

# STUDENT EVALUATIONS: HELP OR HINDRANCE?

Tom Dulz, Frostburg State University  
Paul Lyons, Frostburg State University

## Abstract

*Our basic assumption is that instructor classroom behavior is strongly influenced by student evaluations. With a review of the literature and our empirical findings, we try to answer the following questions: What instructor behaviors are encouraged by evaluations? What do we know about the connection to student outcomes? Do students feel their needs are being served? Is there a connection between student evaluations and student learning?*

*We find that contemporary student evaluations most highly value the traditional model of the “sage on the stage”. Elements reflecting student behavior are valued less. Outside of test taking performance there seems to be little known about the relationship between evaluations and other student outcomes such as learning, development, growth and change. Our sample of students report that their evaluation system serves administrator’s needs much more than student needs.*

*Based on our findings, we argue that contemporary student evaluations act as a hindrance to teaching improvement.*

## I. INTRODUCTION

Cashin (1995) reports more than 1,500 references dealing with research on student evaluations of teaching. The November 1997 issue of the *American Psychologist* had a “Current Issues” section devoted to student evaluations. The topic of student evaluations is alive and well.

Student evaluations are important. They feed into the assessment and evaluation of instructor performance. Evaluations typically serve a dual purpose — as a key input into personnel decisions (for example, promotion, pay, tenure) and an additional (and often secondary) purpose of “instructor development”. A vast number of college and university instructors are rated by students in every course, every semester or term.

Our interest in this topic stems from our belief that instructor classroom behavior is strongly influenced by student evaluations. What behaviors are encouraged by evaluations? What do we know about the connection to student outcomes? Do students feel their needs are being served? Is there a connection between student evaluations and student learning? The purpose of this paper is to examine these questions.

## II. WHAT GETS MEASURED?

Although the term “student evaluation” is regularly used, it is more realistic to talk about student *rating*, because that is typically what each student does. That is, the student chooses a number using a rating scale of some sort. There is little evidence in the literature of any wide use of student-volunteered evaluation criteria. In other words, students respond to the instrument presented to them. Evaluations are not often conducted by students

No single rating item or set of related rating items is claimed to be effective for all purposes of teaching. (see, for example, Abrami & d’Apollonia, 1990; Marsh & Dunkin, 1992). Yet, evaluations seem to share common elements. Cashin (1995) reports that reviews of factor analytic studies of evaluation instruments find six common factors. They are: 1) Course organization and planning; 2) Clarity, communication skills; 3) Teacher - student interaction; 4) Course difficulty, workload; 5) Grading and examinations; 6) Student self-rated learning.

For the most part, these factors emphasize instructor related attributes. It is interesting to note that relatively little attention is given to student-related issues.

Overall, a picture of the “ideal” instructor emerges — a “sparkplug”, a motivator, a source of knowledge, the center, the linchpin of the learning process. Of lesser importance are “student” or “process” centered behaviors. The “sage on the stage” seems valued more than the “guide on the side”.

As Caplow, et.al. (1995) point out, items commonly included on course evaluation instruments are based on “teacher characteristics/behaviors that *students* ( our emphasis ) identify as effective or ideal.” (1995:225). It is curious that the only topic on which students are considered to be experts is teaching effectiveness.

## III. VALIDITY AND INTERPRETATION OF MEASURES

In most studies, validity and outcomes aren’t discussed. A critical reading of these studies shows that effective teaching is assumed to be what students say it is. As discussed earlier, a common set of factors shows up in most evaluation instruments. The student’s ideal teacher, among other things, knows a lot, lectures in tandem with the textbook and is very entertaining. Not surprisingly, student’s definitions of “effective teaching” are idealized descriptions of what they are used to — a teacher centered environment. Less value is placed on student centered factors with which they have less experience, such as encouragement to develop their own ideas. (Caplow, et.al., 1995) Evaluation instruments tend to reflect this bias toward a teacher centered ideal.

Interestingly, in those studies where validity is addressed, the discussion almost always includes statements such as Cashin’s: “Unfortunately there is no agreed upon definition of ”effective teaching” nor any single all-embracing criterion.” (1995:3) Yet in spite of this lack of agreement, evaluations show a remarkable similarity. In fact, it is not considered strange to use a common evaluation instrument across the board for all courses in a department, program, school or even the entire university.

Reviewing the literature, we find that, in many cases, validity is not explicitly considered. Studies which do attempt to assess validity ( Greenwald (1997) reviews a number of these.) usually consist of comparing evaluation scores of different instructors teaching the same course with student scores on

final exams. Students who scored higher on exams tended to give their teachers higher ratings. This is congruent with the finding discussed earlier which shows that students place a higher value on instructor behavior which enables them to achieve higher evaluations from the instructor.

If competency in taking tests is acceptable as an outcome measure of student learning, then contemporary evaluation practices can be said to exhibit a degree of validity. The question, of course, is to what degree does skill in test taking represent learning?

Perkins (1985) examined the reasoning faculties of students who were entering college and compared them to students who were graduating. He found little discernible change. Voss, et. al (1989) found very little difference in the ability to solve problems about economic issues between those who had studied economics and those who had not. Gardner (1991) describes a number of studies where students who had successfully passed courses in a particular subject area performed poorly when asked to apply the knowledge they had supposedly learned outside of a classroom setting. He discusses examples from physics, biology, mathematics, statistics, history, literature and the arts. In every case, students who were successful test takers did poorly when asked to draw on the knowledge they had ostensibly developed through several or more semesters of study.

These results shouldn't be too surprising. After all, learning is an *active* process. Listening and reading are but one component of learning. By themselves, they are not sufficient preparation for the application of learning. But, in the typical classroom, this is as far as it goes. We know that very little is retained when we hear a long string of words. Yet lecturing remains the favorite method of college teaching.

We were unable to find reported instances of student outcomes other than enhanced test-taking related to teacher centered behavior.

#### **IV. VALIDITY FROM THE STUDENT'S PERSPECTIVE**

During a classroom discussion of issues in process measurement, some students expressed dissatisfaction with the university course evaluation instrument. Given the popularity of course evaluations in academia in general, and especially the heavy reliance on course evaluation scores at this university, we were somewhat concerned with these student's attitudes.

These were students in an evening MBA program. They were very familiar with the evaluation process. For most, this would be their tenth course in the program, and consequently, the tenth time they had been presented with the same format. While we were concerned with validity — was there a connection between the university evaluation instrument and learning — student's classroom comments made us wonder about a more basic question. What were student's perceptions of validity of the process in general?

To test whether the sentiment expressed in this class was an anomaly, or representative of a deeper issue, we decided to administer a "quick and dirty" set of questions to students. Toward the end of the semester, a week before the "official" instructor and course rating forms were distributed, we asked four basic questions to graduate students enrolled in three sections of the same course (n = 41). One of these sections was the one described above where students expressed dissatisfaction.

These are the questions:

1. How important is course evaluation to you? (4.0)
2. To what degree do you feel the present evaluation system serves student needs? (2.4)

3. To what degree do you feel the present evaluation system serves faculty needs? (3.5)
4. To what degree do you feel the present evaluation system serves university administration needs? (4.2)

Students were asked to respond to each question using the following scale:

Not at all					Very Much	
1	2	3	4	5	6	7

Following each of the questions, in parentheses, is the mean response value. Assuming the value of 4 is the mid-point of all of the ratings, the results do not lean toward the rating, “very much,” on any of the items.

## V. INTERPRETING THE RESPONSES

In general, it seems that the evaluation of courses and instructors is of middling importance to these students (4.0 on a 7-point scale). However, the system with which they have had the most experience, receives a very low value (2.4 on a 7-point scale). This is somewhat alarming when we realize that students are (or should be) the most important participants in the whole teaching-learning-assessment continuum. This sample of students seems to feel that the evaluation system in use barely serves their needs, if at all. They feel that faculty needs (3.5) are met more than student’s (2.4), while the present system serves university administration needs most of all (4.2).

## VI. EVALUATION FROM THE STUDENT’S PERSPECTIVE

If the questions on the standard evaluation instrument were not those students would ask, we decided to try to discover what they *would* ask, if they had the chance. Students in six MBA courses were asked to evaluate the course they had just completed. The following is based on these results.

Reviewing the their responses we found two basic themes. The first, and overall, most common, was the issue of *relevancy*. Were personal objectives met? Were useful skills learned? Could they be used on the job or in life? Altogether students seemed most interested in what they felt they could “take away” from the course.

The other theme we found reflects interest in the quality of *relationships* with other *students*: the ability to share what was learned, the freedom to express ideas and the opportunity to learn from other students. What seems noteworthy is what they don’t mention. They don’t focus on the instructor as the source of knowledge, their emphasis is on learning from other students. A small minority disagreed, they feel their learning is the responsibility of the instructor.

It is important to note here that these students were mature ( average age in the thirties ), part-time graduate students with a substantial work history. Their responses contrast quite sharply with the instructor-centered characteristics found by Sherman, et. al., (1987) described earlier.

## VII. THE INFLUENCE OF EVALUATIONS

We opened with an assumption that instructor classroom behavior is influenced by student evaluations and posed several questions regarding these influences. Here we will briefly address these questions.

*What behaviors are encouraged by evaluations?* Student learning, skill development, performance and other student-related outcomes are usually not assessed. The behavior that typically gets measured and encouraged is the instructor's. Common examples are instructor presentation, organization, communication, knowledge and the like. A pretty clear picture of the ideal type emerges: that of the "sage on the stage" — the instructor as the center of learning. Any instructor who doesn't fit this mold will receive less positive evaluations. It is not unreasonable to assume instructors will be reluctant to engage in any behaviors that deviate from this model.

*What do we know about the connection of ratings to student outcomes?* Unfortunately, we don't know much. Many studies don't actually directly address student outcomes. In studies where student outcomes are considered, they are often represented by scores on objective tests. Some studies correlate instructor ratings with student scores on examinations, others with student grades.

While test scores are a popular method of gauging student performance, it is not a complete measure of learning. When other indicators of learning such as understanding or the ability to apply concepts are used, college students across the board do poorly. (Gardner, 1991; Perkins, 1985; Voss, et al, 1989) The best that can be said is that it seems that instructors who rate well tend to produce students who test well.

*Do students feel their needs are being served?* From our survey we conclude that the response to this question cannot be a resounding "yes." In our sample, students see the administration as the primary beneficiary of the process. They do not seem to feel that course (and instructor) evaluation serves their needs. They report that there is little connection regarding their participation in the ratings process and what happens to them. As this survey only captured responses of students in one program, an unresolved question is whether these sentiments are typical of students in general.

*Is there a connection between student evaluations (ratings) and student learning?* From the information we reviewed, it seems that student ratings of instruction are only marginally related to student learning or to student needs. Ratings do provide administrators with objective-appearing indicators of instructor performance often translated to numerical scores. Taking the scores to multiple decimal places further asserts the scores' "objectivity". However, do the scores speak to the learning, development, growth, and change of students? We have not found evidence that they do.

## **VIII. TEACHING EVALUATIONS AS AN IMPEDIMENT TO CHANGE**

The overarching impression one gets from reviewing the literature is the general agreement on valued instructor characteristics. The picture that emerges shows the instructor as the keystone to the whole learning process.

If there were general satisfaction with student learning outcomes, this approach could be applauded. However, as we noted earlier, a number of studies have questioned the quality of student learning, at least at the undergraduate level. More specifically, at the MBA level, there have been a number of critical voices, from Porter and McKibbin (1988) to Elliott, et. al. (1994) and Eberhardt et. al., (1997) which reflect similar concerns. In a nutshell, a common theme in all of these reports is the student's inability to apply knowledge in unstructured situations.

One response to these issues has been a pedagogical shift from a concern for teaching to an emphasis on learning. (Savery & Duffy, 1995; Stinson & Milter, 1996; Wheeler, 1998) The issues involved are too complex to be discussed in detail here, but the overall theme is that, to be successful,

the role of the teacher must change. “He or she spends very little time lecturing and transferring information to students. Rather the role becomes a combination of both learning manager, tutor, and coach.” (Stinson, 1997). Notice that this is the antithesis of the role valued in most evaluations.

Faculty resistance to change has often been noted. (Witt, 1994; Mullins and Fukami, 1996) Clearly, there are a number of potential causes of resistance, but a major one, we believe, is the method of faculty evaluation presently in common use. If we hope that student outcomes will change, then teacher behavior will have to change. New and different roles must be developed. We are not arguing here for or against any particular method or approach. We do however, wish to point out that it is unreasonable to expect faculty to experiment with new approaches when low student evaluation scores can result in negative consequences for pay or promotion.

## IX. CONCLUSION

There is an old management saying to the effect that management gets, not what it expects, but what it inspects. “Official” measures, sanctioned by organization administrators and managers, contribute to insuring conformity and similarity of behavior. Imbedded in the use of any measure is the notion of comparison — a comparison between what is measured and what is desired.

It is possible, then, to infer management desired behaviors by examining management sanctioned measures. As we have described, student evaluations used in colleges and universities across the nation display a remarkable commonality. Using what is measured as a guide to what is desired, we note the following.

The implication is that the instructor is assumed to be the expert and the primary source of knowledge. “What “ is to be learned is the instructor’s responsibility. Evaluation instruments are silent on the subject of student objectives, roles or responsibilities in learning. The primary emphasis is on *teacher* led behaviors, *student* led behaviors are generally ignored. Vaill (1996) calls this style “institutional learning”.

As we have noted, there are a number of critics who argue that this approach does not prepare students adequately for the contemporary world, a world of change described by Vaill as “permanent whitewater”. In Vaill’s words: “The permanent white water in today’s systems is creating a situation in which institutional learning patterns are simply inadequate to the challenge. Subject matter is changing too fast. Learner’s demographics are changing too rapidly. Learners are interweaving their learning with work responsibilities and expecting their learning to be directly relevant to these responsibilities.” (1996:41) The results of our student surveys reinforce Vail’s viewpoint. Students’ central concern is personal relevancy, not instructor behavior. They don’t find conventional evaluation instruments to be of any great value to them.

If we are satisfied with present outcomes, then the present situation is adequate. However, if we accept the idea that the instructor’s role must change in order to better help students develop more effectiveness in their changing worlds, then it seems fair to ask the question: “Do contemporary student evaluations help or hinder this change process?” Based on our observations, we find more evidence of a hindrance than a help.

References upon request.