2011

Using Student Course Evaluations to Design Faculty Development Workshops

Raymond Benton
Loyola University Chicago

USING STUDENT COURSE EVALUATIONS TO DESIGN FACULTY DEVELOPMENT WORKSHOPS

Raymond Benton, Jr., Loyola University Chicago

ABSTRACT

Current practice is to administer end-of-course student evaluations and to use the results as part of a faculty member's annual teaching performance evaluation. Since the administration collects the data it ought to use it to help faculty improve their course evaluation scores. This may seem self-defeating but satisfied students not only rate the professor higher but likely rate the program and the university higher. In this era of external and public rankings of programs, this is important. Factor analysis can help administrators analyze student course evaluations and identify problem areas that can then be the targeted for faculty development programs and workshops.

INTRODUCTION

Teaching consumes fifty percent or more of a professors time (Bowen and Schuster, 1986), yet professors are tenured, promoted and evaluated more on the basis of their research and scholarly activities than on their teaching. It may be too much to say that institutions of higher learning "have paid lip service" to the importance of teaching, or that "Policies, procedures and criteria for the evaluation and promoting of faculty in higher education contribute to the marginalization of teaching" (Davidovitch and Soen, 2006, p. 351). It is curious, however, why the activity that consumes so much time, and is seen by many outside the academy as the overarching objective of a college or university (namely, to educate students), is often of lesser importance when evaluating faculty performance.

It may be, at least in part, due to the reward structure outside of colleges and universities. As Kai Peters (2005, p. 150) wrote in a letter to the editor of the *Harvard Business Review*,

> business schools, through their accreditation systems, are driven to adhere to a common academic model that heavily emphasizes the number of articles their faculty members publish in first tier journals rather than the impact the research might have on practitioners. Opting out of this system carries high penalties for those institutions—possible loss of credentials, of degree awarding powers, of access to government funding.

It may also be because research and scholarly activity is easier to evaluate than is teaching. Most institutions count journal articles, consider the quality of the journals (often

using published rankings), how often articles are cited, how many conference presentations are made, how many funding grants have been applied for and received, and so on. This is not all that difficult, either conceptually or in practice.

Assessment of a professor's teaching effectiveness requires, as Graeme Decarie (2005) stated, "some standard measure of what students know before the course and what they know after." It may be too much to say, as Decarie then opined, "No one has the faintest idea how to do that." We do know how to do it: have some idea what is to be accomplished in the class before hand, administer a pre-test, administer a post-test, and compare the results. There may be professors, schools, colleges or universities that do something like this, but certainly outcomes based measures are not the standard procedure for evaluating a professor's teaching effectiveness. And even at just this, it certainly would be more involved than the current standard procedure for evaluating scholarly activities.

Instead, the current standard procedure at most institutions is to rely on one form or another of end-of-course student evaluation as an indicator of faculty teaching performance. As Seldin (1993) opined, "student ratings have become the most widely used – and, in many cases, the only – source of information on teaching effectiveness" (see, also, Wilson 1998 for a similar observation). And student evaluations are not outcomes based measures; they are largely satisfaction surveys.[1]

Using student course evaluations as input into personnel decisions about who to hire, hire back, tenure, and promote is controversial.[2] The purpose of the present paper is not to further contribute to the large literature regarding the validity and reliability (or lack thereof) of student evaluations, but to suggest that since we do administer them, and since there is zero likelihood that we will stop administering them, department chairs, program directors, deans and those responsible for faculty development programs should use the information collected for formative purposed. The student voice, while impacted by any number of variables, does say something regarding the instruction they have received and it ought not be ignored. While we should not mistake student course evaluations as an assessment of teaching effectiveness, we should fully appreciate that satisfied students may learn more but they certainly evaluate professors higher and, likely, have a higher opinion of the program, the school, the college or the university. In this age of external and public ranking of institutions, this should matters a great deal, and not only to faculty but to department chairs, program directors, deans, university provosts and presidents.

FORMATIVE USE OF STUDENT EVALUATIONS

While most of the literature on student course evaluations focuses on their summative use, Centra (1993, Ch 4) does discuss their formative use. His focus is on how individual faculty members, striving to improve their own classroom instruction, can use the information provided by student evaluations. Centra emphasizes, however, that a professor may glean something from

course evaluations, believe the information credible, and be motivated to use the information, yet not know how to make changes called for by students.

There is evidence that those faculty that receive help make more progress than those that go it alone (Cohen 1980; Cohen and McKeachie 1980; Williams and Cici 1997). But even here the evidence is ambiguous. For example, Davidovitch and Soen (2006) evaluated their institution's attempt to promote quality instruction, as measured by student evaluations, by investigating a range of variables for their impact on student evaluation scores. One relationship they were interested in was the relationship between faculty participation in teaching workshops and the end-of-course student evaluation scores, something that had only recently been introduced at their institution.

They found, over a five-semester period, that there was significant improvement in student evaluation scores. They also found no correlation between participation in teaching workshops and scores on the student evaluations of teaching. In short, improvements in teaching "were not related to instructors' participation in teaching workshops" (p. 373).

Davidovitch and Soen discussed several possible reasons for these surprising and certainly disappointing findings. One possible reason not discussed was that the topics for the teaching workshops were unrelated to what students were being asked to evaluate on their teacher and course evaluations.

HOW WORKSHOP TOPICS ARE SELECTED

Like many colleges and universities, my institution conducts faculty teaching workshops. I asked one of the organizers in charge of a recent round of workshops how the themes or topics for workshops are chosen. I was told they "ask faculty what they want," that they "monitor IT help desk calls to identify problem areas," and that they "pay attention to 'hot topics' (for example, a current hot topic is digital copyright)." They also "sometimes have focus groups" with students.

Each of these approaches will probably provide a workshop that will be interesting and informative. But will they improve student opinion of, and satisfaction with, their classes? Not necessarily and only accidentally if the workshops are unrelated to what students are being asked to evaluate? Conducting focus groups with students is an appropriate strategy, but why collect new and original data from students when virtually every institution already and regularly surveys students about how professors perform and how well and what they like and dislike about their classes? The data are already collected; department chairs, deans, and those charged with faculty development activities should use it. Unfortunately, current practice at far too many institutions is to collect the data, calculate summary statistics, and provide these summary statistics and sometimes the raw data and the written comments to the faculty member, who is then left to do with them as he or she sees fit.

STUDENT EVALUATION FORMS

Most student evaluation forms ask students to numerically rate a list of 15, 20, sometimes 30 classroom teaching performance traits. Some items are fairly specific (Instructor puts outline of lecture on board); others are more general (Class sessions are well planned). Student evaluation forms almost always include a general or overall evaluation of the instructor and/or of the course, and they almost always provide space for the student to write comments about the course and the way it was taught.

If instructors look at their course evaluations at all, they often turn to the overall evaluation items first and then to the written comments. Faculty look at the written comments for anecdotal insights and, as often as not, for confirmation of their own great performance. What they less carefully consider are the multiple individual items rated by students. Looking at 15, 20 or 30 items, rated by 20, 60 or more students, to ascertain how students rated various aspects of a professor and his or her course is much more difficult and time consuming than scanning the written responses for a quick sense impression.

The obverse is true when a department, school, college or division within a university is looking at several thousand evaluations for several hundred courses. Reading, coding, and making sense of the written comments would be a daunting task; statistically analyzing a series of rating scales is much easier.

STATISTICALLY ANALYZING COURSE EVALUATIONS

The statistical analysis of student course evaluations that I have seen are limited to the calculation of the number and proportion of responses in each response category for each item on the form and the calculation of the average response for each item. These are presented to the instructor, sometimes accompanied by the same calculations for the department or for the school. Occasionally they are even accompanied by results from peer schools if the evaluation forms are administered and analyzed by an outside vendor.

A recent analysis I received for a course I taught at another university during summer 2008 will serve as an illustration (see Table 1, below).

Table 1: Instructor Score Analysis							
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Number of Responses	Average Response
1. Instructional methods enhanced my analytical problem solving skills	0	1 (5.88%)	5 (29.41%)	9 52.94%)	2 (11.76%)	17	3.71
2. The instructional methods enhanced my critical thinking skills	0	0	2 (11.76%)	10 (58.82%)	5 (29.41%)	17	4.18
	Very Poor	Poor	Neutral	Very Good	Excellent	Number of Responses	Average Response
7. Instructor's effectiveness in conducting the class	0	0	4 (23.53%)	9 (52.94%)	4 (23.53%)	17	4.00
10. Instructor's knowledge of material and subject	0	0	1 (5.88%)	9 (52.94%)	7 (41.18%)	17	4.35
	1	2	3	4	5	Number of Responses	Average Response
11. Rate the degree to which the course met your expectations	1 (5.88%)	1 (5.88%)	2 (11.76%)	7 (41.18%)	6 (35.29%)	17	3.94

Had I been a regular member of the faculty, I would have also received a summary average representing my own history of ratings for each of the thirteen items on their form, a similar average for the school in which the course was taught, and a similar average for the division of the university within which the school was housed.

What is an instructor to do with this data? Presumably one can look at one's performance on any one item and compare it with the performance of others or even with one's own historical performance. Do you do better than others? Do you do worse? Are you getting better? Are you getting worse? How this information can be used for self-improvement is not obviously clear. As Centra pointed out, faculty members often do not know how to make the changes called for by the students?

Presently far too many institutions use such simple data analysis of student course evaluations, and often considering only the overall evaluation score(s), as an indication of teaching performance and as input into personnel decisions. This paper suggests that administrations – department chairs, program administrators, deans – can use the information already collected, by way of student course evaluations, to help plan and design faculty development activities and workshops that will actually help improve scores on student course evaluations. A more sophisticated analysis of the data is necessary, however.

USING FACTOR ANALYSIS

Factor analysis is well suited for exploring the interrelatedness between multiple questions asked on a typical course evaluation instrument. By applying an advanced form of correlation analysis to the responses received, a list of 15, 20 or 30 items can be reduced to just a few characteristics that students might, themselves, have difficulty identifying.

The adage in correlation analysis is that correlation does not imply causation. This helps to conceptualize what is at work in factor analysis. Correlation does not imply causation because a third variable may be the unmeasured (or latent) cause of the observed fluctuation and variation in the two measured variables. Factor analysis is a way to identify that third, unmeasured variable (or factor).

As an analytical technique, factor analysis relies on overlapping correlations, searching for patterns of co-variation among the variables. If an instrument has eleven questions, and the responses to five of them co-vary together, the idea is that they each measure the same underlying construct, or "factor." If the other six co-vary together, they are measuring another underlying construct. Thus, eleven "variables" are reduced to two "factors." Examining the items that co-vary together, that "load" on a "factor," for what they have in common provides an understanding of the underlying construct. When applied to 15, 20 or 30 variables, the process "reduces" the many to a few. The end result is easier interpretation and action.

It must always to be remembered that factor analysis is an exploratory tool. Further, it works only on the questions that have actually been asked. If critical questions are not on the course evaluation form, or if the wrong questions have been asked, factor analysis cannot identify characteristics that would have been identified if a different set of questions had been asked. Based on the actual questions asked of students, it identifies what sub-groups of questions are tied together, and, in the minds of the students, what ties them together.

The problem at hand is to analyze student course evaluations such that the student voice is heard and faculty development workshops can be planned that actually address student issues and, thereby, help faculty improve their student evaluation scores. If students are metaphorically screaming answers to 15, 20 or 30 different questions, it will be hard for a faculty development office to hear what they are saying. If students will slow down and consolidate their thoughts into fewer "factors," it will be easier for a faculty development office to understand. That, in essence, is what applying factor analysis to student course evaluations attempts to do, after the fact.

THE ANALYSIS

For the present analysis and illustration, course evaluation data from my School of Business Administration was used. At the time of this study our course evaluation instrument was administered as a pencil-and-paper questionnaire using a Scantron form for their reply. It consisted of eighteen ungrouped statements (see Table 2, below). Although the instrument is now administered online, it consists of the same eighteen ungrouped statements. Using a 5-point scale, anchored with Strongly Disagree (1) and Strongly Agree (5), students indicate the extent to which they agree or disagree with each statement. These eighteen items are followed by two general overall evaluation questions. The first is an overall evaluation of the instructor; the second an overall evaluation of the course. The overall ratings use a 5-point ordinal scale

(Excellent, Good, Satisfactory, Poor, and Very Poor) to record the student response. Because each of these five response categories is presented in association with a number (Excellent = 5, etc.), they are treated by my institution as interval measures.

Table 2*
Items 1-18 are rated on a five-point scale with 1=Strongly Disagree and 5=Strongly Agree.
1. The goals of the course were clearly expressed at the beginning of the term.
2. What was actually taught was consistent with the goals of the course.
3. The course syllabus clearly explained the basis for determining grades.
4. The instructor followed the stated basis for determining grades.
5. The instructor communicated in a clear, effective way.
6. The instructor was organized and prepared for class.
7. The instructor presented the material in an interesting, thought-provoking way.
8. The text and/or assigned readings contributed to my understanding of the subject.
9. Other assignments (papers, projects, homework, etc.) contributed to my understanding of the subject.
10. I received useful and timely feedback on my performance.
11. The amount of work demanded for this course was appropriate and reasonable.
12. The instructor used appropriate methods to evaluate my performance.
13. The instructor was fair in grading my performance.
14. The instructor was sensitive to students' varying backgrounds and academic preparations.
15. The instructor was caring and respectful of students.
16. The course stimulated my interest in the subject area.
17. The course helped me to develop intellectual skills, such as critical thinking or problem solving.
18. I have achieved my education goals for this course.
Items 19-20 are rated on the following scale: 5=Excellent 4=Good 3=Satisfactory 2=Poor 1=Very Poor.
19. Overall rating of instructor.
20. Overall rating of course.
* The first 20 items are followed by two additional overall ratings, one for library resources and one for computer resources. These are then followed by standard census items. There are an additional four questions pertinent only to laboratory and clinical courses. Questions 21-31 are not relevant to this analysis so their exact wording and response structure is omitted.

The initial data set consisted of two years of course evaluations. There were 701 classes and 20,877 evaluation forms, both from undergraduate and graduate programs and from all departments. Although many faculty teach in both programs, only undergraduate evaluations were included in the analysis because the overall evaluation scores differ markedly between undergraduate and graduate classes. In addition, removed from the data set were all independent study classes, all classes with less than 10 students, and all classes in which fewer than half of the enrolled students completed a course evaluation form.

Since the problem at hand is one of using student course evaluations to aid in designing faculty development workshops, it was further decided to focus on those sections which students indicated were most in need of help. Quartile scores for each of the two overall ratings were calculated and only those courses that were in the fourth quartile on both the overall evaluation of the instructor and the overall evaluation of the course were selected for analysis. These are the instructors and courses that students evaluated lowest and, presumably, are the instructors and courses most in need of help (from the students' point of view). The final data set includes 3,146 evaluations, representing 103 sections. Because listwise deletion of variables was employed in the analysis, the final sample size was 3,017 student evaluations. The mean response to each of the eighteen variables in presented in Table 3, below.

Table 3: Descriptive Statistics

	Mean	Std. Dev	Analysis N
ITEM 1 Goals of course were clearly expressed	4.03	1.018	3017
ITEM 2 Material taught was consistent w/goals	3.91	1.067	3017
ITEM 3 Syllabus clearly explained basis for determining grades	4.05	1.084	3017
ITEM 4 Followed stated basis for determining grades	4.09	1.036	3017
ITEM 5 Instructor communicated in a clear, effective way	3.36	1.291	3017
ITEM 6 Instructor was organized and prepared for class	3.97	1.112	3017
ITEM 7 Material presented interestingly and thought-provokingly	3.13	1.332	3017
ITEM 8 Text or readings contributed to my understanding	3.63	1.245	3017
ITEM 9 Other assignments (papers, projects, homework) contributed	3.63	1.210	3017
ITEM 10 Student received useful and timely feedback	3.78	1.171	3017
ITEM 11 Amount of work was appropriate and reasonable	4.01	1.039	3017
ITEM 12 Instructor used appropriate methods for evaluation	3.85	1.142	3017
ITEM 13 Instructor was fair in grading performance	3.94	1.114	3017
ITEM 14 Instructor was sensitive to students' varying backgrounds	3.92	1.169	3017
ITEM 15 Instructor was caring and respectful of students	4.11	1.114	3017
ITEM 16 Course stimulated interest in the subject matter	3.22	1.359	3017
ITEM 17 Helped develop intellectual skills	3.46	1.252	3017
ITEM 18 Student achieved educational goals	3.47	1.258	3017

Because the intent of the analysis is to reduce the set of measured variables (the 18 items on the course evaluation form) to a smaller set of underlying dimensions for the sake of parsimony and conceptual simplicity, Principal Components Analysis (PCA) was used to extract the factors. Because it is believed the resulting factors will be independent and because the desire is to produce a solution in which measured variables substantially load on only one factor rather than on several factors, verimax rotation was employed.

In the final solution, discussed below, five factors were kept. This number was arrived at through an iterative process. The initial analysis applied Kaiser's criterion that only factors with an eigenvalue of 1.0 or more be retained. This initial solution retained two factors, one of which can only be described as a global factor. Eleven of the eighteen items substantially load on it (.500 or greater). This factor was very difficult to interpret and did not provide much guidance for the practical problem at hand: developing faculty development workshops that address the issues in the minds of the students.

Subsequent iterations increased the number of factors to be extracted and rotated. In this iterative process an eye was kept on the stability of the factors with each iteration. The 3-factor solution split the largest factor of the 2-factor solution into two separate factors; the smaller of the two original factors remained stable. The 4-factor iteration removed two variables from the untouched smaller factor of the original 2-factor solution, producing a fourth factor. In all subsequent iterations this two-variable factor remained stable. The 5-factor iteration segregated two variables from one of the two factors generated in the 3-factor solution, creating a second two-variable factor; in all subsequent iterations this two-variable factor also remained stable. The 6-factor and the 7-factor solution each extracted one additional variable from the previous 4-factor solution, creating two additional one-variable factors.

The 5-factor solution was settled on for the present purposes. The "themes" or "factors" in the minds of the students that emerged follow:

* Whether or not the professor is stimulating, interesting, and thought provoking. (Communication Skills)

* Whether or not the course goals and the basis for determining grades are clear and followed. (Course Organization)

* Whether or not the actual workload and grading was fair and appropriate. (Evaluation)

* Whether or not the instructor was caring and respectful. (Personality)

* Whether or not the texts, readings and assignments contributed to student understanding. (Assignments)

The final rotated solution is presented in Table 4, below.

Table 4: Rotated Component Matrix

	1	2	3	4	5
ITEM_16 Course stimulated interest in the subject matter	.836	.171	.229	.162	.201
ITEM_7 Material presented interestingly and thought-provokingly	.775	.284	.093	.251	.217
ITEM_17 Helped develop intellectual skills	.772	.210	.316	.114	.265
ITEM_18 Student achieved educational goals	.719	.250	.388	.184	.198
ITEM_5 Instructor communicated in a clear, effective way	.624	.503	.131	.333	.172
ITEM_1 Goals of course were clearly expressed	.302	.740	.246	.189	.161
ITEM_3 Syllabus clearly explained basis for determining grades	.112	.732	.455	.076	.127
ITEM_2 Material taught was consistent w/goals	.399	.712	.243	.186	.196
ITEM_6 Instructor was organized and prepared for class	.315	.691	.087	.299	.243
ITEM_4 Followed stated basis for determining grades	.130	.680	.512	.193	.124
ITEM_13 Instructor was fair in grading performance	.260	.334	.711	.325	.152
ITEM_12 Instructor used appropriate methods for evaluation	.325	.331	.705	.279	.192
ITEM_11 Amount of work was appropriate and reasonable	.283	.251	.601	.275	.261
ITEM_10 Student received useful and timely feedback	.302	.355	.507	.269	.240
ITEM_15 Instructor was caring and respectful of students	.239	.257	.312	.798	.108
ITEM_14 Instructor was sensitive to students' varying backgrounds	.278	.237	.346	.753	.143
ITEM_8 Text or readings contributed to my understanding	.289	.183	.173	.096	.838
ITEM_9 Other assignments (papers, projects, homework) also contributed	.344	.278	.284	.165	.680

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 8 iterations.

At this point, the issue facing those responsible for developing faculty development workshops for which of these five factors do they develop a faculty workshop? The answer lies in the evaluation scores given by students to each of the five factors. A simple averaging of the evaluation scores in Table 3 for each item in each factor is presented in Table 5, below. Students are clear. Faculty most need to make their courses stimulating, interesting and thought provoking. Following that are issues involving the selection and use of texts, readings and other assignments.

Table 5: Averaged Scores for Items in Each Factor

Factor 1	Communication Skills	3.33
Factor 5	Selection of Texts and Assignments	3.63
Factor 3	Evaluation of Students	3.90
Factor 2	Course Organization	4.01
Factor 4	Instructor Personality	4.02

Of course, the preceding is based on the actual items contained on an actual course evaluation form. Ask different questions and a different analysis will result.

CONCLUSION

Information obtained from course evaluations is almost universally used for personnel decisions: who to hire, promote, tenure and reward with a pay raise. The information ought to be used, as well or instead, to help faculty improve their course evaluation scores. If the objective is to improve student satisfaction as measured by course evaluation instruments, then department chairs, program directors, deans, and those responsible for faculty development would be wise to skip "hot button issues" like digital copyright, as important as they may be, and focus, instead, on what students are telling them in their end-of-term courses evaluations. Since the data are collected, they ought to be used for formative purposes as well as for summative purposes. They should be used, that is, to improve student satisfaction. The faculty member benefits, the program benefits, and the college or university benefits.

In the present example, students are saying that faculty should focus on fundamentals, with communication skills on top. It might be desirable, before proceeding, to further investigate, by way of focus groups with students, what it is about classroom communication skills that is lacking and what it is about the texts, the readings, and the assignments they find disagreeable. But at least then the focus group with students will be targeted and not simply a fishing exhibition.

This much having been accomplished, the next step is clearly to provide faculty with the opportunity to attend a targeted faculty development workshop or series of workshops and then monitor future student course evaluations to determine if the workshops have the desired impact and outcome. What little there is in the literature suggests, as indicated above, that those faculty that receive help make more progress than those that go it alone. A particularly interesting case is that reported by Williams and Cici (1997).

Ceci, a seasoned and respected psychologist, was invited by his university's faculty development program to participate in a teaching effectiveness workshop. He used this opportunity to conduct a naturalistic experiment to "test" whether or not oral presentation skills, alone, can make a difference. He taught a class in the fall, participated in the workshop conducted by a media consultant over the winter break, and then taught the same class the following spring. He used the same syllabus, presented the same lectures (he had independent observers watch video taped sessions from the two semesters and confirmed they presented the same content), had the same schedule, at the same time, used the same book, and gave the same assignments and the same exams. All that changed from the fall semester to the spring semester was the manner in which he presented the material in class: greater pitch variability in his voice, more hand gestures, etc. His course evaluation scores improved on every aspect of the student

evaluation form, including items such as instructor's knowledge, organization, accessibility, the quality of the textbook, and fairness in grading.

ENDNOTES

[1] Instructional effectiveness is about more then just measuring student satisfaction. As Merritt states, "At a very minimum thoughtful evaluation of teaching requires time and attention" and "takes more time than traditional student evaluations" (2007, p. 281, 283). McLaughlin and Bates (2004) discuss an approach for obtaining reflective and deliberative input from students via the Delphi method and Merritt (2007, pp. 281-286) describes a Small-Group Instructional Diagnosis scheme.

[2] Research into and debate about the validity, reliability, and utility of student course evaluations blossomed soon after the practice of using them for administrative decisions began. The literature on the adequacies and inadequacies of student course evaluations is now voluminous. Extensive reviews can be found in each of the following: Deborah J. Merritt (2007), "Bias, the Brain, and Student Evaluations of Teaching," *St. John's Law Review* 82: 235-287, provides an informative discussion of much of it, as well as extensive references. Dennis E. Clayson and Mary Jane Sheffet (2006), "Personality and the Student Evaluation of Teaching," *Journal of Marketing Education* 28 (2): 149-160 covers much of the same territory and also offers extensive references. Additional discussion and references can be found in Philip C. Abrami, Les Leventhal and Raymond P. Perry (1982), "Educational Seduction," *Review of Educational Research* 52 (3): 446-464; Peter Seldin (1993), "The Use and Abuse of Student Ratings of Professors," *The Chronicle of Higher Education* Vol 39, Issue 46, 21 July, p. A-40; Mary Gray and Barbara R. Bergmann (2003), "Student Teaching Evaluations: Inaccurate, Demeaning, Misused," Academe Online September October, http://www.aaup.org/AAUP/pubsres/academe/2003/SO/Feat/gray.htm; Charles R. Emery, Tracy R. Kramer and Robert G. Tian (2003), "Return to Academic Standards: A Critique of Student Evaluations of Teaching Effectiveness," *Quality Assurance in Education* 11 (1): 37-46; Nitza Davidovitch and Dan Soen (2006), "Using Students' Assessments to Improve Instructors' Quality of Teaching," *Journal of Further and Higher Education* 30 (4): 351-376; and Robin Wilson (1998), "New Research Casts Doubt on Value of Student Evaluations of Professors," *The Chronicle of Higher Education* 44 (19): A12-A14.

REFERENCES

Bowen, H. R. and J. H. Schuster (1986). *American Professors: a National Resource Imperiled*. New York: Oxford University Press.

Centra, J. A. (1993). *Reflective Faculty Evaluation: Enhancing teaching and Determining Faculty Effectiveness*. San Francisco (Jossey-Bass Publishers).

Clayson, D. E. and M. J. Sheffet (2006). Personality and the Student Evaluation of Teaching. *Journal of Marketing Education* 28(2), 149-160.

Cohen, P. A. (1980). Using Student Rating Feedback for Improving College Instruction: A Meta-Analysis of Findings. *Research in Higher Education* 13, 321-341.

Cohen, P. A. and W. J. McKeachie (1980). The Role of Colleagues in Evaluation of College Teaching. *Improving College and University Teaching* 28(4), 147-154.

Davidovitch, N. and D. Soen (2006). Using Students' Assessments to Improve Instructors' Quality of Teaching. *Journal of Further and Higher Education* 30(4), 351-376.

Decarie, G. (2005). AT ISSUE: Course evaluation is 'a good idea gone terribly bad'. *Concordia's Thursday Report* 30(3). Retrived on October 8, 2008 from http://ctr.concordia.ca/2005-06/oct_13/04/ on October 8, 2008.

Gray, M. and B. R. Bergmann (2003). Student Teaching Evaluations: Inaccurate, Demeaning, Misused. *Academe Online* (Sept/Oct). Retrived on September 28, 2008 from http://www.aaup.org/AAUP/pubsres/academe/2003/SO/Feat/gray.htm

McLaughlin, F. S. and H. L. Bates (2004). Using the Delphi Method in Student Evaluations of Faculty. *Academy of Educational Leadership Journal* 8(2), 29-43.

Merritt, D. (2007). Bias, the Brain, and Student Evaluations of Teaching. *St. John's Law Review* 82, 235-287.

Peters, K. (2005). How Business Schools Lost Their Way. *Harvard Business Review* 83(9), 97-104.

Seldin, Peter (1993). The Use and Abuse of Student Ratings of Professors. *The Chronicle of Higher Education* 39(46), A40.

Williams, W. M. and S. J. Ceci (1997). How 'm I doing? *Change* 29(5), 13-24.

Wilson, R. (1998). New Research Casts Doubt on Value of Student Evaluations of Professors. *The Chronicle of Higher Education* 44(19), A12-A14.