From the Learners' Eyes:
Student Evaluation of Online Instruction

Melinda Hess
mhess@tempest.coedu.usf.edu

Ann E. Barron
Lou Carey
Amy Hilbelink
Kris Hogarty
Jeffrey D. Kromrey
Ha Phan
Shauna Schullo

University of South Florida

Key Words: Online Course Evaluation, Distance Learning, Assessment of Instruction, Instructional Technology, Assessment Research

Paper presented at the National Educational Computing Conference, Philadelphia, Pennsylvania, June 27-30, 2005.

Introduction

In higher education, student evaluation of instruction provides data that serve a variety of purposes including the revision of courses and programs, improvement of instruction, institutional accreditation, and tenure decisions about faculty (Dunkin & Barnes, 1986). When instruction is delivered online, student evaluation becomes notably more complicated, as issues of technology and pedagogy intertwine (Cohen, 2003). Delivering effective instruction on the Internet tends to be more complex due, at least in part, to concerns about the technology skills of students, availability of universally-accessible resources, and clarity of expectations and requirements. Educators, administrators, and institutions need tools and methods to evaluate whether their courses and programs meet the requirements of accreditation, policy-making, and funding agencies in addition to meeting the needs of their students and faculty.

The primary purpose of this study was to develop a comprehensive and feasible evaluation system that institutions of higher learning, as well as other educational institutions, can adopt or adapt to evaluate courses and programs delivered online. In addition to describing the system development, this paper also reports results of initial investigations into the reliability and validity of scores obtained from this evaluation system. This research was conducted in support of a five-year technology project at a large, metropolitan research university and builds upon previously conducted research (see Hogarty, Kromrey, Barron, Hess, & Schullo, 2004).

Student Evaluation of Instruction: A Controversial History

Researchers tend to agree that the evaluation of teaching must encompass a variety of measures (Arreola, 2000). Students' evaluations of courses and faculty continue to be used as important measures despite years of controversy and research into the reliability and validity of student ratings. Many institutional policies imply that student ratings represent an important, if not the best, dimension of evaluation (Greenwald & Gillmore, 1997; Cohen, 1997; Abrami, 2005). The ongoing controversy has been centered in three main areas including the influence of corrupting factors in the evaluation, the quality of student rating instruments and evaluation procedures, and the tension between the multiple purposes for faculty evaluation. Across many campuses today, rather than being resolved, the controversy is intensifying.

Corrupting factors in evaluation. Historically, tension has been evident between administrators who need to evaluate college teaching and faculty who argue that student ratings lack validity due to corrupting influences beyond faculty control (Costin, Greenough, & Menges, 1971). The controversy related to corruption is even more complex today because of the rapid increase in distance and blended instruction. One such corrupting factor impacting students' online course experience is the technology infrastructure within the university, including server availability, speed, and quality. This category also

includes the quality of features such as ease of inputting information and graphics, student assessment and data analysis, and communications. Another factor is the team of people who are often involved in developing and delivering a web-based course, which can include instructional designers, content developers, video/sound production directors and editors, programmers, computer system administrators, and computer lab assistants. Link these factors to issues with students' readiness for a distance course in terms of their prior distance experience, technology skills, and equipment, and the difference between the potential influence of corrupting factors in the traditional classroom and web-based environments becomes quite clear.

Psychometric quality. The second area of controversy relates to the quality of typical instruments used to gather students' ratings of faculty and courses, and the associated evaluation procedures. Measurement specialists lament that most student rating forms lack proven psychometric qualities related to validity and reliability (Costin, Greenough, & Menges, 1971; Arreola; 2005). These qualities may be traced to the origins of most rating forms, typically having been developed by student or faculty committees who do not have the expertise to create valid instruments or to follow through with demonstrating their psychometric viability, causing Arreola to describe most instruments used across the country as "homemade." A related factor that exacerbates problems with evaluation quality is the misuse of student ratings wherein instructors are compared with one another using numerical averages, a practice considered inappropriate because courses have different goals, styles of instruction, student expectations, and resources. Across a comprehensive college or campus, such comparisons are analogous to comparing apples, ant hills, and automobiles.

Yet another measurement quality issue is related to the dimensionality of instructional effectiveness. Teaching is a complex and multidimensional task (Howard, Conway, & Maxwell, 1985; Marsh & Roche, 1997) because faculty may exhibit different strengths in interpersonal, organizational, and content skills. An important question to ask is whether *students* also perceive instruction as multidimensional. To investigate the dimensionality of students' perceptions, Marsh (1984) reviewed over 30 published factor analyses of student rating forms and identified nine factors of instruction including learning/value, instructor enthusiasm, organization/clarity, group interaction, individual rapport, breadth of coverage, examinations/grading, assignments/reading, and workload/difficulty. In contrast, Abrami and d'Apollonia (1991) reanalyzed the factor structure of the instruments used by Marsh and identified only two significant factors, a global instructional skill factor and a course difficulty and workload factor. They then analyzed seven other rating forms and found similar evidence supporting a global skill factor. Similar results were obtained by Grinnel, Carey, and White (1999) who factor analyzed distance students' course ratings across multiple courses and identified only two factors: a

global factor called *instruction* that included everything from accessibility through satisfaction and a second smaller factor named *student effort*.

Another issue in this category is that most student rating forms focus almost completely on conventional classroom teaching and assess instructional features such as classroom interaction that cannot be rated accurately because there is little or no classroom experience from which distance students can judge (McKeachie, 1997; Sorenson & Johnson, 2003; Franklin, 2005). For example, Greenwald and Gillmore (1998) suggested there was useful information on a professor's "bedside manner" that could be garnered from student ratings. In examining factors related to positive student ratings, Moore, Masterson, Christophel, and Shea (1996) found that instructor immediacy (physical and psychological closeness) had a significant influence. Faculty characteristics such as closeness and bedside manner would be difficult indeed for distance students to assess.

Multiple purposes of course evaluation. The third main area of controversy relates to the multiple purposes served by faculty evaluations. Institutionally sponsored faculty evaluation systems have two contradictory purposes: summative evaluation related to promotion and tenure decisions and formative evaluation related to faculty development (Palmer, 1983). Many researchers believe that a single faculty evaluation instrument, such as student ratings of teaching, cannot be used for both these purposes because gathering evidence for dismissal is in conflict with a climate of support, communication, and professional growth (Rifkin, 1995).

Current movements in higher education as well as the exponential complexity of faculty and course evaluation in distance and blended instruction are resulting in efforts to resolve these ongoing controversies or at least update them given new realities in university teaching. Redirecting the focus in instructional evaluation away from "faculty as the center piece" and more toward innovative instructional methods such as student centered learning and innovative delivery systems, where students may not even know their faculty, will result in new approaches to faculty and course evaluation. Other factors from regional accreditation agencies and professional societies that suggest new approaches to course evaluation are needed include the current focus on learning outcomes, the use of achievement and attitudinal data for continuous course and program improvement, and overall instructional accountability in higher education. These innovations and new standards serve as catalysts for beginning research in the new instructional arena.

The purpose of this study was to develop a comprehensive student evaluation system appropriate for distance and blended instructional context that are based on theories used to develop effective instruction in any context and principles of instructional design. In constructing the instruments, attempts were made to avoid common pitfalls in current faculty evaluation instruments related to corrupting factors in the interpretation of effectiveness, instrument quality and evaluation focus.

Method

Large scale evaluation planning and implementation should be grounded in formal models of program evaluation. A consideration of the variety of standards, theories and models for evaluation, within the context and nature of the specific program to be evaluated often suggests a hybrid model or framework that will provide appropriate direction for evaluation planning. The foundations of this research were built on the work of Eaton (2002); Phipps, Wellman, and Merisotis (1998); and Moore and Kearsley (1996); and were guided by the standards of ISTE (2002) and NCATE (2000). Drawing upon evaluation models delineated by Baker (2004); Cohen (2003); Gunawardena, Lowe and Carabajal (2000); and Bonk, Cummings, Hara, Fischler, and Lee (2000), our team of measurement and instructional technology specialists developed and field-tested the student evaluation system.

Previous research that focused on both evaluation models and methods of evaluating online courses were thematically analyzed to identify the primary domains that underlie an effective evaluation system (see for example, Baker, 2004; Cohen 2003; Gunawardena, Lowe & Carabajal, 2000). During our analysis of the various models and systems, a taxonomy was developed to guide the analysis of the components of the various systems. Important indicators that were central to our investigation included the presence of evaluation tools and methods, the discussion of learning theories and models, the delineation of a variety of interactions and roles (i.e., student, instructor, practitioner), attention to student satisfaction and engagement, and course implementation. Further, we were equally concerned with technology use and skill, technology design and support, student learning, specificity of context and administration/management/institutional commitment and concern.

In addition to the issues described above, Bonk, et al., (2000) contend that the degree of web integration is an important consideration in the development of an evaluation system. For example, at the lowest levels of integration, the use of technology in course delivery may be so sparse that traditional methods of evaluation are entirely satisfactory. At the higher levels, however, the differentiation between pedagogy and technology becomes more difficult and the need for an improved student evaluation system becomes more compelling. Finally, considerations of the utility of an improved student course evaluation system suggest the need for a system that is applicable across a broad range of web-based courses.

The literature review was augmented with an analysis of current instruments used by instructors of online courses and commercially available online evaluation software. The themes identified from the models and examples of student evaluation instruments (for example, see Hogarty, Kromrey, Barron, Hess, & Schullo, 2004) provided the framework for the development of this evaluation system.

The results of the aforementioned analyses suggested seven primary domains that should be addressed in student evaluation of online courses: online design and organization, instructional design and

organization, student assessment, technological support, communications, interactions, and student characteristics (see Table 1). The first domain, *online design and organization*, is comprised of questions regarding the look and feel of the course (i.e., aesthetics), and the accessibility and usability of the interface. The second domain, *instructional design and delivery*, contains questions designed to measure the clarity of expectations (e.g., course objectives and assignments), organization and the utility of resources. Items in this section query students about the logical organization of lessons, the utility and clarity of examples and non-examples used to elucidate instruction, opportunities for practice, and the difficulty level and the clear articulation of course assignments. Further, students are asked about the utility of links to other sites or resources, quizzes and tests, online help, the online grade book, online presentations and submission of assignments and homework.

The third section, *student assessment*, is concerned with the clarity of assignments, the integration of assessments with instruction, and both the quality and the timeliness of formative feedback. In the fourth section, *technological support*, items were drafted to glean information regarding hardware and software requirements and the provision of contacts for technical support. With respect to *communications*, our aim was to query students regarding the flexibility and variety of options for communicating with their instructors and peers. In contrast, when we examined the *interaction* domain, we were concerned with the quality and the quantity of both instructor and peer interactions. Lastly, a series of items designed to glean information about the students themselves posed questions regarding technological capabilities and proficiencies, reasons for taking online courses and time commitments in the online environment. Questions within this domain query students regarding their history of taking web-based or Internet courses, their current course load, reasons for taking online courses, and their level of proficiency using various software applications such as web browsers, e-mail, chat, word processing, spreadsheets, software for creating web pages, presentation software and audio/video programs.

An eighth domain (*evaluation of the quality of course content*) was suggested by the review of literature but was not included in the evaluation system. Although this domain is appropriate and necessary for curricular review by content experts, students enrolled in a post-secondary course are unlikely to possess the expertise needed to garner meaningful data about the course from such a domain.

Once the domains of primary interest were identified, items were drafted, revised and organized into two online student surveys. The initial student survey, designed to be administered during the first half of the semester, contained items that asked students about the difficulty/ease of performing tasks related to accessing the course online, the extent that they experienced technical problems, issues related to technical support, communication with instructors and peers, proficiency with various software, course load and employment status. The end of term survey contained questions regarding the course instruction, communication, assessments and assignments, and preference for online courses. Lastly, a

section was added to provide flexibility for instructors who wish to assess students' perceptions regarding the fulfillment of course objectives.

Both instruments were pilot tested in the spring of 2005. A convenience sample of approximately 400 students within five courses, at both the undergraduate and graduate level, participated in the pilot study. For the initial student survey, administered at the midpoint of the semester, data were available for 374 students. For the end-of-term survey, 397 students responded to the instrument.

Table 1.
Domains of Student Course Evaluation

Domain	Content Description
Online design and organization	Aesthetics (course look and feel) Accessibility Usability
Instructional design and delivery	Clarity of objectives Organization of materials Utility of resources
Student assessment	Clarity of assignments Integration of assessments with instruction Quality of formative feedback
Technological support	Hardware requirements Software requirements Technical support contacts
Communications	Flexibility of communication vehicles
Interactions	Instructor and peer interactions Quality and quantity
Student characteristics	Technological capabilities and proficiencies Reasons for taking online course Time commitments

Results

Student Characteristics

Student background characteristics included previous experiences with online courses, proficiency with software, number of credit hours enrolled and employment while completing the course. Twenty-five percent of the respondents (n = 94) reported that the current course was their first online course experience and 20% (n = 73) reported only one previous online course. Ten percent of the participants (n = 38) reported 5 or more previous online courses.

Thirty-eight percent of the students (n = 140) were enrolled in more than 12 hours of coursework during the semester of data collection and an additional 36% (n = 134) were enrolled in 10-12 hours of coursework. A small percentage of the students (16%, n = 59) were enrolled in fewer than 7 credit hours. The majority of students were employed while they completed the course. Forty percent of the students (n = 149) reported working more than 30 hours per week, and another 40% (n = 150) reported employment of fewer than 30 hours per week.

Student ratings of proficiency with software are presented in Table 2. More than half of the participants rated themselves as 'Advanced' with email (77%), word processing (73%), instant messaging/chat (62%) and web browsers (54%). In contrast, only 34% rated themselves as 'Advanced' with spreadsheets and only 22% with presentation software. Finally, very few students rated themselves as 'Advanced' with audio/video programs (16%) and web page creation software (8%).

Table 2.
Student Self-Reported Proficiency with Software

	Percentage of Students		
Type of Software	Beginner	Intermediate	Advanced
Email	2%	22%	77%
Word Processing	3%	24%	73%
Instant Messaging/chat	10%	27%	62%
Web Browsers	6%	39%	54%
Spreadsheets	16%	49%	34%
Presentation Software	28%	44%	22%
Audio/Video Programs	35%	41%	16%
Web Page Creation	57%	21%	8%

Note. N = 374.

Course Access and Problems Encountered

The percentage of students reporting 'Easy' or 'Very Easy' access to the course are presented in Table 3. Evident in these data is that course access was a problem for only a small proportion of the students.

Table 3.
Percentage of Students Reporting Easy or Very Easy Access

Item	Percentage of Students
Learn about course availability	90%
Connecting to internet	97%
Accessing the course	95%

Note. N = 374.

A larger proportion of students experienced some technical problems in their completion of the course. Table 4 presents the proportion of students who reported that each of seven types of technical problem occurred 'Rarely' or 'Not at all'. Although substantially more than half of the students reported rare or no problems (ranging from 64% for problems with opening files, to 74% for problems with navigating the course), a sizeable minority reported these problems occurring 'Sometimes' or 'Frequently'. The positive correlations between these items suggested that students having one type of technical problem are also likely to have others. This was verified by a factor analysis of these items which yielded a single factor, with high internal consistency of these items in constructing a single, summative scale (Cronbach's alpha = 0.87).

Table 4.
Percentage of Students Reporting Technical Problems Rarely or Not at All

Type of Problem	Percentage of Students
Links	68%
Graphics	73%
Audio	72%
Video	69%
Uploading Files	64%
Opening Documents	68%
Navigating the Course	74%

Note. N = 374.

For problem resolution, only 15% of the students (n = 53) reported that technology support was available only 'Rarely' (2%, n = 8) or 'Sometimes' (13%, n = 45). Similar percentages reported on the

success of the technology support. Two percent of the students (*n* = 9) reported that technology support solved their problems 'Rarely' and 10% (*n* = 38) reported problem solutions only 'Sometimes'.

Communication Channels

The communication channels used by students are reported in Table 5. More than half of the students communicated with their instructors via email (87%) and messaging through the Blackboard system (55%). Approximately one-third communicated via other bulletin boards (35%), instant messaging (32%) or chat (29%), while very few students communicated with their instructors via telephone (6%), listserve (6%) or fax (6%). Interestingly, only 26% of the students reported using face-to-face meetings with their instructors as a vehicle for communication. Bivariate correlations between reported use of communication vehicles (e.g., phi coefficients) were all positive (ranging from r = .01 to r = .41), suggesting that students who communicate using one method also communicate using others (i.e., a general tendency to communicate with the instructor).

Table 5.
Percentage of Students Using Communication Vehicles

	Percentage Using Vehicle to Communicate with	
Communication Vehicle	Instructor	Other Students
Email	87%	52%
Blackboard Messaging	55%	42%
Bulletin Board	35%	30%
Instant Messaging	32%	14%
Chat	29%	24%
Face-to-Face Meeting	26%	20%
Telephone	6%	11%
Listserve	6%	8%
Fax	1%	3%

Note. N = 374.

Students' reported communications with other students paralleled those of the reported student-instruction communication, with email (52%) and Blackboard messaging (42%) being the most popular communication vehicles. The percentage of students reporting telephone communication with other students (11%) is nearly twice that of the student-instructor use of telephones (6%), but still represents a small proportion of the students. As with the instructor communications, the phi coefficients calculated between reported methods of communication with other students were all positive (ranging from r = .01 to r = .44).

Student Evaluation of Course Quality

The end of course survey provided the opportunity for students to rate the quality of the course on 15 items. An exploratory factor analysis was conducted on these item responses to identify the underlying dimensions of the students' perceptions. The results suggested that three factors accounted for 100% of the common variance and 63% of the total variance in the item responses. A procrustean rotation was performed to allow for correlations between the factors, and moderate to strong inter-factor correlations were obtained (r_{12} = .52, r_{13} = .63, r_{23} = .58). The resulting pattern and structure coefficients are presented in Table 6. The structure coefficients are simply the Pearson correlation between the items and factors, while the pattern coefficients are correlations between items and each factor with the other factors statistically controlled.

The first factor 'Assessment' consisted of items related to quizzes, exams, projects and assessments. These items were focused on the perceived relationships between assessments and course objectives, the extent to which the activities reinforced or allowed demonstration of learning, and the extent to which they were challenging. This factor accounted for 62% of the common variance in the item responses. After adjusting for the other factors, this factor accounted for 18% of the unique variance (e.g., common variance in the items that is not accounted for by other factors).

The second factor 'Communications' contained items related to instructor communications and encouragement of communications with other students. These items included both the timeliness and the constructiveness of responses to students' questions. In addition, this factor included items related to the timeliness and the quality of instructor feedback provided. The communications factor accounted for 57% of the common variance in item responses. After adjusting for variance shared with other factors, this factor accounted for 18% of the unique variance.

This third factor 'Instructional Design' was comprised of items related to examples provided (both quantity of examples and quality of examples), lesson organization, and opportunities for practice. The instructional design factor accounted for 61% of the common variance in the item responses, and 13% of the unique variance.

After verifying the underlying dimension of the student responses, factor score estimates were computed by calculating the mean response to the items related to each factor. Estimates of the internal consistency of each factor score estimate (Cronbach's alpha) were high. Coefficient alpha for the assessment factor was 0.87, alpha for the communication factor was 0.84, and alpha for the instructional design factor was 0.87.

The extent to which the five courses could be differentiated on these factor score estimates was evaluated by examining differences in student responses across the courses and testing for mean differences using analysis of variance (ANOVA). Box-and-whisker plots of the within-course

distributions are provided in Figures 1–3. Although these data indicate substantial overlap in the distributions of student ratings, as well as substantial within-course variability in student perceptions, the results suggested statistically significant differences across courses in the mean responses for each of the factor scores. Further, the effect sizes associated with the differences in mean responses across courses $\left(\hat{f}\right)$ represent medium effect sizes according to Cohen's (1988) guidelines. For these data, the effect sizes were 0.23, 0.24, and 0.19 for assessment, communications, and instructional design, respectively.

Table 6.
Factor Analysis Results for Student Assessment of Course Quality

	Factor 1: Assessment		Factor 2: Communications		Factor 3: Instructional Design	
Item	Pattern	Structure	Pattern	Structure	Pattern	Structure
Quizzes and practice activities relevant to course goals	86	81	---	---	---	---
Practice tests reinforced important concepts and skills	76	74	---	---	---	---
Exams, projects, assignments enabled demonstration of learning	67	78	---	---	---	---
Exams, projects, assignments were aligned with objectives	65	73	---	---	---	---
Activities/Assignments facilitated understanding	53	80	---	---	---	---
Assignments were appropriately challenging	46	65	---	---	---	---
Instructor responded to questions in timely manner	---	---	89	88	---	---
Instructor responded to questions in constructive manner	---	---	88	89	---	---
Instructor provided timely feedback	---	---	70	76	---	---
Instructor provided constructive feedback	---	---	62	75	---	---
Email/Discussion with peers was encouraged	---	---	35	51	---	---
Examples were clear and easy to follow	---	---	---	---	89	89
Sufficient examples/non-examples	---	---	---	---	83	84
Organizations logical/easy to follow	---	---	---	---	57	75
Sufficient opportunities to practice	---	---	---	---	51	72
	Total	Unique	Total	Unique	Total	Unique
Variance Accounted For	62%	18%	57%	18%	61%	13%

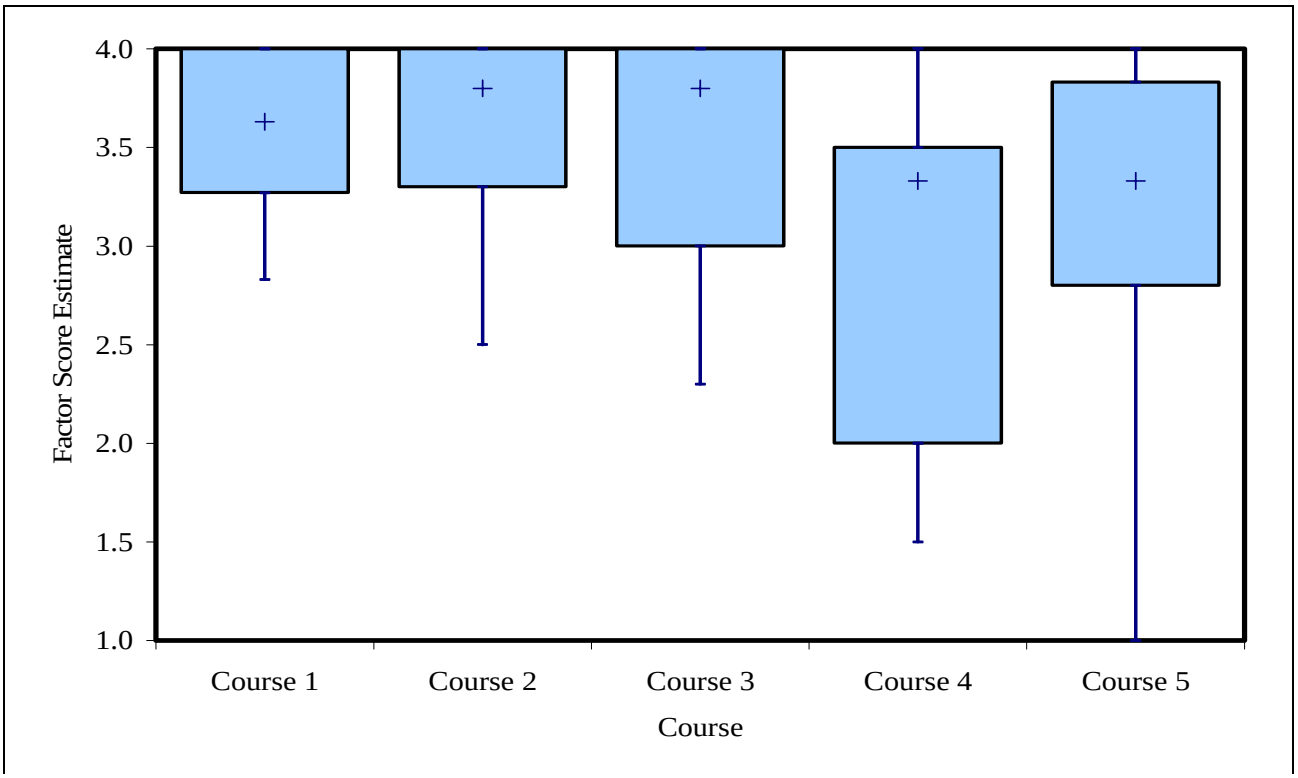

Figure 1. Within-Class Distributions of Assessment Factor Score Estimates.

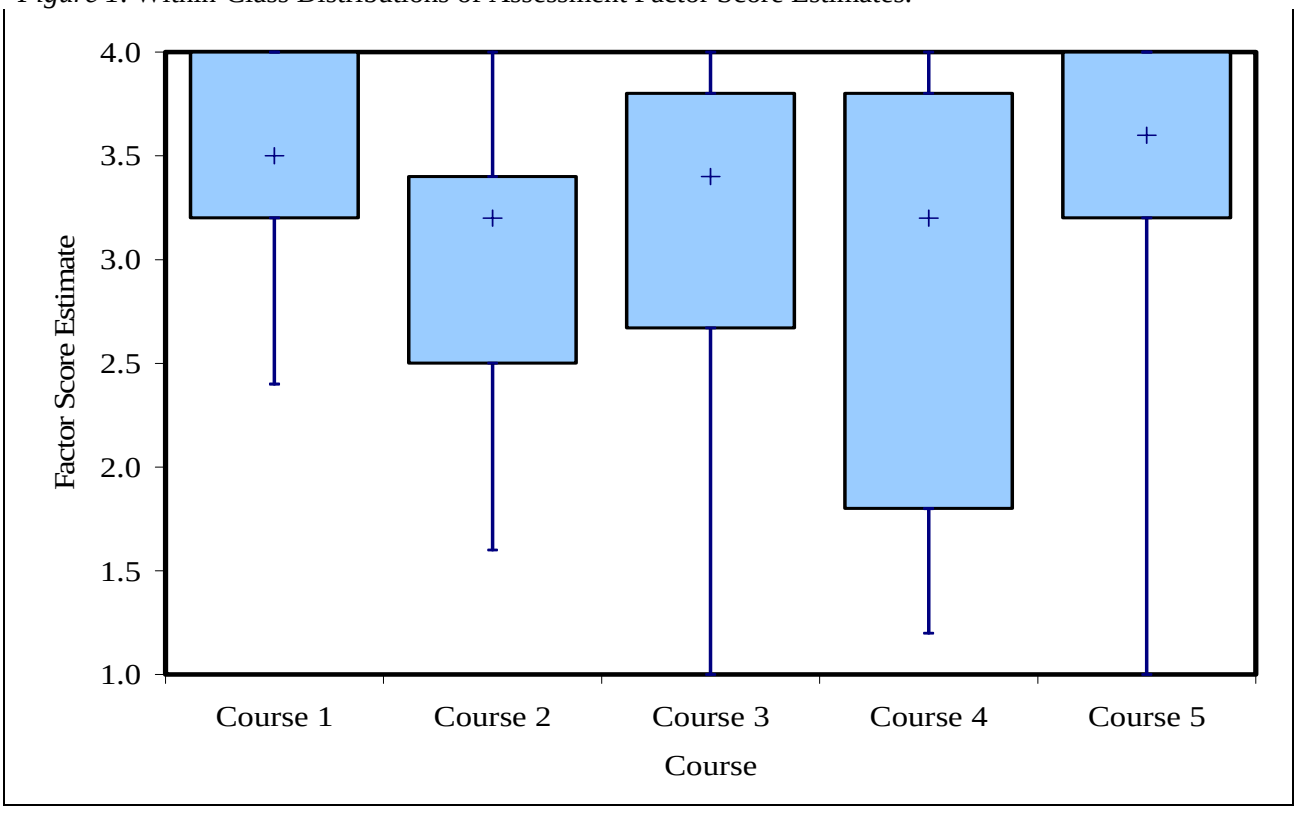

Figure 2. Within-Class Distributions of Communication Factor Score Estimates.

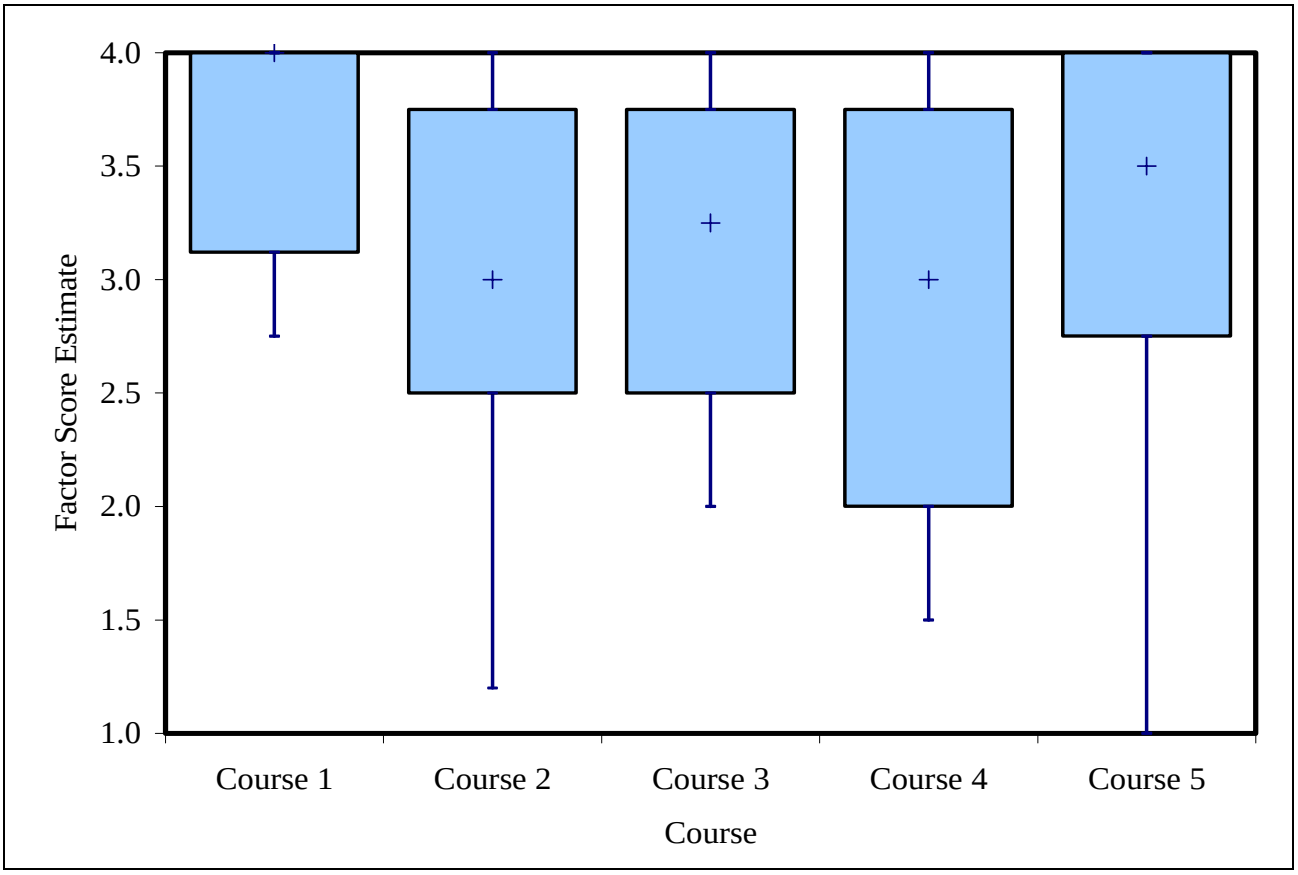

Figure 3. Within-Class Distributions of Instructional Design Factor Score Estimates.

Thematic Analysis

Open-ended questions for the end-of-term survey underwent a thematic analysis in order to identify common strengths and concerns across courses as well as those unique to a specific course. A total of 397 students from five courses answered the following three questions:

1. If given an option, based upon your experience with this course, how would you prefer to take this course?
2. What did you like most about this course?
3. If you could change one thing about this course, what would it be?

Of the five courses evaluated, one was a completely online web design course, one was a completely online human anatomy/physiology course, one was a hybrid business systems course and two were online measurement methods courses that involved the students coming to campus for three face-to-face sessions.

Interrater Reliability

An instrument for coding student responses was developed via an iterative process in which one rater developed an instrument with possible categories for analysis. These initial categories were based on a review of all entries and identification of themes that appeared to reoccur across courses. This instrument was then used to review the first ten statements made by students in each of the five courses. Two raters used the tool to determine where inconsistencies lay and which categories could be combined or eliminated. After this initial pass with the instrument, a second coding instrument was developed with fewer, but more comprehensive categories. The same two raters used this instrument to review the first 20 statements for each question in each of the five courses. The interrater reliability for the second instrument was greater than 80% for all categories. After discussion and agreement on how to interpret some of the finer nuances found within the comments, a third, and final coding protocol was developed. This instrument was very similar to the second protocol but with adjustments that served to enhance the effectiveness of the reliability of the coding. Independent coding of a final sample of the responses resulted in an interrater reliability of greater than 90% for all themes within the three questions, at times as high as 99%. Table 7 provides a summary of the interrater reliability for the three questions within each of the themes.

Table 7.
Summary of Percent Agreement for Themes within Each Question

Theme	Percent Agree (%)
If given an option, based upon your experience with this course, how would you prefer to take this course?	
Preference	
Face-to-face	95.2
Online	94.3
Combination	91.9
Explanation	
Pace/time	94.0
Communication with Students	99.4
Communication with Faculty	93.1
Course Concepts	95.5
Assignments	97.3
Instructor	97.3
What did you like most about this course?	
Organization/Structure of Course	95.5
Course Activities/Assignments	91.1
Course Content/Material	92.9
Time/Pace	96.4
Assessment	99.1
Communication with Students	98.5
Communication with Faculty	97.9
If you could change one thing about this course, what would it be?	
Organization of Course	95.1
Assignments	95.4
Assessment	97.6
Course Materials	96.9
Time/Pace	98.8
Group Work	97.6
Communication with Students	99.1
Communication with Faculty	99.4
Instructor	99.7
Technology	99.7

Question One Results

A number of themes evolved from the qualitative thematic analysis of the first question, "***If given an option, based upon your experience with this course, how would you prefer to take this course? Please explain***" (See Table 8 for preferences and explanations). Statements given for preferring the web design course suggest the majority of students (75%) enjoyed the online structure of the course, which included online testing and quizzes, and they were comfortable with the pace (25%) of the course. This may be because the students enrolled in this course were instructional technology students who may be more comfortable with technology in general. Students enrolled in the EDF6492 online measurement course clearly stated that they preferred a combination approach (66%) to the course as opposed to online (0%). This appears to be consistent with the tendency for the majority of these students (92%) to find the concepts of this course "*difficult to grasp*" and "*challenging*". Many students stated that they required more face-to-face time to understand these complex concepts.

Students enrolled in the online anatomy/physiology course stated that they preferred an online structure (41%) over either face-to-face (24%) only or a combination approach (29%). Students were comfortable with the pace of the schedule (29%) however; many did suggest they would like to have had more direct communication with the faculty (24%). Within the hybrid business course, students were split as to a preference for online (33%) or a combination (32%). Many were comfortable with the pace (18%) but enjoyed the direct communication with the faculty (10%).

Table 8.
Preference for Delivery Mode

Preference	All	EME6939	HSC2933	EDF6432	EDF6492	ISM3011
Face-to-face	43	1 (6%)	4 (24%)	5 (21%)	2 (2%)	31 (12%)
Online	110	12 (75%)	7 (41%)	2 (8%)	0 (0%)	89 (33%)
Combination	104	2 (12.5%)	5 (29%)	4 (16%)	8 (66%)	85 (32%)
Explanation						
Pace/time	61	4 (25%)	5 (29%)	3 (12.5%)	0 (0%)	49 (18%)
Communication with Students	12	1 (6%)	2 (12%)	0 (0%)	1 (8%)	8 (3%)
Communication with "Faculty	34	1 (6%)	4 (24%)	2 (8%)	0 (0%)	27 (10%)
Course Materials	39	3 (19%)	1 (6%)	5 (21%)	2 (2%)	28 (10%)
Course Concepts	37	2 (12.5%)	1 (6%)	12 (50%)	11(92%)	11 (4%)
Assignments	15	1 (6%)	1 (6%)	2 (8%)	2 (2%)	9 (3%)
Instructor	4	0 (0%)	0 (0%)	2 (8%)	0 (0%)	2 (.7%)
Other	34	0 (0%)	0 (0%)	0 (0%)	0 (0%)	34 (13%)

Question Two Results

When students were asked "***What did you like most about this course?***", there were some notable trends, both from the perspective of what students identified as things they liked and the areas that were not mentioned by many respondents. The majority (71%) of those enrolled in EME6939, the web design course, enjoyed the content and material of the course, while fewer (35%) liked the assignments. Within the human anatomy/physiology course, 42% liked the pace of the course. Obtaining support (as noted by responses that fell in the *'Other'* category), either online or face-to-face from the instructor or Teaching Assistants was seen as a common theme, particularly for those courses whose students were not as comfortable with technology. In addition to the students in the web design course that identified '*Course Content'* as an area that they liked, a number of students in the other courses also identified this area as a strength. For example, almost half (48%) of those in one of the measurement courses (EDF6432) also cited this as one of their favorite elements of the course. Very few students cited the communication aspects of the course as one of their favorite elements. The business course was the only one with more than one student citing communication with either faculty or students as an element of the course that they liked the most. Assessment was also an area that students did not tend to cite as something they liked. Only respondents in the Anatomy course (HSC2933) had over 10% of respondents citing this as something they liked about the course (e.g., *'The quizzes and tests related directly to the text'*), with no students in two of the courses (EME6939 and EDF6492) noting this as an area that they liked.

Table 9.
What Students Liked Most

Preference	All	EME6939	HSC2933	EDF6432	EDF6492	ISM3011
Organization of Course	27	2 (12%)	2 (8%)	1 (6%)	0 (0%)	22 (8%)
Course Activities/Assignments	126	6 (35%)	6 (25%)	4 (24%)	2 (20%)	108 (41%)
Course Content/Material	88	12 (71%)	9 (38%)	8 (48%)	3 (30%)	56 (21%)
Time/Pace	57	2 (12%)	10 (42%)	1 (6%)	3 (30%)	41 (15%)
Assessment	29	0 (0%)	9 (38%)	1 (6%)	0 (0%)	19 (7%)
Communication with Students	15	0 (0%)	1 (4%)	0 (0%)	0 (0%)	14 (5%)
Communication with Faculty	18	0 (0%)	0 (0%)	0 (0%)	1 (10%)	17 (6%)
Other	125	2 (12%)	6 (25%)	7 (41%)	4 (40%)	106 (40%)

Question Three Results

When asked "***If you could change one thing about this course, what would it be?***", again strong themes emerged (see Table 10). Students enrolled in the online web design course, EME6939, expressed a desire to have more and varied examples of course content (40%), however they appreciated the course organization and structure (33%). Conversely, those in the HSC2933, human anatomy/physiology course, suggested they would like to change the group work (35%) and/ or the assignments (40%). The following quotation from a student captures this thought; *"I would eliminate group projects and/or papers. It is too difficult to combine efforts with people exclusively via the web, especially when no one has ever met face-to-face before"*. Students enrolled in the online measurement courses seemed to feel that the courses were poorly structured (25%) for an online course. Many (33%) also felt that assessments should have been delivered online rather than to require students to come to campus to take them. As stated by one student, *"If this is an online course, the mid-term and final exams should be online."* The one thing students seemed to want to change in the business course were the assessments (42%). Many seemed to feel the examinations were too difficult, and, as in the measurement courses, that they should not have been required to travel to campus to take them (32%).

Table 10.
What One Thing Should Be Changed

	All	EME6939	HSC2933	EDF6432	EDF6492	ISM3011
Organization of Course	41	5 (33%)	2 (10%)	5 (25%)	3 (25%)	26 (9%)
Assignments	101	3 (20%)	8 (40%)	2 (10%)	3 (25%)	85 (32%)
Assessment	125	2 (13%)	6 (30%)	3 (15%)	4 (33%)	110 (42%)
Course Materials	39	8 (40%)	2 (10%)	3 (15%)	3 (25%)	23 (8%)
Time/Pace	7	0 (0%)	1 (5%)	2 (10%)	0 (0%)	4 (1%)
Group Work	26	1 (6%)	7 (35%)	0 (0%)	0 (0%)	18 (7%)
Communication with Students	9	0 (0%)	3 (15%)	1 (5%)	0 (0%)	6 (2%)
Communication with Faculty	14	0 (0%)	4 (20%)	4 (20%)	3 (25%)	3 (1%)
Instructor	1	0 (0%)	0 (0%)	1 (5%)	0 (0%)	0 (0%)
Technology	1	1 (6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Other	84	0	0	0	0	84 (32%)

Discussion

Providing an online system for student ratings of instruction is becoming much more common. A study conducted in 2000 revealed that only 2% of the "most wired" institutions reported institution-wide, evaluation systems on the web (Hmieleski & Champagne, 2000). In March, 2005, the Faculty Center at Brigham Young University conducted an extensive web search and communicated with over 100 institutions of higher education. The results indicate that at least 16% of the institutions have implemented online evaluation campus-wide, and more than 33% incorporate an online system to rate online and distance courses (Online Student Evaluation of Teaching in Higher Education, 2005).

There are many advantages of conducting course evaluations online, such as the immediacy of feedback, convenience for students, and opportunities for data warehousing (McGourty, Scoles, & Thorpe, 2002). However, there are a few issues that must be considered prior to implementing an online assessment, including the goals of the evaluation and student response rates.

In many cases, when institutions transition to an online course evaluation system, they simply convert their paper-based forms to web delivery (Hmieleski & Champagne, 2000). Although this may be a good starting point for the requisite ratings of instructors for promotion and tenure purposes, an online evaluation system can offer much more. For example, instead of conducting evaluations only when a course is ending, input from students can be collected throughout the semester, providing opportunities for formative reviews and adjustments.

The instruments designed in this study were implemented in two separate settings – one at mid-semester, and the other at the end of the semester. Depending on an institution's goals, one survey could be conducted for program/course review (and the results go directly to the faculty member), and another survey could be used at the end of the semester (and the results go anonymously to the administration). In other cases, it may not be practical or advisable to administer two surveys in each course.

Traditional, paper/pencil student ratings for university courses are generally conducted in classrooms at the end of the semester. In this environment, the response rate is generally high (Johnson, 2002). However, when the evaluation is conducted on the web, students have more "freedom" to decide if and when they will complete the form, and the response rate may fall to 30%- 40% (Hmieleski, 2000). This can be especially true in institutions where anonymity is important and extra incentives are not allowed for student evaluations. For example, a study completed by Johnson found that 87% of the students completed online ratings if extra points were given; 77% completed the form if it was a course assignment; 32% completed the form if it was encouraged, but not a formal assignment; and only 20% completed the evaluation if it was not specifically mentioned (2002).

Another concern is the bias that might exist based on non-respondents of online evaluations. Thorpe (2002) found that female students were significantly more likely to complete an online evaluation

form than male students. In addition, student who were low achievers were less likely to respond online. However, Hmieleski and Champagne assert: that "if faculty are 'on-board' and eager to use the information provided by a good evaluation, students see changes resulting from their feedback, and both parties recognize that the instrument measures what it is supposed to measure, then return rates will be high" (p. 3, 2000).

Conclusion

Concomitant with the increasing prevalence of distance learning as a means of delivering instruction throughout the educational system; comes the attendant importance of ensuring that these courses are both effective and useful. As stated earlier, issues and concerns related to the evaluation of online courses differ markedly from factors that have been traditionally examined in classroom settings As such, educators, administrators, and institutions need tools and methods to confirm that the courses and programs they offer not only meet the requirements of governing accreditation, policy-making, and funding agencies but also meet the needs of their students and instructors.

This study provides a foundation for developing sound practices for student evaluation of instruction in an online environment, both with respect to conceptual frameworks of evaluation and the nature of instruments and methods that should be employed. This research provides not only concrete examples of instrumentation that can be directly used or adapted by individuals, institutions, and programs; but also a concise reference guide to a host of previously constructed instruments and the development of new evaluation and assessment systems and instruments. The results of this research should provide instructors with easily accessible resources to gather information that will help them meet the needs of their students. In this vein, we encourage our colleagues to examine, adopt, or adapt the instruments that we have developed. Copies of the instruments and other reports of our research are available at http://sirocco.coedu.usf.edu/itt/website/instruments.htm.

References

Abrami, P.C., (2005, April). *Using teacher rating forms to evaluate teaching: Doing a better job with what we've got*. Symposium presented at the annual meeting of the American Educational Research Association. Montreal, CA. (http://www.cedanet.com/meta/AERA2005valid.pdf)

Abrami, P.C. & d'Apollonia, S. (1991). Multidimensional students' evaluations of teaching effectiveness; The generalizability of "N = 1" research: Comment on Marsh (1991). *Journal of Educational Psychology*. *83*, 411-415.

Arreola, R.A., (2000). *Developing a comprehensive faculty evaluation system: A handbook for college faculty administrators on designing and operating a comprehensive faculty evaluation system*. (2nd Ed.). Bolton, MA: Anker Publishing.

Arreola, R.A., (2005, April). *Validity like beauty is.....* Symposium presented at the annual meeting of the American Educational Research Association, Montreal, CA. (http://www.cedanet.com/meta/AERA2005valid.pdf)

Cohen, P.A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research, 51*, 281-309.

Cohen, S. (1997). The mismeasure of the college professor. *College Student Journal*, 31, 293-300.

Costin, F., Greenough, W., and Menges, R. (1971). Student ratings of college teaching: Reliability, validity, and usefulness. *Review of Educational Research*, 41, 511-536.

d'Apollonia, S. and Abrami, P. (1997). Navigating student ratings of instruction. *American Psychologist*, 52, 1198-1208.

d'Apollonia, S. and Abrami, P. (1999). Navigating student ratings of instruction. *American Psychologist*, 54, 519-520.

Dunkin, M. J. & Barnes, J. (1986). Research on teaching in higher education. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp 754 – 777).NY: Simon & Schuster Macmillan.

Franklin, J. (2005, April). *Validity, research, and reality: Student ratings of instruction at the crossroads*. Symposium presented at the annual meeting of the American Educational Research Association. Montreal, CA. (http://www.cedanet.com/meta/AERA2005valid.pdf)

Greenwald, A. and Gillmore, G. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, 52, 1209-1217.

Greenwald, A. and Gillmore, G. (1998). How useful are student ratings? *American Psychologist*, 53, 1228-1229.

Grinnell, L.D., Carey, L.M. & White, J.A (1999, November). *Modifying end-of-course student ratings for distance and web-enhanced courses: Do students discriminate among content, delivery, and infrastructure support?* Paper delivered at the annual meeting of the Florida Educational Research Association.

Hmieleski, K. (2000). Barriers to online evaluation: Surveying the nation's top 200 most wired colleges. Report prepared by the Interactive and Distance Education Assessment Laboratory at Rensselaer Polytechnic Institute, Troy, NY.

Hmieleski, K., and Champagne, M. V. (2000, September/October). Plugging into course evaluation. *The Technology Source*. (http://distance.wsu.edu/facultyresources/savedfromweb/pluggingin.htm)

Howard, G, Conway, C., and Maxwell, S. (1985). Construct validity of measures of college teaching effectiveness. *Journal of Educational Psychology*, 77, 187-196.

Johnson, T. (2002). Online student ratings: Will students respond? Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Marsh, H.W. (1984). Student evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76 (5). 707-754.

Marsh, W. (1987). Student evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal for Educational Research*, 11, 253-388.

Marsh, W. and Roche, L. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52, 1187-1197.

McGourty, J, Scoles, K., & Thorpe, S. (2002, June). Web-based student evaluation of instruction: Promises and pitfalls. Paper presented at the 42nd Annual Forum of the Association for Institutional Research, Toronto, Canada.

McKeachie, W. (1997). Student ratings: Validity and use. *American Psychologist*, 52, 1218-1225.

Moore, A., Masterson, J. T., Christopher, D. M., & Shea, K. A. (1996). College teacher immediacy and student ratings of instruction. *Communication Education*, 45, 29-39.

Online Student Evaluation of Teaching in Higher Education. (2005). Faculty Center at Brigham Young University. (http://onset.byu.edu)

Palmer, J. (1983). Sources and information: Faculty and administrator evaluation. In A. Smith (Ed.), *New Directions for Community Colleges*, No.41. San Francisco: Jossey-Bass.

Rifkin, T. (1995). ERIC Review: Faculty evaluation in community colleges. *Community College Review*, 23(1), 63-72.

Sorenson, D.L., & Johnson, T.D. (2003). Online student rating of instruction. *New Dimensions for Teaching and Leaning*, # 96. San Francisco: Jossey-Bass.

Sorenson, D.L., & Reiner, C. (2003). Charting the uncharted seas of online student ratings of instruction. *New Dimensions for Teaching and Leaning*, # 96. San Francisco: Jossey-Bass.

Thorpe, S. W. (2002, June). Online student evaluation of instruction: An investigation of non-response bias. Paper presented at the 42nd Annual Forum of the Association for Institutional Research, Toronto, Canada.

Acknowledgements

This work was supported, in part, by the University of South Florida and the Fund for the Improvement of Postsecondary Education, under Grant No. P339Z000006. $2,774,950 in federal funds were provided for the project, representing 50% of the total project costs. The remaining 50% of the project costs ($2,774,950) were financed by nonfederal sources. The opinions expressed are those of the authors and do not reflect the views of the United States Department of Education or the University of South Florida.