The Chalkboard Versus the Avatar: Comparing the Effectiveness of Online and In-class Courses

Kelly Bergstrand[1] and Scott V. Savage[2]

Abstract

Increasingly, colleges and universities are relying on fully online classes to teach students. This article investigates how students evaluate online courses in comparison to more traditional face-to-face courses. Data come from undergraduate student evaluations of 118 sociology courses, and results of a series of hierarchical linear models indicate that students feel they have learned less in online courses, believe they are treated with more respect in in-class courses, and rate online courses less highly than in-class courses. Findings also suggest that the negative effects of teaching online are not universal for instructors, as the switch to online classes actually results in better evaluations for teachers who typically perform poorly in the classroom. These findings caution against the broad use of online sociology classes as a strategy for coping with increasing enrollments and shrinking budgets and suggest educators should select the course format that best complements their teaching strengths and skills.

Keywords

online teaching, student evaluations of instructors, learning outcomes

Increasingly, instructors in higher education are teaching classes that are entirely Web-based as colleges and universities use online courses to deal with larger enrollments and smaller budgets (Finkelstein and Scholz 2000). A survey of 2,500 U.S. colleges and universities found that 5.6 million students were enrolled in at least one online class (Allen and Seaman 2010). Despite the proliferation of online courses, universities and researchers have only begun to investigate systematically what they can and cannot teach effectively online (Coughan 2012; Jaffee 1997; Van Gundy et al. 2006).

This study adds to the existing research on the effectiveness of online classes by comparing students' evaluations of online sociology classes to more traditional, face-to-face classes. Relying on a unique data set of aggregate student assessments for 118 college sociology courses at a large public university in the Southwest, we find that students typically evaluate online teachers as less effective than teachers in face-to-face classes, view online teachers as less respectful, indicate that they learn less in online classes, and rate online classes more negatively than face-to-face classes. We also find that these effects are not uniform across instructors, indicating that some instructors are better at teaching Web-delivered courses while others excel

[1]University of Arizona, Tucson, AZ, USA
[2]University of California, Riverside, CA, USA

Corresponding Author:
Kelly Bergstrand, Department of Sociology, University of Arizona, Social Sciences Building Room 400, 1145 E. South Campus Dr., Tucson, AZ 85721, USA.
Email: kellyjb@email.arizona.edu

in the traditional classroom environment. These findings inform discussion about how mode of course delivery matters for teaching sociology and is a topic warranting both further scholarly investigation and scrutiny in implementation.

LITERATURE REVIEW

Student Evaluations: Advantages and Shortcomings

Student evaluations of teachers and courses are a fixture of higher education in the United States, with studies reporting that more than 90 percent of schools use them (e.g., Simpson and Siguaw 2000). As such, they provide unique insight into students' perceptions of the educational experience and teacher effectiveness. Further, student evaluations are an institutionalized component of judging college instructors and often weigh heavily in tenure, promotion, and pay decisions (Murray 1984).

Nevertheless, student evaluations have been criticized for being biased by factors unrelated to learning; for example, expectations of good grades can increase evaluations of an instructor (e.g., Ewing 2012; Johnson 2003), and instructors rated as likable receive a boost in ratings of teaching ability, even when associated with a decrease in student perceptions of learning (Delucchi 2000). Researchers have also studied the effects of instructors' traits on student evaluations. Generally, they have not found substantial evaluation differences by gender, although when differences exist they tend to favor male instructors (Wachtel 1998). The effects of instructors' race or ethnicity on evaluations has been surprisingly understudied (Wachtel 1998), although Shapiro (1990) found that any differences between white and nonwhite instructors disappeared after controlling for other variables. Additionally, attributes of the course, such as class size, may affect evaluations, with larger classes receiving lower evaluations (Feldman 1984). Similarly, student perceptions of course difficulty can affect evaluations, with courses perceived as very easy or very difficult receiving lower evaluations (Centra 2003). Finally, low response rates for course evaluations have the potential to introduce bias by not adequately representing students' sentiments. This might be

particularly relevant for online courses, as one study found that online evaluations had lower response rates than did in-class evaluations, although the course evaluations still had comparable results (Stowell, Addison, and Smith 2012).

Despite these concerns, numerous reviews have demonstrated that student evaluations are generally reliable and valid instruments for measuring learning outcomes (d'Apollonia and Abrami 1997; Marsh 1987; Marsh and Roche 1997; Seldin 1993). For example, d'Apollonia and Abrami (1997) found that almost half of the variation in student learning can be explained by student perceptions of teaching effectiveness. Scholars also defend student ratings by noting that other methods of evaluating teaching, such as classroom visitations and self-ratings, are beset with their own problems (Aleamoni 1981). In fact, research has shown that students' ratings have greater validity coefficients than ratings from colleagues, trained observers, or self-reports (Howard, Conway, and Maxwell 1985). For these reasons, many instructors and administrators support using student ratings to evaluate instruction (Seldin 1993).

Despite the potential shortcomings of student evaluations and the fact that they only speak to actual teaching effectiveness to the extent that student perceptions reflect reality, they do provide insight into instruction quality and are important indicators of students' satisfaction with their education. Whether students believe they are receiving a quality education can affect the success of sociology departments, such as their ability to attract and retain students. Scholars recognize the importance of public opinion for the survival of any profession (e.g., Abbott 1988). For a profession like sociology, one of the primary ways to influence public opinion is through teaching. If switching to online classes hurts our ability to do this, we hurt the profession.

Comparing Web-delivered and Traditional Courses

Web courses, defined as "asynchronous computer-mediated courses" (Schulte 2004:6), lack direct, face-to-face interaction between instructors and students, making the fully online class qualitatively different from the traditional classroom

setting. Proponents of online classes contend that fully online courses cater to the unique experiences of many young adults living in an increasingly global economy and thus might be an effective medium for teaching. Little, Titarenko, and Bergelson (2005), for instance, described how online classes can foster cross-cultural learning by connecting students all over the globe. Online classes also provide platforms for students to engage course material on their own time, which might encourage participation and learning (Schulte 2004). This might be particularly true for students who have difficulties attending a traditionally scheduled class because of work responsibilities, family obligations, or chronic illness. In this way, online classes make higher education accessible for students who might otherwise be unable to attend college.[1] Thus, as Clark-Ibanez and Scott (2008) have noted, many students like online classes for the flexibility and anonymity they afford them and even the savings they experience from not having to pay for parking and gasoline. An affinity for the online format might in turn lead students to rate such classes more positively than traditional classroom-based courses.

Opponents of online classes disagree, worrying that online classes place too much of the burden on students and may adversely affect education quality. Students have more discretion over how they manage their time. While such flexibility is typically viewed as a benefit, it could be a drawback if students fall behind on course material. Cramming for tests is a notoriously bad method for learning (Bloom and Shuell 1981), and if online courses breed this type of preparatory behavior, students may view them as less effective. The lack of face-to-face interaction with the instructor also may negatively affect how students think about their online instructors and the education they receive. Onwuegbuzie and his colleagues (2007:146) noted that college students often identify "the interpersonal context as the most important indicator of effective instruction." This is a problem for online classes, as emotions are more difficult to express electronically. Students may feel that written feedback or recorded lectures lack the enthusiasm or substance of verbal feedback or real-time lectures. These potential shortcomings could result in lower student evaluations of courses and teachers.

The disagreement between proponents and opponents of online education has spawned research on the effects of online classes. Results are decidedly mixed. A number of studies have found no significant difference between traditional classes and Web-based courses in student achievement (e.g., van Schaik, Barker, and Beckstrand 2003; Waschull 2001). Still, others have indicated that they do differ, with some studies finding students in online classes perform better (e.g., Tucker 2001) and some finding the opposite (e.g., Urtel 2008). Meta-analyses have confirmed these mixed results, consistently finding that quasi-experimental studies on the whole show no significant difference between the two formats (Means et al. 2009; Sitzmann et al. 2006), but that this is in part due to the great variability in findings across studies (Bernard et al. 2004).

Closer inspection of these studies helps explain the discrepant results. A majority of studies comparing online to more traditional classes were quasi-experimental where one course was taught in both an online and in-class format (Logan, Augustyniak, and Rees 2002; Tucker 2001; Urtel 2008; van Schaik et al. 2003; Waschull 2001). Several of these studies had small sample sizes, with classes as small as 11 students (Logan et al. 2002). Additionally, the author of the study and the instructor were frequently the same person, begging the question of whether instructors introduced bias into the studies by altering their teaching strategies, consciously or not. For example, by artificially making online and in-class courses equivalent, these studies could be missing out on key differences that emerge between the two forms in practice, such as if instructors design syllabi differently depending on the mode of course delivery. Lastly, by limiting the study to one instructor, these studies were unable to evaluate other relevant factors and traits that vary across instructors and could affect course evaluations. On those occasions when researchers have relied on other methodological approaches to study the effects of online classes, new problems arise. For instance, Kelly, Ponton, and Rovai's (2007) study compared Web courses to traditional courses without controlling for other variables that may affect student evaluations, like course difficulty and course size. Similarly, Carle's (2009) study comparing online

and in-class courses also did not control for these two variables. Because student perceptions of course difficulty and course size can affect teaching evaluations (Centra 2003; Feldman 1984), failing to control for these variables may conceal the true effect of course format.

Thus, our research fills a unique niche by examining multiple courses and instructors, none of whom taught under the assumption they would be studied. Moreover, this research incorporates a number of important contextual variables to the analysis, such as class size, student perceptions of course difficulty, and instructor traits. Finally, our study provides a broad portrait of the effects of online courses and traditional courses. Although a number of studies have relied on the quasi-experimental design to compare traditional and online sociology courses (e.g., Driscoll et al. 2012), we are unaware of any study that compares the broad implementation of online courses across an entire sociology department. This is particularly important given the mixed results of previous studies, leaving an open question as to whether online or traditional courses are perceived as more effective. As such, our study addresses the practical issue of what the wide-scale use of online teaching means for students in sociology.

Ongoing debates over the utility of online courses also tend to focus on the course format at the expense of the professionals who teach them. This omission suggests a uniform effect of online classes on teaching effectiveness, but such a view may be inaccurate. Perhaps whether the shift to an online format has a positive or negative effect on teacher effectiveness depends on the teacher. Casual observation tells us that not all teachers are the same. While some thrive in front of the classroom, others struggle with presenting material and engaging in discussion. Online teaching, therefore, may have different effects for different teachers, with some instructors having communication and performance abilities that allow them to excel in classroom settings and others having skill sets better suited for Web-delivered courses, such as strong writing skills.

In addition to assessing perceived teaching effectiveness, we also address whether students felt that they were treated with respect in the course. While studies have focused on the effects of mode of course delivery on evaluations of teaching effectiveness or learning outcomes, rarely do they assess whether an online format might negatively affect interpersonal interactions between students and instructors. Because online courses do not include face-to-face interaction and primarily rely on electronic, written communications that lack facial and vocal cues that can convey sympathy or emotion, it is possible that students will perceive online instructors as being less respectful toward, or concerned about, students. Thus, in our analyses we examine whether online classes affect whether students believe they were treated with respect, a measure that has been found to be part of a valid construct assessing how students perceive interactions with instructors (Roberts and Clifton 1992). This is supported by factor analyses and undergraduate perceptions that have identified a "respect for student" construct emerging out of whether instructors are polite, sympathetic, helpful, accessible, and hold generally favorable attitudes toward students (Shank, Walker, and Hayes 1996). Consequently, a course evaluation question on respect can reveal whether students perceive their interactions with instructors to be negative and may tap into whether students view instructors as unfair, discriminatory, sarcastic, disdainful, or patronizing toward students. Examining this is important because favorable student evaluations of instructors' respect or concern for students are positively correlated with student achievement and learning in courses (Feldman 1989).

This review of the literature on Web courses reveals a number of shortcomings that indicate how little we know about what their implementation means for student evaluations in sociology. By comparing student evaluations of 118 undergraduate sociology courses taught at a large public research university, we examine whether students deem online classes to be as effective as more traditional in-person classes. We also examine whether the shift to online classes affects some instructors more than others. Our results are particularly noteworthy because previous studies have linked student evaluations to actual learning outcomes (d'Apollonia and Abrami 1997).

DATA AND METHODS

To compare online to in-class courses, we rely on teacher-course evaluation data collected and published online by a large public research university in the Southwest, with more than 400 sociology majors. Because course evaluation data are publicly available to the larger university community, we downloaded it to create a data set, with the approval of the university's Institutional Review Board. Our data set consists of aggregate undergraduate student course evaluations for sociology courses taught between the fall of 2009 and the summer of 2011. The sample only includes evaluations of courses taught by graduate student instructors, as no faculty members taught online courses. The data do not provide information about individual undergraduate students, so we do not control for individual student characteristics or for the possibility that certain types of students are self-selecting into online courses, which could affect evaluations. Our analytic sample includes aggregate evaluations for 118 sociology courses taught by 21 instructors, with an average of 5.6 course evaluations per instructor.[2]

In our sample, the majority of instructors were white ($N = 18$), and there were 8 male instructors and 13 female instructors. On average, the instructors had been enrolled in the graduate program for approximately 5.5 years. Graduate students must complete both a master's degree and a course on teaching before they can teach classes. The course on teaching provides a general overview of undergraduate instruction and a section on Web-delivered instruction; however, this class is not specifically designed for training in online teaching methods. There are also university resources available to help with the technical aspects of designing and implementing online courses, such as online instructional support liaisons.

Graduate instructors teach undergraduate courses at all levels (100–400) and can choose the format of courses (online or in-class). They are assigned to teach courses based on the needs of the department as well as the expertise and interests of the instructor. Graduate instructors have freedom in designing and teaching their courses; however, all new syllabi for courses must be approved by the department, and faculty members periodically observe classes to evaluate graduate student instruction. Regardless of whether a course is taught completely online or in the classroom, graduate student instructors typically rely on the Desire to Learn (D2L) Web platform. D2L has a number of important functionalities, including a course homepage, discussion pages, chat rooms, online quizzes, drop boxes, and content pages where instructors can post video lectures, Power-Point slides, and written memos. We do not have survey data on the online teaching practices of graduate students; however, common practices include posting lecture memos or PowerPoint slides to the course Web site and utilizing discussion boards. It is rare for an instructor to provide video lectures or live lectures through videoconferencing, although some instructors do provide audio lectures.

Dependent Variables

The four dependent variables for this study are average undergraduate perceptions of the amount learned in the course, the teaching effectiveness of the instructor, the overall rating of the course, and whether the student felt he or she was treated with respect. These come from questions on course evaluation forms administered by the university. The unit of analysis is the course, and each of these four variables is measured at the aggregate course level. To do so, for each dependent variable we calculate a class average by multiplying the number of students who selected an answer choice by the relevant value assigned to the category (e.g., 1–5) and then dividing that sum by the total number of survey respondents.

The dependent variables were measured using students' responses to the following questions and prompts: "How much do you feel you have learned in this course?" "What is your overall rating of this course?" "What is your overall rating of this instructor's teaching effectiveness?" and "I was treated with respect in this class." The answer choices for each variable ranged from 1 (most negative evaluation) to 5 (most positive evaluation) and were phrased as follows: The amount learned ranged from "almost nothing" to "an exceptional amount," overall course rating varied from "one of the worst" to "one of the best," teaching

effectiveness varied from "almost never effective" to "almost always effective," and whether students felt like they were treated with respect in the class ranged from "strongly disagree" to "strongly agree."

Course-level and Instructor-level Independent Variables

The primary independent variable of interest was whether the course was an in-class course or a Web course (in-class = 0, online = 1). Also of interest was whether some instructors benefited by teaching online classes and whether some instructors were hurt by teaching online classes. To evaluate this, we included two variables. First, we include a measure of each instructor's average effectiveness for in-class courses only. This measure serves as a proxy for instructors' in-class skills, such as public speaking or the ability to foster class discussion. Second, we interacted this with the online variable. This allowed us to evaluate whether strong classroom instructors are disproportionately hurt by teaching in a completely online environment and whether weak classroom instructors are helped by teaching in the online environment.

We also controlled for other variables that might affect course evaluations. These independent variables were the gender of the instructor (0 = female, 1 = male), the number of students enrolled in the course, and the response rate of how many enrolled students returned the course evaluation form. We also included a measure of the perceived difficulty of each course, which came from a question on the course evaluation forms asking, "The difficulty level of the course is," with answers ranging from 1 = "extremely easy" to 5 = "extremely difficult." We then created a course average by summing together the products of the number of students who selected an answer choice by its assigned value and dividing this summation by the total number of respondents.[3]

Hierarchical Linear Models

Because it is likely that the teaching abilities of individual instructors are affecting course evaluations, it is necessary to control for the fact that often multiple course evaluations cluster within the same instructor. Consequently, we employ hierarchical linear modeling (HLM) to analyze the data (Raudenbush and Bryk 2002; Snijders and Bosker 1999). HLM allows researchers to evaluate data that are nested in larger units; for example, students can be nested in schools, employees can be nested in organizations, and families can be nested in neighborhoods. When such clustering occurs, it is likely that observations are no longer independent from one another, violating an assumption in traditional regression methods. HLM accounts for such relations among observations. It also has the ability to incorporate independent variables at both the individual level (e.g., students, employees) as well as at the cluster level (e.g., school, company) as well as include interactions across levels. Specifically, it allows for a multilevel model in which there are level 1 units (in our data, the course evaluations) and level 2 units (in our data, the instructors) and the slope and the intercept from level 1 data are allowed to vary across level 2 units (Raudenbush and Bryk 2002). When an intercept or slope does not vary across units, this is termed a *fixed effect*. Fixed effects represent average values across units and are interpreted as regression coefficients (Snijders 2005).

For this study, we use HLM to assess the effects of course-level variables on course evaluations while also accounting for any potential effects from instructor-level variables. We use a full maximum likelihood estimation and conduct the analyses using Stata software. The level 1 variables include the predictors relevant to course evaluations, such as whether it is online or in-class, the course evaluation response rate, the number of students enrolled in the class, and the perceived difficulty of the course by students. The level 2 variables pertain to characteristics of the instructors, such as the gender of the instructor.[4] Of the level 1 variables, we expect the effect of course format and course difficulty to vary depending on the instructor. While instructors have less control over the number of students enrolled in the class, the evaluation response rate, or instructors' gender, instructors can choose to teach online or in-class or to make their courses more or less difficult. Thus, we include course mode of delivery and course difficulty as random effects in the model.

An intraclass correlation coefficient (ICC) measures the proportion of variance in the outcome that is between units (Raudenbush and Bryk 2002). The ICC values for each dependent variable when grouped by instructor and with no independent variables included in the model are as follows: overall course rating, .204; amount learned, .159; teaching effectiveness, .345; and respect, .173. These ICCs indicate that clustering is evident and that it is appropriate to use HLM. Additionally, a conservative likelihood ratio test comparing the multilevel model to linear regression was significant below the .05 level for most models, indicating HLM is the preferred analytical method.

Differences in Online and Traditional Course Evaluations

Our first set of models assesses the effects of course mode of delivery, the gender of the instructor, the number of enrolled students, the response rate, and the perceived difficulty of the course on each of the four dependent variables (teaching effectiveness, overall course evaluation, amount learned, and treated with respect), grouped by instructor. We initially included course format and perceived difficulty as random effect variables, but indicators of model fit and between-instructor variance revealed that only course format should be estimated as a random effect.

Our second set of models tests the hypothesis that the perceived effectiveness of online courses depends on the skill sets of individual instructors. For these models we add the variable that assesses instructors' average in-class teaching effectiveness and an interaction term between course format and this variable. To facilitate the interpretation of interaction effects, the in-class effectiveness variable is grand mean centered, such that each instructor's in-class effectiveness score is subtracted from the mean of all instructors' in-class effectiveness scores. Because instructors might vary in terms of their preferences for online courses, course difficulty, and the effectiveness of in-class teaching methods, we allow these effects to have a random component. However, as in the first set of models, only the course format variable proved to have a significant random component and was the only random effect included in final models.

RESULTS

Comparing Online Courses to Traditional Courses

Undergraduate students tend to evaluate online courses more negatively than in-class courses (see Table 1). When providing overall evaluations of a course, students gave online courses significantly lower ratings than traditional courses ($\beta = -.238$, $p = .032$). They also indicated that they learned significantly less in online courses ($\beta = -.232$, $p = .022$) and felt that they were treated with less respect in online courses ($\beta = -.244$, $p = .002$). Results further show that online courses have a marginally significant negative effect on undergraduate perceptions of teaching effectiveness ($\beta = -.191$, $p = .076$). Thus, moving from an in-class format to an online format results in a change of approximately $-.20$ for course evaluation outcome variables that scale from 1 to 5. These coefficients are not large; however, the standard deviations of the means of the outcome variables are fairly small, between .30 and .50, indicating that many instructors have similar scores on the dependent variables. Therefore, the negative effect of online course delivery may in fact be a relatively substantial effect.

In addition, we found several other factors affected course evaluations. Larger class sizes and increased perceptions of course difficulty were associated with more negative course evaluations, and higher response rates were associated with more positive course evaluations. Perceptions of course difficulty had significant negative effects for three of the dependent variables: teaching effectiveness ($\beta = -.260$, $p = .015$), the overall course rating ($\beta = -.377$, $p = .001$), and the respect variable ($\beta = -.244$, $p = .000$). Class size negatively affected both teaching effectiveness ($\beta = -.004$, $p = .011$) and the overall rating of a course ($\beta = -.003$, $p = .092$). The coefficients for class size are close to zero, but given that online and in-class courses average 42 students, a one unit increase is likely to have a small effect on the dependent variable. Conversely, the response rate had positive effects on teaching effectiveness ($\beta = .457$, $p = .043$) and the overall course rating ($\beta = .462$, $p = .051$). The coefficients for response rate are larger, but this variable ranges between zero and one.

Table 1. Hierarchical Linear Model on Students' Course Evaluations.

	Overall course rating		Amount learned		Teaching effectiveness		Respect	
	β	SE	β	SE	β	SE	β	SE
Fixed effects								
Course evaluation level								
Online	−.238*	.111	−.232*	−.101	−.191^	.108	−.244**	.079
Class size	−.003^	.002	−.002	.002	−.004*	.002	−.001	.001
Response rate	.462^	.236	.107	.226	.457*	.225	.156	.138
Course difficulty	−.377***	.109	−.108	.104	−.260*	.107	−.244***	.063
Instructor level								
Gender	.078	.114	.059	.108	.097	.128	.097	.068
Constant	4.99***	.408	4.32***	.391	5.00***	.397	5.48***	.238
Random effects								
Online	.043	.036	.015	.024	.050	.040	.060	.026
Constant	.016	.021	.022	.018	.037	.028	.000	.000
Residual	.142	.023	.132	.020	.120	.021	.048	.007
Sample size	118		118		118		118	
−2 log likelihood	126.44		114.12		116.93		11.24	
Bayesian Information Criterion	169.38		157.05		159.87		49.40	

Notes: All regression coefficients are unstandardized. Random effects estimates are variances. SE = standard errors.
^$p \leq .10$. *$p \leq .05$. **$p \leq .01$. ***$p \leq .001$ (two-tailed tests).

Thus, moving from a hypothetical response rate of zero, where no one submits a course evaluation form, to one in which all enrolled students fill out a course evaluation form increases evaluations of teaching effectiveness and the overall course rating by approximately .45 on a 1 to 5 scale.

Instructor In-class Effectiveness and Differences in Course Evaluations

Our second set of models tests the hypothesis that the perceived effectiveness of online courses depends on the skill sets of individual instructors. These results are in Table 2. We find a significant interaction between the instructors' average in-class teaching effectiveness and course format for two outcome variables: (1) overall course rating and (2) the amount learned in the class. Likelihood ratio tests for these two variables were significant ($p = .000$ for both), indicating improved model fit when adding the in-class teaching effectiveness variable and the interaction term to the first set of models. In regard to the overall course rating, for an instructor with an average in-class teaching

effectiveness score, online courses are rated more negatively than in-class courses ($\beta = -.322$, $p = .002$). To evaluate the interaction effect, we entered values that would typify instructors with a high in-class effectiveness score, an average effectiveness score, and a low effectiveness score.[5] We find that for teachers with the highest in-class effectiveness score, the effect of online courses is the most strongly negative, while for instructors with the lowest in-class effectiveness score, online courses actually have a positive effect on overall course evaluations.

Moreover, for instructors with an average in-class teaching effectiveness score, students feel that they have learned less in online classes when compared to in-class courses ($\beta = -.303$, $p = .001$). We find that in regard to the amount learned in online courses, the instructors with the highest rates of in-class effectiveness experience the most negative decrease in evaluations, while instructors with low rates of in-class effectiveness experience an increase in positive evaluations. These results suggest that the relative effectiveness of online and in-class courses is not uniform across instructors

Table 2. Hierarchical Linear Modeling with Interaction Effects on Students' Course Evaluations.

	Course rating		Amount learned		Respect	
	β	SE	β	SE	β	SE
Fixed effects						
Course evaluation level						
Online	−.322**	.104	−.303***	.094	−.260***	.080
Class size	−.002	.002	−.002	.002	−.001	.001
Response rate	.118	.210	−.179	.204	.086	.143
Course difficulty	−.296***	.091	−.027	.088	−.234***	.063
Instructor level						
Gender	.038	.093	.024	.085	.095	.068
In-class effectiveness	.802***	.126	.770***	.124	.128	.084
Cross-level interaction						
Online × In-class effectiveness	−.632***	.183	−.624***	.165	−.107	.145
Constant	4.943***	.342	4.238***	.330	5.49***	.235
Random effects						
Online	.069	.036	.037	.023	.063	.027
Constant	.000	.000	.000	.000	.000	.000
Residual	.104	.015	.101	.015	.046	.007
Sample size	118		118		118	
−2 log likelihood	92.54		82.24		8.95	
Bayesian Information Criterion	140.25		129.95		56.66	

Notes: Regression coefficients are unstandardized. Random effects estimates are variances. SE = standard errors.
^$p \leq .10$. *$p \leq .05$. **$p \leq .01$. ***$p \leq .001$ (two-tailed tests).

and is contingent on the individual skill sets of instructors.

DISCUSSION

The use of Web-delivered courses in higher education is becoming more common, with almost 30 percent of all college and university students having enrolled in at least one online course (Allen and Seaman 2010). Some have even speculated that online courses will revolutionize higher education, with a shrinking market for residential colleges and low-cost online programs offered by elite institutions outcompeting other types of educational experiences (Perez-Pena 2012). But what does this shift mean for teaching sociology? Research relying on quasi-experimental methods shows that a well-designed Web-delivered sociology course can be as effective as its more traditional counterpart (e.g., Driscoll et al. 2012). This research, however, does not examine the consequences of large-scale implementation of Web-based

sociology courses across a department. To our knowledge, no other study has looked at more than 100 courses and 20 instructors in the field of sociology to provide a broad portrait of the effect of online courses for sociology. Thus, our study contributes to debate on an important topic in teaching sociology by examining the consequences of the mass implementation of completely online sociology courses.

Our research indicates that undergraduates rate online sociology classes more negatively than in-class courses. Online courses receive lower overall course ratings, and students state that they learn less in online courses. This is worrisome because research reveals a high correlation between student ratings and actual achievement (Abrami, d'Apollonia, and Cohen 1990; d'Apollonia and Abrami 1997). Thus, contrary to previous research (e.g., Kelly et al. 2007; Little et al. 2005), the general implementation of online sociology courses at this university detrimentally affected student perceptions about the quality of education.

Our results also reveal that undergraduates rate online sociology instructors as treating students with less respect. Respect is an important dimension of effective teaching (d'Apollonia and Abrami 1997). Students who feel that their teachers respect them tend to like and perform better in school (Hallinan 2008). While the data preclude our ability to state with certainty why this is the case, a number of possibilities exist. For instance, by separating students from teachers in space, online classes prevent the face-to-face interactions critical to the student-teacher relationship. The written electronic format may also make it difficult for instructors to deliver negative criticism or feedback in a supportive manner, as instructors are unable to soften criticism with facial or vocal expressions. Similarly, it is possible that electronic forms of correspondence may lead to greater confusion if students misinterpret sarcasm or humor or misunderstand messages. Future research should investigate these possibilities.

Another important contribution of this article is that it identifies a potential factor affecting whether online courses are received positively or negatively by students. It is possible that one of the reasons why previous studies on this topic have mixed findings is that the effectiveness of online courses depends on the instructor teaching the course. It makes sense that the amount of time and energy put into an online course by an instructor, as well as the types of resources utilized by an instructor, will impact the effectiveness of Web-delivered courses. Our research also suggests that the skill sets and traits of instructors could make some individuals better suited to teach online courses. Results show that instructors with above-average levels of in-class effectiveness experience the greatest drop in perceptions regarding overall course ratings and the amount learned in the course. Instructors who excel in methods associated with in-class course presentation, such as public speaking or generating class discussion, may not have been able to recreate effectively these skills in an online environment. Conversely, instructors with the lowest ratings of in-class effectiveness actually experienced a positive boost in course evaluations from the online setting. Thus, the finding that online courses have a negative

effect on course evaluations must be conditioned by an understanding that different instructors may be better suited for different forums and methods of teaching.

Before instructors agree to teach online courses, they should think carefully about what their strengths and weaknesses are and whether the shift to the online format will enhance or detract from their strongest instructional abilities. In-class courses reward public speaking skills, an ability to respond quickly and orally to student questions, and a confidence in social settings. Online courses remove the pressures of public speaking and allow for more time to formulate responses to student questions. Additionally, online courses often require more writing, such as written class notes, blogs, or lectures, as well as responses to discussion boards and to students through e-mail. This could create a distinct advantage to instructors with strong and clear writing skills. For these reasons, we caution against concluding that Web courses universally result in more negative student evaluations. For some instructors, they offer a viable alternative.

The aforementioned findings hold even when controlling for factors other than the mode of delivery thought to affect course evaluations. Consideration of these control variables points to several interesting results. For example, courses that students deem more difficult tended to receive lower evaluations, indicating a negative reaction to more challenging courses. The first set of models also suggests that increased class size may have a small negative effect on course evaluations; this could reflect the increased difficulty instructors experience in trying to give quality feedback and individual attention to students in large classes. Lastly, there is some evidence that an increased response rate increases positive course evaluations, perhaps reflecting a trend in which students who like a course are more willing to take the time to fill out a course evaluation form than students who are neutral toward or dislike the course.

Despite the merits of this research, we recognize its limitations. This research was performed at a large public research university, so it is possible that other types of institutions, such as liberal arts colleges or community colleges, will have different

experiences with Web-delivered courses. We rely on students' interpretations of the course evaluation questions, which could vary. Also, we do not have data on the teaching practices of instructors, students' characteristics, or learning outcomes, all of which would be useful for comparing online and in-class courses and suggest avenues for future research. Moreover, our sample only includes graduate student instructors. Research suggests, however, that teaching evaluations for an instructor do not change significantly over time, indicating that new instructors do not differ dramatically from more advanced instructors (Marsh 2007). Thus, while we acknowledge the possibility that the results of this study might differ for faculty members—and indeed could even be more pronounced if faculty members have developed instructional skills better suited for in-class courses than Web-delivered courses—it is also reasonable to expect that the results of this study are applicable across both types of instructors.

CONCLUSION

While some have argued that undergraduate students prefer the self-paced style of Internet courses and the ability to use novel forms of technology, our study indicates that at least in terms of evaluations of the course overall and the amount learned in the course, online sociology classes are being rated by undergraduates as less effective when compared to in-class courses. Additionally, this negative evaluation extends to ratings of the instructor, with undergraduates rating online instructors as being less effective teachers and as treating them with less respect than do their in-class counterparts. Our study, then, cautions against the broad use of online sociology classes as a strategy for coping with increasing student enrollment or budget constraints. Web-based courses do not have to come at the expense of student evaluations, but this requires technological infrastructure as well as motivated instructors who use Web-based courses to provide unique learning opportunities unavailable in more traditional course settings (Little et al. 2005). Furthermore, as our findings suggest, some instructors may even benefit from a move to online courses. Consequently, it is likely that online courses are not universally "superior" or "inferior"

to traditional courses but depend on the training and resources provided by institutions, as well as on the decisions and teaching strategies of instructors. Thus, administrators and instructors at institutions of higher education should continue to strive to ensure that undergraduate students enrolled in online courses receive the same quality of education as their peers in traditional classroom environments.

ACKNOWLEDGMENT

NOTES

Reviewers for this manuscript were, in alphabetical order, Melinda Messineo and Daphne Pedersen.

Contributions of the authors are equal; authors are listed in alphabetical order.

1. The caveat here is that online classes require access to a computer and the Internet, which may be a barrier for students of low socioeconomic status.

2. Two cases had incorrect data on one of the variables and were dropped from the analysis, resulting in the total number of observations changing from 120 to 118. We remove from the sample required theory, methods, and statistics courses because they are not taught online. We also limit the sample to instructors who have taught both in-class and online.

3. Sensitivity analyses reveal that on average students were only slightly more likely to view upper-level courses as more difficult than introductory-level courses (3.4 to 3.2), suggesting that variations in student perceptions of course difficulty have more to do with the instructor than with course level.

4. Initial models included instructors' race, but inclusion of this variable did not improve model fit nor was it significant. It is also possible that instructors are themselves also grouped into course topical areas; however, our analyses did not show evidence that a third level of clustering by course type was having a significant effect on the analysis.

5. For these calculations, we entered 0 or 1 for online, and for in-class effectiveness, which was mean centered, the minimum of our data to represent a low score (–1.77), the average (0), and the maximum to represent a high score (0.53). We then entered these values into an equation that retained the coefficients of relevant variables pertaining to online format,

in-class effectiveness, and the interaction term, as well as any other relevant significant variables (for these we entered the variable's mean) and the constant. For both overall course rating and amount learned, the highest in-class effectiveness instructors' scores dropped ≈ -0.65, the average in-class effectiveness instructors' scores dropped ≈ -0.30, and the lowest in-class effectiveness instructors' scores increased 0.8 when changing from an in-class to online format.

REFERENCES

Abbott, Andrew. 1988. *The System of Professions: An Essay on the Division of Expert Labor*. Chicago: University of Chicago Press.

Abrami, Philip C., Sylvia d'Apollonia, and Peter A. Cohen. 1990. "Validity of Student Ratings of Instruction: What We Know and What We Do Not." *Journal of Educational Psychology* 82(2):219–31.

Aleamoni, Lawrence M. 1981. "Student Ratings of Instruction." Pp. 110–45 in *Handbook of Teacher Evaluation*, edited by J. Millman. Beverly Hills, CA: Sage.

Allen, I. Elaine and Jeff Seaman. 2010. *Class Differences: Online Education in the United States, 2010*. Retrieved July 25, 2012 (http://sloanconsortium.org/publications/survey/class_differences).

Bernard, Robert M., Philip C. Abrami, Yiping Lou, Evgueni Borokhovski, Anne Wade, Lori Wozney, Peter Andrew Wallet, Manon Fiset, and Binru Huang. 2004. "How Does Distance Education Compare with Classroom Instruction? A Meta-analysis of the Empirical Literature." *Review of Educational Research* 74(3):379–439.

Bloom, Kristine C. and Thomas J. Shuell. 1981. "Effects of Massed and Distributed Practice on the Learning and Retention of Second-language Vocabulary." *Journal of Educational Research* 74(4):245–48.

Carle, Adam 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 and Education* 53(2):429–35.

Centra, John A. 2003. "Will Teachers Receive Higher Student Evaluations by Giving Higher Grades and Less Course Work?" *Research in Higher Education* 44(5):495–518.

Clark-Ibanez, Marisol and Linda Scott. 2008. "Learning to Teach Online." *Teaching Sociology* 36(1):34–41.

Coughan, Sean. 2012. "MIT Launches Free Online 'Fully Automated' Course." *BBC News*. Retrieved March 5, 2012 (http://www.bbc.co.uk/news/education-17012968).

d'Apollonia, Sylvia and Philip C. Abrami. 1997. "Navigating Student Ratings of Instruction." *American Psychologist* 52(11):1198–208.

Delucchi, Michael. 2000. "Don't Worry, Be Happy: Instructor Likability, Student Perceptions of Learning, and Teacher Ratings in Upper-level Sociology Courses." *Teaching Sociology* 28(3):220–31.

Driscoll, Adam, Karla Jicha, Andrea N. Hunt, Lisa Tichavsky, and Gretchen Thompson. 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(4):312–31.

Ewing, Andrew M. 2012. "Estimating the Impact of Relative Expected Grade on Student Evaluations of Teachers." *Economics of Education Review* 31(1):141–54.

Feldman, Kenneth. 1984. "Class Size and College Students' Evaluations of Teachers and Courses: A Closer Look." *Research in Higher Education* 21(1):45–116.

Feldman, Kenneth A. 1989. "The Association between Student Ratings of Specific Instructional Dimensions and Student Achievement: Refining and Extending the Synthesis of Data from Multisection Validity Studies." *Research in Higher Education* 30(6):583–645.

Finkelstein, Martin J. and Bernhard W. Scholz. 2000. "What Do We Know about Information Technology and the Cost of Collegiate Teaching and Learning?" Pp. 3–34 in *Dollars, Distance, and Online Education: The New Economics of College Teaching and Learning*, edited by M. J. Finkelstein, C. Frances, F. I. Jewett, and B. W. Scholz. Phoenix, AZ: The Oryx Press.

Hallinan, Maureen T. 2008. "Teacher Influences on Students' Attachment to School." *Sociology of Education* 81(3):271–83.

Howard, George S., Christine G. Conway, and Scott E. Maxwell. 1985. "Construct Validity of Measures of College Teaching Effectiveness." *Journal of Educational Psychology* 77(2):187–96.

Jaffee, David. 1997. "Asynchronous Learning: Technology and Pedagogical Strategy in a Distance Learning Course." *Teaching Sociology* 25(4):262–77.

Johnson, Valen E. 2003. *Grade Inflation: A Crisis in College Education*. New York: Springer-Verlag.

Kelly, Henry F., Michael Ponton, and Alfred P. Rovai. 2007. "A Comparison of Student Evaluations of Teaching between Online and Face to Face Courses." *Internet and Higher Education* 10(2):89–101.

Little, Craig B., Larissa Titarenko, and Mira Bergelson. 2005. "Creating a Successful International Distance-learning Classroom." *Teaching Sociology* 33(4):355–70.

Logan, Elisabeth, Rebecca Augustyniak, and Alison Rees. 2002. "Distance Education as Different Education: A Student-centered Investigation of Distance Learning Experience." *Journal of Education for Library and Information Science* 43(1):32–42.

Marsh, Herbert W. 1987. "Students' Evaluations of University Teaching: Research Findings, Methodological Issues, and Directions for Future Research." *International Journal of Educational Research* 11(3):253–388.

Marsh, Herbert W. 2007. "Do University Teachers Become More Effective with Experience? A Multilevel Growth Model of Students' Evaluations of Teaching over 13 Years." *Journal of Educational Psychology* 99:775–90.

Marsh, Herbert W. and Lawrence A. Roche. 1997. "Making Students' Evaluations of Teaching Effectiveness Effective." *American Psychologist* 52(11):1187–97.

Means, Barbara, Yukie Toyama, Robert Murphy, Marianne Bakia, and Karla Jones. 2009. *Evaluation of Evidence-based Practices in Online Learning: A Meta-analysis and Review of Online Learning Studies*. Washington, DC: Office of Planning, Evaluation, and Policy Development, U.S. Department of Education.

Murray, Harry G. 1984. "The Impact of Formative and Summative Evaluation of Teaching in North American Universities." *Assessment and Evaluation in Higher Education* 9(2):117–32.

Onwuegbuzie, Anthony J., Ann E. Witcher, Kathleen M. T. Collins, Janet D. Filer, Cheryl D. Wiedmaier, and Chris W. Moore. 2007. "Students' Perceptions of Characteristics of Effective College Teachers: A Validity Study of a Teaching Evaluation Form Using a Mixed-methods Analysis." *American Educational Research Journal* 44(1):113–60.

Perez-Pena, Richard. 2012. "Top Universities Test the Online Appeal of Free." *The New York Times*, July 17. Retrieved August 10, 2012 (http://www.nytimes.com/2012/07/18/education/top-universities-test-the-online-appeal-of-free.html).

Raudenbush, Stephen W. and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd ed. Thousand Oaks, CA: Sage.

Roberts, Lance W. and Rodney A. Clifton. 1992. "Measuring the Affective Quality of Life of University Students: The Validation of an Instrument." *Social Indicators Research* 27(2):113–37.

Schulte, Aileen. 2004. "The Development of an Asynchronous Computer-mediated Course: Observations on How to Promote Interactivity." *College Teaching* 52(1):6–10.

Seldin, Peter. 1993. "When Students Rate Professors." *The Chronicle of Higher Education*, July 21. Retrieved August 2, 2012 (http://chronicle.com/article/When-Students-Rate-Professors/73120/).

Shank, Matthew D., Mary Walker, and Thomas Hayes. 1996. "Understanding Professional Service Expectations: Do We Know What Our Students Expect in a Quality Education?" *Journal of Professional Services Marketing* 13(1):71–89.

Shapiro, Gary. 1990. "Effect of Instructor and Class Characteristics on Students' Class Evaluations." *Research in Higher Education* 31(2):135–48.

Simpson, Penny and Judy Siguaw. 2000. "Student Evaluations of Teaching: An Exploratory Study of the Faculty Response." *Journal of Marketing Education* 22(3):199–213.

Sitzmann, Traci, Kurt Kraiger, David Steward, and Robert Wisher. 2006. "The Comparative Effectiveness of Web-based and Classroom Instruction: A Meta-analysis." *Personnel Psychology* 59(3):623–64.

Snijders, Tom A. B. 2005. "Fixed and Random Effects." Pp. 664–65 in *Encyclopedia of Statistics in Behavioral Science*, edited by B. S. Everitt and D. C. Howell. New York: Wiley.

Snijders, Tom A. B. and Roel J. Bosker. 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Thousand Oaks, CA: Sage.

Stowell, Jeffrey R., William E. Addison, and Jennifer L. Smith. 2012. "Comparison of Online and Classroom-based Student Evaluations of Instruction." *Assessment and Evaluation in Higher Education* 37(4):465–73.

Tucker, Sheila. 2001. "Distance Education: Better, Worse, or as Good as Traditional Education?" *Journal of Distance Learning Administration* 4(4). Retrieved July 15, 2012 (http://www.westga.edu/~distance/ojdla/winter44/tucker44.html).

Urtel, Mark G. 2008. "Assessing Academic Performance between Traditional and Distance Education Course Formats." *Technology and Society* 11(1):322–30.

Van Gundy, Karen, Beth A. Morton, Hope Q. Liu, and Jennifer Kline. 2006. "Effects of Web Based Instruction on Math Anxiety, the Sense of Mastery, and Global Self-esteem: A Quasi-experimental Study of Undergraduate Statistics Students." *Teaching Sociology* 34(4):370–88.

van Schaik, Paul, Philip Barker, and Scott Beckstrand. 2003. "A Comparison of On-campus and Online Course Delivery Methods in Southern Nevada." *Innovations in Education and Teaching International* 40(1):5–15.

Waschull, Stefanie. 2001. "The Online Delivery of Psychology Courses: Attrition, Performance, and Evaluation." *Teaching of Psychology* 28(2):143–47.

Wachtel, Howard K.1998. "Student Evaluation of College Teaching Effectiveness: A Brief Review." *Assessment and Evaluation in Higher Education* 23(2):191–212.

AUTHOR BIOGRAPHIES

Kelly Bergstrand is a PhD candidate at the University of Arizona and specializes in social movements, environmental sociology, and social psychology. Her current research projects include examining the role of grievances in mobilization, investigating activism on the U.S.-Mexico border, and mapping the community impact of environmental disasters.

Scott V. Savage is an assistant professor at the University of California, Riverside. He is a social psychologist and organizational theorist interested in issues of trust, commitment, and influence. His current research investigates job retention and occupational mobility as well as extensions of status characteristics theory.