A Comparison of Student Attitudes, Statistical Reasoning, Performance, and Perceptions for Web-augmented Traditional, Fully Online, and Flipped Sections of a Statistical Literacy Class

Ellen Gundlach
Purdue University

K. Andrew R. Richards
Northern Illinois University

David Nelson
Purdue University

Chantal Levesque-Bristol
Purdue University

Journal of Statistics Education Volume 23, Number 1 (2015),
www.amstat.org/publications/jse/v23n1/gundlach.pdf

Key Words: Attitudes toward statistics; Blended; SRA; SATS-36; Flipped classroom; Statistical reasoning; Technology-enabled statistics education; Statistical literacy.

Abstract

Web-augmented traditional lecture, fully online, and flipped sections, all taught by the same instructor with the same course schedule, assignments, and exams in the same semester, were compared with regards to student attitudes; statistical reasoning; performance on common exams, homework, and projects; and perceptions of the course and instructor. The Survey of Attitudes Toward Statistics-36 (SATS-36) instrument and eight questions from the Statistical Reasoning Assessment (SRA) were given both at the beginning and end of the semester to measure change. The students selected their own sections, but the students in the sections were similar demographically, with similar pre-course college grade point averages. The SATS-36 showed increases in affect, cognitive competence, and perceived easiness and decreases in value, interest, and effort from beginning to end of the semester for all sections. Only affect and perceived easiness showed any differences for section, with traditional higher than online on average for both. Results from the SRA questions showed an increase in correct statistical reasoning skills and decrease in misconceptions for all sections over the semester. Traditional students scored higher on average on all three exams, but there were no significant differences between sections on homework, the project, or on university evaluations of the course or instructor. Results are contextualized with prior educational research on course modalities, and proposals for future research are provided.

1. Introduction

A 16-week, undergraduate-level, introductory course on statistical literacy was redesigned to be taught in three different formats (web-augmented traditional, fully online, and flipped) as part of a university-sponsored course redesign initiative which targeted large, introductory, service courses at a research-intensive university in the American Midwest. Redesign is best conceptualized as an iterative process and requires instructors to work carefully and intentionally toward improving their approach to instruction over time (Ramsden 1992; Toohey 1999). Faculty in this initiative were encouraged to evaluate their existing course critically, consider desired learning outcomes, and restructure their course to improve student learning. This process encouraged a shift from instructor-focused pedagogies to those which prioritized the students' needs (Kember 1997). In the last 30 years, the teaching and learning literature has highlighted the importance of involving students in the learning process and attending to their learning preferences (Chickering and Gamson 1987; Trigwell, Prosser, and Taylor 1994; Haidet, Morgan, O'Malley, Moran, and Richards 2004). This approach to redesign also aligns with the Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report's recommendations for introductory statistics classes, which emphasizes conceptual understanding and active learning (Aliaga et al. 2005). For the statistical literacy course, we were interested in comparing student attitudes, statistical reasoning, performance, and perceptions in web-augmented traditional, fully online, and flipped sections that were implemented after the redesign.

1.1 Course Delivery Methods

Online and flipped methods of teaching are becoming increasingly popular alternatives to traditional lecture classes, but the research into which method is most effective has not provided a clear answer (Allen, Seaman, and Garret 2007; Bowen, Chingos, Lack, and Nygren 2012). In the current study, we compared students' opinions and success in web-enhanced traditional, online, and flipped sections of a statistical literacy course. In the absence of clear definitions of the three different delivery methods (Zieffler, Garfield, Alt, Dupuis, Holleque, and Chang 2008), we draw from the literature to define the courses in the context of this study.

A *web-augmented traditional* course "uses web-based technology to facilitate what is essentially a face-to-face course. Uses course management system (CMS) or web pages to post the syllabus and assignments, for example" (Allen, et al. 2007, p. 5). Today's traditional classes rarely involve only lecture; most incorporate pedagogical tools such as small and large group discussion, in-class assignments, formative assessments, and instructional technologies (Corrigan

2013). In the current study, we will abbreviate web-augmented traditional as simply "traditional."

In a *fully online* course, "most or all of the content is delivered online. Typically have no face-to-face meetings" (Allen et al. 2007, p. 5). In contrast, *flipped classes* "are those in which students receive content from technology (i.e., technology-transmitted) and apply knowledge with help from an instructor (i.e. instructor-mediated)" (Margulieux, Bujak, McCracken, and Majerich 2014, p. 8). The course design literature uses the terms blended, flipped, hybrid, and inverted interchangeably, but definitions are beginning to be formalized. Blended is becoming the generalized term for all of these methods. Flipped (synonymous with inverted) is becoming a more specialized term which defines the method of lecture delivery and application of students' knowledge (Strayer 2012; Margulieux et al. 2014). In our course, the delivery method more accurately corresponds to the specialized moniker of "flipped," which is why it appears throughout this paper.

1.2 Research Comparing Course Delivery Methods

The research comparing course delivery methods has been contradictory and inconclusive. While some studies have noted benefits of one method over others, several studies have found no differences, and there is little consistency with regards to a "preferred" method. There exist few studies that examine comparisons of delivery methods for large, introductory courses at public universities, or studies that have applied sufficiently rigorous research design or random assignment of students to classes with different delivery methods (Ward 2004; Parkhurst, et al. 2008a). Only one study (Somenarain, Akkraju, and Gharbaran 2010) was found that utilized all three delivery methods (traditional, online, and flipped), but different instructors taught different sections in the memorization-heavy, medical terminology course.

A well-designed study by Bowen et al. (2012) compared traditional and flipped introductory statistics course sections taught using the machine-guided Carnegie Mellon online course for the flipped section's online instruction at six public universities in seven departments. Each department offered both the traditional and flipped sections, although not necessarily with the same instructor teaching both sections. A unique feature of this study was that students in each department were randomly assigned to the traditional and flipped sections, which removed a common source of bias associated with self-selection. Bowen et al. used the Comprehensive Assessment of Outcomes in First Statistics Course (CAOS; del Mas, Garfield, Ooms, and Chance 2007) for pre- and post-tests to measure statistical reasoning, final exams with common questions, and course pass rates. No significant differences in these reasoning or performance measures were found for flipped and traditional students. There was also no significant difference between the amount of time that flipped and traditional students reported spending on coursework.

When used effectively, traditional course delivery methods can provide an efficient way to teach large numbers of students in introductory-level courses and are also well-suited for introverted students, students who do not prepare reliably for class sessions, or those who are not willing to participate fully in class discussions (Burgan 2006; Walthausen 2013). However, while some evidence has supported the traditional mode of delivery, Shachar and Neumann's (2010) meta-analysis of student academic performance in traditional, online, and flipped courses found that in 70% of the studies, online and flipped (jointly considered in one group) students performed better than traditional students with an overall small effect size (d = 0.247), and the overall effect size steadily increased from 2000-2009. Other researchers have found significant benefits of flipped courses with regards to fostering active learning (Froyd 2007, 2008).

Flipped and blended classrooms have been associated with some documented benefits over traditional and online approaches, such as higher grades, more evidence of learning, greater motivation and engagement, stronger sense of community, better attitudes, more opportunities to integrate active learning methods, and more frequent classroom attendance (Rovai and Jordan 2004; Babb, Stewart, and Johnson 2010; Deslauriers, Schelew, and Weiman 2011; Pearl et al. 2012; Tishkovskaya and Lancaster 2012; McLaughlin et al. 2014). Flipped classes may also include superior "external connections with the material by reference to a variety of experiences and perspectives," and opportunities for both stronger and weaker students to strengthen their skills and knowledge (Giraud 1997, p. 2). Ward's (2004) study of traditional and flipped sections of an introductory statistics course taught by the same instructor with the same assignments showed that students in the flipped section were more positive about the instructor and the presentation of the material.

In other studies, particularly those in statistics education, comparisons of students' academic success across course delivery methods using course grades or CAOS (delMas et al. 2007) have found little difference among sections (Mills and Raju 2011; Bowen et al. 2012). Most of these studies compared traditional sections to either an online or a flipped format, but none have compared all three methods concurrently. It is also important to note that some course delivery methods may lead to greater gains in certain areas, while different delivery methods significantly increase gains in others. For example, Parkhurst et al. (2008b) found that students in a traditional section showed greater improvement in conceptual learning, but students in an online section showed greater improvement in factual learning. A more recent comparison of traditional vs. flipped sections of a statistics course, with the traditional section taught in an earlier semester than the flipped section, showed that the flipped section students performed better on the statistics scale of the Psychology Area Concentration Achievement Test, but not on the other nine scales when this test with all ten scales was taken a year after completing the course (Winquist and Carlson 2014).

The literature related to course delivery method in statistics education is still in its infancy, and further research is necessary to investigate differences across traditional, online, and flipped course sections (Griffith, Adams, Gu, Hart, and Nicholas-Whitehead 2012). This investigation adds to the literature by measuring both student performance and attitude outcomes across traditional, online, and flipped sections of a statistical literacy course all taught by the same instructor in the same semester.

1.3 Student Attitudes Toward Statistics

Numerous studies have demonstrated a connection between student attitudes and success (e.g., DeVaney 2010; Vanhoof, Kuppens, Castro Sotos, Verschaffel, and Onghena 2011; Griffith et al. 2012; Finney and Schraw 2003). Unfortunately, statistics courses do not always have a positive reputation among students who are not majoring in science, technology, engineering, and mathematics (STEM) disciplines. Non-STEM students often have a significant amount of anxiety related to statistics courses, which can negatively affect their performance (Baloglu 2003; Finney and Schraw 2003; Onwuegbuzie and Wilson 2003; Verhoeven 2006; Dempster and McCorry 2009; Vanhoof et al. 2011; Tishkovskaya and Lancaster 2012). As statistics educators, we should be interested in the attitudes our students have about statistics as a field, statistics as a class, and their own confidence in their ability to learn statistics.

The Survey of Attitudes Toward Statistics (SATS; Schau 2003a, b) is one instrument measuring students' attitudes that has consistently demonstrated acceptable levels of validity and reliability (Pearl et al. 2012). Griffin, et al. (2012) note that SATS has "solid theoretical underpinnings…based on a number of popular theories including expectancy value, attribution, social cognition, and goal theories" (pp. 46). Both a 28-item (SATS-28) version with only four subscales and an updated 36-item (SATS-36) version of the SATS instrument have been developed. The SATS-36 instrument has six subscales: *affect* (students' feelings concerning statistics), *cognitive competence* (students' attitudes about their intellectual knowledge and skills when applied to statistics), *value* (students' attitudes about the usefulness, relevance, and worth of statistics in personal and professional life), *difficulty* (students' attitudes about the perceived easiness of statistics as a subject; referred to hereafter as "perceived easiness"), *interest* (students' level of individual interest in statistics), and *effort* (amount of work the student expends to learn statistics).

SATS has been widely adopted as a measure of attitudes toward statistics (e.g., Bond, Perkins, and Ramirez 2012; Ramirez, Schau, and Emmioglu 2012). A meta-analysis of 17 studies examining the correlations between four SATS-28 and SATS-36 components and statistics achievement (primarily student grades) showed larger effect sizes for American students than international students in all components, with the strongest (medium) effect sizes between statistics achievement and affect and cognitive competence in the United States (Emmioglu and Capa-Aydin 2012). Hannigan, Gill, and Leavy (2013) measured students' statistical knowledge and attitudes using both the CAOS concept inventory and the SATS-36, respectively, but found no significant correlations between CAOS test score and any of the six attitude subscales. Some researchers have noted that attitudinal dimensions measured by the instrument may not change markedly from the beginning to the end of one academic semester (Gal and Ginsburg 1994; Gal, Ginsburg, and Schau 1997; Zieffler et al. 2008). A large study of 2,200 students taking the SATS-36 at the beginning and end of 101 introductory statistics service courses across the United States showed an average decrease in the value, interest, and effort subscales but no change in affect, cognitive competence, or perceived easiness (Schau and Emmioglu 2012).

Three recent studies have used the SATS to explore student attitudes in classes taught in the online or flipped formats. DeVaney (2010) used the SATS-28 to compare traditional and online sections of a graduate statistics course taught by different instructors. From pre- to post-semester, the researcher found significant decreases in anxiety and corresponding increases in positive attitudes by students in the online course, but noted no significant change for students in the traditional course. Initial levels of anxiety for online students were significantly higher than for traditional students, but there were no significant differences in the final levels of anxiety between the sections. The traditional students felt that learning statistics was significantly easier than the students in the online section. Overall, only small effect sizes were seen for the changes in SATS-28 subscales over the semester. Gibbs and Tayback (2014) provided SATS-36 scores for their blended introductory statistics course of 38 students using a Massive Open Online Course (MOOC) for the online lectures. Their students showed an average increase in affect and effort. White (2014) compared SATS-36 and CAOS scores for traditional, team-taught, face-to-face sections at two different colleges and a flipped introductory statistics course which also incorporated Powerpoint with voice-over online lectures and resources from a Massive Open Online Course (MOOC). The response rate for CAOS and SATS-36 questions was small (only 8% for SATS-36), and sizeable differences between the traditional and flipped sections were not found.

1.4 Statistical Reasoning

In the statistics education literature, the preferred method for evaluating comprehension is the use of a previously validated and standardized assessment instrument, such as CAOS (delMas et al. 2007), the Statistics Concept Inventory (SCI; Allen 2006), or the Statistical Reasoning Assessment (SRA; Garfield 2003). However, previous researchers note that there is a shortage of high-quality, standardized assessments for measuring statistical literacy (Garfield and Gal 1999; Watson 1997; Gal 2002; Pearl et al. 2012; Tishkovskaya and Lancaster 2012), possibly because low-level statistical literacy classes are not as common as the typical introductory statistics courses, which also emphasize data analysis.

While there is not yet an ideal instrument for measuring statistical reasoning in an introductory statistical literacy course, the SRA may come closest. The SRA contains 20 multiple-choice and choose-all-that-apply questions designed to assess not simply right or wrong answers, but correct reasoning and common misconceptions (Garfield 2003). The questions and possible answers are sorted into eight correct reasoning and eight misconception categories. The SRA has shown appropriate levels of test-retest reliability and content validity (Zieffler et al. 2008), though several studies have failed to find strong correlations between the SRA and measures of course performance, such as exams, or even strong intercorrelations between items (Garfield 2003; Tempelaar, Gijselaers, and van der Loeff 2006).

Scholars have analyzed SRA data by examining both individual correct reasoning and misconception category scores as well as aggregate correct reasoning and aggregate misconception scores (Garfield 2003). To date, no published research has incorporated the SRA at both the beginning and end of a semester as a measure of growth in statistical reasoning ability. Typically, the SRA is administered at the end of the semester (Garfield 2003), but Tempelaar et al. (2006) implemented it at the beginning of the semester to measure preconceptions independent of the introductory-level quantitative methods course.

1.5 Student Performance

Johnson and Kuennen (2006), Zieffler et al. (2008), and Pearl et al. (2012) note that using instructor-created measures of student learning is not ideal because readers have no way of knowing the specific material covered and the required level of understanding in an individual instructor's course. However, exam, homework, and projects can serve as a comparative tool between sections with different delivery methods, especially when the instructor and assessments are consistent across semesters and sections (Johnson and Kuennen 2006; Shachar and Neumann 2010; Emmioglu and Capa-Aydin 2012). Tempelaar et al. (2010) used project, final exam, quiz, and homework grades to measure course performance in their introductory quantitative methods course in conjunction with the SRA and the SATS-28. In order to compensate for the limitations of individual instructor-created measures of performance, the current study gathered performance data from multiple sources. The three exams were written by the instructor, the online homework was written by an outside company, and the online project involved students posting and discussing articles and videos that reflected their own interests.

1.6 Student Perception of the Course and Instructor

Course evaluations are one method for measuring student perceptions towards a course and an instructor. Stark (2013a) highlights problems with course evaluations, including non-response, small sample sizes, use of averages for categorical variables rating quality, failure to account for spread, and the inappropriateness of comparing different courses across campus using the same questions. Stark (2013b) summarizes recent studies that indicate course evaluations, while reliable and consistent among students, may not evaluate teaching effectiveness but instead correlate with student enjoyment and grade expectations. These evaluations can also be biased toward personal characteristics of the instructor (Stark 2013b; Voeten 2013). Nevertheless, course evaluations are still relevant as they are integrated into faculty promotion, retention, and compensation decisions at many institutions of higher education (Gravestock and Gregor-Greenleaf 2008). In the context of this paper, course evaluations were compared in multiple sections for only one instructor; therefore, biases such as those discussed by Stark (2013a, b) and Voeten (2013) should be limited.

1.7 Purpose and Research Questions

To date, no research has been published comparing the attitudes and success of students in traditional, online, and flipped sections of an introductory undergraduate statistics course taught by the same instructor during the same semester with the same assignments and assessments. With the growing popularity of teaching using modalities other than traditional lecture, instructors and students should have more information about how these modalities compare. As such, the purpose of this study was to examine differences in student attitudes and success among traditional, online, and flipped versions of an introductory course in statistical literacy taught by the same instructor during the same semester. Specifically, the following research questions (RQs) guided our inquiry:

- Do changes in attitudes toward statistics tend to vary across sections?
- Do changes in statistical reasoning tend to vary across sections?
- Does student performance tend to vary across sections?
- Do student perceptions of the course and instructor tend to vary across sections?

2. Method

2.1 Overview of the Statistical Literacy Course

The course we studied was a 16-week, undergraduate-level, introductory statistical literacy course at a research-intensive university in the American Midwest. This course placed greater emphasis on narrative interpretation of statistical claims than on calculations and detailed statistical analysis. Students were asked to "explain, judge, evaluate, and make decisions about the information" (Rumsey 2002) and become "educated citizens [who] understand basic statistical concepts and interpret and critically evaluate statistical messages so that they can detect any misuse of statistics by policymakers, physicians and others" (Tishkovskaya and Lancaster 2012, pp. 5-6). The required textbook was *Statistics: Concepts and Controversies* (Moore and Notz 2009). Topics covered in the course included: experimental and sampling design, basic summary statistics and graphing, probability, normal distributions, basic confidence intervals, correlation and regression, and two-way tables. Course materials and activities for all sections were planned using Guidelines for Assessment and Instruction in Statistics Education (GAISE; Aliaga et al. 2005), incorporating active learning, real data and stories, conceptual understanding, technology where appropriate for learning and data analysis, and authentic assessment. The course emphasized discussion of statistical literacy concepts in articles, videos, advertisements, medical advice, and legislation.

The course was offered in traditional, fully online, and flipped sections (See Table 1). In the flipped section, students viewed the online lectures outside of class and met once per week in an innovative, open-design classroom to work in groups on problem-solving tasks and hands-on activities. Attendance in these weekly meetings was required, and the lead instructor and teaching assistant were available during class to work with individual students or groups to answer questions as necessary. As recommended by Utts, Sommer, Acredolo, Maher, and Matthews (2003) and Brame (2013), students in the flipped section took weekly quizzes at the end of their class time to motivate them to learn the material.

Table 1. Overview of the structure of the traditional, online, and flipped sections of the coordinated statistical literacy course in the spring 2013 semester.

	Traditional	**Online**	**Flipped**
Number of students enrolled	330	74	56
Structure	Meeting in large lecture hall with the lead instructor twice per week. Weekly recitations with a teaching assistant.	Everything except exams done online.	Lectures are watched online outside of class. Meeting once per week in an open concept class space for active learning and group work.
Delivery of Course Content	Lectures delivered in face-to-face meetings. Lectures also provided online.	Lectures provided online.	Lectures provided online.
Homework	Perdisco Online Homework system. www.perdisco.com StatsPortal Learning Curve for additional practice.		
Online Discussion Board	Online statistical literacy discussion project.		
Exams	Two midterms and one final – all sections took the same exams together in a large lecture hall on the university campus.		
Quizzes	Given weekly in recitations.	No quizzes.	Given weekly in class.
Class Participation	Classroom response system (i>Clicker) questions in lecture. (http://www1.iclicker.com/)	Check-in free-response surveys on the course website about how the course is going.	Participation in group activities during class.

The traditional and flipped students both had weekly in-class participation activities to give the instructor an opportunity to check on the students' immediate questions and concerns. In lieu of in-class participation activities, the online students were required to complete free-response surveys with 9-15 questions each, four times throughout the semester. These surveys allowed the instructor to monitor student progress and make critical adjustments and announcements. Typical questions from these surveys included: "What will you do to study for the exam next week?" and "What is the toughest topic for this exam? Why? How can I help you with this topic?" These required surveys sought to promote personal, individual conversations with the instructor about the course and any concerns at regular intervals.

While the format of the classes differed, all sections followed the same basic schedule, and used the same textbook, exams, and online homework. Students across all sections were also assigned the same online statistical literacy project, which involved sharing and discussing links to articles and videos related to concepts learned in class. Following recommendations from Jaki (2009), the traditional students were given access to the same online lectures provided to flipped and online students. The in-class traditional lectures used the same Powerpoint slides as the online lectures. Finally, all students had access to approximately 10 office hours per week where they could meet with the instructor and teaching assistants in the department's help center.

2.2 Participants

Participants in this study were undergraduate students who were registered for the statistical literacy course in the spring 2013 semester. Data were collected from the instructor gradebook, an online survey, and the university end-of-semester course evaluation survey. All students who completed the course were included in the instructor gradebook. However, not all students elected to complete the online survey and the end-of-semester course evaluation survey. Available descriptive statistics are provided for the students contributing to each data source.

2.2.1 Instructor Gradebook

A total of 462 students (331 traditional, 75 online, 56 flipped) completed the class. Most of the students were sophomores (N=169; 36.5%) and freshmen (N=147; 31.7%), with fewer holding status as juniors (N=93; 20.2%) and seniors (N=53; 11.5%). The average participant was 20.00 years old (SD=1.95) and there were more females (N=298; 64.5%) than males (N=164; 35.5%). Most students were from the College of Liberal Arts (N=242; 52.6%) and the College of Health and Human Sciences (N=143; 31.1%). Most of the participants were Caucasian (*N*=340; 73.6%) with fewer participants reporting African American (*N*=14; 3.0%), Asian American (*N*=11; 2.4%), Hispanic (*N*=24; 5.2%), Mixed Race (*N*=9; .1.9%), Native American Indian (N=2; .4%), and Other (*N*=64; 13.9%) ethnicities. Most of the students were domestic (N=404; 87.4%) with fewer reporting international status (N=58; 12.6%).

2.2.2 Online Survey

A total of 261 students (193 traditional, 43 online, and 25 flipped) responded to the survey at both pre- and post-survey administration. These numbers translate to a response rate of 58.5% for students in the traditional section, 58.1% of students in the online section, and 47.2% of students in the flipped section. The majority of the students were freshmen (*N*=104; 39.8%) and sophomores (*N*=83; 31.8%), with fewer students classified as juniors (*N*=44; 16.5%) and seniors (*N*=31; 11.9%). Most respondents were female (*N*=186; 71.3%), with fewer males (*N*=73; 28.0%) and two who did not state their gender (*N*=2; .8%). Participants reported Caucasian (*N*=196; 75.1%), African American (*N*=6; 2.3%), Asian American (*N*=10; 3.8%), Hispanic (*N*=19; 7.3%), Mixed Race (*N*=3; .1.1%), Native American Indian (N=2; .8%), and Other (*N*=25; 9.6%) ethnicities. Most of the participants were domestic students (*N*=234; 89.7%), with fewer reporting international student status (*N*=27; 10.3%). The course was taken primarily by students who were not majoring in STEM fields.

2.2.3 Course Evaluation Survey

Due to the anonymous nature of the student evaluations, only limited demographic information is available. A total of 374 students provided responses to the anonymous course evaluation survey. Among the respondents, 273 (82.7%) were enrolled in a traditional section, 62 (83.8%) were enrolled in an online only section, and 39 (69.6%) were enrolled in a flipped class. (Response rates are from each respective section's total enrollment.)

2.3 Overview of Research Procedures and Data Collection

This study design is quasi-experimental, since students chose their own sections when they registered for classes and random assignment to class sections was not possible. Student performance and attitudinal data were gathered from different sources including official university and instructor records, an online survey delivered using Qualtrics external online survey engine (www.qualtrics.com) and an end-of-semester course evaluation survey. Table 2 provides a summary of the research questions and associated data sources, as well as the number and percent of students who contributed to each data source. The research procedure was approved by the Institutional Review Board (IRB) at the researchers' university.

Table 2. Overview of sources, collection, and participants, for data collected.

Research Questions	Data Sources	Data Collected	When during semester?	Student participants Number (and percent response) from each section	
RQ3	Instructor gradebook	Exam, homework, and project grades	Throughout	Traditional	331
				Online	75
				Flipped	56
				Total	460 (100%)
RQ1 RQ2	Online survey	SATS-36 8 SRA questions	Beginning and end of semester	Traditional	193 (58.5%)
				Online	43 (58.1%)
				Flipped	25 (44.6%)
				Total	261 (56.7%)
RQ4	Course evaluation survey	Overall course and instructor ratings	End of semester	Traditional	273 (82.7%)
				Online	62 (83.8%)
				Flipped	39 (69.6%)
				Total	374 (81.30%)

2.3.1 Online Survey

Students completed an online survey at the beginning and end of the semester that measured both statistical reasoning and attitudes toward statistics. Students were given a few class participation points (worth approximately 1% of their final grade) for completing the surveys.

2.3.1.1 Student Attitudes toward Statistics

The SATS-36 was administered as a measure of students' attitudes toward statistics. Students in all three sections completed the survey at the beginning and end of the semester. Prior research validated a six-factor structure that measures student attitudes across the domains of affect, cognitive competence, value, difficulty, interest, and effort (see Table 3 for a description of the subscale and example questions). Participants were asked to respond to the SATS-36 items using a seven-point, Likert-type scale anchored by strongly disagree (1) and strongly agree (7). In the current study, internal consistency ranged from acceptable to excellent (Cronbach's α ranged from 0.74 to 0.93).

Table 3. The six SATS-36 subscales.

SATS-36 subscale	Definition	Example Question	# of Questions
Affect	Students' feelings concerning statistics	"I will like statistics."	6
Cognitive competence	Students' attitudes about their intellectual knowledge and skills when applied to statistics)	"I can learn statistics."	6
Value	Students' attitudes about the usefulness, relevance, and worth of statistics in personal and professional life	"Statistical skills will make me more employable."	9
Perceived easiness	Students' attitudes about the perceived easiness of statistics as a subject	"Statistics is a subject quickly learned by most people."	7
Interest	Students' level of individual interest in statistics	"I am interested in being able to use statistics."	4
Effort	Amount of work the student expends to learn statistics)	"I plan to work hard in my statistics course."	4

2.3.1.2 Statistical Reasoning

The SRA was the most appropriate validated instrument to measure the statistical reasoning for students in this course, though some SRA topics were not covered in the course. While it would have been preferable to use the entire instrument, we elected to administer the eight questions that were the most relevant to the material covered in the course. In addition to administering questions not related to the course materials, survey fatigue was also a concern, especially since the SRA questions were combined with the SATS-36 questions in the online survey. These eight questions provide some idea of statistical reasoning changes over the semester. The SRA is scored with points given for correct reasoning and misconceptions based on which answer choices are selected. Therefore, more points are available than the total number of questions. Table 4 provides an overview of the eight correct reasoning and eight misconception categories included in the SRA. This table also indicates the categories from which points were available in the current study.

Table 4. Overview of the SRA categories used in the current study

Correct Reasoning Skills	# of possible points from survey questions	Misconceptions	# of possible points from survey questions
CR1: Correctly interprets probabilities	N/A	MC1: Misconceptions involving averages	1
CR2: Understands how to select an appropriate average	1	MC2: Outcome orientation misconceptions	4
CR3: Correctly computes probability	N/A	MC3: Good samples have to represent a high percentage of the population	2
CR4: Understands independence	3	MC4: Law of small numbers	2
CR5: Understands sampling variability	1	MC5: Representativeness misconception	3
CR6: Distinguishes between correlation and causation	1	MC6: Correlation implies causation	2
CR7: Correctly interprets two-way tables	N/A	MC7: Equiprobability bias	N/A
CR8: Understands the importance of large samples	1	MC8: Groups can only be compared if they are the same size	N/A

Note. CR = Correct Reasoning Skills, MC = Misconceptions, N/A = No items from this category were included in the current study

2.3.2 Instructor Gradebook

Three exam grades, average homework scores, and project grades were used as measures of student performance in the course. The exams for all sections were held at the same time in a large room on campus. There were two evening midterm exams and a final exam. Each exam's structure was a mixture of multiple-choice questions about statistical concepts (often pulled from the instructors' exam question manual that accompanied the textbook; e.g., "What makes a margin of error narrower?"), article reading/written response to practice statistical literacy in real-world situations (e.g., "What are some possible lurking variables?" and "Do you think the researchers showed causation? Why or why not?"), and show-your-work calculation problems (e.g., "Calculate a 95% confidence interval for…"). The ten homework scores for the semester were averaged for each student, with the lowest score dropped. Homework questions included conceptual questions and data analysis questions by hand or using Excel. The project grade consisted of points for the student posting an article on an experiment or observational study, discussing the details of that article using statistical terms, and then making statistically intelligent comments on the articles posted by four peers.

2.3.3 Ratings of the Instructor and Course

Official university evaluations were used to compare student ratings of the same instructor and course for the three different sections. The anonymous end-of-course evaluations are conducted online at the end of each academic semester for all courses on campus. Emails are sent to all students in the courses several times during a two-week period with a link to a password-protected survey. Each student can only complete one survey per class in which they are enrolled. Students are asked to respond to a variety of questions (some of which are course-specific) on a five-point, Likert-type scale ranging from very poor (1) to excellent (5). The two questions that were of most interest to the current study prompted students to evaluate the instructor ("overall, I rate this instructor as…") and the course ("overall, I rate this course as…").

2.4. Data Analysis Procedures

All statistical analyses were conducted using SPSS 21.0 (IBM Corporation 2012). The data analysis began with standard procedures for data screening recommended for inferential statistics (Tabachnick and Fidell 2007). This process determined that the data were appropriate for statistical analysis and that the basic assumptions of ANOVA, the primary data analysis technique employed, were satisfied (i.e., scores on the dependent variable approximate an interval level of measurement, scores on the dependent variable are normally distributed, observations are independent, homogeneity of variance). After determining that the data were appropriate for use with inferential statistics, we created indices averaging the variables associated with each of the subscales of the SATS-36 using the instructions provided on the SATS website (http://www.evaluationandstatistics.com/Final36scoring.pdf). For the SRA, total correct reasoning (CR) and misconception (MC) points were calculated (see Garfield (2003) for details on scoring the SRA). With this revised version of the SRA, the points were summed for each CR or MC component and scaled to 1. Then, a total CR score and a total MC score were computed by adding all the scaled CR components and all the scaled MC components. CR total had a maximum of five points, and MC total had a maximum of six points since there were five and six components, respectively.

A pre-analysis comparison was performed in order to examine differences in student demographics between the sections. Pearson χ^2 tests were used to determine if gender, nationality (domestic or international), and class rank varied by course section. A one-way ANOVA using expected mean squares (EMS) estimates and Type III sums of squares was used to examine whether or not the students were academically different across the sections by comparing their GPA prior to the start of the course. Prior GPA was available for 444 (N=316 traditional, N=54 flipped, N=74 online) of the 462 students (96.1%) who participated in the study.

Following the pre-analysis comparison of demographic variables, descriptive statistics were calculated for all variables. Changes in statistical reasoning (RQ1) and attitudes toward statistics (RQ2) were examined using 2 x 3 time x section mixed ANOVAs conducted using EMS estimates and Type III sums of squares. The ANOVAs were mixed because they included a within subjects factor (two administrations) and a between subjects factor (course section). These mixed ANOVAs allowed examination of changes of statistical reasoning over time and attitudes toward statistics while controlling for section type (traditional, online, flipped). We examined differences in student performance (RQ3) and end-of-course evaluations (RQ4) using one-way ANOVAs, calculated with EMS and Type III sums of squares. These analyses allowed the researchers to determine if student end-of-course evaluations varied according to course section (traditional, online, flipped).

To account for the three different sections included in the analysis, a Bonferroni adjustment was made for multiple comparisons in interpreting the results of analyses related to the section variable. Partial-η^2 and η^2 were used as an estimate of effect size for mixed ANOVAs and one-way ANOVAs, respectively (Warner 2012). In the case of a significant time x section interaction effect in a mixed ANOVA, paired-samples *t*-tests were used as follow-up tests for simple effects. When *t*-tests were used, Cohen's *d* was used as an estimate of effect size (Cohen 1992). In addition to the statistical measures of effect size, practical measures are provided for significant tests using mean difference scores and corresponding 95% confidence intervals (CI). Specific to SATS, a mean difference of greater than 0.5 on the original Likert-type scale is considered practically significant (Millar and Schau 2010; Schau and Emmioglu 2012).

There is controversy in the literature about whether the Bonferroni correction to the significance level should be used for the six SATS-36 subscales (α=0.05/6), as recommended by Millar and Schau (2010), with additional corrections for comparing sections or instructors. Some articles (e.g., Emmioglu and Capa-Aydin (2012), Posner (2011)) have used this correction, while others (e.g., DeVaney (2010), Tempelaar, et al. (2006)) have not. Carlson and Winquist (2011) use α=0.01, but they do not explain the reasoning for this choice of alpha. The Bonferroni correction is quite conservative, and Type II error is a concern. DeVaney's work is the most similar to our work since his research compared pre/post SATS scores for traditional vs. online sections. DeVaney did not use the Bonferroni correction, and we will model our analysis on his when discussing our results.

3. Results

3.1 Pre-Analysis Comparison of Demographic Variables by Section

The students appear to be similar prior to taking the course with the exception of a lower proportion of freshmen in the online class than the traditional and flipped classes. Gender differences between the sections were not significant, although female students comprised the majority of all types of classes. Pearson χ^2 tests were used to examine how gender, nationality, and class rank varied according to each of the course sections (see Table 5). For class rank, the χ^2 test was significant, because hardly any freshmen were in the online class, but freshmen and sophomores comprised the majority of the students in the traditional and flipped classes. The online class exhibited a nearly even divide between freshmen/sophomores and junior/seniors. The χ^2 test for nationality was not significant, giving no evidence of a difference in the way that international and domestic students were distributed among the classes.

A one-way ANOVA examining differences across the three sections of GPA prior to taking the course was not significant, $F(2,441)=1.77$, $p=0.172$, $\eta^2=0.008$, indicating that students in the traditional (M=2.95, SD=0.59), online (M=2.94, SD=0.44), and flipped (M=2.80. SD=0.61) classes were academically similar at the beginning of the semester.

Table 5. Results of Pearson χ^2 tests of demographic variables by course section.

Demographic Variable		Traditional	Online	Flipped	Pearson χ^2
Gender	Male	114 (34.4%)	31 (41.3%)	19 (33.9%)	$\chi^2(2)=1.34$, $p=0.512$.
	Female	217 (65.6%)	44 (58.7%)	37 (66.1%)	
	Total	331 (100%)	75 (100%)	56 (100%)	
Class Rank	Freshmen	125 (37.8%)	5 (6.7%)	17 (30.4%)	$\chi^2(6)=33.21$, $p<0.001$.
	Sophomores	119 (36.0%)	34 (44.0%)	16 (28.6%)	
	Juniors	58 (17.5%)	21 (29.3%)	14 (25.0%)	
	Seniors	29 (8.8%)	15 (20.0%)	9 (16.1%)	
	Total	331 (100%)	75 (100%)	56 (100%)	
Nationality	Domestic	291 (87.9%)	64 (85.3%)	49 (87.5%)	$\chi^2(2)=0.37$, $p=0.830$.
	International	40 (12.1%)	11 (14.7%)	7 (12.5%)	
	Total	331 (100%)	75 (100%)	56 (100%)	

3.2 Research Question 1: Differences in Student Attitudes toward Statistics

Student affect generally increased from pre- to post-semester, with effects differing among section types. A series of 2x3 (time x section) mixed ANOVAs were conducted to examine changes in the SATS-36 subscales from pre- to post-semester while accounting for differences in the traditional, online, and flipped classes (see Table 6). Results indicated a statistically significant time effect for the affect subscale. There was also a significant main effect for section. Follow-up tests using a Bonferroni adjustment for multiple comparisons indicated that, on the seven-point, Likert-type scale underlying the SATS-36, the traditional students generally averaged 0.38 points higher on the affect subscale than the online students (95% CI=0.01, 0.74). Both main effects were qualified by a significant time x section interaction effect (see Figure 1a). Paired-samples t-tests to investigate simple effects indicated that the increase over time was significant for students in the traditional section, $t(192)=8.28$, $p<0.001$, $d=0.85$, who on average score 0.67 points higher at the end of the semester (95% CI=0.51, 0.83). The increase over time was also significant for students in the flipped section, $t(24)=2.13$, $p=0.043$, $d=0.63$, who on average scored 0.42 points higher at the end of the semester (95% CI=0.01, 0.83). The change over time was not significant for students in the online section.

Cognitive competence generally increased from pre- to post-semester. There was a significant time effect on the cognitive competence subscale. While the main effect for section was not significant, there was a significant time x section interaction effect (see Figure 1b). Follow-up tests for simple effects using paired-samples t-tests indicated that there was a significant increase in cognitive competence among the students in the traditional section, t(192)=7.32, p<0.001, d=0.75, who on average scored 0.56 points higher at the end of the semester (95% CI=0.41, 0.70). Changes over time for the online and flipped students were not significant.

Value generally decreased from pre- to post-semester. Results related to the value subscale indicated that there was a significant main effect for time. The main effect for section and the time x section interaction effects were not significant; therefore, mean change from pre- to post-semester was examined for all sections in aggregate. Students generally scored on average 0.16 points lower at the end of the semester than they did in the beginning of the semester (95% CI= -0.27, -0.06).

Table 6. Summary statistics and 2x3 (time x section) mixed ANOVA results for SATS-36 subscales

Subscale Time	Course Section			ANOVA Statistics			
	Traditional M(SD)	Online M(SD)	Flipped M(SD)	Factor	F	P	Partial-η^2
Affect				Time**	20.68	<0.001	0.074
Pre	4.28(1.05)	4.14(1.04)	4.13(1.01)	Section*	3.69	0.026	0.028
Post	4.92(0.97)	4.36(1.08)	4.55(1.28)	Interaction*	3.22	0.041	0.024
Cognitive				Time**	7.38	0.007	0.028
Pre	4.98(0.98)	5.10(0.72)	4.83(0.71)	Section	2.79	0.063	0.021
Post	5.54(0.90)	5.05(0.98)	5.05(1.20)	Interaction**	6.46	0.002	0.048
Value				Time**	7.70	0.006	0.029
Pre	5.21(0.87)	5.20(0.92)	5.10(1.03)	Section	0.496	0.609	0.004
Post	5.07(0.92)	4.95(1.17)	4.86(1.25)	Interaction	0.438	0.646	0.003
Easiness				Time**	34.87	<0.001	0.119
Pre	3.95(0.70)	3.75(0.67)	3.79(0.61)	Section**	7.60	0.001	0.056
Post	4.61(0.82)	4.08(0.90)	4.07(0.93)	Interaction*	4.57	0.011	0.034
Interest				Time**	15.86	<0.001	0.058
Pre	4.85(1.09)	4.72(1.15)	4.87(1.04)	Section	1.18	0.310	0.009
Post	4.64(1.19)	4.24(1.44)	4.48(1.32)	Interaction	1.32	0.269	0.010
Effort				Time**	42.15	<0.001	0.140
Pre	6.49(0.58)	6.45(0.54)	6.61(0.96)	Section	0.372	0.690	0.003
Post	6.01(0.88)	6.03(0.85)	6.11(0.96)	Interaction	0.110	0.896	0.001

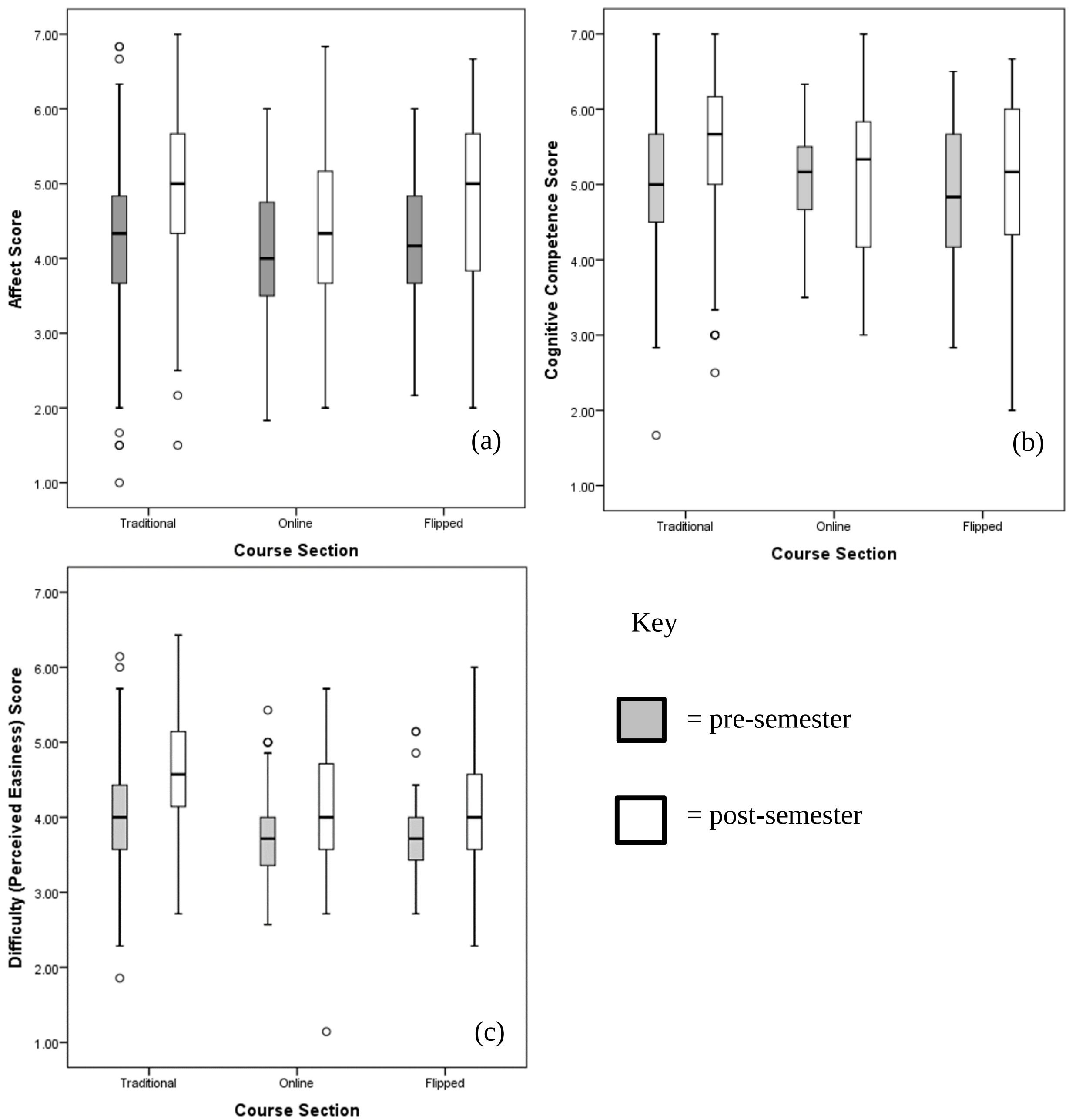
Note. All subscales of the SATS-36 were measured on a seven-point, Likert-type scale ranging from strongly disagree (1) to strongly agree (7). Cognitive=Cognitive Competence, Easiness=Perceived Easiness, Traditional (N=193), Online (N=43), Flipped (N=25). *p<0.05, **p<0.01.

Students generally increased in their perceived easiness rating from pre- to post-semester, with traditional students experiencing the greatest increase. Results indicated that there was a significant effect for time on perceived easiness. There was also a significant main effect for section. Follow-up tests using a Bonferroni adjustment for multiple comparisons indicated that the traditional students averaged 0.37 points higher on the perceived easiness subscale than the online students (95% CI=0.10, 0.63), and 0.35 points higher than the hybrid students (95% CI=0.02, 0.69). Both main effects were qualified by a significant time x section interaction effect (see Figure 1c). Follow-up paired-samples t-tests to examine simple effects indicated perceived easiness significantly increased among students in the traditional section, $t(192)=11.13$, $p<0.001$, $d=1.14$, who, on average, scored 0.66 points higher at the end of the semester (95% CI=0.54, 0.78). Perceived easiness also increased among the students in the online section, $t(42)=2.55$, $p=0.015$, $d=0.56$, who on average scored 0.33 points higher at the end of the semester (95% CI=0.07, 0.58). Changes over time for students in the flipped section were not significant.

Results related to the interest subscale indicated that there was a significant time effect. Interest generally decreased among students across all sections. Since the main effect for section and the time x section interaction effects were not significant, the mean change from pre- to post-semester was examined for all sections in aggregate. Generally, students scored on average 0.27 points lower at the end of the semester than they did in the beginning of the semester (95% CI= -0.40, -0.15).

Effort generally decreased over time. The main effect for section and the time x section interaction effect were not significant. Therefore, mean change from pre- to post-semester was examined for all sections in aggregate. Generally, students scored on average 0.47 points lower at the end of the semester than they did in the beginning of the semester (95% CI=-0.57, -0.37).

Figure 1. Side-by-side box plots displaying significant time x section interaction for the (a) affect, (b) cognitive competence, and (c) difficulty (perceived easiness) subscales of the SATS-36.



3.3 Research Question 2: Differences in Statistical Reasoning Skills

A series of 2x3 (time x section) mixed ANOVAs were conducted to examine changes in SRA scores from pre- to post-semester while also accounting for differences in traditional, online, and flipped classes (see Table 7). There was a significant effect for time on correct reasoning, indicating a general increase from pre- to post-semester. The main effect for section and the time x section interaction effect were not significant. Therefore, the mean change from pre- to post-semester was examined for all sections in aggregate. Generally, students scored on average 0.24 points higher on correct reasoning in total (out of 5 points) at the end of the semester than they did in the beginning of the semester (95% CI=0.08, 0.39).

The results relative to misconceptions revealed a significant time effect, indicating a general decrease in statistical reasoning misconceptions from pre- to post-semester. The main effect for section and time x section interaction effect were not significant, so the mean change from pre- to post-semester was examined for all sections in aggregate. Generally, students scored on average 0.12 points lower on misconceptions in total (out of 6 points) at the end of the semester than they did in the beginning of the semester (95% CI=-0.23, -0.02).

Table 7. Summary statistics and 2x3 (time x section) mixed ANOVA results for SRA correct reasoning and misconceptions subscale totals

Subscale	Course Section			ANOVA Statistics			
Time	Traditional M(SD)	Online M(SD)	Flipped M(SD)	Factor	F	P	Partial-η^2
CR				Time*	6.60	0.011	0.025
Pre	2.73(1.12)	2.52(1.03)	2.52(1.19)	Section	0.64	0.530	0.005
Post	2.92(1.07)	2.98(1.22)	2.72(1.21)	Interaction	0.78	0.461	0.006
MC				Time*	5.72	0.018	0.022
Pre	1.87(0.84)	1.89(0.77)	2.08(0.74)	Section	1.14	0.320	0.009
Post	1.79(0.84)	1.58(0.82)	1.92(0.89)	Interaction	1.25	0.288	0.010

Note. CR=Correct Reasoning, MC=Misconception, Traditional (N=193), Online (N=43), Flipped (N=25).
*p<0.05

3.4 Research Question 3: Differences in Student Performance

The traditional section was associated with significantly higher scores on Exam 1 than the online and flipped sections. One-way ANOVAs were used to examine differences in the three exams, average homework scores, and final project grades across the traditional, online, and flipped sections (See Table 8). For Exam 1, a post-hoc test using a Bonferroni adjustment for multiple comparisons indicated that out of 100 possible exam points, students in the traditional section scored on average 6.52 points higher than the online students (95% CI=3.57, 9.47), and 5.22 points higher than the hybrid students (95% CI=1.90, 8.53).

A post-hoc test using a Bonferroni adjustment for multiple comparisons indicated that the traditional section was associated with higher grades on Exam 2 than the online and flipped sections. However, there were no significant differences between the flipped and online sections. Out of 100 possible points, students in the traditional section scored on average 9.19 points higher than the online students (95% CI=5.26, 13.12), and 4.64 points higher than the hybrid students (95% CI=0.23, 9.06).

Table 8. Descriptive statistics and results of ANOVA tests for differences in exam, homework, and project scores.

Assessment	Section			ANOVA Statistics		
	Traditional M(SD)	Online M(SD)	Flipped M(SD)	F	P	η^2
Exam 1	83.41(9.56)[a]	76.89(9.74)[b]	78.20(9.27)[b]	18.45	<0.001	0.075
Exam 2	87.84(12.11)[a]	78.65(14.96)[b]	83.20(12.74)[b]	17.12	<0.001	0.071
Exam 3	73.26(14.42)[a]	68.40(12.21)[b]	70.93(12.32)[a,b]	3.91	0.021	0.017
Homework	82.60(21.75)[a]	86.92(16.73)[a]	80.35(19.98)[a]	1.85	0.159	0.008
Project	87.80(27.99)[a]	85.04(30.63)[a]	84.64(29.10)[a]	0.50	0.608	0.002

Note: Groups sharing the same letter are not significantly different from one another

A post-hoc test with a Bonferroni adjustment for multiple comparisons indicated that the traditional section was associated with higher Exam 3 scores than the online section, but the flipped section was not significantly different from the other sections. Out of 100 possible points, students in the traditional section scored on average 4.85 points higher than the online students (95% CI=0.54, 9.16).

Analyses indicated that there were no significant differences relative to the course sections for either the average homework score or project grade.

3.5 Research Question 4: Differences in Student End-of-Course Evaluations

One-way ANOVAs indicated that there were no significant differences for either overall instructor rating or overall course rating among the three sections (see Table 9).

Table 9. Descriptive statistics for overall course and instructor ratings during spring 2013

Evaluation	Teaching Level			ANOVA Statistics		
	Traditional M(SD)	Online M(SD)	Flipped M(SD)	F	P	η^2
Instructor Rating	4.56(0.69)	4.48(0.59)	4.54(0.60)	0.34	0.714	0.002
Course Rating	4.21(0.75)	4.15(0.65)	4.31(0.69)	0.59	0.554	0.003

Note. Overall instructor and course ratings was measured on a five-point, Likert-type scale ranging from very poor (1) to excellent (5).

4. Discussion

The purpose of this study was to investigate whether the type of teaching method (traditional, online, or flipped) made a difference with regards to statistical reasoning, attitudes toward statistics, course performance, and student ratings of the course and instructor. No other published studies on course design method were found to report on results from a single instructor who taught coordinated traditional, online, and flipped sections during the same semester.

4.1 Comparison of Our Results to the Literature

Our SATS-36 results showed mean increases in affect, cognitive competence, and perceived easiness with decreases in value, interest, and effort from beginning to end of the semester for all sections. These findings are particularly noteworthy given that prior researchers (e.g., Gal and Ginsburg 1994; Gal et al. 1997; Zieffler et al. 2008) have discussed the difficulty in eliciting changes in the SATS-36 subscales in the course of one semester. These findings also contrast with the findings of Bond et al. (2012), who saw decreases in all subscales except perceived easiness. Schau and Emmioglu's study (2012) similarly found decreases in value, interest, and effort from beginning to end of the semester, but their other subscales showed no change over the semester.

In our study, only affect and perceived easiness showed any differences for section, with the traditional sections on average higher than online for both. A mean increase from pre- to post- of 0.5 or higher is interpreted as practically significant (Schau and Emmioglu 2012), and affect, cognitive competence, and perceived easiness all showed practically (as well as statistically) significant increases from beginning to end of the semester for the traditional section only. Perhaps students consider learning statistics to be easier in the traditional lecture because their classroom experience is more passive and because of increased contact time with their instructor. The question of contact time is interesting because the traditional students see their lead instructor or teaching assistant for 150 minutes a week, but 100 of those minutes are in a large lecture hall with hundreds of other students. The flipped section students see their lead instructor and teaching assistant for 75 minutes a week, but that occurs while in a smaller-sized class with more active learning and more time for one-on-one contact and peer discussions.

The reasons for the differences in affect and perceived easiness would be interesting to explore. DeVaney's (2010) traditional students also perceived learning statistics as easier than his online students did. As noted by DeVaney (2010), the SATS-36 has never been validated specifically for online or flipped section students, only for traditional section students. As online and flipped classes become more prevalent, many of our attitudes and reasoning assessments will benefit from validation in multiple learning modalities.

While some of the results highlighted positive changes in student attitudes, other findings related to the SATS-36 were less encouraging. Students were less interested in learning more about statistics (interest), and felt statistics to be of lower relevance to them (value) at the end of the semester than at the beginning, although they did feel better about their own abilities to do statistics. As noted by Bond et al. (2012), final exam stress and burnout may create greater negative attitudes toward statistical inquiry at the end of a semester. While measuring students' attitudes several weeks after the end of a semester might minimize the potential negative effect related to the timing of final exams upon responses, such an approach would also likely engender a lower response rate, as the students are no longer in the instructors' course and participation incentives are more difficult to provide.

Our study is the first published report to compare SRA results from beginning to end of the semester, and analysis showed increases in correct reasoning and decreases in misconceptions for all sections. Much like Bowen et al. (2012) found no differences in CAOS statistical reasoning skills for their multi-school comparison of flipped and traditional sections, we did not see any differences between the statistical reasoning skills of the students in the three sections or any interaction between section and time.

Traditional students scored higher on average on all three exams, but there were no significant differences between sections on homework or the project. This contrasts with Shachar and Neumann's (2010) meta-analysis of courses from diverse fields showing online and flipped sections performing better than traditional sections 70% of the time. Course components such as exams, homework, and projects are less than ideal measurements of student learning since they are specific to the course, but when combined with the increase in correct reasoning and decrease in misconception measures from the SRA questions, they provide evidence that the students were learning statistical reasoning concepts in the course. Readers may note that the Exam 3 (final exam) scores were lower than Exam 1 or Exam 2 scores for all sections. The instructor attributes this to posting in the course management system, in the week before final exam, grade columns that allowed students to see the minimum score needed on the final exam to earn a particular grade in the course. Many of the best students, seeing that they did not require more than a C on the final exam to keep their A status in the course, self-reported that they did not feel the need to prepare for the final exam and instead focused on exams for other courses. The instructor will not post these grade columns in future semesters to encourage all students to perform to their maximum ability on the final exam.

There was no significant difference in how the students rated the course or instructor in the end-of-course evaluations. There are many reasons for educators to be wary of official university evaluations, but since they are metrics often used in job evaluations and promotion, it is important to know that the specific course modality will not necessarily aid or penalize an instructor. Based on the results of this study, it appears as if instructors can feel confident in choosing the course modality that they feel works best for their own teaching style and specific context.

4.2 Limitations

One limitation of the design of the study that should be taken into consideration when interpreting results is that the authors were not able to assign students randomly to the various sections, though student demographics were fairly similar across the sections, including the pre-semester GPA. Ideally, the students carefully considered their preferred learning style and chose the section that best fit that style. However, it is likely that other factors played a role in this decision. Registration dates for courses at this university are based on seniority, with more advanced students choosing their schedules first. The online section was the first to reach capacity, which could explain why fewer freshmen were in this section compared to the traditional and flipped sections. The traditional lecture classes were held twice weekly, early in the morning, which are not typically preferred times for college students and might have made the online and flipped (late morning, once weekly) sections look more attractive. It is also possible that many students and even their advisors lacked clarity on the flipped class designation during registration. Flipped classes were fairly new to this university in the Spring 2013 semester, and the registrar often changed the course catalog designation for this type of listing. Therefore, it is doubtful that learning style preference was the primary reason students selected the type of section.

The lack of validated instruments designed specifically for statistical literacy courses also limited this study in 2013. With eight of the twenty SRA questions used, only internal comparisons of statistical reasoning between the sections can be made instead of comparing to research done by others. A new statistical literacy assessment called Basic Literacy in Statistics (BLIS, Ziegler 2014) recently became available, and this instrument appears to be a better match to the concepts and level taught in the statistical literacy course discussed in this paper.

4.3 Final thoughts

It is possible that student success and attitudes in the flipped section will improve as flipped courses increase in number at the university. For some students, the flipped class structure represents unfamiliar territory. Many of the students who enrolled in the flipped section in the spring 2013 semester reported on the first day of class in a group discussion that they (and their academic advisors) had no idea what a flipped class was, but they registered for the course because it had available seats or accommodated their schedule. Since students are accustomed to taking more traditional-style courses in high school and college, it is possible that they experience a learning curve the first time they take a flipped course. However, more departments across campus are beginning to offer flipped courses, so more students who will take this course in the future may already have experience with this learning modality. This familiarity could have a positive effect on student performance. In addition to student comfort, instructor experience teaching various course delivery modalities is likely to impact student performance. Spring 2013 was only the second semester the instructor had taught a flipped section, compared to 6 years teaching online and 15 years teaching a traditional course. Both instructor and student experience should be examined as potential moderating variables in future research. As Winquist and Carlson (2014) note, the difference in student performance may depend on the particular instructor, and more research will need to be done with many other instructors to determine a clear answer as to whether there is an overall "best" pedagogical approach to section design.

While all three sections had the same lecture material, online homework system, project, and exams, there were differences between the sections in time spent interacting with other people (including the instructor and teaching assistants) and on formative feedback opportunities. The traditional section students saw their classmates and teacher three times a week; the flipped section students saw their classmates and teacher once a week; the online students only saw their classmates and teacher during proctored exams. All students were invited to attend office hours each week if they needed additional in-person help, but not many students from any of the sections used these times. The traditional and flipped section students had in-class group work and quizzes each week; the online students did not have the opportunity to do either of these. The instructor felt that having additional weekly group work problem solving and quizzes for the online students would have been logistically onerous. The online quizzes could not have been proctored, which means they would have been treated as additional online homework problems by the students. The online students did have one-to-one interactions with the instructor using the online surveys and responses from the instructor four times over the semester, but these are not a direct substitute for the in-class group activities and quizzes over the material. Therefore, any conclusions about the differences in results for the sections should take into account whether the type of section alone is the cause or if simply having fewer formative assessments has a bigger role.

With previous research in STEM courses showing that traditional lecture is inferior to online or flipped classes (Freeman et al. 2014), why did our results show the traditional section students have superior attitudes and performance? The difference may come from how the traditional section is being taught. We used a web-augmented traditional section, which means web-based technology (including online homework, an online project, a course management system, online study tools, and online lectures) were available to the traditional students as well as to the online and flipped students. The traditional students did have the material presented to them in a large lecture hall with hundreds of other students, but those lectures included many content-based i>Clicker questions to help them stay focused, and the weekly recitations included active learning and problem-solving with peers and their teaching assistant. Perhaps this method of teaching a traditional lecture section is actually a “super” traditional section with the best elements of the other sections combined with traditional lecture presentations. The literature on other traditional sections is not clear on how these traditional sections were taught (e.g., large lecture, small classes, web-augmented, group work, etc.). We need clarity in how to define a “traditional” section so that we are comparing apples to apples when deciding which section is preferable. This lack of clarity in defining what a “traditional” section is may also explain why the statistics education literature does not show much of an advantage to online and flipped sections over traditional sections—perhaps many statistics educators already following GAISE (Aliaga et al. 2005) guidelines are using sufficient technology and active learning in their traditional sections to make them more than what we consider “traditional.”

Institutions often look for opportunities to save time and/or money when deciding how various courses should be taught. From an instructor’s point of view, which type of section is the most time-consuming to plan and teach? Since these sections were planned at the same time and share so many of the same resources (lectures, homework, exams, project), it is difficult to separate out time per section in planning. It would be difficult to teach these three coordinated sections without having all of the materials written before the semester starts. Actual teaching time during the semester for the traditional section involves standing in front of a large group of students to give the lecture twice a week with the teaching assistants leading group-work recitations once a week. In the traditional section, i<Clicker grades are collected automatically in lecture (although these grades have to be uploaded by the instructor to the course management system), and the weekly recitation quizzes require manual grading. The flipped section’s weekly meeting necessitates that the instructor and a teaching assistant walk around the room answering questions while the students do group work. In the flipped class, the teaching assistant records class participation points and grades the quizzes. In the online class, the instructor responds to all of the individual surveys four times a semester and answers many questions by e-mail. For all sections, online homework is automatically graded but needs to have scores uploaded into the course management system, exams and projects are graded by teaching assistants, and office hours are available.

If the comparison of students' attitudes, reasoning, performance, and perception in these sections show few differences or small advantages to the traditional section, is it even worth offering students three options for how to take this course? People learn best when they are in an environment that gives them the opportunity to feel competence, relatedness, and autonomy (Deci and Ryan 2000). In other words, students benefit when they show what they can do, interact with others in various ways, and have choices in how they learn. By offering students three modalities for this statistical literacy class, the students have choices about how to interact with their instructor and peers to learn and demonstrate their knowledge. It is possible that the students in the online and flipped sections would have worse results if they were forced to be in the traditional section without any other options and that the few differences between the sections should be perceived as a success. The instructor has also found teaching the same material in three different ways can be energizing and creatively challenging. The diversity of interactions with the students means that new types of questions lead to new explanations and course activities. Teachers need to be able to share what they know about their subject, feel connected to their students, and have choices about how they teach.

References

Aliaga, M., Cobb, G., Cuff, C. Garfield, J., Gould, R., Lock, R., et al. (2005), "Guidelines for Assessment and Instruction in Statistics Education College Report," American Statistical Association. Available at http://www.amstat.org/education/gaise/.

Allen, I. E., Seaman, J., and Garrett, R. (The Sloan Consortium) (2007), "Blending In: The Extent and Promise of Blended Education in the United States," The Sloan Consortium. Available at http://sloanconsortium.org/publications/survey/blended06.

Allen, K. (2006), *The Statistics Concept Inventory: Development and Analysis of a Cognitive Assessment Instrument in Statistics.* (Unpublished Doctoral Dissertation), University of Oklahoma.

Babb, S., Stewart, C., and Johnson, R. (2010), "Constructing Communication in Blended Learning Environments: Students' Perceptions of Good Practice in Flipped Courses," *Journal of Online Learning and Teaching* [online], 6. Available at http://jolt.merlot.org/vol6no4/babb_1210.htm.

Baloglu, M. (2003), "Individual Differences in Statistics Anxiety among College Students," *Personality and Individual Differences,* 32, 855-865.

Bond, M., Perkins, S., and Ramirez. C. (2012), "Students' Perceptions of Statistics: an Exploration of Attitudes, Conceptualizations, and Content Knowledge of Statistics," *Statistics Education Research Journal,* 11, 6-25.

Bowen, W., Chingos, M., Lack, K., and Nygren, T. (2012), "Interactive Learning Online at Public Universities: Evidence from Randomized Trials," *ITHAKA S+R* [online]. Available at http://www.sr.ithaka.org/research-publications/interactive-learning-online-public-universities-evidence-randomized-trials

Brame, C. (2013), "Flipping the Classroom," Vanderbilt University Center for Teaching [online]. Available at http://cft.vanderbilt.edu/teaching-guides/teaching-activities/flipping-the-classroom.

Burgan, M. (2006), "In Defense of Lecturing," *Change* [online]. Available at http://www.ltrr.arizona.edu/~katie/kt/COLLEGE-TEACHING/topic%203%20-%20syllabus/In%20Defense%20of%20Lecturing.pdf.

Carlson, K., and Winquist, J. (2011), "Evaluating an active learning approach to teaching introductory statistics: A classroom workbook approach," *Journal of Statistics Education*, 19(1). Available at http://www.amstat.org/publications/jse/v19n1/carlson.pdf

Chickering, A. W., and Gamson, Z. F. (1987), "Seven Principles for Good Practice," *AAHE Bulletin,* 39, 3-7.

Cohen, J. (1992), "A power primer," *Psychological Bulletin*, 122, 155-159.

Corrigan, P. (2013), "To Lecture or Not to Lecture?," *The Atlantic* [online]. Available at http://m.theatlantic.com/education/archive/2013/12/to-lecture-or-not-to-lecture/282585/.

Deci, E., and Ryan, R. (2000), "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being," *American Psychologist,* 55(1), 68-78.

del Mas, R., Garfield, J., Ooms, A., and Chance, B. (2007), "Assessing Students' Conceptual Understanding after a First Course in Statistics," *Statistics Education Research Journal*, 6(2), 28-58.

Dempster, M., and McCorry, N. (2009), "The Role of Previous Experience and Attitudes Toward Statistics in Statistics Assessment Outcomes among Undergraduate Psychology Students," *Journal of Statistics Education* [online], 17. Available at http://www.amstat.org/publications/jse/v17n2/dempster.pdf.

Deslauriers, L., Schelew, E., and Wieman, C. (2011), "Improved Learning in a Large-Enrollment Physics Class," *Science,* 332, 862-864.

DeVaney, T. (2010), "Anxiety and Attitude of Graduate Students in On-Campus vs. Online Statistics Courses," *Journal of Statistics Education* [online], 18. Available at http://www.amstat.org/publications/jse/v18n1/devaney.pdf.

Emmioglu, E., and Capa-Aydin, Y. (2012), "Attitudes and Achievement in Statistics: A Meta-analysis Study," *Statistics Education Research Journal,* 11(2), 95-102.

Finney, S., and Schraw, G. (2003), "Self-efficacy beliefs in college statistics courses," *Contemporary Educational Psychology,* 28(2), 161-186.

Freeman, S., Eddy, S., Mcdonough, M., Smith, M., Okorafor, N., Jordt, H., and Wenderoth, M. (2014), "Active learning increases student performance in science, engineering, and mathematics," *Prcoeedings of the National Academy of Sciences*, 111(23). Available at http://www.pnas.org/content/111/23/8410

Froyd, J. E. (2007), "Evidence for the efficacy of student-active learning pedagogies," *Project Kaleidoscope* [online]. Available at www.pkal.org/documents/BibliographyofSALPedagogies.cfm.

Froyd, J. E. (2008), "White paper on promising practices in undergraduate STEM education" [Commissioned paper for the Evidence on Promising Practices in Undergraduate Science, Technology, Engineering, and Mathematics (STEM) Education Project, The National Academies Board on Science Education]. Available at www7.nationalacademies.org/bose/Froyd_Promising_Practices_CommissionedPaper.pdf.

Gal, I. (2002), "Adults' Statistical Literacy: Meanings, Components, Responsibilities, *International Statistics Review*, 70(1), 1-51.

Gal, I., and Ginsburg, L. (1994), "The Role and Beliefs and Attitudes in Learning Statistics: Towards an Assessment Framework," *Journal of Statistics Education* [online], 2. Available at http://www.amstat.org/publications/jse/v2n2/gal.html.

Gal, I., Ginsburg, L., and Schau, C. (1997), "Monitoring Attitudes and Beliefs in Statistics Education," in *The Assessment Challenge in Statistics Education*, Gal, I. and Garfield, J. B., Amsterdam: The International Statistical Institute, pp. 37-51.

Garfield, J. (2003), "Assessing statistical reasoning," *Statistics Education Research Journal, 2*(1), 22-38. Available at https://www.stat.auckland.ac.nz/~iase/serj/SERJ2%281%29.pdf.

Garfield, J., and Gal, I. (1999), "Assessment and Statistics Education: Current Challenges and Directions," *International Statistical Review,"* 67(1), 1-12.

Gibbs, A., and Tayback, N. (2014), "A new blend: An introductory statistics course integrating a MOOC, active learning sessions, and open online resources, with seasoning from the students' areas of study," poster from The Second Biennial Electronic Conference on Teaching Statistics (eCOTS14), May 19-23. Available at: https://www.causeweb.org/ecots/paywall/9/

Giraud, G. (1997), "Cooperative Learning and Statistics Instruction," *Journal of Statistics Education* [online], 5. Available at www.amstat.org/publications/jse/v5n3/giraud.html.

Gravestock, P., and Gregor-Greenleaf, E. (2008), *Student Course Evaluations: Research, Models and Trends.* Toronto: Higher Education Quality Council of Ontario. Available at http://www.heqco.ca/SiteCollectionDocuments/Student%20Course%20Evaluations.pdf

Griffith, J., Adams, L., Gu, L., Hart, C., Nichols-Whitehead, P. (2012), "Students' Attitudes Toward Statistics Across the Disciplines: A Mixed-Methods Approach," *Statistics Education Research Journal*, 11, 45-56.

Haidet, P., Morgan, R. O., O'Malley, K. O., Moran, B. J., and Richards, B. F. (2004), "A Controlled Tial of Active Versus Passive Learning Strategies in a Large Group Setting'" *Advances in Health Sciences Education,* 8, 9-30.

Hannigan, A., Gill, O., and Leavy, A. M. (2013), "An Investigation of Prospective Secondary Mathematics Teachers' Conceptual Knowledge of and Attitudes Towards Statistics," *Journal of Mathematics Teacher Education*, 16, 427-449.

IBM Corporation (2012), "SPSS 21.0." In. Chicago, IL: Author.

Jaki, T. (2009), "Recording lectures as a service in a service course," *Journal of Statistics Education* [online], 17. Available at www.amstat.org/publications/jse/v17n3/jaki.html.

Johnson, M., and Kuennen, E. (2006), "Basic Math Skills and Performance in an Introductory Statistics Course," *Journal of Statistics Education* [online], 14. Available at http://www.amstat.org/publications/jse/v14n2/johnson.html.

Kember, D. (1997), "A Reconceptualization of the Research into University Academics' Conceptions of Teaching," *Learning and Instruction,* 7, 255-275.

Margulieux, L. E., Bujak, K. R., McCracken, W. M., and Majerich, D. M. (2014), "Flipped, Blended, Flipped, and Inverted: Defining Terms in a Two Dimensional Taxonomy. Paper accepted to the 12th Annual Hawaii International Conference on Education, Honolulu, HI, January 5-9. Available at http://c21u.gatech.edu/sites/default/files/HICE%20Conference%20Proceedings_1556_with%20citation%5B4%5D.pdf

McLaughlin, J., Roth, M., Glatt, D., Gharkholonarehe, N., Davidson, C., Griffin, L., Esserman, D., Mumper, R. (2014), "The Flipped Classroom: A Course Redesign to Foster Learning and Engagement in a Health Professions School," *Academic Medicine*, 89(2). Available at http://journals.lww.com/academicmedicine/Abstract/publishahead/The_Flipped_Classroom___A_Course_Redesign_to.99241.aspx.

Millar, A., and Schau, C. (2010), "Assessing Students' Attitudes: The Good, the Bad, and the Ugly," *Joint Statistical Meeting Section on Statistical Education.*

Mills, J., and Raju, D. (2011), "Teaching Statistics Online: A Decade's Review of the Literature about What Works," *Journal of Statistics Education* [online], 19. Available at www.amstat.org/publications/jse/v19n2/mills.pdf.

Moore, D., and Notz, W. (2009), *Statistics: Concepts and Controversies*, 7th Edition, New York, NY: W. H. Freeman and Company.

Onwuegbuzie, A., and Wilson, V. (2003), "Statistics Anxiety: Nature, Etiology, Antecedents, Effects, and Treatments—a Comprehensive Review of the Literature," *Teaching in Higher Education*, 8, 195-209.

Parkhurst, R., Moskal, B., Downey, G., Lucena, J., Bigley, T., and Elberb, S. (2008a), "Engineering Cultures: Online Versus In-class," *Journal of Online Learning and Teaching*, 4, 438-445.

Parkhurst, R., Moskal, B., Downey, G., Lucena, J., Bigley, T., and Elberb, S. (2008b), "Engineering cultures: Comparing student learning in on-line and classroom based implementations," *International Journal of Engineering Education*, 24 (5), 955-964.

Pearl, D., Garfield, J., delMas, R., Groth, R., Kaplan, J., McGowan, H., and Lee, H. (2012), "Connecting Research to Practice in a Culture of Assessment for Introductory College-level Statistics." Available at www.causeweb.org/research/guidelines/ResearchReport_Dec_2012.pdf.

Posner, M. (2011), "The Impact of a Proficiency-based Assessment and Reassessment of Learning Outcomes System on Student Achievement and Attitudes," *Statistics Education Research Journal*, 10(1), pp. 3-14. Available at https://www.stat.auckland.ac.nz/~iase/serj/SERJ10%281%29_Posner.pdf

Ramirez, C., Schau, C., and Emmioglu, E. (2012), "The Importance of Attitudes in Statistics Education," *Statistics Education Research Journal*, 11, 57-71.

Ramsden, P. (1992), *Learning to teach in higher education*, New York, NY: Routledge.

Rovai, A., and Jordan, H. (2004), "Blended Learning and Sense of Community: A Comparative Analysis with Traditional and Fully Online Graduate Courses," *International Review of Research in Open and Distance Learning* [online], 5(2). Available at file:///C:/Users/K.%20Andrew%20Richards/Downloads/192-2687-2-PB.pdf.

Rumsey, D. (2002), "Statistical Literacy as a Goal for Introductory Statistics Courses," *Journal of Statistics Education* [online], 10. Available at www.amstat.org/publications/jse/v10n3/rumsey2.html.

Shachar, M., and Neumann, Y. (2010), "Twenty Years of Research on the Academic Performance Differences Between Traditional and Distance Learning: Summative Meta-Analysis and Trend Examination," *Journal of Online Learning and Teaching*, 6, 318-334.

Schau, C. (2003a), *Survey of Attitudes Toward Statistics* (SATS-36). Available at http://evaluationandstatistics.com

Schau, C. (2003b, August), "Students' attitudes: The 'other' important outcome in statistics education. Paper presented at the Joint Statistical Meetings, San Francisco, CA.

Schau, C., and Emmioglu, E. (2012), "Do Introductory Statistics Courses in the United States Improve Students' Attitudes?" *Statistics Education Research Journal*, 11(2), 86-94.

Somenarain, L., Akkaraju, S., and Gharbaran, R. (2010), "Student Perceptions and Learning Outcomes in Asynchronous and Synchronous Online Learning Environments in a Biology Course," *Journal of Online Learning and Teaching*, 6, 353-356.

Stark, P. (2013a), "Evaluating Evaluations: Part I". University of California at Berkeley Center for Teaching and Learning, posted on October 9, 2013. Available at http://teaching.berkeley.edu/blog/evaluating-evaluations-part-1.

Stark, P. (2013b), "What Evaluations Measure: Part II". University of California at Berkeley Center for Teaching and Learning, posted on October 17, 2013. Available at http://teaching.berkeley.edu/blog/what-evaluations-measure-part-ii.

Strayer, J. F. (2012), "How Learning in an Inverted Classroom Influences Cooperation, Innovation, and Task Orientation," *Learning Environments Research*, 15(2), 171-193. Available at http://link.springer.com/article/10.1007%2Fs10984-012-9108-4.

Tabachnick, B. G., and Fidell, L. S. (2007), *Using Multivariate Statistics*, 5th Edition, Boston, MA: Allyn and Bacon.

Tempelaar, D., Gijselaers, W., and van der Loeff, S. (2006), "Puzzles in Statistical Reasoning," *Journal of Statistics Education* [online], 14. Available at http://www.amstat.org/publications/jse/v14n1/tempelaar.html.

Tishkovskaya, S., and Lancaster, G. (2012), "Statistical Education in the 21st Century: A Review of Challenges, Teaching Innovations and Strategies for Reform," *Journal of Statistics Education* [online], 20. Available at www.amstat.org/publications/jse/v20n2/tishkovskaya.pdf.

Toohey, S. (1999), *Designing Courses for Higher Education*, Buckingham, UK: The Society for Research into Higher Education and Open University Press.

Trigwell, K., Prosser, M., and Taylor, P. (1994), "Qualitative Differences in Approaches to Teaching First Year University Science," *Higher Education*, 27, 75-84.

Utts, J., Sommer, B., Acredolo, C., Maher, M., and Matthews, H. (2003), "A Study Comparing Traditional and Flipped Internet-Based Instruction in Introductory Statistics Classes," *Journal of Statistics Education* [online], 11. Available at www.amstat.org/publications/jse/v11n3/utts.html on 10/26/13.

Vanhoof, S., Kuppens, S., Castro Sotos, A., Verschaffel, L., and Onghena, P. (2011), "Measuring Statistics Attitudes: Structure of the Survey of Attitudes Toward Statistics (SATS-36)," *Statistics Education Research Journal*, 10, 35-51.

Verhoeven, P. (2006), "Statistics Education in the Netherland and Flanders: An Outline of Introductory Courses at Universities and Colleges," *ICOTS-7 Conference Proceedings.*

Voeten, E. (October 2, 2013), "Student evaluations of teaching are probably biased. Does it matter?" *The Washington Post*. Available at http://www.washingtonpost.com/blogs/monkey-cage/wp/2013/10/02/student-evaluations-of-teaching-are-probably-biased-does-it-matter/.

Walthausen, A. (2013), "Don't Give Up on the Lecture," *The Atlantic* [online]. Available at http://www.theatlantic.com/education/archive/2013/11/dont-give-up-on-the-lecture/281624/.

Ward, B. (2004), "The Best of Both Worlds: A Flipped Statistics Course," *Journal of Statistics Education* [online], 12. Available at www.amstat.org/publications/jse/v12n3/ward.html.

Warner, R. M. (2012), Applied statistics: From Bivariate through Multivariate Techniques, 2nd Edition, Los Angeles, CA: Sage.

Watson, J. (1997), "Assessing Statistical Thinking Using the Media," *The Assessment Challenge in Statistics Education,"* Gal, I. and Garfield, J. (Eds.). Amsterdam: IOS Press and The International Statistical Institute, 107-121.

White, B. (2014), "Exploring a 'Flip' in Introductory Statistics," poster from The Second Biennial Electronic Conference on Teaching Statistics (eCOTS14), May 19-23. Available at: https://www.causeweb.org/ecots/paywall/25/

Winquist, J. and Carlson, K. (2014), "Flipped Statistics Class Results: Better Performance Than Lecture Over One Year Later," *Journal of Statistics Education*, 22(3). Available at: http://www.amstat.org/publications/jse/v22n3/winquist.pdf

Zieffler, A., Garfield, J, Alt, S., Dupuis, D., Holleque, K., and Chang, B. (2008), "What Does Research Suggest about Teaching and Learning of Introductory Statistics at the College Level? A Review of the Literature," *Journal of Statistics Education* [online], 16. Available at www.amstat.org/publications/jse/v16n2/zieffler.html.

Ziegler, L. (2014), "Reconceptualizing statistical literacy: Developing an assessment for the modern introductory statistics course." Unpublished doctoral dissertation available at: http://iase-web.org/documents/dissertations/14.LauraZiegler.Dissertation.pdf

---

Ellen Gundlach
Department of Statistics
Purdue University
250 N. University Street
West Lafayette, IN 47907-2066
gundlach@purdue.edu

K. Andrew R. Richards
Department of Kinesiology and Physical Education
Northern Illinois University
Anderson Hall
520 Garden Rd., Rm. 229
DeKalb, IL 60115-2828
karichar@niu.edu

David Nelson
Center for Instructional Excellence
Purdue University
Hall for Discovery and Learning Research
207 S Martin Jischke Drive, Suite 341
West Lafayette, IN 47907-1971
davenelson@purdue.edu

Chantal Levesque-Bristol
Center for Instructional Excellence
Purdue University
Hall for Discovery and Learning Research
207 S Martin Jischke Drive, Suite 341
West Lafayette, IN 47907-1971
cbristol@purdue.edu
