

The Counterintuitive Nature Of Student Evaluations Of Faculty, Or, Raise Your Teaching Effectiveness Rating By Canceling Class!

Michelle L. Slagle, University of South Alabama
Marjorie L. Icenogle, University of South Alabama

Student evaluation of faculty teaching effectiveness is commonly used in faculty performance evaluations. To evaluate a new instrument, statistical analysis was undertaken of data collected for a semester, a sample of over 3000 responses. Regression analysis and analysis of variance found some relationships as predicted by proponents of these tools, but others stood expected relationships on their heads. This paper identifies those findings, discusses them, and offers some ideas for examining and perhaps explaining them through future research.

I. INTRODUCTION

The Outcomes Assessment Committee of a College of Business at a regional public university in the Southeast developed an instrument for “Student Evaluation of Teaching Effectiveness” based on information available in the literature and examination of instruments in use at other institutions. This project was undertaken with utmost seriousness of intent, required months of effort, and involved the entire faculty of the college in refining and creating the final version before it reached an acceptable form. Analysis of student responses to the instrument, however, presents some results that are contrary to what one would hope to get from a diligent student body.

II. METHODOLOGY

Data Collection

Data were collected in any of a number of ways. The student surveys were distributed to the faculty with instructions to administer them at some time within a one-week time frame (which instructions may or may not have been followed, depending on the course schedule, or if the faculty member remembered). While faculty are encouraged to have someone else administer the survey, this is also not always done for various reasons. The instructions specifically state that the instructor is to leave the room while students are filling out the survey, but not all do. Someone other than the instructor is to take the completed surveys to the designated collection location; this is also not always the case. One other item of note: students are given no instructions other than to fill out the forms.

III. ANALYSIS AND RESULTS

Analysis

Regression and ANOVA were used to analyze a sample size of 3134 completed course evaluation instruments, although because of missing values the full sample size was not always available for every computation. In no case, however, did the number of observations fall below 3000.

At this point, the reader may be heard to groan, “Oh, no! Not another number-cruncher on the same-ole, same-ole!” Don’t bail yet! While this paper presents some of the more interesting results of the statistical analysis, the authors use that analysis as a springboard to discuss topics such as: 1) Is it reasonable to assume that the concept of “teaching effectiveness” (T.E.) means the same thing to students as it does to faculty and administrators (assuming agreement between faculty and administrators, which may be a stretch)? 2) What other terms in the survey possibly suffer from double, or even multiple meanings, depending on the mind set of the reader? 3) Could, and if so how might, the Expectancy Theory of Motivation, for example, have a role to play in how students rate their instructors’ teaching effectiveness? 4) What other theories may help to explain counterintuitive empirical results? The MBTI has been utilized to address different teaching and learning styles - could, for example, extroverted instructors be more likely to receive higher T.E.s than their introverted counterparts, thus making such instruments biased against introverts? (Do we smell *lawsuits*???) What can Piaget’s Theory of Cognitive Development add? As well as others? -- to the on-going discussion and continuing debate over the validity and reliability of student evaluations of faculty, and the serious questions about their (mis-?) use in faculty tenure, promotion and merit pay decisions.

Results

Teaching Effectiveness (T.E.) as a function of Expected Grade, Attendance, Out of Class Preparation Time Spent, and Perceived Course Difficulty

As happens toward the end of each term, students arrived at class one day to be confronted with an opportunity to evaluate the instructor. On the new instrument, students were specifically asked: “How would you rate this instructor’s teaching effectiveness?” and five choices, with 1 = Significantly Below Average, 3 = Average, 5 = Significantly Above Average.

Using the REG procedure in SAS, regression analysis found several not-so-surprising correlations: 1) the higher the expected grade ($t < .0001$); and 2) the easier the course was perceived to be relative to other business courses ($t > .0001$), the higher the T.E. The interesting result was that 3) the fewer hours *the student spent preparing* for class ($t = .0015$); the higher the *instructor* was rated on T.E. There was no significant relationship between class attendance and T.E., so it looks as though the key to being an effective teacher, at least that semester, was leading your students to expect high grades, to feel that the course was easy, not requiring them to spend much time out of class on the course -- while not making them spend much time **in** class, either. If you’re finding it difficult to relate that to your effectiveness as a teacher, don’t feel alone!

But wait! Examination of the survey instrument reveals that the two questions preceding the “number of hours spent preparing for this course” have answers with 1 = a small number to 5 = a large number, while the question on “out-of-class preparation time” has 1 = “10 hours or more” to 5 = “less than 1 hour.” Could the change in the pattern of answers have had an effect on the responses? Are we assuming too much about the attention span of our audience, since the item of interest is on page 3, and is item number 22? One would hope that college students could handle this, but what about all those stories about the “TV Generation”?

In examining covariance of the independent variables, there was no significant relationship between hours of out-of-class preparation and expected grade, although the relationship between expected grade and difficulty of the course was positive and significant, as was the difficulty of the course and number of preparation hours reported.

Teaching Effectiveness as a function of Clear Policies, Follows the Syllabus, Dependable in Meeting Class, Well-prepared for Class, Careful & Precise in Answering Questions, Course Materials Well-Organized, Assignments Enhance Learning, Exams Reflect Objectives & Material, Stimulates Interest in the Subject, Available Outside Class, Motivates Students to Best Work, Increases Knowledge of Students

The above twelve items from the survey were analyzed to determine their relationship to teaching effectiveness. Most results followed common sense expectations - except one. There was a strong significant, **negative** relationship where it made no sense; between the statement “The instructor is dependable in meeting scheduled classes” and Teaching Effectiveness ($t < .0001$).

As we puzzled over these results, several thoughts came to mind. First was the possibility that meeting class confused a statistically significant number of students, thus reducing T.E. Next was the possibility that they didn’t understand one (or both) of the original questions. In the face of these mind-boggling possibilities, the authors decided to conduct a bit more research.

Ideal vs. Typical Student

We faculty and administrators would like to believe that all our students are interpreting and responding to the concepts used and questions included in such faculty evaluation instruments as “the ideal student” would, yet we know that many of our students, and just about all of them on some days, are more “typical” and less “ideal” than we are willing to admit. So what kind of information **are** we getting from these ubiquitous student surveys?

In a decidedly *unscientific* survey, the authors used a convenience sample ($n=14$), class of seniors, and gave them a series of printed, open-ended questions. These questions were passed out one at a time and collected sequentially. Responses have been selected to demonstrate the wide-ranging variability in student interpretation of questions.

Students were asked: “When you rate instructors on the following item, **what characteristics or behaviors of the instructor** do you consider in your rating? *How would you rate this instructor’s effectiveness?*”

- “First I take into consideration the amount of knowledge I have gained while taking the course. ... If the course is dull, then I believe the effectiveness of the teacher will be lower than if it were an interesting course.
- Instructor’s willingness to accept feedback and answer questions without making student feel ignorant. Ability ... to keep students interested.

- Does he/she keep me awake, or does he/she put me asleep (sic).
- How well the instructor accomplished the class purpose as defined in the syllabus.
- The last thing I use is my overall attitude towards the teacher about how I felt about the material I learned.”

Has the thought, “Have they read the question?” crossed your mind?

Remember the interesting negative correlation between “dependable in meeting class” and Teaching Effectiveness? Wait until you read some of these! Students were directed: “When rating instructors on *The instructor is dependable in meeting scheduled classes*, explain how you interpret this item.” Here are some of the responses:

- “If we meet for class and the teacher is there and on time every time then I would consider the instructor dependable. Considerations are taken for special situation and crazy traffic days.
- The instructor will show up for the class unless she/he has informed the class in advance of the absence.
- The instructor arrives to class on time and does not miss class unexpectedly often (sic).
- I see this not as classes, but is available for assistance. (*emphasis added*)
- This means that the instructor is showing up for class. They are not canceling classes on a regular basis.
- The instructor comes to class when expected. I do not consider tardiness.”

If this doesn’t raise questions of inter-rater reliability, what will?

Another issue that concerned us came from comments by students to the effect that “Nobody pays any attention to the evaluations anyway.” So we asked the question: “How do you think the administrators in the College of Business use the Student Evaluations of Teaching Effectiveness?” Here are some responses (out of the mouths of babes):

- “They probably use this in evaluating performance annually, or whenever a professor is due for a pay increase. It only makes a small percentage of the overall instructor evaluation.
- I really do not know. I do not think that it’s weighted very much because you cannot trust some evaluations. Some students may give a teacher a bad eval because they are doing poorly in the class.
- Teacher evaluations. Effectiveness measurements. Student satisfaction survey. Something to use their \$200 paperweight on.
- I don’t think they use it at all. Because I know teachers that have gotten (& deserved) bad evaluations from me & others, and I have seen or heard no difference in their teaching.
- I don’t know, but most of the teachers have tenure so there’s not much they can do. I really don’t think they use it much.
- I would hope that they use it for purposes like: giving raises, whether or not some instructors remain here or are let go. Although I doubt that many instructors are let go because of a bad evaluation.”

Finally, we asked the students, “Are there any factors that you believe are important that aren’t included on the evaluation form?” They had some interesting suggestions!

- “If you had to take this course again, would you take the same instructor? Why?”
- Did the professor seem genuinely concerned about students learning?

- The personality of the instructor. Some instructors have some really crappy personalities.
- Should have on there what the evaluation is used for. Most students can't believe that they are used for anything."

Perhaps it is time for the administrators to engage in some active listening.

IV. CONCLUSIONS AND IMPLICATIONS FOR FUTURE RESEARCH

From the above examples, it is very clear that wide discrepancies exist among students in their understanding of concepts and terminology used in the faculty evaluation instrument tested here. Further, they have no common understanding of why they are asked to complete the survey, and how or even whether it will be used. These are significant issues.

Referring to the questions posed previously; first, it is **not** reasonable to assume that the concept "Teaching Effectiveness" means the same thing to students as it does to faculty and administrators. In fact, it means quite different things to different students.

Second, there **is** confusion about the meaning of other terms, and in the interpretation of other questions on the survey. This is shown clearly in the responses to the open-ended questions posed to seniors, who are veterans of the process.

Third, in a very simplistic application of the Expectancy Theory of Motivation to this particular College of Business and its "Student Evaluation of Teaching Effectiveness," we can anticipate little motivation on the part of many, perhaps most, students to take the evaluation seriously, and thus to exert much effort or attention to complete it, since they have little expectation that anything of value will come of their effort.

Fourth, as far as other theories such as Piaget's Model of Cognitive Development, clearly one's ability to think abstractly will affect one's perception of teaching effectiveness in a more theoretical course. Any truly meaningful model of the learning process will incorporate multiple theoretical perspectives, and any truly meaningful evaluation of teaching effectiveness will recognize that learning is the ultimate measure of teaching effectiveness.

Finally, given the laxity in data collection, such "data" (if it can even be called that) would not be acceptable for a term paper. Yet it is being utilized to evaluate the performance of professionals. What more is there to say.

V. REFERENCES

- Cashin, W.E. (1995) "Student Ratings of Teaching: The Research Revisited." *IDEA Paper No. 32*. Manhattan: Kansas State University, Center for Faculty Evaluation and Development.
- Marsh, Herbert W. (1987) "Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research." *International Journal of Educational Research*, 11(3), 253-388.
- Read, William J., Dasaratha V. Rama and K. Raghunandan. (2001) "The Relationship Between Student Evaluations of Teaching and Faculty Evaluations." *Journal of Education for Business*, v. 76 I 4, p.189.