student ratings are negative (Greenwald, 1997). Based on Greenwald's (1997) electronic search for publications over the last 25 years, more publications favored validity of student ratings of teaching than invalidity. However, the research activity on the validity of student ratings has declined noticeably since 1980.

According to McKeachine (1996), in the 1950s there was a debate about students' evaluations of all courses. Some faculty state that the student ratings of teaching affected the respect faculty deserved from their students. Others stated that teaching is an art and hence "it is impossible to evaluate in terms of some form of measurement" (McKeachine, 1996: 2).

Because of students' leniency (tendency to mark the rating scale toward the high end) and the halo, i.e, the "tendency to allow overall impressions to guide completion of the evaluation instrument rather than scoring each item independently," student ratings are said to be imperfect (Cook, 1989:32). Some other arguments are: student ratings are not reliable; student ratings are highly correlated with grades; students who receive good grades in a course give a positive evaluation to the instructor and the course; students' appreciation of a good course usually are seen after graduation from colleges and universities; and colleagues are more qualified than students in rating faculty performance (Chau, 1997; Schmelkin, Spencer, & Gellman, 1997).

Students may not have the necessary knowledge to judge how well they are learning; students give high ratings to courses with low standards and undemanding faculty members; teaching is not equally effective for every student, and what works for one class may not work for others (McKeachine, 1996). Students' evaluations of teaching are difficult to validate because no single criterion of effective teaching is sufficient (Marsh & Roche, 1997). Other researchers (Chandler, 1978; Sheehan, 1975) also question the validity of student ratings and oppose its use for administrative decision-making purposes.

Others suggest that both faculty and the administration abuse and misuse student ratings. For example, faculty may impact student ratings by bringing in cookies and soft drinks on the day of the evaluation, standing near students or circulating in the room while the students are filling in the forms, or giving leading comments prior to evaluation. Similarly, users of the ratings lack the knowledge to use them appropriately and there is over reliance on some of the data, and student ratings are used for ranking of faculty (Schmelkin, Spencer, & Gellman, 1997). McKeachine (1997) also argued that there is no basic problem with the student ratings but rather with *the lack of sophistication of those using (student ratings of teaching) for personnel purposes* (p. 1218).

One of the recent arguments about student evaluation of faculty is that of Haskell (1997). Haskell questions the validity of students' evaluation of faculty by saying that: a) there are conflicting data, b) different instruments of evaluation are used by different institutions, c) student evaluations are not

systematically interpreted, and d) students give negative comments that affect the tenure of a faculty.

Student ratings of teaching were meant to help faculty to improve their performance and understand the needs of their students. However, student ratings have become an indirect "administrative intrusion into the classroom" and are also used as an "instrument of intimidation forcing conformity to politically correct standards...[and student ratings] do not eliminate poor or below average teachers but instead increase poor teaching practice" (Haskell, 1997:3). Therefore, student ratings of teaching are said to contribute to lowering course standards, lenient grading, overall grade inflation, changing teachers' instructional behavior, and interfering with academic freedom (Haskell, 1997).

Another argument deals with the method of providing a faculty member with individualized feedback. Although feedback has been exercised it has not been successful, mainly because many who feed back the information to the teacher are not trained in giving feedback. In addition, less attention is given to the process by which the instructor receives the information than to the kind of information that is fedback (Brinko, 1993). Therefore, it is suggested that the importance of a consultant is to separate important from superficial information, to assist in minimizing the anxiety and hopelessness and to provide an encouragement to the instructor, and to suggest alternative methods of teaching other than those used in the past (McKeachine, 1987).

Feedback to the teacher can enhance the collaborative relationship, especially if the feedback informs the instructor what the students are thinking and feeling as conveyed in their own words. This kind of feedback stimulates a conversation. If the teacher and students are participating in real conversation, in which feedback flows in both directions, the effect on teaching performance is even more dramatic and hence can have a beneficial effect on student attitudes and achievement (Tiberius, et al., 1989). It is suggested that reciprocal feedback, in which teachers and students engage in a two-way, face-to-face conversation, has a great impact on teaching improvement; in addition, administrators need to consider data related to the nature of the course, instructor's skills in organization, enthusiasm, clarity, and so on in personnel decisions (Tiberius, et al., 1989; Kaplan, Mets, & Cook, n.d).

Nature of the Course and Student Ratings

Some faculty criticize the practice by arguing that one can get good student ratings by lowering standards, cutting down the amount of work of a course or giving higher grades to students. This may bias the student's evaluation of the instructor and hence affect the validity of the evaluation (Greenwald, 1997). However, according to McKeachine (1996) courses that require more work or are more difficult and are intellectually challenging, but result in more learning, are rated higher. In addition, student ratings do not fluctuate with grades, i.e., do not go up with higher grades or down with lower grades automatically (Bain, 1996).

Teachers of non-required or elective courses, and the courses they teach, received higher ratings than those required courses and their teachers (Haskell, 1997). This difference in ratings may be directly accounted for by examining differences among teachers themselves in conjunction with differences in the size of the course rather than by the fact that the course is elective (Feldman, 1978).

Ratings in higher level courses tend to be higher than those in lower level courses. And, lower ratings are given to courses in business, computer science, mathematics, engineering, and physical sciences as compared to the fine arts and humanities, biological and social sciences. Considering class size, larger classes have more reliability of class ratings than smaller classes; however, smaller classes tend to receive higher ratings as compared to the larger ones (McKeachine, 1996; Braskamp & Ory, 1994; Feldman, 1978). According to Marsh and Roche (1997), given a sufficient number of students in one class, the reliability of class-average students' evaluations of teaching compare favorably with that of the best objective test.

Instructor and Student Ratings

In multi-section courses instructors of those sections achieving higher scores on classroom examinations are rated higher than those teachers whose students have not earned as much (Mckeachie, 1996). If an instructor focuses his or her course on the top students, those top students rate higher than the rest of the class (Renner, 1981). On the other hand, if an instructor aims at the poorer

students those students rate the instructor higher. Instructors that are witty, enthusiastic, theatrical or engaging receive higher ratings. According to Renner (1981) instructors who want to win at the statistical ratings game must cater for the middle range of most numerous students.

Enthusiasm is considered an essential element of effective teaching. In addition, instructors that address students by their names, use examples, stress important points, ask and encourage questions, speak expressively or emphatically, show an interest in student ideas and show concern for student progress, move about during lecture and use more gestures, are said to make active in learning are rated highly (Mckeachie, 1996; Kaplan, Mets, & Cook, n.d).

Feldman (1983) suggested that rank, age, and years of experience are gen unrelated to student ratings. According to Feldman (1983), in a stuc Blount, Stallings, and Gupta (1978) where the data excludes ass professors, "teaching assistants" were found to receive higher ratings Illinois Course Evaluation Questionnaire than either assistant professors professors. However, teachers of higher academic rank · as compa teachers of lower academic rank typically receive either equivalent evaluations or some what better ones. Therefore, ratings are more pos' the instructor is a professor rather than a teaching assistant (Feldman, 1983; Braskamp & Ory, 1994).

The teacher's age is a less direct measure of instructional experience; however, it is useful as additional information. The academic rank of the teacher is believed to be an indirect measure of teaching experience, but it obviously is a fairly direct measure of seniority (Feldman, 1983). According to Goldberg and Callahan (1991) adjunct (part-time) business course instructors receive higher student ratings as compared to full-time business faculty.

An instructor's skill in organizing and managing his or her course requirements, i.e., clearness of assignment, fairness of tests, puts outline of lecture on board, and good use of class time, has less impact on teaching effectiveness than personality factors (Sherman & Blackburn, 1976). Feldman (1986) noted that personality traits of college teachers affected the overall evaluations of teachers.

In terms of the broad divisions of academic endeavors, teachers of courses in the humanities, fine arts and languages tend to receive somewhat higher ratings than do teachers of social science or of physical science, mathematics and engineering (Mckeachine, 1996). Furthermore, it is reported that female students rate female teachers higher than male teachers (Renner, 1981).

Male and female college teachers do not differ in the ratings they receive from their students. However, according to Feldman (1993), when statistically significant differences are found, female teachers are more favored than male teachers. In addition, Feldman (1993) stated that female teachers have greater contact time with students and invest greater amount of time in advising and counseling activities; and students tend to rate same-gender teachers a little higher than opposite-gender teachers. However, other studies done earlier tend to indicate that male instructors are perceived as more competent than female instructors by male students, whereas female students show no sex bias (Basow & Silberg, 1987).

Administration of the Evaluation

It is suggested that every evaluation form should contain some standard open ended questions and each instructor should have the opportunity to add one or more questions, the results of which could go only to the instructor (Bain, 1996). Examples include questions that indicate students' perceptions of teaching methods such as: the primary teaching strengths of the instructor, the primary weaknesses of the instruction, and inviting the students to offer suggestions for improvement (Bain, 1996). Besides these, others that indicate the learning process and that can be reported to individual instructors need to be included. For example, questions that show the number of hours per week a student spends on the course (Bain, 1996), help the instructor to direct his or her method of teaching.

As to the time of administration, ratings administered during final exams are generally lower than those given during class. Therefore, student ratings forms should be administered in the classroom during regular class hours (not during informal get-together, too) (Braskamp & Ory, 1994). Furthermore, differences may be found in ratings of courses and teachers, among classes that are grouped by their meeting time, i.e., time of day when the class is taught.

However, in general there are no differences in ratings related to class meetings.

Differences in ratings occur when the directions in the forms indicate that the results will be used for personnel decisions rather than when they indicate that the results will go only to the instructor. Ratings are higher if the use of the ratings is stated (for promotion purposes) and if the teacher being rated remains in the evaluation room. Midterm ratings are reported to be less reliable; this may be due to the fear that the student raters can be identified. However, it is preferable that student ratings be administered during the last two or three weeks of the semester rather than during examination times (Braskamp & Ory, 1994).

Interpretation and Analysis

Although there is no easily available method of evaluating university teaching effectiveness, most colleges and universities have adopted students' evaluations of teaching as a measure of teaching effectiveness. Colleges and universities are administering some type of student rating forms, but they are often carried out at the end of the semester and are quantified for faculty ranking purposes rather than for faculty development. Therefore, such kinds of procedures are not offering the faculty an opportunity to make adjustments while the students are still involved.

Based on the literature review on the evaluation of teaching effectiveness, it seems there is an absence of recent experimental data to validate the fairness of student ratings. The results obtained during the 1970s and 1980s also look different. Students' evaluations of teaching have not been free from controversy and have not been as successful as required. This is because there is less agreement on their reliability and validity and less attention is also given to the processes by which instructors receive their feedback than the attention given to the kind of information that they receive.

However, some researchers noted that under appropriate conditions, students' evaluations of teaching are said to be reliable, stable, and valid when compared with a variety of indicators of effective teaching. Moreover, feedback targeted

at specific problems identified by student ratings can result in teaching improvement.

The heavy reliance on student-ratings as the primary measure of teaching effectiveness rather than on other indicators of effective teaching therefore, seems to stem from the lack of more valid indicators of teaching effectiveness. Students' evaluations of teaching contain important information, and hence, are being used for student guidance in choice of courses, improvement of teaching, and evaluation of teaching for use in personnel decisions like tenure and salary increases.

Students' evaluations of teaching effectiveness could enhance teacher-student collaboration if there is feedback that contains students' own essays that show their opinion about the course and their instructor. In addition, if feedback focuses on formative rather than summative evaluations and a teaching consultant or an expert helps the faculty in interpreting the feedback, then it is likely that this may help to the improve teaching effectiveness (Arubayi, 1987; Mckeachie, 1987; Brinko, 1993; Schmelkin, Marsh & Roche, 1997; Spencer, & Gellman, 1997). This kind of feedback stimulates a conversation between students and the teacher and favors strong faculty-student relations.

Conclusions and Recommendations

Students' evaluations of teaching have been increasingly used in personnel decisions. However, in most colleges and universities, open discussions about the strengths and weaknesses of instructors and courses and a clear orientation to students about the impact of faculty ratings on teaching improvement and the faculty's job are missing.

If teachers and students are engaged in an open face-to-face conversation, in which feedback proceeds in both directions, the effect on teaching performance would be more positive and hence could have a beneficial effect on students' attitudes and achievement, and overall improvement of the teaching-learning process. And, the primary purpose of students' evaluations of teaching should be to help faculty improve their teaching performance rather than primarily to make administrative decisions. Furthermore, those who currently summarize and give the student ratings, the feedback information, to the faculty are not

trained in feedback giving practice; hence, student ratings need to be coupled with appropriate consultation.

In addition to this, administrators should not only look for an instructor's certain statistical mean on students' ratings of teaching. More attention needs to be given to the factors that influence student ratings such as the nature of the course taught, the nature of the students, the method used to administer the evaluation, and the overall environment at which the instructor works. Student ratings of teaching need to be used in conjunction with other evaluation methods.

Finally, the reliability and validity of students' evaluation of teaching effectiveness for administrative purposes is still controversial and is construed as intervening on academic freedom and tenure. And, currently, there is a heated debate on the allegation that student evaluation of teaching measures popularity rather than teaching effectiveness. Hence, continuos research has to be conducted on the reliability and validity of students' evaluation of teaching effectiveness.

Reference

Abrami, P. (1989). *Seequing the Truth about Student Ratings of Instruction* **Educational Researcher.** 18:1, 43-45.

Arubayi, E. (1987). *Improvement of Instruction and Teacher Effectiveness: Are Students Ratings Reliable and Valid?* **Higher Education.** 16:3, 267-277.

Bain, K. (1996). *Teaching and Learning Issues.* [Online] Available: **http://www.nwu.edu/scfte/White.html**

Basow, S., & Silberg, N. (1987). *Student Evaluations of College Professors: Are Female and Male Professors Rated Differently?* **Journal of Educational Psychology.** 79:3, 308-314.

Benson, D., & Lewis, J. (1994). *Students' Evaluation of Teaching and Accountability: Implications from the Boyer and the ASA Reports.* **Teaching Sociology.** 22, 195-199

Boyer, E. (1990). **Scholarship reconsidered: Priorities of the professoriate.** Princeton, N.J.: The Carnegie Foundations for the Advancement of Teaching.

Braskamp, L.A. & Ory, J.C. (1994). **Assessing Faculty Work: Enhancing Individual and Institutional Performance.** San Francisco: Jossey-Bass.

Brinko, K. (1993). *The Practice of Giving Feedback to Improve Teaching: What is Effective?* **Journal of Higher Education.** 64:5,574-93.

Chandler, T.A. (1978). *The Questionable Status of Student Evaluations of Teaching.* **Teaching of Psychology.** 5:150-152.

ɩ, Chak-Tong (1997). *A Bootstrap Experiment on the Statistical Properties of Students' Ratings of Teaching Effectiveness.* **Research in Higher Education.** 38:4,497-516.

ːn, P. (1980). *Effectiveness of Student-rating Feedback for Improving College Instruction: A Meta-Analysis of Findings.* **Research in Higher Education.** 13:4, 321-341.

ː, S. (1989). *Improving the Quality of Student Ratings of Instruction: A Look at Two Strategies.* **Research in Higher Education.** 30:1, 31-44.

Divoky, J., & Rothermel, M (1989). *Improving Teaching Using Differences in Student Course Ratings.* **Journal of Education for Business.** 65:3, 116-119.

Dressel, P. (1977). *Evaluation.* In A.S. Knowles (Ed.), **The International Encyclopedia of Higher Education**, (vol. 26, pp. 1480-1493). San Francisco, Washington and London: Jossey-Bass.

Feldman, K. (1978). *Course Characteristics and College Students' Ratings of their Own Teachers: What We Know and What We Don't.* **Research in Higher Education.** 9, 199-242.

Feldman, K. (1983). *Seniority and Experience of College Teachers as Related to Evaluations They Receive from Their Students.* **Research in Higher Education.** 18:1, 3-124.

Feldman, K. (1986). *The Perceived Instructional Effectiveness of College Teachers as Related to their Personality and Attitudinal Characteristics: A Review and Synthesis.* **Research in Higher Education.** 24:2, 139-213.

Feldman, K. (1993). *College Students' Views of Male and Female College Teachers: Part II - Evidence From Students' Evaluations of Their Classroom Teachers.* **Research in Higher Education.** 34:2, 151-211.

Goldberg, G., & Callahan, J. (1991). *Objectivity of Student Evaluations of Instructors.* **Journal of Education for Business.** 66:6, 377-378.

Greenwald, A. (1997). *Validity Concerns and Usefulness of Student Ratings of Instruction.*

American Psychologist. 52:11, 1182-1186.

Haskell, R. (1997). *Academic Freedom, Tenure, and Student Evaluation of Faculty: Galloping Polls in the 21st Century.* [Online] Available: **Http://olam.ed.asu.edu/epaa/v5n6.html**

Kaplan, M., Mets, L., & Cook, C. (n.d). *Questions Frequently Asked About Student Ratings Forms: Summary of Research Findings.* [Online] Available: **http://www.umich.edu/~crltmich/crlt.faq.html**

Macmillan. Murray, H. (1984*). Classroom teaching Behaviors Related to College Teaching Effectiveness.* In J.G. Donald and A.M. Sullivan (Eds.), **Using Research to Improve teaching**. San Francisco: Jossey-Bass.

Marsh, H. (1987a). *Students Evaluations of Teaching.* In M.J. Dunkin (Ed), **The International Encyclopedia of Teaching and Teacher Education**. (vol. 26, pp. 181-187). New York: Pergamon.

Marsh, H. (1987b). *Students' Evaluations of University Teaching: Research Findings, Methodological Issues, and Directions for Future Research.* **International Journal of Educational Research.** 11, 253-388.

Marsh, H. & Roche, L. (1993). *The Use of Students' Evaluations and An Individually Structured Intervention to Enhance University Teaching Effectiveness.* **American Educational Research Journal.** 30:1 217-251.

Marsh, H. & Roche, L. (1997). *Making Students' Evaluations of Teaching Effectiveness Effective.* **American Psychologist.** 52:11, 1187-1197.

McKeachine, W. (1987). *Can Evaluating Instruction Improve Teaching?* In L.M. Aleamoni (ED.), **Techniques for Evaluating and Improving Instruction**. (pp 3-7) San Francisco: Jossey-Bass.

McKeachine, W. (1996). *Student Ratings of Teaching.* [Online] Available: **http://www.acls.org/op33.html**

McKeachine, W. (1997). *Student Ratings: The Validity of use.* **American Psychologist.** 52:11, 1218-1225.

Medley, D.M. (1992). *Teacher Evaluation. In M.C. Alkin.* (Ed.), **Encyclopedia of Educational Research,** (vol. 26, pp. 1345-1352). New York:

Park, S. (1996). *Research, Teaching, and Service: Why Shouldn't Women's Work Count?* **Journal of Higher Education,** 67:1, 46-83.

Renner, R. (1981). *Comparing Professors: How Student Ratings Contribute to the Decline in Quality of Higher Education.* **Phi Delta Kappan.** 63:2, 128-130.

Rice, R.E. (1991). *Toward a Broader Conception of Scholarship: The American Context.* In I.T. Whitson and R.C. Geiger (Eds.), **Research and Higher Education: The United Kingdom and the United States**. Bristol, Pa.: Society for Research into Higher Education and Open University Press.

Schmelkin, L. Spencer, K., & Gellman, E. (1997). *Faculty Perspectives on Course and Teacher Evaluations.* **Research in Higher Education**. 38:5, 575-592.

Sheehan, D.S. (1975). *On the Invalidity of Student Ratings for Administrative Personnel Decisions.* **Journal of Higher Education.** 46: 687-700.

Sherman, B., & Blackburn, R., (1976). *Personal Characteristics and Teaching Effectiveness of College Faculty.* **Journal of Educational Psychology.** 67:1, 124-131.

Stake, J. (1997). Response to Haskell: *Academic Freedom, Tenure, and Student Evaluation of Faculty. Education Policy Analysis Archives*, 5 (8). [Online] Available: **http://olam.ed. asu. edu/epaa/v5n8. Html**

Stevens, J. (1987). *Using Student Ratings to Improve Instruction.* In L.M. Aleamoni (Ed.), **Techniques for Evaluating and Improving Instruction**. San Francisco: Jossy-Bass.

Tiberius, R., et al. (1989). *The Influence of Student Evaluative Feedback on the Improvement of Clinical Teaching.* **Journal of Higher Education.** 60:6, 665-681.

Williams, W., & Ceci, S. (1997, Sept/October). *Problems With Student Ratings of Instructors.* **Change**, p. 13-23.