

# Evaluation Of College Faculty: What Do Accounting Students Really Think?

Paul J. Carruth, (e-mail: [pcarruth@selu.edu](mailto:pcarruth@selu.edu)), Southeastern Louisiana University  
Ann K. Carruth, (e-mail: [acarruth@selu.edu](mailto:acarruth@selu.edu)), Southeastern Louisiana University

## Abstract

*Student opinion surveys of teaching are widely implemented at the university level for many reasons. Faculty members use the input from students to improve their teaching while administrators often use the information for merit, and tenure and promotion decisions. The extensive use of student ratings is in part the result of the increased interest in improving teaching and the growing demand for greater accountability on the part of colleges and universities. At issue is the concern that students are asked to evaluate faculty on components of teaching deemed important by divisions of institutional research and not necessarily by faculty themselves. Final approval of student opinion surveys may be prematurely granted without sufficient input from consumers. The purpose of this paper is to explore issues surrounding the wide spread use of student opinion of teaching surveys for evaluative purposes and to describe the role of student ratings in judging teaching effectiveness. An extensive review of literature was used to construct a survey administered to 165 students enrolled in undergraduate and graduate accounting courses. Students were asked to rank effective teaching practices. The results of this survey guide faculty and administrators to understand, from students' perspective, effective teaching practices. Conclusions are drawn concerning the appropriate use of the information obtained from student evaluations and the impact the student ratings have on the quality of instruction in colleges and universities.*

## Introduction

*S*ystematic student evaluation of teaching is widely used to support curriculum and personnel decisions in higher education. In recent years it has become just as likely for students to evaluate teachers as it is for teachers to evaluate students. This is due in large measure to a reform movement in higher education that demands faculty be more accountable to the student population (Seldin, 1993). This has led to the acceptance and widespread use of student evaluations as an index of instructional quality. In fact, more than 70 percent of all institutions use student opinions of teaching (Ory, 1991). This type of teacher evaluation is believed to encourage effective teaching and to increase accountability.

Teaching effectiveness has been defined as the teacher's ability to facilitate students to achieve their highest level of independent thinking. It therefore involves more than merely presenting information in an organized, interesting format. Effective teaching requires a meaningful interaction and dialogue between teacher and student. Faculty attributes that have been associated with teacher effectiveness include: enthusiasm, clarity, preparation/organization, and love of knowledge (Sherman, Armistead, Fowler, Barksdale, & Reif, 1987). Facilitative teaching behaviors also include good communication skills, positive attitude toward students, and the ability to encourage students to think for themselves. On the assumption that these attributes are measurable, students are routinely asked to evaluate teaching effectiveness. This has led to a long standing debate over the reliability and validity of instruments used to measure student opinions. Perhaps one of the reasons for the debate over the validity of student evaluation instruments is that researchers have not asked students which teacher attributes they believe are important to learning. Therefore, the purpose of this study was to assess, from the students' perspective, the importance of indicators used to measure teacher effectiveness. To characterize what is viewed as important by students may help explain variables that bias the evaluation process.

## **Review of Literature**

In recent years, student evaluations have played a greater role in administrative personnel decisions. Faculty salary, promotion, and tenure decisions are increasingly influenced by student ratings. A perceived misplaced emphasis on the results of student surveys has led many faculty to reject outright their validity and usefulness. One unfortunate result of this is reluctance on the part of some faculty to accept the usefulness of student ratings for other purposes, such as faculty improvement and development (Emery, Kramer, & Tian, 2003). This negative view of student surveys is reinforced by the opinion of many faculty that student ratings do not provide a true reflection of teaching ability. This has led to evaluations being referred to as a “necessary evil” by many academicians (Benson & Lewis, 1994; Headrick, 1991).

Much debate exists in the literature concerning whether student evaluation instruments are reliable and valid. As a measure of reliability, Marsh (1984) found that evaluations at the end of the course were highly correlated with the evaluations by the same students 1-3 years later. More recently, Rinderman & Schofield (2001) concluded that instructor variables were reliable when courses were taught by the same instructor. Other researchers, while accepting the reliability of student ratings, maintain there is only limited evidence of their validity (Craig, Redfield, & Galluzzo, 1986; Shevlin, & Banyard, Davies, & Griffiths, 2000). A number of extraneous variables are viewed as contributing to this lack of validity. These variables include instructor personality and popularity, sex and rank of the professor, class size, class level, and expected grade (Rotem & Glassman, 1979). For example, ratings were found to be marginally higher but not statistically significant in small classes, discussion classes, classes in the humanities (Seldin, 1993), and electives (Arden, 1989). However, no consistent relationship has been established between student ratings and the instructor’s rank, sex, or research productivity (Seldin, 1993; Tanner, Manakyan, & Hotard, 1992). Furthermore, little or no relationship has been found between student ratings and student demographic data such as age, year in college, sex, grade-point average, or academic ability.

Many instruments that evaluate teaching effectiveness focus principally on classroom instruction. Giving a student the opportunity to evaluate the quality of teaching is generally not objectionable to most faculty. There is general agreement that students are in the best position to assess a faculty member’s ability to communicate ideas in a logical, understandable way. In addition, students are able to judge ethical and professional behavior in the classroom, to assess faculty rapport with students, and to evaluate an instructor’s ability to stimulate interest in the subject (Seldin, 1993). However, it is not reasonable to expect students to be able to evaluate whether a teacher has an adequate command of subject matter or whether the content is appropriate for the level of students enrolled in the course. Nevertheless, while some believe students have limited ability to evaluate teaching effectiveness, others do not believe students should evaluate faculty on any criteria.

An important aspect of student evaluations that is often overlooked relates to the identification of factors deemed by students to be important or relevant to their learning. How often do universities ask students to validate whether items on the questionnaire contribute to their successful learning? Furthermore, if recent curriculum changes are designed to encourage creativity, independent thinking, strong communication skills, and expertise in developing logical thinking, then universities should consider revising current evaluation instruments to incorporate appropriate indicators reflective of a wide range of teaching/learning strategies. These evaluation instruments should reflect learning skills needed by students to survive in a competitive world. Perhaps faculty should be evaluated on their ability to teach students the ability to communicate, analyze, reflect, integrate, understand, and appreciate. The lack of research from the students’ perspective generated the research question for this study: To what extent do students believe frequently used indicators are important to learning?

## **Methodology**

Data were collected from a convenience sample of 165 undergraduate and MBA students enrolled in accounting courses. The questionnaire used as the data collection instrument was developed for this study and consisted of 45 items generated from an extensive review of the literature. Items included reflect teaching effectiveness, course effectiveness, and student behaviors believed to contribute to successful learning. Students were assured confidentiality and that their standing in the course, college, or university would not be jeopardized if

they choose not to participate. Instructions were given to rate the importance of indicators without any one professor or course in mind. Possible responses ranged from 5 (very important) to 1 (not important at all).

Among the examined demographic characteristics of respondents were age, gender, and class standing. Over half of the respondents (66%) were traditional students between the ages of 18-25 years. The majority of respondents (56%) were female. Sixty-one percent of the sample were undergraduate accounting majors while 39% were MBA students taking an accounting course. Both undergraduate and MBA courses were required in their respective degree programs.

### **Data Analysis**

Because teaching effectiveness is a multidimensional construct, it was deemed appropriate that each item be used for comparison between groups. Mean ranks using Mann-Whitney  $\bar{U}$  were used for three post-hoc comparison groups: traditional vs. nontraditional students; undergraduate vs. graduate students; and male vs. female students. Mann-Whitney  $\bar{U}$  was used because it does not require normally distributed data but is sensitive to the central tendency and the distribution of the scores (Munro & Page, 1993). To protect against a Type I error, a Bonferroni correction was employed by dividing the level of significance by the number of comparisons. Therefore, a significance level of  $p \leq 0.02$ , not 0.05 was used for this study.

In Table 1, mean ranks according to traditional vs. nontraditional students are summarized. Only those items that were significant are reported. First of all, traditional students rated teacher concern for student progress significantly higher than did nontraditional students. Perhaps older, nontraditional students are more mature and self-reliant and thus need less reassurance from the teacher than do younger, traditional students. In addition, three indicators indirectly related to grades were rated significantly more important by traditional students. These indicators were that bonus points should be given, tests should come from class material, and that tests should not be too long. This suggests that younger students are interested in learning only a finite amount of material to pass the course. Nontraditional students, on the other hand, felt course work that encourages creative and original thinking should be encouraged. One explanation for this may be that older students may have a perspective that allows them to appreciate the benefits of courses that require thinking rather than mere rote memorization. Also, older students may have more life experiences to draw upon when asked to apply new information to various situations.

In Table 2, mean ranks according to gender are summarized. Female students rated the desire for clear explanations of subject matter, learning factual information, and courses that enhance their career-related skills as more important than did their male counterparts. This may suggest that female students are generally more focused on their school work and career-related preparation than are male students. Furthermore, female students rated the teachers' concern for student progress and the teacher's responsiveness to student needs as significantly more important than did male students. This may indicate that, for females, establishing a relationship with faculty is important to their learning.

In Table 3, mean ranks comparing undergraduate and graduate students are summarized. Interestingly, teacher attributes were rated consistently more important by undergraduate students when compared to graduate students. For example, undergraduates rated the item "teacher demonstrates respect for students as individuals" as significantly more important than did graduate students. Furthermore, undergraduate students, compared to graduate students, felt it was very important that teachers provide clear explanations of the course material as well as provide students the opportunity to review their exams. Undergraduates seemed to express greater expectations of faculty than graduate students regarding these issues. In addition, compared to graduate students, undergraduates tend to be focused on the more immediate aspects of school. High ratings by undergraduates on issues such as classes beginning and ending on time and course objectives being made clear demonstrate this. The greater emphasis by undergraduates on the short-term aspects of schooling is further revealed by the importance they place on having test questions come from material covered in class, having test that are not too long, and receiving a good grade in the course.

Undergraduate students (apparently with one eye on the future) rated the item “the course enhances my career-related skills” as being more important than did graduate students. However, when given the opportunity to rate the importance of items such as course requirements that assist in the development of speaking and creative thinking skills, undergraduate students rated these items as significantly less important than did graduate students. The higher ranking of items related to speaking and critical thinking skills by graduate students may be due in part to their having a broader perspective concerning what knowledge and learning skills are necessary for success in today’s job market.

Several indicators not significantly different among the three comparison groups are worth noting. Table 4 summarizes in rank order of importance indicators included in the questionnaire that were not rated significantly different between the group comparisons. Not surprisingly, indicators rated most important by all respondents relate to student grades, while indicators that relate to students being intellectually challenged were rated much lower in importance by the students. For example, items such as clear and unambiguous test questions, fair grading process, and a clear basis for determining grades received the highest rating of importance to students. On the other hand, courses that require assignments that assist in the development of writing skills, intellectual effort greater than that of most courses, and out-of-class homework assignments receive much lower ratings of importance by students. Further, the respondents do not seem to highly value courses that have a challenging pace, challenging examinations, or teachers that ask thought-provoking questions. Nevertheless, perhaps not even surprisingly, students ranked as highly important a course that affords the opportunity to learn a great deal!

One criticism often made of student evaluations is that they are essentially popularity contest that favor teachers with a good personality or sense of humor. The results of this study only partially support this assertion. While students did rate a friendly personality as important, a sense of humor was not rated as important as many other indicators. “Having a friendly personality” may speak to the development of rapport with students, which also was highly rated among indicators.

A number of indicators were rated much lower than anticipated. For example, one might assume that business students would expect faculty to dress professionally. However, this item was not rated as important as other indicators. This may suggest that students are able to evaluate faculty teaching based upon factors more important than how instructors dress. Perhaps more surprising, however, was the fact that the item “the teacher curves the grades” was rated so low on the importance scale. And finally, the use of teaching aids, such as audio and video presentations, was rated as having the least amount of importance of all factors. Future studies should try to determine which, if any, audio/visual aides are viewed by students as beneficial to learning. This is particularly important given the fact that many business schools are increasing their use of technology in the classroom.

## **Discussion**

Student evaluation scores generated to document improvement in teaching effectiveness is one way faculty speak to instructional quality for the purposes of the tenure and promotion. Marsh (1984) suggested the most widely accepted criterion of effective teaching is student learning. This assertion was confirmed in this study by the fact that students rated the opportunity to learn as one of the most important indicators. Consistent with other studies, the questionnaire developed for this study used instructional indicators to assess the quality of teaching (Orpen, 1980). These indicators of instructional quality included items such as teachers writing test questions that are clear and unambiguous, being fair in grading practices, motivating students to do their best work, and communicating the expectations of the course.

The importance of various indicators of teaching effectiveness differed among students in the study sample. For example, undergraduate students rated assignments and tests that require original or creative thinking as significantly less important than graduate students. Additionally, undergraduate students rated the development of speaking skills as less important than did graduate students. These findings underscore the reason many argue against allowing students to evaluate effective teaching. If students do not value the types of activities that faculty believe help students learn, students may in turn rate demanding faculty lower than less demanding faculty who require less rigorous course work (Carey, 1993; Crumley, Henry, & Kratchman, 2001; Seldin, 1993). Another

argument against student ratings is that some student groups want learning to be “easy”. For example, traditional students rated giving bonus points, testing the material covered in class, and giving tests that are not too long as significantly more important than did nontraditional students. What happens to student ratings for those faculty who teach traditional students and do not engage in the types of activities these students view as important? One concern is that faculty may subconsciously or even knowingly change their teaching strategies in the hopes of getting higher student evaluations.

Giving students what they want can erode instructional quality in institutions of higher learning in a number of ways. Faculty may feel more pressure to give less challenging exams, which can result in grade inflation. It is reasonable for students to feel test questions should be clear and unambiguous, but is it reasonable for students to believe tests should not be challenging? Additionally, when faculty know students do not value certain course attributes, such as a challenging pace with a demanding workload or thought provoking questions, they may be more inclined to lower standards. This may be especially true for faculty teaching traditional, undergraduate students. It is especially difficult for untenured faculty not to feel they must solicit favors and even lower their academic standards in order to obtain job security (Cholakian, 1994). This may be one reason why Young (1993) argues that future generations will look back to this time in university history and wonder why this generation allowed “predominately immature, undisciplined, naive, and academically weak students to wield such enormous influence in determining what they should be taught, as well as how and by whom it should be done.”

In recent years, the quality of the indicators used by administrators to determine teaching effectiveness has come into question. The relatively recent widespread use of student ratings to judge “excellence” in teaching, has resulted in a great deal of faculty frustration associated with the process of trying “to get the ratings up.” This frustration has been compounded by the fact that as the emphasis of education moves toward active learning and away from passive learning, current instruments used to evaluate teaching may not be able to adequately measure teaching effectiveness.

The dichotomy between administrators’ emphasis on using student evaluation scores to judge teaching effectiveness and faculties’ belief that these scores are not valid as an indicator of teaching effectiveness is growing in the academic community. Much has been written arguing the merits and limitations of using student generated scores to measure teaching effectiveness. For example, Cholakian (1994) points out that with so much emphasis placed on the “performance” of the teacher, young faculty may become motivated to concentrate on style over substance. The results of this study support this argument. For example, female students rated teacher responsiveness to their feelings and needs as significantly more important than did males. Additionally, undergraduate students felt strongly that classes should begin and end on time. Also, all students felt faculty should demonstrate respect for them as individuals. A positive teacher-student relationship should be expected from every faculty member. However, should one faculty member who is better than another at generating higher scores on these particular indicators be rated by administrators as a more effective teacher? The point is that teaching effectiveness is multidimensional and several indicators should be used for evaluation.

### **Summary and Conclusion**

The fact remains that if a student-teacher model is adopted where students are considered active, adult learners, then student opinions of teaching will continue to play a role in faculty evaluations. From a faculty’s perspective, the real issue is not whether student evaluation instruments are valid or whether they can be used to improve teaching. The issue is whether or not too much emphasis has been placed upon the student ratings by administrators for purposes of tenure, promotion, and merit raise decisions. Because of the emphasis on students’ opinion of teaching, faculty may feel that they have to change their teaching strategies in order to “please” students. This loss of integrity may be insidious, and ultimately may lead to grade inflation that erodes the quality of higher education. In fact, Vasta and Sarmieto (1979) found that liberal grading on the part of an instructor does improve student evaluations, but not student performance.

Student evaluations of teaching can provide valuable insights that assist faculty in improving their teaching effectiveness. However, faculty who receive low ratings and critical student comments may be more likely to

experience anxiety, frustration, and ultimately, diminished enthusiasm for teaching (Seldin, 1993). Therefore, it is important that universities provide a strong faculty development program to assist faculty and administrators in interpreting and improving the results. This will ensure that student ratings are not used only as an evaluation tool, but as a means to improve faculty teaching effectiveness while at the same time maintaining high standards of academic excellence. Furthermore, because a number of variables can bias the responses of students, student ratings should not be used as the sole criteria by which teaching effectiveness is evaluated. Suggested that in addition to stressing accountability in the classroom, administrators should also encourage and support a culture of academic leadership to facilitate teaching expertise among faculty (Hendry & Dean, 2002).

**Table 1. Indicators of Teaching Effectiveness by Age**

Item	Mean Rank for Traditional Students between 18-25 years of age	Mean Rank for Nontraditional Students > 25 years of age	p
The teacher displays concern for student progress	90.05	69.29	.002
The teacher gives bonus points	89.18	70.97	.01
Tests questions come from material covered in class	90.22	68.94	.0001
Tests are not too long	91.18	67.08	.0008
The course work includes assignments and tests that require original or creative thinking	75.46	97.68	.002

**Table 2. Indicators of Teaching Effectiveness by Gender**

Item	Mean Rank for Males	Mean Rank for Females	p
The teacher provides clear explanations of subject matter	74.50	89.74	.004
The course requirements involve learning factual information	73.47	90.57	.009
The course enhances my career-related skills	72.36	91.45	.002
The teacher displays concern for student progress	72.94	90.97	.007
The teacher is responsive to my feelings and needs	70.03	93.29	.0009

**Table 3. Indicators of Teaching Effectiveness by Class Standing**

Item	Mean Rank for Undergraduate Student	Mean Rank for MBA Student	p
The teacher demonstrates respect for students as individuals	88.7	74.02	.01
The teacher provides clear explanations of the course subject matter	88.3	74.64	.01
The teacher provides students an opportunity to review their exams	90.65	70.92	.003
Classes begin and end on time	92.71	67.8	.0005
Course objectives are made clear	89.26	73.13	.01
Test questions come from material covered in class	88.10	74.95	.01
Tests are not too long	89.56	72.64	.01
Receiving a good grade in the course	89.35	72.98	.01
The course enhances my career-related skills	91.59	69.44	.0004
The course requirements include assignments that assist in the development of speaking skills	72.18	100.08	.0001
The course work includes assignments and tests that require original or creative thinking	75.5	94.8	.007

Table 4. Rank Order of Importance of Indicators Not Significantly Different Among Comparison Groups

Item	M (SD)
The teacher gives examination questions which are clear and unambiguous	4.8 (.46)
The grading process is fair	4.8 (.57)
The teacher makes clear the basis for determining student grades	4.7 (.52)
The course affords me the opportunity to learn a great deal	4.6 (.55)
The teacher is aware when students are confused	4.6 (.65)
The teacher makes clear my responsibilities for success in the course	4.5 (.60)
Agreement exist between stated course objectives and what is actually taught	4.5 (.64)
The teacher relates the course content to real world situations	4.5 (.71)
The teacher is articulate and expressive in manner of speech	4.4 (.62)
The teacher has good rapport with students	4.4 (.67)
Tests provide me a fair opportunity to demonstrate my knowledge	4.4 (.68)
Tests are returned within a reasonable period of time	4.4 (.69)
The teacher exhibits enthusiasm about the subject matter	4.4 (.71)
The teacher is readily available outside of class	4.3 (.65)
The teacher has a friendly personality	4.3 (.69)
The teacher has the ability to motivate me to do my best work	4.3 (.75)
The teacher encourages student comments and questions	4.2 (.74)
Regular feedback about my progress is provided	4.2 (.74)
The course requirements involve the learning of general principles or theories	4.1 (.78)
The teacher challenges me with thought-provoking questions	3.9 (.79)
In relation to other courses, the workload is not too heavy	3.9 (.79)
Examinations are challenging	3.8 (.75)
The teacher has a sense of humor	3.8 (.83)
The course requirements include out-of-class homework assignments that help me achieve the course objectives	3.8 (.94)
The pace of the course is challenging	3.6 (.71)
The teacher curves grades	3.6 (.96)
The teacher is professional in appearance and attire	3.6 (1.06)
The course requires intellectual effort greater than that required by most courses	3.5 (.78)
The course requirements include assignments that assist in the development of writing skills	3.5 (.90)
Teaching aids such as audio and video presentations are used	3.4 (.94)

## References

1. Arden, E. (1989). Who should judge the faculty? *College Board Review*, 152, 36-39.
2. Benson, D. E., & Lewis, J. (1994). Students' evaluation of teaching and accountability: Implications from the Boyer and the ASA reports. *Teaching Sociology*, 22(2), 195-199.
3. Carey, G. (1993). Thoughts on the lesser evil: Student evaluations. *Perspectives on Political Science*, 22, 17-20.
4. Cholokian, R. (1994). The value of evaluating. *ACADEME*, 80(5), 24-26.
5. Craig, J.R., Redfield, D.L., & Galluzzo, G.R. (1986). *Evaluating effective teaching in colleges and universities: how far have we come?* Paper presented at the Annual Meeting of the American Evaluation Association, Kansas City.
6. Crumbley, L., Bryon, H., & Kratchman, S. (2001). Students' perceptions of the evaluation of college teaching. *Quality Assurance in Education*, 9 (4), 197-207.
7. Emery, C. R., Kramer, T.R., & Tian, R.G. (2003). Return to academic standards: A critique of teaching effectiveness. *Quality Assurance in Education*. 11 (1), 37-46.
8. Headrick, C. (1991). Teacher evaluations: A necessary evil? *ACA Bulletin*, 78, 5-7.
9. Hendry, G., & Dean, S. (2002). Accountability, evaluation of teaching and expertise in higher education. *The International Journal for Academic Development*. <http://www.tandf.co.uk/journals>.
10. Marsh, H. (1984). Students; evaluations of university teaching: dimensionality, reliability, validity,

- potential bias and utility. *Journal of Educational Psychology*, 76, 707-754.
11. Marsh, H. (1991). Multidimensional students' evaluations of teaching effectiveness: A test of alternative higher-order structures. *Journal of Educational Psychology*, 83, 285-296.
  12. Munro, B., & Page, E. (1993). *Statistical methods for health care research*. (2nd Ed.). J.B. Lippincott, Co.
  - Orpen, C. (1980). Student evaluation of lecturers as an indicator of instructional quality: A validity study. *Journal of Education Research*, 74, 5-7.
  13. Ory, J. (1991). Changes in evaluating teaching in higher education. *Theory into Practice*, 30 (30), 30-36.
  14. Rindermann, H., & Schofield (2001). Generalizability of multidimensional student ratings of university instruction across courses and teachers. *Research in Higher Education*, 42(4), 377-399.
  15. Rotem, A., & Glassman, N. (1979). On the effectiveness of student's evaluative feedback to university instructors. *Review of Educational Research*, 49, 497-510.
  16. Seldin, P. (1993). The use and abuse of student ratings of professors. *The Chronicle of Higher Education*, 39, A40.
  17. Sherman, T.M., Armistead, L.P., Fowler, F., Barksdale, M.A., & Reif, G. (1987). The quest for excellence in university teaching. *Journal of Higher Education*, 27, 102-106.
  18. Shevlin, M., Banyard, P., Davies, M., & Griffiths, M. (2000). The validity of student evaluation of teaching in higher education: love me, love my lecturer? *Assessment & Evaluation in Higher Education*, 25 (4), 397-405.
  19. Tanner, J., Manakyan, H., & Hotard, D. (1992). Management-faculty research productivity and perceived teaching effectiveness. *Journal of Education for Business*, 67(5), 261-265.
  20. Vasta, R., & Sarmiento, R. (1979). Liberal grading improves evaluations but not performance. *Journal of Educational Psychology*, 71, 207-211.
  21. Young, R. (1993). Student evaluation of faculty: A faculty perspective. *Perspectives on Political Science*, 22(1), 12-16.