

How Well Do We Teach?

Luke Faulkenberry, Carole Goodson, Susan Miertschin, Barbara Stewart

University of Houston, College of Technology

How well do we teach? This question is sometimes answered through scores from student surveys of teaching effectiveness (SSTE) administered to students near completion of a course. Sometimes these survey results are the *only* data compiled systematically and available to use as an indicator of teaching performance of faculty. Thus, it is important to understand proper uses and interpretations as well as limitations of these measures.

The Assessment and Continuous Improvement Committee (ACI) of the College of Technology (CoT) at the University of Houston (UH) was formed, representing faculty in diverse program areas, to do college-wide assessment planning and implementation. A component of assessment planning was to consider issues related to SSTE, both from the perspective of what is contained in the literature about such instruments and with respect to results within the college. In order to accomplish this goal, the ACI committee reviewed reported research results regarding teaching evaluations and analyzed our own instrument, process and results. This paper examines this information, particularly what the literature tells us, the results of our own teaching evaluation implementation, and how SSTE results can be used.

What do Student Surveys of Teaching Effectiveness Measure?

Because of ease of implementation and analysis, student surveys of teaching effectiveness will be with us – at least on the short term – for better or worse. Although not without controversy, they do have validity in that they are a measure of student *reaction* to teaching effectiveness. Reaction is a concept that is defined in the training literature. Specifically, it refers to how those that participate in training as trainees react to the training process. When you measure trainee reaction, you are measuring aspects of customer satisfaction.¹ Similarly, SSTE measure aspects of student satisfaction. As in business, it is important in academia to measure student satisfaction because:

- it gives the faculty member input about how to improve the course,
- administering a satisfaction survey tells the student that what she thinks is important to the faculty and administration,
- it provides quantitative data that can be used by decision makers (including the faculty member in course, curriculum development committees, and administrators), and
- it sets standards for future comparison.

While these surveys are not measures of how much or what was learned by course participants, they are important because a student's motivation to learn is diminished in a situation where their satisfaction with the process is low.

Scriven noted that, "Student ratings add a valuable component to the range of input for the evaluation of teachers."² Some academicians question the validity of ratings based on SSTE; yet who better than the student completing a course is qualified to judge, for example, whether tests covered all the material of the course, whether the instructor presents material in an understandable way, or if the course is perceived as an excellent course?

So what, in fact, do these instruments assess? Research shows that SSTE do measure teaching effectiveness. Other research shows that contextual factors exist that can influence or bias the results of these surveys. McKeachie compiled and analyzed multiple research articles regarding measures of teaching effectiveness and concluded that student ratings were the most valid and practical source of data.³ However, some contextual factors that have been shown to bias results of SSTE include student desire to take the course, major, expected grade, grade point average (GPA), and course level.

These factors were examined by looking at the literature and by analyzing data collected at UH CoT. Teaching evaluations for all courses in the College of Technology in Spring 2002 and Spring 2004 were used in the analysis. Since SSTE are required at UH CoT, this data set covered most of the courses taught in the college by both full and part-time faculty of all ranks. The instrument items have been used for many years, and the survey items have been reviewed and critiqued periodically by faculty committees and evaluation experts. The instrument has been repeatedly commended (internally) for its simplicity and conciseness. Instrument items are in Appendix A.

For the data analysis, 4605 student ratings were averaged over the twelve items related to teacher/course satisfaction. (See Appendix A.) The values were entered into Minitab statistical software along with the corresponding course number, course level, a self reported desire to take the course, student major, and self reported GPA. Examination of boxplots and descriptive statistics indicated that the data violated assumptions of normality and homoscedasticity. Thus, non-parametric methods were used in the analysis to determine whether certain factors biased the ratings of teaching effectiveness.

Motivation. One factor examined was **student desire to take the course**. The UH instrument asks the student to respond Yes or No to the statement “I had a desire to take this course.” Asking the student to respond directly to a survey item is one way to determine the student’s intrinsic motivation to take a course. Another way is to infer a student’s desire to take a course by determining if the course is required or elective and whether the course is in the student’s major. Similarly, whether students expect a course or instructor to be good has been shown to be a weak, yet useful indicator of student desire to take the course.

Both direct and inferential measures of student desire to take a course have been examined as sources of potential bias to SSTE results in prior research. Although some research found that student desire to take a course contributed little to variations in SSTE rating scores, Cashin reports that “student motivation tends to show higher correlations with other student rating items than any other variable” and should be included in interpreting student rating data.^{4,5} Elective courses and those with a higher percentage of students taking the course for general interest were reported to generate higher ratings.^{6,7} Whether a course was within the student’s major has been shown also to yield higher ratings, and students have rated courses in their major fields and elective courses higher than required courses outside of their majors.^{7,8}

A Mann-Whitney test was performed to analyze the UH data; average ratings of teaching effectiveness were compared based on the student’s reported desire to take the course. The analysis showed that students who reported a desire to take the course generally rated the course from 0.5 to 0.59 points higher than students who did not. See Table 1 in Appendix B.

Major. Another factor sometimes reported to bias SSTE results is **field of study** or **major**. Evidence exists that major or field of study may be an influencing factor for student ratings of

instruction.^{7,9,10,11,12,13,14,15} In general, arts and humanities yielded the highest ratings. Interestingly, engineering as well as business, economics, computer science, math, and physical sciences yielded the lowest ratings. The biological and social sciences as well as the professional areas were in the middle.

All results examined at UH were from students enrolled in College of Technology courses [Engineering Technology (18%), Human Development and Consumer Science (25%), Information and Logistics Technology (34%), and Occupational Technology/Other (23%)]. Based on what had been previously reported in the literature, we expected to see Engineering Technology students rating courses and teachers lower than students in the other two departments in the college. In fact, our data showed no significant differences in survey results based on whether the students were Engineering Technology students, Consumer Science and Merchandising students, or Information and Logistics Technology students. See Table 1 in Appendix B.

Grading. Expected grade in the course, actual grade in the course, and GPA have all been examined as sources of bias in SSTE ratings. Findings suggest that expected grade is positively correlated with evaluation of teaching scores.⁶ Interpretation of such findings leads to consideration of whether high grades or high grade expectations represent grading leniency, superior learning, or some other preexisting difference. Sometimes practitioners criticize SSTE methods with the notion that professors can buy ratings by requiring very little work and by easy grading. Kaplan, Mets, and Cook reported that higher expected grades did yield slightly higher ratings on teacher evaluations; yet, they warned that the effect of grades on student evaluations may be overstated.¹⁶ Other research showed that students rated courses with high workloads higher than courses with low workloads.⁷ A more generalized approach is to look at grades via grade point average rather than grade in the course. Several studies show little or no relationship between grade point average and student ratings.^{7,8,10,11} At the same time, other studies yielded low but positive correlations (0.10 to 0.30) between ratings and expected grades.^{4,7,10,12,17}

In examining the relationship between average SSTE scores and self-reported GPA, the UH results showed no significant influence of GPA on SSTE rankings. See Table 1 in Appendix B.

Course Level. The final factor examined was **course level**. Mason, Stegall, and Fabritius reported no difference in evaluation scores based on course level.¹⁸ Neumann and others, however, did report that higher level courses received higher ratings.¹⁴

The UH results showed that higher level courses resulted in higher ratings. Graduate level ratings were the highest while freshmen courses resulted in the lowest ratings. See Table 1 in Appendix B.

Using Student Evaluations of Teaching

SSTE results are used by instructors for the improvement of instruction. Another major application is for administrative decision making and to meet accountability or accreditation requirements of institutions.¹⁹

With respect to the potential for SSTE results to improve teaching, Wachtel noted that while student ratings can help instructors to improve their teaching, some authors who support this use do not believe the student evaluations alone will automatically result in improved instruction without other types of feedback.²⁰ Adams stated a major problem with the use of SSTE for teaching improvement, “The original purpose of student evaluations of faculty, however, was not summative but formative — that is, for the improvement of instructors' teaching. In spite of this original intent, there is no evidence to support the notion that student evaluations of faculty actually improve instruction. The reason for the lack of impact on teaching is simple. In order to improve instruction, the evaluation device has to identify particular difficulties.”⁹

The administrative use of student evaluations of teaching is on the upswing particularly in support of tenure, promotion, and salary determination exercises. If student ratings are being used to influence crucial administrative decisions, it is important that their meanings be well understood.²¹

Thus, the use of SSTE results in administrative decision making needs to be cautionary, making sure that the data clearly match the decision criteria. For example, if learning gains is the decision criterion for a particular administrative action, SSTE results do not provide the answers, even though there may be an indirect relationship between learning gains and SSTE results if one

subscribes to the notion that a positive reaction is required to provide the motivation to learn the course content.

Adams summarized both the attractiveness and some problems with the use of student evaluations of teaching for administrative purposes.

For administrators, the attractiveness of student evaluations of faculty is that they provide an easy, seemingly objective assessment of teaching that does not require justification. The ease of student evaluations comes in reducing the complexities of teaching performance to a series of numbers, particularly when commercial forms are used... One of the problems with this type of student evaluation of faculty is that few administrators are trained to interpret the numbers.⁹

Is there a middle ground that justifies using SSTE results in support of administrative decisions? Perhaps this use can be justified provided SSTE results are monitored over long-term periods. Improvement of teaching effectiveness is a life-long process. There is no perfect course nor is there a perfect instructor. Hence, there is never a time when SSTE are useless. Thus, one approach would be for instructors to graph their performance on a number of items over time, with the focus being on quality control and continuous improvement.

Hall and Fitzgerald suggested a code of practice that the UH CoT ACI committee endorses. In applying SSTE results in support of administrative decisions, embrace the following guidelines.²²

- SSTE should be used for summative decision making.
- SSTE should not be used in isolation.
- Those entrusted with the task of making tenure, promotion and similar decisions should be skilled in interpreting and drawing together different sources of information, including data from SSTE.
- SSTE should be based on a representative range of a person's teaching.
- SSTE instruments should be soundly based in teaching/learning theory, i.e. survey items should address teaching behaviors that are known to foster student learning.

References

- ¹ Kirkpatrick, D. *Evaluating Training Programs: The Four Levels*. Berret-Kohler Publishers, Inc. 1998.
- ² Scriven, M. Student ratings offer useful input to teacher evaluations. *Practical Assessment, Research & Evaluation*. 4 (7). 1995. Retrieved from <http://PAREonline.net/getvn.asp?v=4&n=7> on December 13, 2004.
- ³ McKeachie, W. Student ratings, the validity of use. *American Psychologist* 52 (11). 1997. 1216-1225.
- ⁴ Cashin, W. and Downey, R. Using global student ratings for summative evaluation. *Journal of Educational Psychology*, 84. 1992. 563-572.
- ⁵ Cashin, W., Downey, R. and Sixbury, G. Global and specific ratings of teaching effectiveness and their relation to course objectives: Reply to Marsh (1994). *Journal of Educational Psychology*, 86. 1994. 649-657.
- ⁶ Perttula, B. Notes on Student Ratings of Instructors: Selected findings from the Literature. Retrieved from <http://online.sfsu.edu/~perttula/ratings/> on February 11, 2003.
- ⁷ Marsh, H. and Duncan, M. Students' evaluations of university teaching: A multi-dimensional perspective. *Higher Education: Handbook on Theory and Research*. J. C. Smart (Ed.). New York: Agathon Press. 143-234.
- ⁸ McKeachie, W. Student ratings of faculty: a reprise. *Academe*, 6S (6). 1979. 384-397.
- ⁹ Adams, J. Student evaluations: The ratings game. *Inquiry*, 1 (2). 1997. 10-16.
- ¹⁰ Braskamp L. and Ory, J. *Assessing Faculty Work: Enhancing Individual and Institutional Performance*. San Francisco, CA: Jossey-Bass. 1994.
- ¹¹ Centra, J. *Reflective Faculty Evaluation: Enhancing Teaching and Determining Faculty Effectiveness*. San Francisco, CA: Jossey-Bass. 1993.
- ¹² Feldman, K. Course characteristics and college students' rating of their teachers: What we know and what we don't. *Research in Higher Education*, 9. 1978. 199-242.
- ¹³ McKeachie, W. and Kaplan, M. Persistent problems in evaluating college teaching. *AAHE Bulletin*, (February). 1996. 5-9.
- ¹⁴ Neumann, R. Communicating student evaluation of teaching results: Rating interpretation guides (RIGs). *Assessment & Evaluation in Higher Education*, 25 (2). 2000. 121-134.

¹⁵ Sixbury, G. and Cashin, W. Comparative data by academic field. *IDEA Technical Report, vol. 10*. Center for Faculty Evaluation and Development, Manhattan, Kansas. Kansas State University.

¹⁶ Kaplan, Mets, Cook. 2003.

¹⁷ Dooris, M. An analysis of the Penn State student rating of teaching effectiveness. 1997. Retrieved from <http://www.psu.edu/president/cqi/cqi/srte/analysis.html> on February 11, 2003.

¹⁸ Mason, P., Steagall, J. and Fabritius, M. Student evaluations of faculty: A new procedure for using aggregate measures of performance. *Economics of Education Review*, 14 (4). 1995. 403-416.

¹⁹ Davies, W. and Lunt, H. Literature Review. 2004. Retrieved from http://tlu.ecom.unimelb.edu.au/papers/davies_lit_review.pdf on 29 December 2004.

²⁰ Wachtel, H. Student evaluation of college teaching effectiveness: a brief review. *Evaluation in Higher Education*, 23 (2). June 1998. 191.

²¹ Nerger, J. and Viney, W. Student ratings of teaching effectiveness: use and misuse. *Midwest Quarterly* 38 (2). Winter, 1997. 218.

²² Hall, C. and Fitzgerald, C. Student summative evaluation of teaching: code of practice. *Assessment and Evaluation In Higher Education*, 20 (3). Fall 1995. 307.

Appendix A

Student Survey of Teaching Effectiveness

The students responded using the following scale:

Strongly Agree Agree Neutral Disagree Strongly Disagree

COURSE EVALUATION

1. The texts used in this course were useful.
2. This course contributed to my professional growth and development.
3. This course met my expectations.
4. This course was relevant.
5. I would recommend this course to others.
6. This was an excellent course.

INSTRUCTOR EVALUATION

7. The instructor presents material in an understandable way.
8. The instructor welcomes student questions.
9. The instructor is available to help students outside of class.
10. The instructor is organized.
11. The instructor is effective.
12. Grading is fair.

EVALUATION

13. I had a desire to take this course.
14. The instructor remained in the room while I filled out the evaluation.

Appendix B

Table 1 Summary of Statistical Results

Contextual Factors	Statistical Test	Null Hypothesis	p	Conclusion
Motivation	Mann Whitney	Ho: Median of Motivated = Median of Non-motivated	$p < 0.001$	Reject Ho and conclude medians are not equal.
Major	Kruskal-Wallis	Ho: The population medians are all equal based on major.	$p = 0.66$	Do not reject Ho and conclude that there are no differences in medians based on major.
GPA	Spearman Rank	Ho: There is no correlation between GPA and course rating.	$p = 0.91$	Do not reject Ho; the correlation is not significant.
Course Level	Kruskal-Wallis	Ho: The population medians are all equal based on course level.	$p < 0.001$	Reject Ho and conclude at least two medians differ.

Carole E. Goodson is Professor of Technology at the University of Houston. She has developed continuous improvement plans for technology programs. She has also developed web-based courses in Statistics and Evaluation in Training. An active member of ASEE, she is a member of the Academy of Fellows, a past editor of the *Journal of Engineering Technology*, past Chair of the ERM Division, and past Chair of the Gulf Southwest Section.

Lucas M. Faulkenberry is an Associate Professor and Coordinator of the Electrical Power Technology program at University of Houston. He earned a B.S. degree in Physics from University of Texas at Arlington and M.Ed. and Ph.D. in Industrial Education from Texas A&M University.

Susan L. Miertschin is an Associate Professor in the Information Systems Technology program at University of Houston. She earned a B.S. degree in Mathematics and a M.Ed. degree in Mathematics Education, both from the University of Houston. She is a member of the American Society of Engineering Education (ASEE), active in the Engineering Technology Division and the Association of Computing Machinery (ACM). She is also active in the Information Technology Education SIG. She has served as Editor of the *Journal of Engineering Technology*.

Barbara L. Stewart is Associate Professor and Coordinator of the Consumer Science and Merchandising, program at University of Houston. She earned a B.A. degree in Clothing and Textiles from Brigham Young University, a M.S. in Consumer and Home Economics Education from Utah State University, and an Ed.D. in Curriculum and Instruction from Brigham Young University.