

2006-297: STUDENT EVALUATION OF TEACHING - MYTHS AND REALITIES

Warren Hill, Weber State University

Student Evaluations of Teaching - Myths and Realities

Every semester, in almost every college and university across the country, many if not all students are asked to fill out evaluation forms about their instructors. The questionnaires that are used for this purpose frequently have very similar questions and often use a Likert type of scale for the responses. The results from these surveys are then tabulated for each instructor for each class in which that instructor is being evaluated. Many times these surveys also contain places for comments from the students that are not related to specific questions. The results of these evaluations are used for both summative and formative purposes. But how good is the data that comes back from these surveys? Is this information of sufficient quality such that we can or should be depending upon it to make decisions on such things as promotion and tenure? This paper will examine the myths and realities regarding student evaluations and suggest how the data from these should and should not be used. Also the paper will discuss the limitations of student evaluations in terms of what they can tell you and what they can't tell you.

Faculty Evaluation

Student evaluations of instruction are really part of the larger topic of faculty evaluation where usually faculty are assessed in the three areas of teaching, scholarship and service. This paper is only going to look at the area of evaluating teaching and more specifically, the use of student evaluations to evaluate teaching. In attempting to evaluate teaching, we are attempting to evaluate a very complex and difficult task. Some of the factors that should be included in the evaluation of teaching include learning styles, teaching styles, class size, classroom setting, required courses versus elective courses, and lecture courses versus discussion courses just to name a few. A further consideration in such evaluations is whether or not these evaluations are to be used for summative or formative purposes.

Defining Teaching

It should be clear that to talk about teaching, one must first talk about learning. The desired outcome of the teaching process is student learning. In the handbook Developing a Comprehensive Faculty Evaluation System, Arreola¹ describes learning as "*...a persistent, measurable, specified change in the behavior of the student resulting from an experience designed by the teacher.*" It is clear from this definition that we are talking about the learning that goes on in the classroom separate and apart from much of the other learning that goes on throughout a student's life. What we are concerned with here is the learning that is specific to the course being taught that meets both the objectives of the course and the broader objectives of the program of which the course is a part. Arreola then goes on to describe teaching as "*...presenting a set of experiences which induces student learning.*" It should be obvious that if learning has not occurred in the context of the above definition, then teaching has not occurred.

Arreola outlines three different perspectives on what defines teaching that match to some extent what Peter Filene in his book, The Joy of Teaching² describes as the teaching process. Arreola's first perspective states that teaching is "*...an interaction between a*

teacher and a student conducted in such a way that the student is provided with the opportunity to learn." This perspective says that the major responsibility for learning rests with the student. His second perspective defines teaching as "*.....an interaction between a teacher and a student conducted in such a way as to enable the student to learn."* Here the primary responsibility for learning still rests with the student but some responsibility is now placed upon the teacher as an enabler. His third perspective states that teaching is "*....an interaction between a teacher and a student conducted in such a way as to cause the student to learn."* Here the responsibility rests primarily with the teacher to create learning. In the author's experience, in any given class all three of these perspectives happen, frequently simultaneously.

In Arreola's first perspective, all that is required of the teacher is content area knowledge and the act of teaching would be that of the sage on the stage. To paraphrase the old mathematical expression, content area knowledge is a necessary but not sufficient condition to be a good teacher. If instead, we accept Arreola's second perspective of teaching, besides content knowledge, the teacher must also provide an enabling environment that is conducive to learning. Finally, accepting his third perspective requires that we evaluate teaching in quite a different light, namely that a good teacher is one where he or she produces student learning.

These two definitions then lead to three broad aspects of teaching that are interrelated. The first of these is subject matter expertise, a prerequisite for the teaching enterprise. The second is instructional delivery skills and expertise and the third is instructional design skills. To this list, Arreola adds a fourth aspect, course management skills. It is this set of four different skills then that we want to assess when we talk about the evaluation of teaching.

Arreola describes each of these skill elements as follows. First, he defines subject matter expertise as "*....that body of skills, competencies, and knowledge in a specific subject area in which the faculty member has received by advanced experience, training or education."* It is generally agreed that students are not competent to evaluate faculty in this area although they can make a judgment as to whether or not the faculty member appears to be content area knowledgeable. Ideally, a faculty member would be both knowledgeable and appear to be knowledgeable.

In the instructional delivery skills area, he defines these as "*....those human interactive skills and characteristics which (1) make for clear communication of information, concepts, and attitudes, and (2) promote or facilitate learning by creating an appropriate affective learning experience."* These skills relate to the particular attributes listed by Filene, namely the presentation of ideas with clarity, engagement with the students, and laying the groundwork for stimulating students. In this dimension, students are probably the best people to evaluate a faculty member's skills.

Arreola then defines the instructional design skills dimension as "*....those technical skills in (1) designing, sequencing, and presenting experiences which induce student learning and (2) designing, developing, and implementing tools and procedures for assessing*

student learning outcomes." Remember that what we are looking for is teaching that causes learning. Thus the faculty member must create a change in student behavior which persists and is measurable and hence must have the necessary expertise to provide instruction that causes such a change to occur. These include skills in test design, developing learning objectives, developing appropriate supporting materials, using appropriate instructional technology, and having good organizational skills regarding classroom activities. While students are not necessarily knowledgeable about what constitutes good instructional design, they do know if a test is or is not related to the course objectives or if the course is too easy or too hard.

Finally, he defines course management skills as "*...those bureaucratic skills in operating and managing a course, including, but not limited to, timely grading of examinations, timely completion of drop/add and incomplete grade forms; maintaining published office hours; and generally making arrangements for facilities and resources required for teaching the course.*" Here again, students are in a position to evaluate certain aspects of course management, particularly the aspect of timely grading.

Filene, in his book, makes several points about what teachers should be doing as part of the teaching process. Specifically he states that "When you teach, you are engaging in a relationship with your students." He goes on to say that teachers and their students should aspire to create an environment conducive to learning. As part of doing these things, he then states that good teachers display five basic characteristics:

- They are enthusiastic about their subject
- They present their ideas with clarity
- They are engaged
- They lay the groundwork for stimulating their students
- They care about their students

Finally, Arreola links evaluation with development. If the institution is to evaluate teaching and reward faculty, at least partially, on the basis of good teaching, then they must also have in place the resources and assistance to help faculty overcome their classroom shortcomings and become better at teaching. He goes on to say that "A successful evaluation system...is one that *provides information which faculty, administrators, and, where appropriate, students consider important and useful.*"

Student Evaluations

For better or for worse, student evaluations of faculty are in wide use throughout higher education in the United States for both formative and summative evaluations. It is therefore critical for both the faculty and the institution to note that such evaluations can improve teaching provided several criteria are met. According to Bhada³ et. al., these criteria include:

- an evaluation tool that is both reliable and valid
- an evaluation that provides norming data

- that the institution has services available to assist faculty with their teaching
- that the institution rewards good teaching

With respect to student ratings which have been extensively studied over the past 75 years, several misperceptions persist. Some of the more common ones include the following:

1. Students cannot make consistent and reliable evaluations of faculty as they lack the maturity and experience to do so.
2. Student evaluations are nothing more than a popularity contest with the best liked instructors getting the highest ratings.
3. Student evaluations are neither reliable nor valid.
4. Class size, gender makeup, time of day, course level, or required vs elective all affect the ratings.
5. Grades the students receive are highly related to the rating of the course and the instructor.

The research shows that all of these are false, again given that the evaluation instrument has been tested for both reliability and validity. This means that most home grown evaluation forms may end up doing more harm than good as they usually do not undergo any kind of psychometric evaluation. An example of such a locally developed form is shown in Appendix A.

There are a number of commercially available evaluation tools that have undergone rigorous testing for both reliability and validity. Some of the more well known include the Instructional Development and Assessment (IDEA) system from Kansas State University, the Student Instructional Report (SIR) from Educational Testing Service, and the Course/Instructor Evaluation Questionnaire (CIEQ) from Comprehensive Data Evaluation Services, Inc. At Weber State University in the College of Applied Science and Technology, we have been using the CIEQ instrument for over ten years and have been very satisfied with the results. The CIEQ questionnaire is shown in Appendix B.

Finally a note regarding open ended student comments or responses to opened ended questions on the evaluation form. At the author's institution, these comments are collected by a secretary and typed such that the students' identities remain anonymous. The collected comments from each class are then given directly to the instructor and are not part of the evaluation process. In Student Ratings of Instruction: Issues for Improving Practice⁴ edited by Theall and Franklin, Arreola and Alemoni state that "Deans and department heads should be made aware of the need to use comparative data to interpret results, rather than relying on the more subjective and highly unreliable written and oral comments of students."

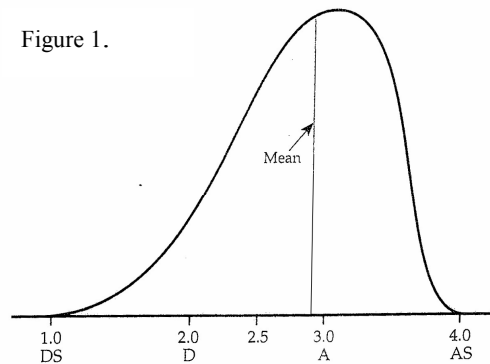
One dimension that is frequently neglected in student evaluations is some aspect, however indirect it may be, of student learning. Since this aspect is the critical element of the teaching process, it might be useful actually ask students about the learning experiences in their courses as part of the evaluation. John Centra in Reflective Faculty

Evaluation⁵, discusses the SEEQ (Student's Evaluation of Educational Quality) evaluation form which specifically asks about learning and the academic value of the course.

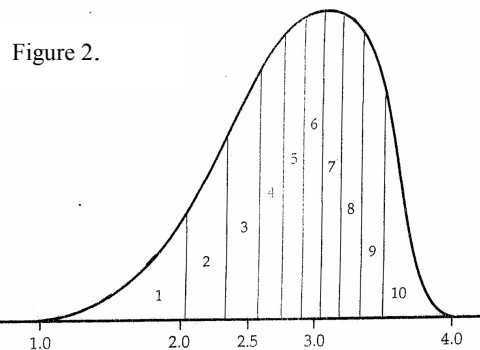
Even with a good instrument, a note of caution is in order. The results from a single course should never be used as a measure of instructional effectiveness, good or bad. A number of factors can influence the results from a single course such as it being the first time the course was taught, the instructor being new or relatively new, or the instructor having multiple preparations that term. The results from these surveys need to be examined over time and the longer term trends used for evaluative purposes.

Interpretation of Results

The research by Arreola and Alemoni shows that when using a four point Likert scale ranging from Disagree Strongly to Agree Strongly, there is a definite positive response bias. A typical rating response curve for such questions is as shown in Figure 1.



To use this information for improvement, it is useful to break the responses up into deciles and then look at decile groupings to see if there is a need for improvement with respect to a given question or group of questions. Such a grouping is shown in Figure 2 where a response falling in deciles one through three, indicates a definite need for improvement, a response in deciles four and five need some improvement, a response in deciles six and seven need very little improvement, and a response in deciles eight through ten need no improvement.



This information coupled with the student responses to some leading open ended questions such as those on the back of the CIEQ form should provide the faculty member with sufficient information to help them improve their teaching.

Further Sources of Information for the Evaluation of Teaching

Besides students' evaluations which can give both the faculty and the administration information about instructional delivery skills and instructional design skills, there are other sources that can provide information about content expertise and course management skills. One source is the faculty member himself through a self-report of what he has been doing in the area of teaching. A form of self-reporting that is becoming rather prevalent is the use of a teaching portfolio. In the book Improving College Teaching⁶ edited by Peter Seldin in an article by John Zubizarreta, the author discusses the use of the teaching portfolio as a way to improve instruction.

In this article, Zubizarreta suggests some things that should be included in a teaching portfolio such as a statement of course philosophy for each course, methods and strategies used for each course taught, a description of the various course materials, long term and short term course goals, evidence of student's progressive learning in each course, and descriptions and analyses of student and peer reviews. Given this data base, he then suggests that the faculty member look at such things as the effectiveness of the various materials and whether or not the desired course outcomes are being met. The results can then be used to examine ways to improve each of their courses.

The content of the teaching portfolio as described by Zubizarreta posits another source of information about teaching. This source consists of peer reviews or as John Centra suggests, colleague reviews. Centra prefers the term colleague since peer implies equals and the group doing the review may not be equals to the faculty member being evaluated. Using the best source principle discussed by Arreola and Aleamoni, colleague reviews are the best source of information about the faculty member's content knowledge and also are the best source for determining such things as appropriate content, course objectives, instructional materials, and the use and construction of evaluative instruments such as quizzes, exams, reports, and written assignments.

Again, peer reviews should be done in the context of how the instructor facilitates learning through the use of the materials provided the reviewers. The information being reviewed should, in part, examine how the course materials and assessment methods used in a course assist in the learning process. For example, do the quizzes or exams adequately measure what the students have learned about a particular aspect of the course or the material in the course up to the point of the examination.

Another approach for gathering information about teaching effectiveness is the use of either classroom visits or taping classroom presentations for review by colleagues or the department head at a later date. While some additional information can be gleaned from classroom visits, research by Centra has shown that while individual raters demonstrate consistency, there is little agreement between raters. In addition, the time required for

training and doing the visits does not appear to justify the little information that is obtained. The author suggests that videotaping be done for the use of the individual instructor for improvement but that it not be part of the process of teaching evaluation. The author also suggests that groups which assist people with public speaking such as Toastmasters can be very helpful to faculty in improving their classroom presentations.

Finally there is the department head as a source of information about teaching effectiveness. Again using the best source principle, the department head is probably in the best position to know about a faculty member's course management skills. For example, does the faculty member provide in a timely manner a copy of his or her syllabi for the department file, does he or she get their grades in on time, do they keep their office hours, and they make the appropriate arrangements that are necessary for the effective delivery of their courses.

Conclusions

The research shows that student evaluation of instruction can be a very useful tool in determining certain aspects of teaching performance. However for a more complete picture, other aspects should also be considered including such things as content expertise and course management skills. It is also very important is to examine how teaching affects learning. Regardless of how teaching effectiveness is measured and what measures are used, the end goal should be to help faculty become better at the craft of teaching to the end of improving student learning.

Bibliography

1. Arreola, R. A. (1995). *Developing a Comprehensive Faculty Evaluation System*, Bolton, MA; Anker Publishing Company Inc.
2. Filene, P. (12/17/2004). *The Joy of Teaching*, Chronicle of Higher Education,
3. Bhada, Y., Brightman, H. J., and Elliott, M. (1992). in *Evaluating Faculty Performance: The State of the Practice*, Proceedings of the First Annual CEDA Conference on Practical Issues in Faculty Evaluation, Memphis, TN; Center for Educational Development and Assessment
4. Theall, M. and Franklin, J. (1990). *Student Ratings of Instruction: Issues for Improving Practice*, New Directions for Teaching and Learning, San Francisco, CA; Jossey-Bass
5. Centra, J. A. (1993). *Reflective Faculty Evaluation*, San Francisco, CA; Jossey-Bass
6. Seldin, P. (1995). *Improving College Teaching*, Bolton, MA; Anker Publishing Company, Inc.

Appendix B

ALEAMONI COURSE/INSTRUCTOR EVALUATION QUESTIONNAIRE (CIEQ) (FORM 76) COMPREHENSIVE DATA EVALUATION SERVICES, INC. © LAWRENCE M. ALEAMONI, 1975

CODING INFORMATION COURSE CODE

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

STUDENT INFORMATION ARE YOU A ... <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>		ARE YOU TAKING THIS COURSE AS A ... <input type="radio"/> <input type="radio"/>		ARE YOU TAKING THIS COURSE FOR PASS/FAIL ... <input type="radio"/> <input type="radio"/>		COURSE INFORMATION RATE EACH OF THE FOLLOWING ... COURSE CONTENT: E VG G F P VP MAJOR INSTRUCTOR: E VG G F P VP COURSE IN GENERAL: E VG G F P VP	
ARE YOU A ... <input type="radio"/> <input type="radio"/>		YOUR EXPECTED GRADE IN THIS COURSE IS ... <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>		THIS COURSE IS WITHIN YOUR ... <input type="radio"/> <input type="radio"/>		SEMESTER ... <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	
MARKING INSTRUCTIONS AS IF YOU AGREE STRONGLY WITH THE ITEM D IF YOU DISAGREE MODERATELY WITH THE ITEM A IF YOU AGREE MODERATELY WITH THE ITEM DS IF YOU DISAGREE STRONGLY WITH THE ITEM MARK ONLY ONE RESPONSE PER ITEM USING PENCIL ONLY ERASE CHANGED ANSWERS CLEANLY AND COMPLETELY				SAMPLE MARK AS ● D DS		SEMESTER ... <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	

SPECIAL CODE

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

	STANDARD ITEM SECTION	OPTIONAL ITEMS	
		SECTION I	SECTION II
1	It was a very worthwhile course.	AS A D DS	AS A B C D E
2	I would take another course that was taught this way.	AS A D DS	A B C D E
3	The instructor seemed to be interested in students as individuals.	AS A D DS	A B C D E
4	The course material was too difficult.	AS A D DS	A B C D E
5	It was easy to remain attentive.	AS A D DS	A B C D E
6	NOT much was gained by taking this course.	AS A D DS	A B C D E
7	I would have preferred another method of teaching in this course.	AS A D DS	A B C D E
8	The course material seemed worthwhile.	AS A D DS	A B C D E
9	The instructor did NOT synthesize, integrate or summarize effectively.	AS A D DS	A B C D E
10	The course was quite interesting.	AS A D DS	A B C D E
11	The instructor encouraged development of new viewpoints and appreciations.	AS A D DS	A B C D E
12	I learn more when other teaching methods are used.	AS A D DS	A B C D E
13	Some things were NOT explained very well.	AS A D DS	A B C D E
14	The instructor demonstrated a thorough knowledge of the subject matter.	AS A D DS	A B C D E
15	This was one of my poorest courses.	AS A D DS	A B C D E
16	The course content was excellent.	AS A D DS	A B C D E
17	Some days I was NOT very interested in this course.	AS A D DS	A B C D E
18	I think that the course was taught quite well.	AS A D DS	A B C D E
19	The course was quite boring.	AS A D DS	A B C D E
20	The instructor seemed to consider teaching as a chore or routine activity.	AS A D DS	A B C D E
21	Overall, the course was good.	AS A D DS	A B C D E

PLEASE PRINT THE NAME AND NUMBER OF THIS COURSE IS

(Name) _____ (Number) _____

Appendix B, continued

C I E Q

PLEASE USE THIS SIDE OF THE FORM FOR YOUR PERSONAL COMMENTS ON TEACHER EFFECTIVENESS AND GENERAL COURSE VALUE. YOUR INSTRUCTOR WILL NOT SEE YOUR COMPLETED EVALUATION UNTIL AFTER FINAL GRADES ARE IN FOR YOUR COURSE.

PLEASE DO NOT WRITE IN THIS AREA	PLACE ALL WRITTEN COMMENTS IN THIS AREA ONLY
<p><u>COURSE CONTENT</u></p> <p>PLEASE GIVE YOUR COMMENTS ON THE COURSE CONTENT, SUBJECT MATTER AND ANY PARTICULAR RELEVANCE THIS COURSE HAS HAD TO YOUR EDUCATION.</p>	
<p><u>INSTRUCTORS</u></p> <p>WHAT ARE YOUR GENERAL COMMENTS ABOUT THE INSTRUCTOR(S) IN THIS COURSE?</p>	
<p><u>COURSE/INSTRUCTIONAL OBJECTIVES</u></p> <p>PLEASE COMMENT ON HOW CLEARLY THE OBJECTIVES OF THE COURSE WERE STATED AND FOLLOWED.</p>	
<p><u>TEXTBOOK, HOMEWORK AND PAPERS</u></p> <p>COMMENT ON THE VALUE OF BOOKS, HOMEWORK AND PAPERS (IF ANY) IN THIS COURSE.</p>	
<p><u>EXAMINATIONS</u></p> <p>COMMENT ON THE EXAMINATIONS AS TO DIFFICULTY, FAIRNESS, ETC.</p>	
<p><u>GENERAL</u></p> <p>1. WHAT IMPROVEMENTS WOULD YOU SUGGEST FOR THIS COURSE?</p>	
<p>2. WHAT IS YOUR EVALUATION OF THIS COURSE BASED UPON</p> <p>(A) YOUR SATISFACTION WITH WHAT YOU GOT OUT OF THIS COURSE AND</p> <p>(B) WHETHER IT WAS A VALUABLE EDUCATIONAL EXPERIENCE OR A DISAPPOINTMENT?</p>	

PLEASE FILL OUT THE OTHER SIDE