STUDENT PERCEPTIONS AND INSTRUCTIONAL EVALUATIONS: A MULTIVARIATE ANALYSIS OF ONLINE AND TRADITIONAL CLASSROOM SETTINGS

BILLY R. BROCATO, TEXAS A&M UNIVERSITY, email: billbrocato@gmail.com

ALESSANDRO BONANNO, SAM HOUSTON STATE UNIVERSITY, The e-mail address, telephone and fax numbers of the corresponding author

STACY ULBIG, SAM HOUSTON STATE UNIVERSITY, email: ulbig@shsu.edu

FEBRUARY 2013

ABSTRACT

This study examined students' evaluations of faculty performance in traditional and online classes. The study design builds upon prior research that addressed socially relevant factors such as classroom environments, students' learning goals, expected, and received grades, and more importantly, students' ratings of instructors' performance. The sample consists of data from a population of humanities and social sciences faculty from a medium-sized southwest undergraduate university who taught both online and traditional classes during the semester periods Fall 2010 to Spring 2012. In a traditional setting, the evaluation factors (develops rapport with students, stimulates students, challenges student learning, provides timely feedback, and teaches fundamentals), and the external factors – (course level taught and gender)—were found to significantly contribute to faculty summary scores. In an online class, students consistently rank female instructors better. However, the evaluation criteria – develops student rapport, stimulates students, provides timely feedback, and teaches fundamentals (though not 'challenges and involves students in their learning') – mirrored the same affects observed in the traditional classroom evaluations. The finding that "teaches fundamentals" received the largest standardized beta-coefficient in both classrooms further confirms earlier research that university students perceive course mastery as a major indicator of instructor performance regardless of gender or rank. However, the results indicate that students' perceptions are different when attending a traditional versus online classroom setting. This infers that synchronous and asynchronous settings require different teaching styles and different evaluation criteria.

KEY WORDS: ONLINE TEACHING, INSTRUCTOR EVALUATIONS, STUDENT PERCEPTIONS, TRADITIONAL versus VIRTUAL CLASSES.

BACKGROUND

The quality and effectiveness of instruction in online classes versus face-to-face settings continues to foster arguments over the appropriate pedagogy in a qualitatively different setting, i.e., synchronous versus asynchronous environment (Driscoll et al. 2012; Abdous & Yoshimura 2010; Angiello 2010; Milliron 2010; Benigno & Trentin. 2000). However, the tenor of these arguments is changing as student evaluations of college teaching (SET) in both classroom settings become an important (and more controversial) component of the faculty evaluation system (annual merit, promotion and tenure, post-tenure review (Stowell, Addison & Smith 2012; McPherson & Jewell 2007; Algozzine et al. 2004). Clayson (2009) questions the validity of SETs because researchers frequently fail to take into account the different classroom settings students and teachers might face. For example, some researchers argue evaluations in asynchronous settings requires a more robust methodological approach given the increasing importance and difference of students' experiences in that setting (Porter 2011; Kim & Bateman 2010; Clayson 2009; Beuschel, Gaiser, & Draheim 2003; Picciano 2002; Merisotis & Phipps 1999; Jaffee 1997; Marsh 1987).

Given the concerns about the validity of measurements that encourage academic productivity and justify social investments in institutions of higher learning (Fish & Gill 2009; Hannay & Newvine 2006), our study sets out to identify the student biases that affect student evaluations in university courses in both the traditional, brick-and-mortar and the virtual, online class settings. Specifically, we will examine student evaluations in terms of the heuristic biases students bring to bear in their evaluations of instructors' performance. We define students' heuristic biases as those social and institutional characteristics that affect the outcome of student evaluations (positively or negatively) that are related to instructors' pedagogy. Our initial investigation expands on earlier studies of student ratings of instructional performance by controlling for classroom setting, students' expectations regarding course mastery, instructional preferences, and socio-cultural factors outside instructors' locus of control (Klaus & Chagchit 2009; Meyer, 2007; Steiner et al. 2006).

Moreover, we build upon prior research on socially-relevant factors and correlations among classroom environments, students' learning goals, expected and received grades, and more importantly for our study, students' ratings of instructors' effectiveness (Myers & Claus 2012; Pamuk 2012; Smith, Cook, & Buskist 2011; Wilson & Allen 2011; Harris & Danielle 2006; Trevor Sehoole, & Moja 2003; Paris 2000). We take our data from a population of humanities and social sciences faculty from a medium-sized southwest undergraduate university who taught both online and traditional classes during the semester periods Fall 2010 to Spring 2012. Data consist of student class-evaluation scores based on the Individual Development and Educational Assessment (IDEA) rating system. The IDEA evaluation system has a 35-year national history in providing critical research into academic performance criteria and instructor performance assessments linked to the nonprofit's IDEA Center's evaluation services (IDEA 2012). The unit of analysis is the course summary evaluation score for faculty who taught in both classroom settings. The study independent variables are students' responses to questions on the university's course evaluation form. The scores are students' ratings of their instructor's teaching effectiveness, the achievement of learning objectives, and student perceptions of the instructors' contribution to their course mastery. The dependent variable is the adjusted course summary score comprised of the students' evaluation of instructors.

LITERATURE REVIEW

Introduced in the early 1920s, students' evaluations of teaching (SET) remain a mandatory metric of instructor effectiveness and student mastery (Algozzine et al. 2004). Marsh's seminal work (1987) has since called for a more rigorous approach to the study of SETs and their consequent outcomes to quiet the inevitable controversies that have plagued SETs. He writes that "the fairness of students' evaluations…could be better understood if researchers did not concentrate exclusively on trying to interpret background relationships as biases, but instead examined the *meaning* of specific relationships" (p. 311) (italics added). Contemporary studies attempt to show that students' decision-making heuristics, i.e., socially constructed schemas, indeed bias students' satisfaction with their classes because learning emerges from a dynamic social interaction between student and instructor (Driscoll et al. 2012). For example, research demonstrates significant correlations with students' perceptions of their instructors' interaction, course quality comparisons between instructors, course delivery structure, instructors' gender, race, and students' perceived learning outcomes in either a traditional or online classroom setting (Carle 2009; Hannay & Newvine 2006; Johnson, Aragon, Shaik, & Palma-Rivas 2000).

Further, because "teaching is a multidimensional activity, and measuring the effects of instruction is difficult because no single criterion of effectiveness is widely accepted" (Algozzine et al. 2004:135), we propose a student-centered lens might better account for "the great variability surrounding the perception and the measurement of the outcome, achievement, and effectiveness" of virtual settings (Abdous and Yoshimura 2010:735). For example, challenges to the validity of evaluations increases as researchers contrast the "disembedding" that occurs in online settings compared to the social experiences tied to shared physical spaces (Severino and Messina 2011:

66). The literature also reveals that the use of standard evaluation techniques – as a summative metric – remain socially-contested measures of instructors' effectiveness, students' perceived course mastery, and students' satisfaction with course instruction (Benmton & Cashin 2012; Dolnicar & Grün 2009; Phipps, Kidd, & Latif 2006; Aleamoni 1999; Cashin 1995; Abrami, D'Apollonia, & Cohen 1990).

Research criticisms generally point out the following methodological weaknesses in the literature: 1) studies that interpret correlation as causation; 2) a neglect to distinguish between practical and statistical significance; 3) ignoring of the multidimensionality of student ratings; 4) lack of clarity about the selection of the unit of analysis; and 5) a failure to ensure the replicability of findings. Additionally, researchers find that summary measures of course mastery and student satisfaction in either instructional setting vary by geography, student demographics, instructor demographics, and whether the research methodology used was qualitative or quantitative (Kupczynski, Mundy, & Maxwell 2012; Wilson & Allen 2011; Jaggars & Bailey 2010).

Educational researchers also confirm a broad social structure of student evaluations as a function of pedagogical schemas, computer-mediated communication technologies, community ties, students' motivation, and the general experience levels and gender of instructors in either class setting (Burnett 2011; Abdous & Yoshimura. 2010; Lannutti & Strauman 2006; Campbell, Gerdes, & Steiner 2005; Bento & Schuster 2003; Ti & McIssac 2002; Wulff, Hanor, & Bullik 2000; Rowden & Carlson 1996). For example, Harris and Parrish (2006) found that students' practical rationality for taking an online course versus a face-to-face class might play a more substantial role in students' performances and subsequent satisfaction ratings than an instructor's pedagogy. In their study, they found that the participants who received lower grades in the online class were overall less satisfied with their courses because of their initial misperceptions about the course structure and requirements. Hence student decisions to take a course in a particular setting (face-to-face versus online) and their retention rates were significantly affected by their end goal orientation. They found students who selected online courses complained they expected course workloads to be less rigorous than in a traditional setting. Harris and Parrish concluded that online course offerings should make it clear that ""convenient" does not mean "less work" or "less time" (2006:114).

Carle (2009) investigated students' evaluations of instructors' performance and asked if their evaluations improved over time and whether there were differences attributable to traditional and virtual settings. Using a data set of 10,392 classes taught by 1,120 instructors over a three-year period, Carle developed "multilevel growth models" to further examine "whether online vs. face-to-face, tenure, discipline, course level, sex, or minority status affected" student scores (2009:429). His findings support other research demonstrating that student evaluations in either class setting are not statistically significant over time because instructors appeared to improve in either setting (Stowell et al. 2012; Algozzine et al. 2004). Although Carle did find that students' ratings in traditional settings were negatively correlated with instructors' racial minority status, he concluded that the multilevel growth models were unable to explain the underlying score variances sufficiently among the independent variables that purported to distinguish teaching effectiveness.

Carle and other researchers, as the literature review demonstrates, focused on explaining student evaluations of instructors in terms of course mastery and faculty instructor styles across disciplines and classroom environments (Wiesenberg & Stacey 2008). Moreover, a majority of the research conducted has generally focused on methodologically weak descriptions, i.e., correlative studies that describe student experiences with technologies used and faculty instructional preferences and/or the technological expertise of both students and faculty in one setting or the other. For example, in some cases, researchers associate student evaluations with the experiential conditions (stuck in a brick-and-mortar classroom or in the privacy of their home) that comprise their traditional and virtual learning environments. Other studies rush to judgment by correlating students' familiarity with computer-mediated communication technologies to their overall satisfaction and mastery of learning concepts. We believe these studies are employing too simplistic a model.

We do not contest or attempt a thorough critique of past research findings. However, to overcome what appears limited in its explanatory power, we propose adopting a student-centered lens with a multivariate approach that includes a population of faculty instructors in a university setting who have taught in both classroom settings as a more robust account of students' socially-constructed evaluation perspectives. We argue that a multivariate analysis that incorporates the key factors that comprise students' standardized instructor evaluations would provide a more structurally relevant perspective to encourage additional debate regarding student perceptions that affect instructors' evaluation differences in a virtual or face-to-face class setting (Licht 1995). Further, a student-centered approach would extend theoretical understanding in these areas by examining emerging evaluation patterns within the respective institutional classroom settings. We contend that a multivariate analysis would allow a better theoretical model of student heuristics used in their evaluation of instructors' performances.

RESEARCH DESIGN

Given the research concerns mentioned previously, this study incorporates a two-stage nonprobability sampling design to study a southwestern state university college of humanities and social sciences instructors' summary evaluation scores in a face-to-face and virtual classroom environment. In the first stage, an exploratory study examined evaluation scores for a two-semester period at the sample location and found student evaluations were significantly lower for instructors' online classes compared to traditional classrooms (Bonanno & Brocato 2011). In the second stage, an expanded sample of the state university's faculty based on evaluation scores for traditional (N=539) and online (N=166) classrooms from undergraduate courses taught was obtained. The sampling frame included the semester periods Fall 2010, Spring 2011, Fall 2011, and Spring 2012 (excluding summer sessions) to better expand and examine the changes over time between the two classroom settings (Gray 2009).

In this nonprobability sample, the sampling unit approximates all U.S. colleges that use an established system of student evaluations of faculty members in undergraduate traditional or online instructional environments. Although Frankfort-Nachmias (2008) shows that generalizations to a population must use a probability sample to obtain valid statistical estimates of the population parameters, cost and faculty participation concerns limited using a randomized sampling design for the university as a whole. To overcome this methodological weakness, the sampling frame of total course evaluations (N=706) selected represented male (n=356) and female (n = 350) faculty members who taught undergraduate courses within the College of Humanities and Social Sciences at the state university selected. Further, because this study is concerned with student evaluations of instructors in face-to-face and virtual settings, evaluation scores were summed for each instructor and averages obtained for means testing. The aggregate averages obtained for traditional classroom settings were 194 summary scores and 93 online summary scores. The sampling frame's attributes included only those instructors who received undergraduate student evaluations in both a traditional and virtual classroom setting. Finally, to control for individual variability among instructors in the number of evaluations received, the unit of analysis remained course summary scores (dependent variable) in each setting. Moreover, for each summary score, six pedagogical variables (independent variables) and one control variable for student bias were selected. Although some instructors received more than one evaluation per type of classroom setting, we control for inter- and intra-variability in the obtained dependent and independent variables by restricting generalizations to variable effects in the aggregate. To ensure measurement reliability and hypothesis testing, we use Cohen's *d* to determine effect size and sample selection criteria.

Effect Size and Sample Selection

Clark-Carter (1997) warns that research hypothesis testing is significantly affected by sample size that can lead researchers to commit a Type II error – "rejecting the research hypothesis when in fact it is true" (p. 193). To overcome this often overlooked methodological requirement that allows researchers to compare findings independent of sample sizes, Clark-Carter recommends determining "effect size" and calculating the "power" of statistical tests linked to researchers' sampling designs (p. 193) before initial sampling by using the following formula:

$$d = \mu 2 - (\mu 1)/\sigma$$

In this study, μ_1 is the mean for first stage pilot study of evaluation scores for both classroom settings; μ_2 is the mean for second population sampled, σ is the standard deviation for the pilot study population (Bonanno & Brocato 2011). Solving the formula, *d* becomes a measure of how many standard deviations apart the two means are. Commonly referred to as Cohen's *d*, the effect size is similar to a *z score*, where *d* becomes a measure independent of sample size. Thus, in the first stage of this study, student evaluations from their traditional and online courses for a two-semester period had a mean of 4.22 (N=168 evaluation scores) and a standard deviation of 0.25. The second stage population sample over a longer time had a mean of 4.40 (N=706 scores). Solving the equation: [(4.4 – 4.22)/0.25] = *d* = 0.72 = effect size, or the measure of how many standard deviations apart the two sample means are. As the expanded sample is less than a standard deviation from the smaller sample, it is reasonable to infer that the results obtained are representative of the general population of student evaluations based on the IDEA system

Cohen's d is generally interpreted as *d* = 0.2 represents a small effect size, *d* = 0.5 has a medium effect size and *d* = 0.8 represents a large effect size (Clark-Carter 1997: 194-95). Thus, we provide an ad hoc measure that the evaluation summary scores effect size obtained from the population samples at 0.72, which demonstrates overall consistency in the students' evaluations samples, is a reliable measure of the evaluation system-ranking scheme. In other words, the effect size obtained confirms our initial research hypothesis that student evaluations between the two types of classroom experiences are consistent indicators of students' perceptions of instructors' instructor styles in either setting.

Statistical Power Analysis

To establish statistical power and minimize the likelihood of a Type II error, "Cohen and others recommend, as a rule of thumb", a power statistic (β) of 0.8, where the probability of rejecting the research hypothesis when it is in fact true is 1 – power (β) = 0.2 or 20 percent (Clark-Carter 1997:196). With α-level set at 0.05 and a minimal effect or association of 0.2 (correlation coefficient, R = 0.2) for a one-tailed test, the sample size required to obtain statistical power of 0.82 is 160, minimizing committing a Type II error to a level of 18 percent. In our study, the population sample size of average summary scores (N=287) far exceeds the estimated power coefficient sufficiently to reduce the likelihood of a Type-II error.

VARIABLES STUDIED

The dependent variables in this study are instructors' adjusted summary evaluation scores, DV_1, and the difference between an instructor's traditional classroom score and scores earned in online classes (calculated as: traditional classroom summary score *minus* the online evaluation score), DV_2. The independent variables include instructors' professional rankings, gender, course levels taught, semester-year of the course evaluation, and type of classroom setting (traditional or online). As control variables, we include student ratings on six evaluative factors related to the instructor's performance: (1) teaches course fundamentals, (2) inspires/challenges students, (3) introduces stimulating ideas, (4) develops rapport with students, (5) provides sufficient feedback to students, and (6) encourages student involvement in course learning. The inclusion of the control variable – "I really wanted to take this course regardless who taught it" was selected to determine the perceptual factors influencing student evaluations (biases) based on course requirements. We hypothesize that including a bias variable could provide a more transparent indicator of an instructor's received performance assessment (Lannutti & Strauman 2006; Steiner et al. 2006; Campbell et al. 2005; Marsh 1987). Additionally, we hypothesized that student evaluations are harsher when they are required to take a course and cannot select the course and its instructor. Conversely, we would expect that online evaluations would prove less harsh in a university setting because students choose from an array of different courses that better fit their personal needs.

Following Licht's (1995) proposition that a multivariate approach is a useful tool for understanding or explaining "the nature of a phenomenon for purposes of testing or developing theories" (p. 21), we analyze the differences in the relative weighting of the six key pedagogical factors to allow better theoretical interpretation of differences among each variable's influence on students' evaluations.

Research Hypotheses

Guiding the preliminary research are the following hypotheses:

H_1: Summary course evaluation scores will be higher in traditional classes than in online classes.

H_2a: If H_1 is accepted, then differences in summary evaluations are linked to specific pedagogical components that comprise the IDEA evaluation survey.

H_{2b}: Differences between online and traditional classroom evaluations are explainable according to an instructor's gender, and/or professional ranking, or course level.

H_{2c}: Controlling for student course -election bias should demonstrate a significant effect on their faculty evaluations.

H_3: The findings of H_1 through H_2 should provide clear indicators of those evaluations factors that are most predictive of student overall summary evaluation scores.

We would expect the results of H_1 to H_3, to provide a sufficient methodological framework for predicting students' evaluations of instructor performances in the traditional and virtual class settings.

DATA COLLECTION

Undergraduate summary evaluation scores by college, department, and instructor were collected. The total scores collected (N = 706) consist of class evaluations conducted from the Fall 2010 to Spring 2012 semesters, where instructors taught at least one online class and one traditional class. Faculty included in the population frame comprised of full professor, associate professor, assistant professor, and adjunct instructors. However, for comparative purposes, only instructors who taught online and traditional classes during the semester periods were included in the population means tests. Student evaluations consisted of freshman (N = 222; sophomore (N = 198); junior (N = 222); and senior (N = 64). While 48% (N=93) of the courses studied were taught in online settings, the remainder (N=194) took place in the traditional classroom setting.

RESULTS

Statistical analyses followed our three-step hypothesis approach. First, we tested the preliminary hypothesis that there are significant differences in student evaluation scores by type of classroom setting. As shown in Figure 1, an ANOVA comparison of course delivery type demonstrated online course evaluations were significantly lower ($p < .0001$) than traditional class settings (Accept H_1).

FIGURE 1 ABOUT HERE

Having accepted H_1, we examined gender differences by controlling for type of class environment. The results are illustrated in Figure 2 demonstrate that gender differences were significant for total aggregate summary scores (traditional and virtual course settings combined). Female instructors' mean scores were significantly better in the areas of *building rapport*, *involving students in learning*, *challenging students*, and providing *consistent/timely feedback*, and marginally better ($p < .10$) *in providing a stimulating, learning environment*. The remaining independent variable score – *teaching fundamentals* – was not significantly different by gender.

FIGURE 2 ABOUT HERE

Second, we examined instructor summary scores for an online environment as shown in Figure 3. Again, female faculty scored significantly higher for each component of the evaluation instrument ($p < 0.05$ or better) than their male peers.

FIGURE 3 ABOUT HERE

Figure 4 illustrates that for traditional classes, there are not any no significant difference by gender for summary scores or summary variables except for *provides timely feedback* where female faculty scored higher ($p < .05$).

FIGURE 4 ABOUT HERE

Although the statistical results demonstrated that faculty score significantly lower in an online environment, students evaluated female faculty better than they evaluated their male colleagues. Having received support for H_2, the next stage was to test for significance by examining further instructor's gender, the course level (freshman to senior), and the professional status of faculty (adjunct, assistant, associate, or full professor) as shown in Tables 1A and 1B.

TABLES 1A AND 1B ABOUT HERE

We then regress nine variables on summary evaluation scores (dependent variable). The analysis provides a rationale for accepting or rejecting $H_{3:}$ Which of the six evaluations factors are most predictive of student overall summary evaluation scores, controlling for gender, rank, and classroom type.

Traditional Classroom Evaluations

Table 1A shows the results of the regression analyses for the traditional class setting. The regression coefficients show significance for each pedagogical variable ($p < .002$ or better), except *involving students* in their learning ($p < .068$). In contrast to the ANOVA findings, instructor gender loads significantly ($p <.002$), favoring male instructors' (male dummy coded 0; *b* = -.065) in class performance, although an instructor's professional status does not contribute to student evaluations. The positive slope loadings for the pedagogical factors confirm that students evaluate their comfort with the instructor (*builds rapport*, *provides timely feedback*, *stimulates students*) and belief that they are mastering the subject matter (teaches fundamentals) as key issues in a traditional setting. Additionally, instructors who increase students' workloads as measured by *inspired students to set and achieve goals which really challenges* them (b = - .145, $p < .002$), pay a penalty. This is an indicator that students perceive instructors' efforts to challenge them as not critical to their course mastery. The negative slope for course level (coded 1 thru 4) indicates a less favorable or 'tougher' perceptual stance ($b = -0.027$); $p < .015$) on the part of students as they proceed forward in their undergraduate careers.

The significant standardized beta-coefficients (β) found in Table 2A reveal *that teaching fundamentals* has the largest relative ranking (.427) followed by *stimulating students' interest* (.319), *developing rapport with students* (.275), *inspiring students to set and achieve goals, which really challenged* them (-.145), and *providing timely feedback* (.118). The other significant variables – *involving students*, *course level*, and *gender* – had minimal relative influence on the hypothesized perceptual scheme of students' evaluations. As ideal types, the loadings are important indicators of students' perceptual heuristics in a traditional classroom. Overall, college students anticipate class mastery after completing a course and this is certainly a pragmatic, end goal. Students also appear concerned with an instructor's ability to stimulate their interest in face-to-face and virtual class encounters. The students' ratings discussed are representative of their context-dependent communicative cues (Scheff 2006). In a traditional setting, visual and auditory stimuli combine with the temporally structured, physical space occupied between students and instructors. In a virtual classroom, a predominance of interaction occurs simply through visually reading and typing responses in a less temporally structured and anonymous, space. For example, students rely on their subjective interpretations in either setting to evaluate an instructor's approachability, i.e., is the instructor there to help students achieve their end goals. Although students are evaluating whether they can successfully approach an instructor over course materials, the social context that provides the information available to them become qualitatively shifted in its import. As Blumer (1969:20) pointed out people "are caught up in a vast process of interaction in which they have to fit their developing actions to one another" to successfully navigate their social reality. Thus, the communicative success of exchanges in traditional and online setting vary with the context-dependent stimuli available to both parties (students and instructors).

Online Student Evaluations

Table 1B shows a different set of student heuristics at work in the online class setting. In contrast to the in-person class pedagogical factors and exogenous variables examined, only four independent variables significantly affect instructor evaluations; conversely, course level, gender, and instructor ranking are not influential. The significant variables are *building rapport* ($p < .000$), *stimulating students interest* ($p < .020$), *providing timely feedback* ($p < .054$), and *teaching fundamentals* ($p < .000$). The regression coefficients are positive for each variable except *involves students in their learning* ($b = -0.064$) and *challenges students* ($b = -0.048$), both are negative although not significant ($p < 0.202$ and 0.445, respectively). The standardized beta-coefficients for the above-mentioned variables reveal that *teaching fundamentals* remains the strongest relative weighting (.479) as it does in a traditional setting, but *building student rapport* moves up to the second strongest ranking (.326), followed *by stimulating students interest* (.194), and lastly *provides timely feedback* (.107). The relative ranking of the standardized coefficients show a shift in student perceptions from the traditional to the virtual class experience, where in an online setting, the standardized coefficient (*building student rapport*) ranks relatively higher, may indicate students in a virtual setting are more concerned with instructor presence (or absence) via electronic media than in the traditional setting.

As mentioned previously, similar to the findings in the traditional classroom, *challenges* students carries a negative slope, but in contrast to the in-person class, *involves* students in increased workloads, projects, etc. carried a negative slope. This finding, although statistically insignificant ($p < .202$), offers insight: As prior research has well-demonstrated, students who perceive instructors are making them "think they're thinking" are rewarded versus "if you really make them think, they'll hate you" (Clayson 2009:27). However, given the lack of physical cues, personal access to the instructor, and communicative interaction with other students regarding the 'fairness' of assigned work, we believe this indicates students begin online courses with a subjective bias in terms of course workload based on a comparison with their prior in-person class work requirements. As Harris and Parrish's (2006) research revealed, instructors (and colleges) in online classes should stress and clarify course requirements in the beginning to avoid student confusion or frustration because they are biased to believe the online setting should be less demanding given their prior experiences.

Our examination of the bias control factor – *I wanted to take this course regardless who was teaching it* – was not considered in prior equations. Table 2 provides eight regression summary models that examined each of the pedagogical variables and instructor characteristics by adding the student bias control. Overall, the adjusted R^2s obtained confirms that a significant amount of the variance in summary scores is attributed to the selected pedagogical factors. This finding at first glance may seem exceptional, but we remind the reader that the summary score is comprised of the pedagogical factors used in the analysis.

TABLE 2 ABOUT HERE

Based on student ratings – 1 = strongly disagree to 5 = strongly agree – the student bias scores obtained for the traditional and online settings demonstrate a significant negative contributing factor across all models examined ($p < .001$ or better), approximately driving down faculty summary scores (slope $b = -0.20$ or greater). This finding is important for instructors to consider. Because undergraduate students *must* take required courses to obtain their degree, instructors should consider at the outset that students are not necessarily convinced of the course's contribution to their degree plan. This would suggest that instructors might consider offering a better explanation of the course requirements and benefits to students.

DISCUSSION

The ANOVA analyses (refer to Figures 2, 3, and 4) confirmed our preliminary research hypotheses that there are significant differences in summary evaluation scores between online and traditional classroom settings. Instructors tended to receive lower ratings from students for their online teaching. This demonstrates that instructors should consider approaching the online setting differently from the traditional setting. While overall, there were no significant gender differences, female instructors did score higher in online course than their male counterparts. The regression coefficients in Table 1A demonstrated that in a traditional setting, the pedagogical variables (*teaches fundamentals, develops rapport with students, stimulates students, challenges student learning, and provides timely feedback*) and the exogenous variables – (*course level taught and gender*)—significantly contribute to faculty summary scores (excluding instructor ranking and *involves students*). However, the results in Table 1B for online evaluations revealed that gender, course level, and instructor's ranking did not contribute significantly to summary scores, although *teaches fundamentals*, *develops student rapport*, *stimulates students*, and *timely feedback* mirrored the same effects observed in the traditional classroom evaluations (excluding *challenges students* and *involves students in their learning*). The finding that "teaches fundamentals" received the largest standardized beta-coefficient in both classrooms further confirms Steiner et al.'s (2006) research that university students are most concerned with how successful they believe the instructor has transferred knowledge to them in their classes.

Examining the slopes of the independent variables in Table 1A and 1B provided additional insights; for example, in a traditional setting, student perceptions of instructional performance are negative when they *feel challenged* by instructors. The negative slope for course level also demonstrates that students in upper-division courses are likely to be tougher in their evaluations in the traditional setting ($p < .015$). These two variables demonstrate difficult hurdles to overcome as an instructor. Additionally, challenging students to take on more responsibility in their class mastery is a delicate pedagogical style to navigate, especially if students distrust an instructor's classroom strategies. Conversely, in an online class, there were not any significantly negative slopes associated with the summary evaluation variables.

The regression models presented in Table 2 revealed that student bias – whether they wanted to take a course, regardless who taught it – played a significantly negative role in either class environment, irrespective of gender, but the effects are more pronounced for female faculty. As previously mentioned in the results section, this finding suggests that students and faculty find themselves in an ambiguous communicative environment. We believe this question closely resembles a double-barrel survey item. Students are asked if they really wanted to take the course (rate from lowest to highest) and whether the instructor was someone, they wanted to study with (rate from lowest to highest). Because the variable is significant ($p < .0001$) across our models (except for male faculty in the online setting) and the coefficient has a negative slope, it appears students overwhelmingly are not satisfied with their course choices (and/or instructor performance). We suggest that this variable as a survey item needs further operationalization.

In summary, our findings suggest that in face-to-face classes, students rely on social context, social interaction, and normative patterns of evaluating persons of authority. Conversely, in the virtual classroom, gender, race, and course level are not significant contributors to students' expectations of instructors' performance. We posit that this confirms that the student-teacher interaction "is mediated by the social context and set of social relationships" (Koeber 2005:286), suggesting that students views of instructor performance are qualitatively different given the classroom setting. For example, the regression analyses in Table 2 demonstrates that female faculty scores in a traditional setting are more significantly affected by their teaching style on challenging students to do more in their course mastery. However, female instructors do well overall in an online setting, except that involving students in their learning can negatively impact their evaluation. Because the slopes are reversed in the two class settings, it is reasonable to infer a qualitative shift in student perceptions regarding their class work requirements. This could also imply that female instructor communication of work requirements is different by class environment.

Similarly, male faculty in a virtual class should reconsider their attempts to challenge students and involve students, although the variables are not significant, the slope shows a negative effect. It appears that student evaluations of male instructors remain consistent across both classroom platforms. This implies that students might continue to rely on normative behaviors in their evaluations of male instructors, but as noted earlier, overall student perceptions are qualitatively different in an online setting and male faculty might consider this when preparing their course syllabi. As Centra (2003) writes, “the level of difficulty, workload, and pace in a course has a greater influence” on student evaluations than “do expected grades” (p. 508). Additionally, Clayson’s (2009) study revealed that students are less likely to view increased workloads or challenges to their classroom learning as justified.

CONCLUSION

Our study explored the student-instructor relationship and students’ heuristic biases that comprise their end-of-course instructor evaluations in a university setting. Our initial research hypothesis found that instructors receive significantly lower student evaluations in an online versus in-person class setting. We followed this finding with a closer investigation of the dataset and examined student evaluations of instructor effectiveness for significant mean differences by controlling for type of classroom and gender. The ANOVA results demonstrated students’ rate female faculty higher in an online class than their male counterparts, but there were not any significant gender differences in the traditional class setting. Having found that gender differences existed in the online classroom, we next used a multivariate regression model (Table 1A-1B) to examine for gender differences by classroom setting. An analysis of course level, gender, instructor rank, and students’ decision to attend a specific course allowed for testing earlier studies that argued instructors’ gender, race, and age often biased student evaluations (Campbell et al. 2005). The results showed that the exogenous variables were significant contributors to students’ aggregated responses in a traditional classroom, but lost significance in an online setting. We then asked what these contradictory results might indicate.

Overall, our investigation revealed that instructors' teaching styles do not effectively transfer across classroom platforms. However, in both classroom settings, students are most concerned with learning course fundamentals, i.e., gaining course mastery. Students expect to be in a stimulating learning environment as well. Instructors are also responsible for establishing student rapport, i.e., being available to students, gaining their trust, and their confidence. Succeeding at these would appear significantly different given the communicative constraints that exist in the two learning environments. Encouraging students to approach instructors in a traditional class requires a different demeanor than in a virtual setting. Students prize timely feedback. This is a double-edged sword for instructors. How positive or how negative the feedback might be mitigated by temporal and environmental factors. For example, in a virtual setting, do students expect instructors to monitor their email constantly or at a specified time? Moreover, if so, how important is a fast response versus a well-thought out response. As students' progress in their university careers, their evaluations become stricter. Regardless of the setting, the slopes are negative, albeit significantly so for male instructors with evaluations for both classes combined.

We used a multivariate model to bolster our theoretical understanding of the differences in students'' perceptions in traditional and virtual classrooms. The data provided a reasonable description of the qualitative shifts in student evaluation orientations based on the physical and communicative setting they found themselves in. We can conclude that students are indeed concerned with their achieving a meaningful learning experience irrespective of the classroom setting. Students are also concerned with the equity of coursework instructors provide them in both settings. These findings are encouraging from an instructor's perspective. We suggest that female instructors share their teaching styles that bolstered students' online evaluations with their male colleagues in an open discussion forum. It is a significant finding of our research that the evaluation criteria used in a brick-and-mortar class versus a virtual learning environment does not address the principles embedded in faculty evaluations. Teaching students is a life-long learning experience. In addition, with the increased implementation of online courses, experienced and new instructors are required to develop effective delivery techniques in a rapidly changing technological environment. Because instructors remain a critical component of a virtual setting that is a unique social environment filled with ambiguous communicative symbols linked to the context of a student's learning environment (Koeber 2005), we further suggest that the use of the same metric across both platforms appears an unreasonable and unreliable indicator of faculty teaching effectiveness.

REFERENCES

Abdous, M. & Yoshimurra, M. (2010). Learner outcomes and satisfaction: A Comparison of live video-streamed instruction, satellite broadcast instruction, and face-to-face instruction. *Computers & Education* 55: 733–741.

Abrami, P.C., D'Apollonia, S. & Cohen, P. (1990). Validity of student ratings of instruction: what we know and what we do not. *Journal of Educational Psychology* 82:219-231 (http://lib-ezproxy.tamu.edu:2048/login?url=http://search.proquest.com/docview/614383870?accountid=7082).

Aleamoni, L.M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education* 13:153-166 (http://lib-ezproxy.tamu.edu:2048/login?url=http://search.proquest.com/docview/201510895?accountid=7082).

Algozzine, B., Gretes, J., Flowers, C., Howley, L., Beattie, J., Spooner, F. & Mohanty, G. (2004). Student evaluation of college teaching: a practice in search of principles*.* *College Teaching* 52(4): 134-141.

Angiello, R. (2010). Study looks at online learning vs. traditional instruction. *Education Digest* 76: 56.

Benigno, V. & Trentin, G. (2000). The Evaluation of online courses*.* *Journal of Computer Assisted Learning* 16: 259-270.

Bento, R. & Schuster, C. (2003). Participation: the online challenge. In A.K. Aggarwal (Ed.), *Web-based education: learning from experience* (pp. 156-164). Hershey, PA: Information Science Publishing.

Benton, S.L. & Cashin, W.E. (2012). Student ratings of teaching: a summary of research & literature. *IDEA Paper No. 50*, Pp. 22. Manhattan, KS: The IDEA Center. Retrieved from http://www.theideacenter.org/sites/default/files/idea-paper_50.pdf.

Beuschel, W. Gaiser, B., & Draheim, S. (2003). Communication needs of online students. In A.K. Aggarwal (Ed.),*Web-based education: learning from experience* (pp. 203-222). Hershey, PA: Information Science Publishing.

Bonanno, A. & Brocato, B.R. (2011). Issues with online classes. *Report* prepared for the College of Liberal Arts and Science, Sam Houston State University. November.

Blumer, H. (1969). *Symbolic Interactionism: Perspectives and Method*. Berkley, CA: University of California Press.

Burnett, C. (2011). Medium for empowerment or a 'centre for everything': students' experience of control in virtual learning environments within a university context. *Education Information Technologies* 16:245-258.

Campbell, H.E., Gerdes, K., & Steiner, S. (2005). What's looks got to do with it? instructor appearance and student evaluations of teaching. *Journal of Policy Analysis & Management* 24: 611-620.

Carle, A.C. (2009). Evaluating college students' evaluations of a professor's teaching effectiveness across time and instruction mode (online vs. face-to-face) using a multilevel growth modeling approach. *Computers & Education* 532: 429-35.

Cashin, W. (1995). Student ratings of teaching: the research revisited. *IDEA Paper No. 32*, Center for Faculty Evaluation and Development. Manhattan, KS. Retrieved from http://www.theideacenter.org/sites/default/files/Idea_Paper_32.pdf.

Centra, J.A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work?. *Research in Higher Education* 44:495-518.

Clark-Carter, D. (1997). *Doing quantitative psychological research: from design to report.* East Sussex, UK: Psychology Press Limited.

Clayson, D.E. (2009). Student evaluations of teaching: are they related to what students learn?: a meta-analysis and review. *Journal of Marketing Education* 31: 16-30.

Creswell, J.W. (2009). *Research design: qualitative, quantitative, and mixed methods approaches 3rd ed.* Thousand Oaks, CA: Sage Publications Ltd.

Dolnicar, S. & Grün, B. (2009). Response style contamination of student evaluation data. *Journal of Marketing Education* 31:160-172.

Driscoll, A., Jicha, K., Hunt, A.N., Tichavsky, L., & Thompson, G. (2012). Can online courses deliver in-class results?: a comparison of student performance and satisfaction in an online versus a face-to-face introductory sociology course. *Teaching Sociology* 40: 312-331.

Fish, W.W. & Gill, P.G. (2009). Perceptions of online instruction. *The Turkish Online Journal of Educational Technology* 8:53-64.

Frankfort-Nachmias, C. & Nachmias, D. (2008). *Research methods in the social sciences*. New York, NY: Worth Publishers.

Gray, D.E. (2009). *Doing research in the real world, 2nd ed*. Thousand Oaks, CA: Sage Publications Ltd.

Hannay, M. & Newvine, T.. (2005). Perceptions of distance learning: a comparison of online and traditional learning. *Journal of Online Learning and Teaching* http://jolt.merlot.org/05011.htm.

Harris, D.M., & Parrish, D.E. (2006). The art of online teaching: online instruction versus in-class instruction. *Journal of Technology in Human Services* 24: 105-117.

IDEA Center. (2012). About the IDEA center. The IDEA Center Inc. Manhattan, KS. Retrieved from http://www.theideacenter.org/about.

Jaggars, S.S. & Bailey, T. (2010). Effectiveness of fully online courses for college students: response to a department of education meta-analysis. *Community College Research Center*, 18 pps. Columbia University. New York: NY. July. Retrieved from http://ccrc.tc.columbia.edu.

Jaffee, D. (1997). Asynchronous learning: technology and pedagogical strategy in a distance learning course. *Teaching Sociology* 25:262-277.

Jennings, S.E. & Bayless, M.L. (2003). Online vs. traditional instruction: a comparison of student success. *Delta Pi Epsilon Journal* 45.3: 183-190.

Johnson, S.D., Aragon, S.R., Shaik, N., & Palma-Rivas, N. (2000). Comparative analysis of learner satisfaction and learning outcomes in online and face-to-face learning environments. *Journal of Interactive Learning Research* 11: 29-49.

Kim, H.K. & Bateman, B. (2010). Student participation patterns in online discussion: incorporating constructivist discussion into online courses. *International Journal on E-Learning* 9:79-98.

Klaus, T. & Chagchit, C. (2009). Online or traditional: a study to examine course characteristics contributing to students' preference for classroom settings. *International Journal of Communication and Information Technology Education* 5: 14-23.

Koeber, C. (2005). Introducing multimedia presentations and a course website to an introductory sociology course: how technology affects student perceptions of teaching effectiveness. *Teaching Sociology* 33: 285-300.

Kupczynski, L., Mundy, M., & Maxwell, G. (2012). Faculty perceptions of cooperative learning and traditional discussion strategies in online courses. *Turkish Online Journal of Distance Education* 13: 84-95.

Lannutti, P. & Strauman, E.C. (2006). Classroom communication: the influence of instructor self-disclosure on student evaluations. *Communication Quarterly* 54: 89-99.

Licht, M.H. (1995). Multiple regression and correlation. Pp. 19-64 *in Reading and understanding multivariate statistics*, edited by L. G. Grimm and P.R. Yarnold. Washington D.C.: American Psychological Association.

McPherson, M.A. & Jewell, R.T. (2007). Leveling the playing field: should student evaluation scores be adjusted?. *Social Science Quarterly* 88: 868-881.

Merisotis, J.P. & Phipps, R.A. (1999). What's the difference? outcomes of distance vs. traditional classroom-based learning. *Change* 31.3: 12-17.

Meyer, K.A. (2007). Student perceptions of face-to-face and trad and online discussions: the advantage goes to…. *Journal of Asynchronous Learning Networks* 11: 53-69.

Milliron, M.D. (2010). Online education vs. traditional learning: time to end the family feud. *Chronicle of Higher Education* 57.11: B30-B32.

Munch, P.A. (1974). Empirical science and max weber’s verstehende soziologie. Pp. 56-69 in *Verstehen: subjective understanding in the social sciences*, edited by Marcello Truzzi. Reading, MASS: Addison-Wesley Publishing Company.

Myers, S.A. & Claus, C.J. (2012). The relationship between students’ motives to communicate with their instructors and classroom environment. *Communication Quarterly* 60: 386-402. Social Sciences Full Text (H.W. Wilson).

Pamuk, S. (2012). The need for pedagogical change in online adult learning: a distance education case in a traditional university. *University Of Gaziantep Journal of Social Sciences* 11: 389-405.

Paris, D.C. (2000). Is there a professor in this class?. In R.A. Cole (Ed.), *Issues in web-based pedagogy: a critical primer* (pp. 95-110). Westport, Conn: Greenwood Press.

Phipps, S.D., Kidd, R.S., & Latif, D.A. (2006). Relationships among student evaluations, instructor effectiveness, and academic performance. *Pharmacy Education* 6: 237-243.

Picciano, A.G. (2002). Beyond student perceptions: issues of interaction, presence, and performance in an online course. *Journal of Asynchronous Learning Networks* 6: 21-40.

Porter, S.R. (2011). Do college student surveys have any validity? *Review of Higher Education* 35: 45-76.

Rowden, G.V. & Carlson, R. E. (1996). Gender issues and students' perceptions of instructors' immediacy and evaluation of teaching and course. *Psychological Reports* 78, 835-839.

Scheff, T.J. (2006). *Goffman unbound!: a new paradigm for social science*. Boulder, CO: Paradigm Publishers.

Smith, D.L., Cook, P., & Buskist, W. (2011). An experimental analysis of the relation between assigned grades and instructor evaluations. *Teaching of Psychology* 38: 225-228.

Steiner, S., Holley, L.C., Gerdes, K., & Campbell, H.E. (2006). Evaluating teaching: listening to students while acknowledging bias. *Journal of Social Work Education* 42: 355-376.

Trevor S.C. & Moja, T. (2003). Pedagogical issues and gender in cyberspace education: distance education in south africa. *African & Asian Studies* 2: 475-496.

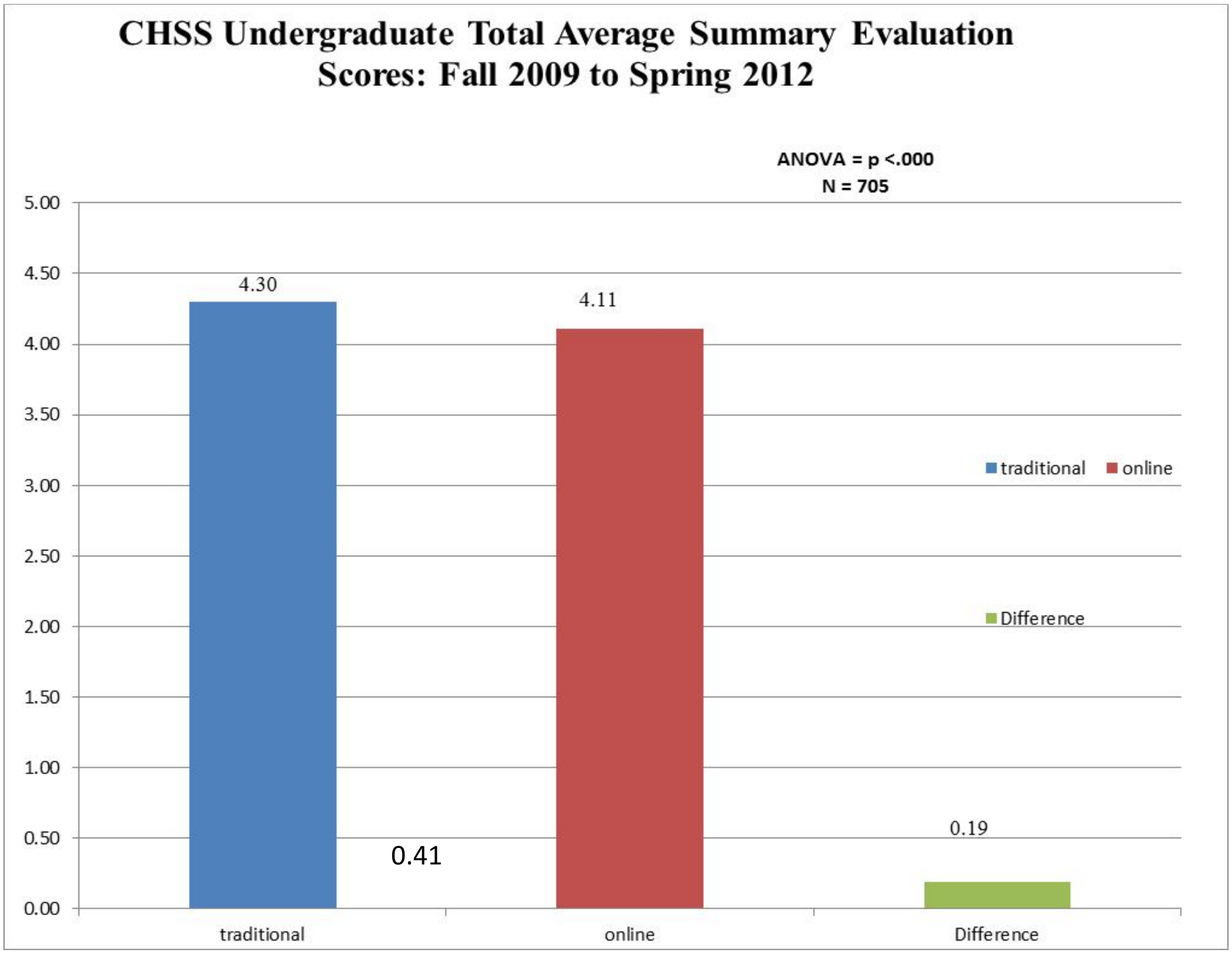
Tu, C.H. & McIssac, M. (2002). The relationship of social presence and interaction in online classes. *The American Journal of Distance Education* 16:131-150.

Wiesenberg F.P. & Stacey, E. (2008). Teaching philosophy: moving from face-to-face to online classrooms. *Canadian Journal of University Continuing Education* 34:63-79.

Wilson, D. & Allen, D. (2011). Success rates of online versus traditional college students. *Research in Higher Education Journal* 14: 1-9.

Wulff, S., Hanor, J., & Bulik, R.J. (2000). The roles and interrelationships of presence, reflection, and self-directed learning in effective world wide web-based pedagogy. In R.A. Cole (Ed.), *Issues in web-based pedagogy: a critical primer* (pp.143-160). Westport, Conn: Greenwood Press.

Figure 1. CHSS Undergraduate Summary Scores and Differences by Semester Period



ANOVA, N=705, $p < .0001$.

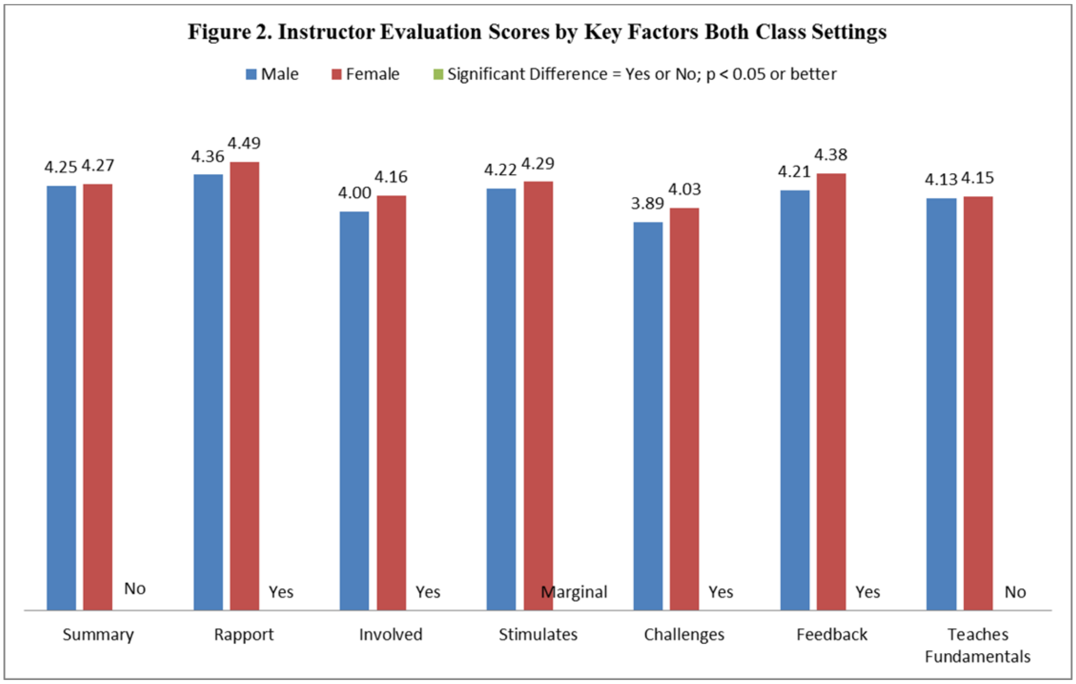


Figure 2. Instructor Evaluation Scores by Key Factors Both Class Settings
Male
Female
Significant Difference = Yes or No; p < 0.05 or better
4.25
4.27
4.36
4.49
4.00
4.16
4.22
4.29
3.89
4.03
4.21
4.38
4.13
4.15
No
Yes
Yes
Marginal
Yes
Yes
No
Summary
Rapport
Involved
Stimulates
Challenges
Feedback
Teaches Fundamentals

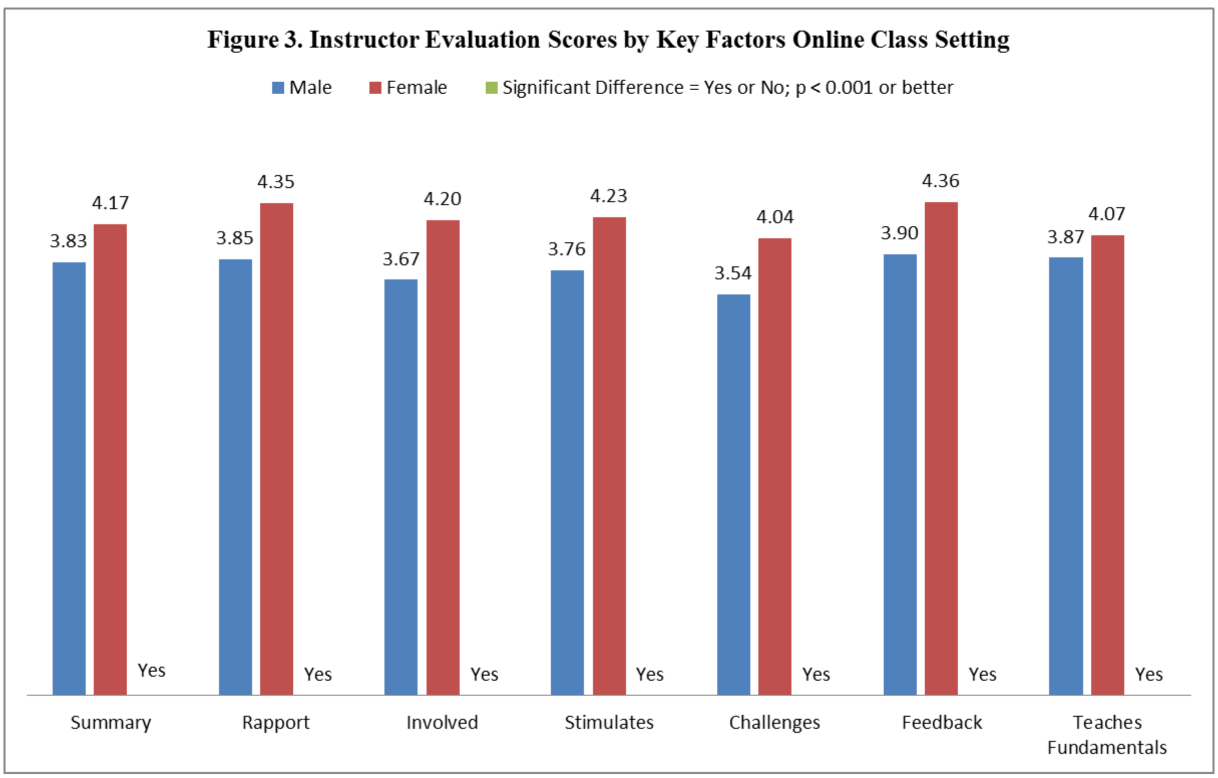


Figure 3. Instructor Evaluation Scores by Key Factors Online Class Setting
Male
Female
Significant Difference = Yes or No; p < 0.001 or better
3.83
4.17
3.85
4.35
3.67
4.20
3.76
4.23
3.54
4.04
3.90
4.36
3.87
4.07
Yes
Yes
Yes
Yes
Yes
Yes
Yes
Summary
Rapport
Involved
Stimulates
Challenges
Feedback
Teaches Fundamentals

Figure 4. Instructor Evaluation Scores by Key Factors Traditional Class Setting

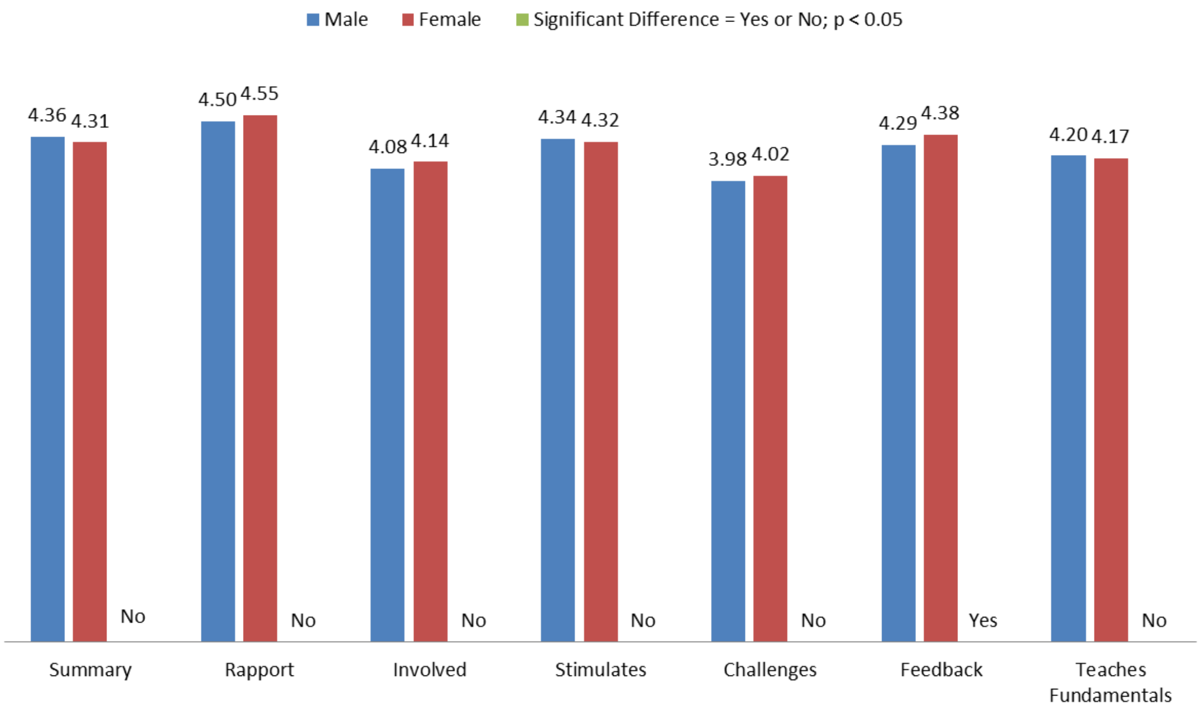


Table 1A. Multivariate Analysis: Traditional Summary Scores (DV): Selected Teaching Factors and Course Characteristics

Independent Variables	unstandardized coefficients	Std. Error	Standardized coefficients	P-value
Constant	-.040	.138		.775
Rapport	.275	.051	.230	.000
Involves Students	.057	.031	.062	.068
Stimulates Students	.319	.047	.286	.000
Challenges Students	-.145	.046	-.145	.002
Timely Feedback	.118	.032	.124	.000
Teaches Fundamentals	.407	.032	.427	.000
Course Level	-.027	.011	-.060	.015
Gender	-.065	.021	-.073	.002
Instructor's Ranking	-.010	.014	-.018	.447

Note: Online class, N = 93, traditional class, N = 194. Male = 151, female = 134.

Table 1B. Multivariate Analysis: Online Summary Scores (DV) and Selected Teaching Factors and Course Characteristics.

Independent Variables	unstandardized coefficients	Std. Error	Standardized coefficients	P-value
Constant	.229	.166		.170
Rapport	.263	.061	.326	.000
Involves Students	-.064	.050	-.080	.202
Stimulates Students	.155	.066	.194	.020
Challenges Students	-.048	.062	-.056	.445
Timely Feedback	.081	.042	.107	.054
Teaches Fundamentals	.519	.055	.479	.000
Course Level	-.003	.019	-.006	.869
Gender	.063	.039	.065	.109
Instructor's Rank	.035	.025	.050	.163

Note: Online class, N = 93, traditional class, N = 194. Male = 151, female = 134.

Table 2. Multiple Regression Models Predicting Student Summary Evaluation Scores for Faculty and Course Characteristics

	Model 1: All Factors	Model 2: Traditional All Factors	Model 3: Both Classes Male	Model 4: Both Classes Female	Model 5: Female In-person	Model 6: Female Online	Model 7: Male In-person	Model 8: Male Online
Constant	.696***	.568***	.809***	.503***	.109	.991	.937***	.257
	(.098)	(.130)	(.139)	(.139)	(.176)	(.255)	(.193)	(.241)
Rank	.010	.009	.011	.005	.004	.023	.012	.020
	(.008)	(.009)	(.011)	(.013)	(.015)	(.026)	(.013)	(.023)
Gender	-.056***	-.061***						
	(.017)	(.019)						
Type Class	.035							
	(.021)							
Course	-.018*	-.020*	-.030*	-.004	-.013	.013	-.026	-.036
	(.009)	(.010)	(.012)	(.013)	(.014)	(.027)	(.014)	(.028)
Fundamental	.426***	.382***	.389***	.501***	.488***	.405***	.335***	.597***
	(.025)	(.028)	(.033)	(.040)	(.046)	(.081)	(.036)	(.069)
Challenge	-.069*	-.094*	-.117	-.021	-.128*	.072	-.074	-.122
	(.035)	(.041)	(.049)	(.046)	(.059)	(.081)	(.057)	(.091)
Stimulates	.321***	.415***	.400***	.205***	.265***	.201*	.473***	.164
	(.037)	(.042)	(.046)	(.052)	(.063)	(.097)	(.056)	(.086)
Rapport	.200***	.210***	.177***	.284***	.323*	.230*	.146*	.176*
	(.036)	(.045)	(.049)	(.050)	(.063)	(.094)	(.067)	(.077)
Feedback	.104***	.106***	.104***	.078*	.103**	.098	.096	.106*
	(.023)	(.028)	(.032)	(.035)	(.040)	(.072)	(.040)	(.049)
Involved	.013	.030	.011	.002	.089*	-.124	-.020	.066
	(.025)	(.028)	(.033)	(.035)	(.046)	(.068)	(.037)	(.069)
Bias	-.208***	-.233***	-.192***	-.234***	-.238***	-.186***	-.222***	-.065
	(.017)	(.019)	(.023)	(.024)	(.027)	(.047)	(.026)	(.051)
R^2	.809	.785	.796	.833	.852	.817	.722	.873
Adjusted R^2	.806	.781	.790	.829	.846	.798	.713	.855

Note: P-values: p < .05*; p < .01**; p < .001***. Online class, N = 93, traditional class, N = 194. Male = 151, female = 134.