

Student Evaluation of Teaching (SET) in Web-based Classes: Preliminary Findings and a Call for Further Research

Karen A. Loveland, *Texas A&M University - Corpus Christi*

Abstract

Student evaluation of teaching (SET) is important to faculty because SET ratings help faculty improve performance and are often used as the basis for evaluations of teaching effectiveness in administrative decisions (e.g., tenure). Researchers have conducted over 2,000 studies on SET during the past 70 years. However, despite the explosive growth in online education during the past decade, researchers have largely neglected the use of SET to evaluate teaching effectiveness in online courses. This exploratory study analyzed the actual SET data collected during a single semester at a large mid-western college that offers over 250 online/Web-based classes. The data included five dependent and eighteen independent measures of teaching effectiveness. The results indicate that average SET ratings in online classes are significantly lower than the average ratings in on-campus classes across all five dependent measures. This finding offers preliminary empirical support for anecdotal evidence cited by earlier authors in this field. Furthermore, regression analysis of the full model for each dependent variable indicated that the independent variables explained a significant portion of the variance in SET ratings. Examination of the standardized beta coefficients revealed that the strength and significance of the independent variables varied across the five dependent measures. Findings also indicate that organization of the course materials had a strong impact on all five measures of overall teaching effectiveness. Other variables including clarity of the instructor's writing, timeliness in providing feedback, and interest in whether students learned were also significant factors in models that measured instructor effectiveness (as opposed to models that measured quality of course content). The paper concludes with a discussion of the implications of this study for administrators, faculty, and researchers.

Background

Almost 44% of all higher education institutions offered distance education courses in 1997, according to the National Center for Education Statistics (NCES) of the U.S. Department of Education (USDE). Steinberg and Wyatt (2000) suggested that four out of five colleges and universities in the United States would offer distance learning by 2002.

Although the studies above focused on the broadest definition of distance education, the NCES report and more recent studies (e.g., Allen and Seaman, 2006; Eastman and Swift, 2001) indicate an explosion of Internet-based distance courses in the past few years. Allen and Seaman (2006) found that, "...for the fall 2002 term, slightly more than 1.6 million students took at least one online course..." in higher education in the US (p. 5). This number grew to just over 3 million the fall of 2006, representing a 35% increase over numbers reported in the 2005 report (Allen and Seaman, 2006).

The rapid growth in the number of online classes poses some challenges for academic administrators including difficulties hiring faculty with online teaching interests and experience, increases in costs associated with technology, training and faculty incentives, and problems associated with the comparison of traditional and online teaching in terms of workload, compensation, and evaluation. Specifically, Loveland and Loveland (2003) suggested that significant differences in student evaluations of teaching effectiveness exist between traditional on-campus classes and online classes.

Student evaluation of teaching (SET) is important to faculty for two primary reasons. First, student evaluations provide data used for administrative decisions such as tenure, promotion, and salary increases. Second, teaching evaluations provide feedback to help faculty improve future teaching performance (Sheehan and DuPrey, 1999). Most higher education institutions use student evaluations as a measure of teaching effectiveness; in many cases, SET is the most important, and sometimes the only, measure of teaching effectiveness (d'Apollonia & Abrami, 1997).

A comprehensive review of the prior research on SET is beyond the scope of this paper. However, it is important to note that Wilson (1998) identified nearly 2,000 studies conducted on SET in the 20th century. The level of attention afforded the reliability and validity of SET over the past 70 years suggests that researchers (faculty) recognize the importance of understanding the factors students use to evaluate teaching effectiveness. A review of these studies prompted Drago, et.al (2002) to state, "One conclusion that can be drawn from this research is that...student evaluations...have considerable validity", and, that "...teaching effectiveness is a multidimensional construct" (p. 71).

Given this level of attention to SET, it is more than a little surprising that multiple researchers haven't replicated prior research in an online setting to study the dimensions of teaching effectiveness in online classes. Several studies in the decade since online courses first emerged have used SET results as measures of quality or outcomes in online courses (e.g., Drago, et.al., 2002). Still other studies (see Sorensen and Johnson, 2004) have investigated the use of online SET to evaluate online and on-campus classes. However, a comprehensive search of multiple databases revealed only one article (Loveland and Loveland, 2003) that specifically addressed the use of SET as a measure of teaching effectiveness in online classes.

One possible explanation for this lack of attention is that the rapid growth of online education prompted online education researchers to focus on measuring the quality of student learning and answering questions about various pedagogical issues. Perhaps the most pressing need during the infancy of online education was to provide information to support course development and respond to critics of online education. Thus, researchers may have simply assumed the validity of SET for online classes to facilitate these studies. However, some prior studies on SET in traditional classes (e.g., DeBerg and Wilson, 1990; Langbein, 1994) suggested that student perceptions of teaching effectiveness can be influenced by the type of class (e.g., lecture, lab, seminar). Furthermore, other prior studies suggested that class size (e.g., Holtfreter, 1991), the time of day (DeBerg and Wilson, 1990) and/or the day of the week (Husbands and Frosh, 1993) the class meets, the day/date of the evaluation (Cronin and Capie, 1986 as cited in Koh and Tan, 1997), and a whole host of instructor characteristics and course outcomes affect SET results in

traditional classes. It seems reasonable to suggest that any of these variables (and possibly some new ones) might affect SET ratings in online classes.

Student Evaluations of Teaching (SETs) in Online Classes¹

The decision to teach online classes may have serious repercussions for a faculty member. Issues include the increased workload associated with online teaching, the impact of increased workload on research productivity, and political problems at the institution from faculty opposed to online education.

Some experts suggest that we should view the Internet merely as a different delivery method. However, the nature of that delivery method would lead many online faculty to say, “Teaching online is not the same as teaching a face-to-face (FTF) class.” The teaching behaviors and strategies required for effective teaching online are not the same as those required for FTF teaching. Unfortunately, in the race to put classes online to meet rapidly increasing demand, some institutions failed to consider these differences in the SET system. Loveland and Loveland (2003) suggested that many institutions may use the same or slightly modified versions of instruments developed for FTF classes, instead of developing and validating separate SET instruments for online classes.

Using evaluation forms designed for on-campus classes to evaluate online classes gives rise to the concern that online teaching may affect faculty negatively in terms of administrative decisions. Some institutions report that SET ratings for online classes are significantly lower than the ratings given in on-campus classes. For example, the institution that provided the data for this study routinely calculates separate mean scores for SET in online and on-campus classes. The institution reported that the mean scores for a recent semester across four global measures of effectiveness were twenty percent lower in online classes than in on-campus classes. Some faculty at this institution report that student evaluations of their online classes are up to one full point lower (on a 6-point scale) than evaluations of the identical classes taught in the classroom.

¹ As noted above, only one prior journal article examined the use of SET in online classes. The lack of prior research in this area prompted the author to base most of the following discussion on this article (Loveland and Loveland, 2003) and on nearly 10 years experience teaching online classes and dozens of conversations about online teaching at conferences and seminars.

These differential results prompted some online faculty to suggest that the criteria used to measure teaching effectiveness in on-campus classes aren't relevant to online teaching. They further suggest that institutions need to reevaluate the operational definitions of teaching effectiveness and develop new student evaluation forms specifically designed to measure teaching effectiveness in online classes. Loveland and Loveland (2003) contended that the criteria measured by traditional student evaluation forms are not the cause of the observed difference in scores. The authors suggested that the same general criteria for effective teaching apply equally to online and on-campus classes. However, the faculty behaviors and characteristics that students consider when assigning a rating and the relative importance of each criterion may differ significantly between online and on-campus classes. Thus, it is important for online faculty to understand how online students might interpret and assess the major criteria for effective teaching.

Purpose of the Study

The main purpose of this study is to identify some of the possible differences between SET in online and on-campus classes to provide a basis for future research. The SET instrument used in this study was subjected to rigorous study to establish the reliability and validity of the scale for use in traditional classes. It is reasonable to expect all the included variables to have a significant impact on ratings. Analysis of actual SET results obtained by using this scale in online classes allows for the identification of variables that have a strong impact on ratings and those that don't appear to be relevant for online classes.

Method

The SET instrument used in this study was a modified form of the University of Washington's Instructional Assessment System (IAS)². The college used IAS Form A for most classes. This form contains four global ratings of course quality/instructor effectiveness and a number of diagnostic questions designed to provide specific feedback for instructional improvement. A

² See the University of Washington Office of Educational Assessment Web site at <http://www.washington.edu/oea/> for more information about the Instructional Assessment System including reports on the reliability and validity of the SET instruments.

college-wide task force consisting of experienced online instructors modified the form (IAS Form A) for use in online classes. College policies mandated the use of the four global ratings for SET in all classes. The committee then modified some of the diagnostic questions to reflect the unique nature of online teaching. For example, existing forms contained a question to assess the “clarity of the instructor’s oral communication”; this question became “clarity of the instructor’s written communication” on the online form.

The final SET form contained the 4 global rating questions and 18 additional questions designed to provide diagnostic feedback on specific aspects of the class and instructor performance. Students rated the course/instructor on these questions using a six-point scale (1 = very poor, 2 = poor, 3 = fair, 4 = good, 5 = very good, 6 = excellent). The form also contained additional questions required by the institution but not relevant to this study.

Administrators use the four global ratings, in some combination, to assess teaching effectiveness in the faculty performance evaluation process at this institution. For example, faculty in the Marketing Department may choose to report results for Q3 or Q4 or an average of both; other departments use the average of all four questions in the annual evaluation guidelines for teaching effectiveness. These questions plus a fifth variable created by taking the average of the four questions served as the dependent measures in this study. Table 1 overleaf a brief discussion of the nature of the dependent measures.

TABLE 1: Dependent Measures

Variable	Question	Discussion
Q1	The course as a whole was:	Broadest measure of overall course quality; based on course content (e.g., topic, materials), outcomes (e.g., student learning) and instructor behaviors (organization, communication skills).
Q2	The course content was:	Primarily a measure of student satisfaction with course topic, materials, and outcomes; in an online course, may also be influenced by instructor behaviors (e.g., organization of online materials may affect student ability to locate/read material) and characteristics (e.g., knowledge of subject).
Q3	The instructor's contribution to the course was:	Primarily a measure of student perceptions of instructor effort; influenced by instructor behaviors (e.g., organization, encouragement given to students/rapport, quality and timeliness of feedback) and instructor characteristics (e.g., knowledge of subject, communication skills).
Q4	The instructor's effectiveness in teaching the subject was:	Primarily a measure of student perceptions of instructor teaching skills; influenced by instructor behaviors (e.g., organization) and instructor characteristics (e.g., knowledge of subject, communication skills).
Average	The average of Q1-Q4	Can be viewed as an overall measure of course quality and instructor effectiveness.

The 18 diagnostic questions measured specific aspects of course quality and instructor effectiveness and served as the independent measures in this study. Table 2 contains a brief discussion of the nature of the independent measures.

TABLE 2: Independent Measures

Variable	Question	Discussion
Q5	Organization of learning materials was:	Student control over order of presentation of materials in online classes, access may be non-linear; design of course Web page/site (e.g., ease of navigation, speed of page loading, page layout, etc...) may affect ratings.
Q6	Clarity of course objectives was:	Related to instructor's written communication skills in online classes; fewer opportunities to clarify objectives through classroom discussion; students must choose to read/review written objectives in online classes.
Q7	Contribution of assignments/activities (e.g., homework, exams, projects, etc.) to understanding of course content was:	Affected by instructor's ability to develop and communicate requirements for assignments; written communication skills important.
Q8	Encouragement given to students to express their own ideas was:	Affects perceptions of "rapport with students"; students carry more responsibility since they must read email and other written material that conveys this encouragement.
Q9	Student confidence in instructor knowledge was:	Some (possibly many) people assume some correlation between knowledge/intelligence and writing skills; any writing errors may affect student confidence. Fewer opportunities in online classes to demonstrate

Variable	Question	Discussion
		knowledge using “off the cuff” examples. Easier for students to learn the age of online materials and compare contents to other sources; inconsistencies or use of outdated material may affect this rating.
Q10	Clarity of instructor’s writing was:	Clear writing critical since most (all) communication is written; spelling, grammar and sentence construction errors can affect clarity of instructions. Style (active versus passive) may also affect clarity.
Q11	Quality of course content (e.g., lectures, handouts, etc.) was:	May be affected by material design (e.g., use of multimedia, graphics to enhance appearance of materials), writing skills, and technical factors (e.g., errors in linked sources on Web page).
Q12	Instructor’s identification of supplementary resources (e.g., movies, books, Web pages, etc.) was:	Providing links to study aids, relevant articles, etc. helps students understand the materials and may be related to ratings of knowledge of subject.
Q13	Instructor’s timeliness in making course materials available was:	Relates to instructor’s “preparation for class”; students may have a self-paced mentality and expect to work ahead. Students may be dissatisfied if they can’t work ahead or if they work ahead and instructor then changes/updates materials before the rest of the class starts working on them.
Q14	Instructor’s encouragement of student participation in the class was:	Lack of FTF interaction makes encouragement more difficult; requires student cooperation (must read email/other material that communicates encouragement); may be more strongly influenced by individual communication.
Q15	Instructor’s timeliness in providing feedback (e.g., grades, responses to questions, etc.) was:	Students may have more control over date/time of assignment submission; better/more motivated students may submit assignments early and wait longer for grades; student cooperation required; may be affected by technical problems.
Q16	User friendliness of online course materials was:	Related to organization; students may mentally compare course site/page design to typical Web sites; may be affected by technical problems.
Q17	Overall quality of learning materials was:	Composite measure of overall course materials including handouts, assignments, lectures. Affected by writing skills, contribution and relevance of material, and more.
Q18	Amount you learned in this class was:	Student assessment of amount learned; single global measure of outcomes.
Q19	Instructor’s interest in whether students learned was:	Related to rapport with students; may be related to enthusiasm for subject and encouragement of participation.
Q20	Relevance and usefulness of course content was:	Student assessment of value of course content; largely beyond the control of instructor but may be influenced by use of supplemental resources and teaching style (e.g., use of experiential exercises, tying class to other courses, etc...).
Q21	Reasonableness of assigned workload was:	Student assessment of course difficulty; may be strongly affected by student skills (lack of computer skills or reading comprehension difficulties may significantly increase workload); may be affected by student expectation for a lighter workload in an online class;
Q22	Clarity of student responsibilities and requirements was:	Related to writing skills and may affect perceptions of fairness in grading.

The SET form for online classes was converted into an HTML form that recorded answers to each question and converted them into a data stream sent to the Office of Institutional Research

(OIR). The form appeared as an institutional announcement the first time students logged in to their online class during the two-week evaluation period near the end of the semester. Students had three choices when the announcement appeared: they could complete the evaluation, they could submit a blank copy, or they could scroll to the very bottom and skip the announcement using the “continue” feature. The announcement appeared once and did not appear again even if students skipped viewing the initial appearance. This procedure is roughly analogous to the procedure used in FTF classes.

The data set for this research consisted of the official SET data collected during a recent semester by the OIR at a large mid-western college. During the semester studied, this institution offered 259 online classes with a total enrollment of 5,304 students. The OIR administered SET in 214 classes with a total enrollment of 4,617 students. A total of 382 (8.3%) students withdrew from the classes before the evaluation period for an effective enrollment of 4,235 students.

One hundred and five of the 2,792 SET forms submitted were blank yielding 2,682 usable responses (63.3% response rate). It should be noted that the institution has a “last grade stands” policy. This policy automatically converts previous grades in a class to an “NC” (“no credit” = “withdrawal”) when a student repeats the class. Several experts on campus suggest that at least half the students that effectively withdraw from a class (i.e., no-shows and students that stop attending before the end of the semester) don’t actually process the forms required to formally withdraw from the class. Excluding those students from the enrollment figures above would increase the response rate to 70% or more.

Findings and Discussion

The five models were analyzed using linear regression. A summary of the results appears in Table 3.

TABLE 3: Summary Regression Results

Model	Variable	R Square	Adjusted R Square	F	P
1	Q1: Course as a whole	.755	.753	365.26	.000
2	Q2: Course content	.720	.718	304.62	.000
3	Q3: Instructor's contribution	.772	.770	398.81	.000
4	Q4: Teaching effectiveness	.784	.782	428.68	.000
5	Average (Q1-Q4)	.858	.857	718.80	.000

All five models were statistically significant ($p = .000$) indicating that variation in the independent measures in this study explained a significant portion of the variation in the dependent measures. Furthermore, the R^2 values for the first four models indicate that the independent variables explained over seventy percent of the variation in the individual global measures. The larger R^2 for the fifth model makes intuitive sense because the fifth dependent variable is a composite of the four dependent variables and each of them clearly captures a different aspect of instructor or course quality. As such, the composite measure is likely to be significantly affected by a larger number of the independent variables.

Table 4 reports the standardized beta values for each independent variable across all five models.

TABLE 4: Standardized Beta Values

Variable	Model 1		Model 2		Model 3		Model 4		Model 5	
	Course as a whole		Course content		Instructor contribution		Teaching effectiveness		Average (Q1-Q4)	
	Std. Beta	p	Std. Beta	P	Std. Beta	P	Std. Beta	p	Std. Beta	P
Q5	.099	.000	.128	.000	.112	.000	.112	.000	.122	.000
Q6	.029	.186	.016	.484	.031	.142	.039	.062	.032	.055
Q7	.114	.000	.064	.004	-.004	.823	.036	.060	.053	.001
Q8	.055	.000	.045	.038	.090	.000	.056	.003	.068	.000
Q9	.008	.702	.053	.020	.145	.000	.130	.000	.097	.000
Q10	.023	.287	-.031	.176	.082	.000	.113	.000	.058	.000
Q11	.062	.006	.077	.001	.063	.004	.104	.000	.083	.000
Q12	.007	.684	-.006	.743	.068	.000	.066	.000	.040	.004
Q13	.013	.493	-.006	.757	-.006	.759	-.046	.009	-.014	.337
Q14	.040	.063	.017	.471	.064	.002	.041	.043	.047	.005
Q15	.009	.601	-.035	.063	.141	.000	.097	.000	.065	.000
Q16	.017	.407	-.047	.031	-.040	.046	-.045	.020	-.032	.042
Q17	.041	.105	.174	.000	.019	.439	.010	.668	.060	.002
Q18	.276	.000	.258	.000	.074	.001	.112	.000	.185	.000
Q19	.065	.004	.000	.986	.240	.000	.252	.000	.165	.000
Q20	.088	.000	.204	.000	-.054	.005	-.055	.003	.036	.016
Q21	.076	.000	.036	.055	-.036	.033	.019	.237	.022	.097
Q22	.020	.836	.013	.554	.014	.499	-.028	.164	-.002	.915

Shaded values were statistically significant at the alpha = .05 level

Examination of the beta values reveals some interesting findings about the independent and the dependent variables:

- Only five of the 18 independent variables were statistically significant in all five models: Organization of course materials (Q5), encouragement given to students to express their own ideas (Q8), quality of course content (e.g., lectures, handouts, etc.) (Q11), amount you learned in this class (Q18), and relevance and usefulness of course content (Q20). Furthermore, the relative importance of these five variables (based on the absolute value of the standardized betas) varied significantly across the five models providing support for the discussion of the differences in the dependent measures in Table 1.

- Two of the independent variables were not statistically significant in any of the models: Clarity of course objectives (Q6) and clarity of student responsibilities and requirements (Q22). However, clarity of instructor's writing (Q10) was statistically significant in three of the five models. Since Q6 and Q22 were both included on the original IAS Form A (a validated scale backed by significant research), and since it doesn't make sense to assume that students in online classes don't want to know what is expected of them, one can conclude that the questions are measuring something that isn't relevant in online classes. Perhaps the written nature of course communication alleviates a problem caused by oral communication of expectations (e.g., recall) or perhaps online students have lower expectations related to clarity of objectives and responsibilities (they expect to be confused and give high ratings to any professor that does anything to alleviate their confusion, something even the least experienced professor would do by providing a syllabus!).
- One variable, user friendliness of online course materials (Q16), was significant in four of the five models. However, the beta was negative in all four models indicating that increasing the user friendliness of your course may decrease the overall ratings of course content (Q2), instructor's contribution to the course (Q3), instructor's effectiveness in teaching the subject (Q4) and the overall evaluation represented by the average of the four individual measures. Perhaps some efforts to increase user friendliness detract from other, more important variables in the models. For example, some instructors might create a "frequently asked questions" (FAQ) page to answer common questions about the course to increase user friendliness. However, if the instructor then refers students to the FAQ when they ask one of the questions via email for example, some students may view the that response as a lack of interest or view the effort involved in finding the question and answer on the FAQ page as part of the workload, which would affect reasonableness of workload ratings.

Further analysis of the beta values suggests that the four dependent measures are capturing different aspects of effective teaching.

Table 5 lists the top five variables in each model based on the absolute value of the standardized beta value.

TABLE 5: Top Five Factors

	Model 1	Model 2	Model 3	Model 4	Model 5
Rank	Course as a whole	Course content	Instructor contribution	Teaching effectiveness	Average (Q1-Q4)
Most important	Amount learned (Q18)	Amount learned (Q18)	Instructor interest in student learning (Q19)	Instructor interest in student learning (Q19)	Amount learned (Q18)
2 nd	Contribution of assignments to learning (Q7)	Relevance/usefulness of course materials (Q20)	Confidence in instructor knowledge (Q9)	Confidence in instructor knowledge (Q9)	Instructor interest in student learning (Q19)
3 rd	Organization of course materials (Q5)	Overall quality of materials (Q17)	Timeliness of feedback (Q15)	Clarity of instructor's writing (Q10)	Organization of course materials (Q5)
4 th	Relevance/usefulness of course materials (Q20)	Organization of course materials (Q5)	Organization of course materials (Q5)	Organization of course materials (Q5)	Confidence in instructor knowledge (Q9)
5 th	Reasonableness of workload (Q21)	Quality of course content (Q11)	Encouragement given to express ideas (Q8)	Amount learned (Q18)	Quality of course content (Q11)

Examination of the factors in Table 5 reveals some interesting results:

- The top five variables in terms of contribution to the dependent variable were different across the five models. This finding indicates that each dependent variable is capturing a different aspect of effective teaching. Models 1, 2, and 5 are affected most by the amount of learning (Q18) while Models 3 and 4 are most strongly affected by instructor's interest in whether students learned (Q19). Furthermore, Models 1 and 2 seem most strongly affected by factors related to the course content (usefulness, quality of materials, workload) while Models 3 and 4 are affected more by instructor behaviors and skills (writing clarity, timeliness of feedback, knowledge). This finding is interesting because it suggests that online students can separate evaluations of the course as a whole from specific instructor behaviors.

- One variable, organization of learning materials (Q5) was in the top five in all five models. This finding suggests that organization of course materials is a critical factor in online classes that can affect SET results across a range of measures of teaching effectiveness.

Implications

The findings in this paper have a number of implications for administrators, online faculty, and researchers. First and possibly foremost, administrators responsible for evaluating online faculty need to be aware that online teaching may result in lower SET ratings, especially for instructors new to the online environment (the learning curve for online teaching is notoriously steep). In addition, existing evaluation forms may not capture all the factors that affect SET ratings and thus put online faculty at a disadvantage because they can't identify the source of the problems with their evaluations.

Most of the training offered to online instructors focuses on the technology. Institutions may need to devote more resources to helping faculty develop effective online teaching behaviors and skills as they do with seminars on effective classroom teaching. Furthermore, institutions may need to provide new forms of support for online faculty including professional editors to improve clarity and quality of written material and IT professionals to improve course design.

Instructors teaching online classes need to be aware that course organization can have a strong impact on SET results. The instructor has nearly complete control over the organization of a traditional class (i.e., the instructor controls the order of presentation); online courses give students a significant amount of control over course organization. Organization ratings may suffer if students have difficulty locating material they need or if they choose to view the course material in a different sequence. Instructors can improve ratings for organization by adding navigational links to make course material accessible from multiple locations/pages, using hyperlinks for navigation of longer documents (e.g., a jump link to "grading scale" at the top of the syllabus page), creating a consistent design for course pages, and using content modules to organize related material in a single location.

Online instructors also can improve SET ratings by devoting more attention to improving written communication. Using proofreaders for online lectures and other material may reduce the number of writing errors that might affect student perceptions of the instructor's effort, knowledge, or effectiveness. Furthermore, using individual communication to encourage participation in class discussions may increase ratings related to instructor's rapport with students, encouragement given to students, and more.

Loveland and Loveland (2003) discussed a large number of suggestions for improving ratings of factors identified as significant in this study including knowledge of subject, timeliness in providing feedback, and more. Please see this article for additional suggestions for improving online SET ratings.

As an exploratory study, the most important contribution of this paper is that it establishes the need for further research on the use of SET in online classes. The SET ratings used in this study were 20% lower than the SET ratings for on-campus classes. Online faculty should find this result alarming and be motivated to investigate the extent and nature of this phenomenon.

The scope and limitations of this study suggest several directions for future research. First, this study focused on identifying factors that affected SET using data from online courses only. A comparison of SET in online and on-campus classes using data from an institution that uses similar rating instruments in both types of courses would allow researchers to identify differences in the relative importance of various factors. For example, is course organization more important in online classes than in on-campus classes? Identifying the significant differences between online and on-campus SET results will help faculty make the transition to online teaching and improve teaching effectiveness.

Second, this study was limited to undergraduate students enrolled in a broad spectrum of classes ranging from freshman composition to the capstone course in business. Future studies could examine SET in online graduate courses and/or courses in a specific field of study. Furthermore, the students at the institution studied are mostly US citizens. Given the global reach of online

classes, a study investigating the influence of non-US cultures on SET in online classes may yield interesting results.

Another stream of research could involve replication of earlier studies investigating relevant factors not investigated by this study. The author performed an exploratory factor analysis on the independent variables in this study. Principal component analysis extracted a single factor that explained 69.3% of the variance in the independent variables. A discussion of this finding is beyond the scope of this paper. However, this finding and the results of additional factor analyses performed on subsets of the data revealed the presence of additional factors that might be influencing the SET ratings in online classes. Additional factor analyses or even structural equation modeling might yield some interesting findings on the dimensions of student satisfaction.

Factors not directly measured by this study include instructor personality characteristics, rapport with students and fairness in grading. The lack of FTF contact in online classes is likely to affect student perceptions related to these and many other factors. Replication of previous studies would allow researchers to determine what factors affect SET in online classes and what instructor behaviors may influence those ratings in online as opposed to on-campus classes.

The importance of SET for faculty evaluation and development should lead to a significant increase in the number of studies on the topic as the number of institutions offering online courses and the number of courses offered at all institutions continues to increase. At least some of this research should parallel the SET research conducted in traditional classes during the last century. Effective online teaching does share some similarities with effective traditional teaching methods. However, the lack of FTF contact and other obvious differences in the delivery method makes it unwise and potentially dangerous to assume that methods and factors validated by previous studies are valid in this new environment.

References

- Allen, I.E. & J. Seaman (2006). *Making the grade: Online education in the United States*. Needham: MA: Babson Survey Research Group, The Sloan Consortium.
- D'Apollonia, S. & P.C. Abrami (1997). Navigating student ratings of instruction. *American Psychologist*, 64, 431-441.
- DeBerg, C.L. & J.R. Wilson (1990). An empirical investigation of the potential confounding variables in student evaluation of teaching. *Journal of Accounting Education*, 8 (1), 37-63.
- Drago, W., J. Peltier, & D. Sorensen (2002). Course content or the instructor: Which is more important in online teaching. *Management Research News*, 25 (6/7), 69-83.
- Eastman, J. & C. Swift (2001). New horizons in distance education: The online learner-centered marketing class. *Journal of Marketing Education*, 23(1), 25-34.
- Holtfreter, R.E. (1991). Student rating biases: Are faculty fears justified? *The Woman CPA, Fall*, 59-62.
- Husbands, C.T. & P. Fosh (1993). Students' evaluation of teaching in higher education: Experiences from four European countries and some implication of the practice. *Assess and Evaluation in Higher Education*, 18(2), 95-114.
- Koh, H.C. & T.M. Tan (1997). Empirical investigation of the factors affecting SET results. *International Journal of Education Management*, 11(4), 170-178.
- Langbein, L.I. (1994). The validity of student evaluations of teaching. *Political Science and Politics*, September, 545-553.
- Loveland, K.A. and J.P. Loveland (2003). Student evaluations of online versus on-campus classes. *Journal of Business and Economics Research*, 1(4), 1-10.
- Sheehan, E.P. & T. DuPrey (1999). Student evaluations of university teaching. *Journal of Instructional Psychology*, 26, 188.
- Sorenson, D.L. and T.D. Johnson, Eds. (2004). *Online student ratings of instruction: New directions for teaching and learning*, No. 96. Hoboken, NJ: Jossey-Bass.
- Steinberg, J. and E. Wyatt (2000, February 13). Boola, boola, e-commerce comes to the quad. *The New York Times*, Section 4, Page 1.
- U.S. Department of Education, National Center for Education Statistics. (1999). *Distance Education at Postsecondary Education Institutions: 1997-98*. NCES 2000-013, by Laurie

Lewis, Kyle Snow, Elizabeth Farris, Douglas Levin. Bernie Greene, project officer.
Washington, DC.

Wilson, R. (1998). New research casts doubt on value of student evaluations of professors. *The Chronicle of Higher Education*, 44(19), A12-A14.