Communications for Statistical Applications and Methods
2020, Vol. 27, No. 3, 269–284

https://doi.org/10.29220/CSAM.2020.27.3.269
Print ISSN 2287-7843 / Online ISSN 2383-4757

Effects of a GAISE-based teaching method on students' learning in introductory statistics

Erik Barry Erhardt[a], Woong Lim[1,b]

[a]Department of Mathematics and Statistics, University of New Mexico, USA;
[b]Graduate School of Education, Yonsei University, Korea

Abstract

This study compares two teaching methods in an introductory statistics course at a large state university. The first method is the traditional lecture-based approach. The second method implements a flipped classroom that incorporates the recommendations of the American Statistical Association's Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report. We compare these two methods, based on student performance, illustrate the procedures of the flipped pedagogy, and discuss the impact of aligning our course to current guidelines for teaching statistics at the college level. Results show that students in the flipped class performed better than students in traditional delivery. Student questionnaire responses also indicate that students in flipped delivery aligned with the GAISE recommendations have built a productive mindset in statistics.

Keywords: GAISE, introductory statistics, student-centered instruction, flipped classroom

1. Introduction

Introductory undergraduate statistics is an important gateway course in statistics education, as it prepares students to succeed in quantitative reasoning and data literacy skills (Gould, 2017; Rumsey, 2002). However, instructors are presented with challenges as students find the course difficult and opt to drop it or do not do well (Hildreth *et al.*, 2018). Over time, the field of statistics education has made efforts to improve the teaching of statistics at the college level and has developed pedagogical guidelines for the introductory statistics course. The Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report was originally published in 2005 and provided a groundbreaking framework for teaching statistics (Franklin *et al.*, 2007). The guidelines were then updated in 2016 to better reflect the advancements in technology, as well as the newly emerging data sets available for contemporary statistics classrooms; however, the core principles were largely unchanged (Carver *et al.*, 2016).

There exists extensive research on the teaching of statistics while incorporating the original GAISE framework. The field of statistics has learned that collaborative and problem-based learning, with appropriate use of technology, shows great potential to improve student achievement in statistics courses (Garfield *et al.*, 2012; Horton, 2015; Zieffler *et al.*, 2008). Current GAISE recommendations emphasize statistical thinking and conceptual understanding as student outcomes in statistics classes; therefore, we need research that explains the nature of a course designed around recommendations and learning environments that can produce desired student learning outcomes. There is a salient need for

[1] Corresponding author: Graduate School of Education, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea. E-mail: woonglim@yonsei.ac.kr

Published 31 May 2020 / journal homepage: http://csam.or.kr

further research about how statistics educators implement these guidelines as integrated curriculum in a single course (for student attitudes in a GAISE-influenced course see Paul and Cunnington (2017)). It is vital to have educators report, in detail, the impact of aligning their course to current guidelines for teaching statistics at the college level. This study describes an effective approach to teaching introductory statistics courses so that the model can be generalized for use by statistics educators at other institutions. In addition, the study seeks to analyze student learning from the GAISE-based, flipped classroom in comparison with a traditional approach.

In this study, we describe a flipped classroom involving preparatory instruction outside the class (videos), that is then reinforced or applied during class time (see Nielsen *et al.* (2018) for a review of definitions). In the flipped learning environment, the teachers implement social- and active-learning strategies and provide a high cognitive, student-centered learning experience, as opposed to conventional, teacher-centered teaching (lectures).

2. Literature review

2.1. Current recommendations for teaching statistics

The American Statistical Association recently revised their report: Guidelines for GAISE (Aliaga *et al.*, 2005; Carver *et al.*, 2016). The six main recommendations are:

1. Teach statistical thinking (**statistical literacy**).
   (a) Teach statistics as an investigative process of problem-solving and decision making.
   (b) Give students experience with multivariable thinking.
2. Focus on **conceptual understanding**.
3. Integrate **real data** with a context and purpose.
4. Foster **active learning**.
5. Use **technology** to explore concepts and analyze data.
6. Use **assessments** to improve and evaluate student learning.

GAISE provides practical examples of activities and projects, assessment items, use of technology, and use of authentic data.

The Adapting and Implementing Innovative Material in Statistics (AIMS) Project has developed materials aligned with GAISE to help realize these recommendations (Everson *et al.*, 2008; Garfield *et al.*, 2008). The project has also provided assessment resource tools known as Assessment Resource Tools for Improving Statistical Thinking (ARTIST). Furthermore, NSF-funded Levels of Conceptual Understanding in Statistics (LOCUS) developed assessments consistent with the five layers of evidence-centered design: domain analysis, domain modeling, conceptual assessment framework, assessment implementation, and assessment delivery (Mislevy *et al.*, 2003). A program wishing to design a new course or renovate current courses has a wealth of assessment materials available.

We explain how our course is designed around the GAISE recommendations. This study has redesigned the introductory statistics course to maintain existing student learning outcomes so that it aligns with updated guidelines (Carver *et al.*, 2016); we restricted innovation to teaching methods in order to make meaningful comparisons with existing traditional courses in our program. Part of

our course design includes implementing a modern model of teaching introductory statistics and incorporating key elements of the pedagogy used in flipped classrooms (Winquist and Carlson, 2014) and project-based learning (Blumenfeld *et al.*, 1991) to promote students' statistical thinking and reasoning as the tool for problem-solving and decision making (Item 1 of GAISE). A large body of literature describes active learning as the most effective method of teaching statistics (Garfield *et al.*, 2012; Gelman and Nolan, 2002; Neumann *et al.*, 2013), emphasizing (a) collaborative learning, (b) concepts over procedure-based skills, and (c) critical thinking skills. Further, it is recommended to get students involved with real-life examples and meaningful projects. In our teaching model, students use technology to explore concepts (Item 5) and engage in active learning (Item 4) using real data (Item 5) with interconnected tasks to create conceptual understanding (Item 2) and increase data literacy. In addition, both the instructor and students used assessment data (Item 6) to diagnose the gaps in knowledge and skill to improve student learning. While the traditional course may meet Items 1 and 6, it may not meet Items 3, 4, or 5 (Johnson and Dasgupta, 2005).

2.2. Evaluating teaching innovation

Chelimsky and Shadish (1997) identified three conceptual frameworks for evaluation: (1) evaluation for accountability, (2) evaluation for development, and (3) evaluation for knowledge. The first framework is useful for measuring the overall impact of our new pedagogy versus traditional pedagogy. One key aspect of substantiating student success is scientifically quantifying changes in student knowledge and skills through learning opportunities in the course (Budgett and Pfannkuch, 2007). Course assessments are therefore used as an instrument to provide evidence indicating student performance. However, course assessment is not about making judgments regarding the success of the pedagogy, but should rather be thought of in terms of student learning and development (Macdonald *et al.*, 2006). The second framework is therefore useful to identify areas for improvement and use the information to refine the course. The units of analysis for meaningful change in the statistics course (Bidgood *et al.*, 2008) include: student academic performance, student engagement, student feedback, or changes in student concepts, skills, applications, reasoning, attitude, or beliefs (Garfield, 1994).

This study conducted course assessment with two goals in mind. First, the study uses student performance as criterion to compare the efficacy of teaching methods versus the traditional method in a single course with multiple sections of students with similar academic backgrounds at the same institution. Second, the study analyzes student feedback and student work as well as explores the relationship between a course design aligned with the current guidelines and the qualitative nature of student learning in introductory statistics courses.

2.3. Related studies on teaching introductory statistics in college

Zieffler *et al.*, (2008) have written a thorough review of the existing literature on the pedagogy of teaching statistics to undergraduate students and concluded that statistics remained a difficult subject for students. The same study (Zieffler *et al.*, 2008) also noted that many undergraduate students approached statistics courses with feelings ranging from little interest to fear and anxiety (i.e., Forte, 1995; Gal and Ginsburg, 1994; Hassad, 2011).

In the 1990s, researchers (Giraud, 1997; Magel, 1998) examined the use of cooperative learning in teaching statistics and reported improvements in student learning. In addition, (Rumsey, 1998) argued that cooperative learning could be useful in the undergraduate statistics classroom to provide active learning experiences for students.

Rinaman (1998) looked into the process of designing and improving a reformed college statistics course and reported on the benefits of statistics courses which focused on concepts and technology,

such as Minitab software, to analyze large datasets. With the advancement of technology available for use in teaching statistics, other researchers (Garfield and Ben-Zvi, 2007; Lane and Tang, 2000) began to investigate the impact and uses of technology on student's statistical reasoning and compared student learning with technology versus student learning through traditional methods (lecture, notes, or textbooks).

Through increased clarity and consensus in regard to the benefits of learning with technology, some researchers have focused on more specific uses of technology to help students understand difficult concepts (for sampling distributions see Garfield *et al.*, 1999). Further, other researchers (Alldredge and Brown, 2006) have explored the relationship between the use of technology in the classroom with non-cognitive factors such as gender, learner beliefs, and student performance.

The field of statistics education has long known that the introductory statistics course can improve the success of students by applying a subset of long-recommended evidence-based statistical education practices. Extensive work has also shown that under-represented and non-underrepresented students learn more in a multidisciplinary project-based statistics course (Die-rker *et al.*, 2016); in addition, similar results were achieved with Biology courses (Haak *et al.*, 2011). However, teaching practices as well as should change. An introductory course needs to teach different material in a different way (Utts, 2015). Gelman and Carlin (2017) also wrote in the context of p-values, "it's not that we're teaching the right thing poorly; unfortunately, we've been teaching the wrong thing all too well." Using a musical analogy, we believe the traditional introductory statistics course teaches the notes out of context, while a data-based statistics course can teach a song (for context and motivation) and then engage students with the notes.

Regarding the course or program design aligned with pedagogical framework, Hall and Rowell (2008) looked into 110 NSF funded projects on developing introductory statistics between 1993 and 2004 which incorporated GAISE guidelines such as: emphasizing statistical literacy and thinking, using real data, stressing conceptual understanding over procedural knowledge, fostering active learning, or using technology. They reported about 95% of the projects met at least one GAISE recommendation, while 65% met more than one. This suggests that we need more research into introductory statistics courses that incorporate multiple recommendations of the GAISE guidelines in a single course as well as provide an interconnected and technologically integrated learning experiences for students. Currently, there is scant research to compare student performance and the nature of learning with student feedback as evidence between traditional pedagogy and the modern pedagogy when offering integrated learning experiences for college students in introductory statistics courses.

3. Methods

3.1. Setting and participants

The setting of this study is a large state university in a southwestern state in the U.S. where 48.5% of the population are Hispanic or Latino, 10.6% are American Indian, 26.7% have a Bachelor's degree or higher, and 19.8% live in poverty (US Census Bureau 2017). The average time to a Bachelor's degree is 4.86 years, and the within 6-year graduation rate for students entering in 2011 for Whites was 54.8% , for Hispanics was 45.3% , and for Native American was 30.7%. Student participants for this study include undergraduate students enrolled in Introduction to Statistics in Spring 2017, a course with 20+ sections and 1200+ students per semester (roughly 55 students per section). The passing rate for the course from Fall 2006 to Spring 2016 ranged from 72% to 81% , with summer semesters ranging from 76% to 89%. Student course evaluations often report that they found the course too theoretical, difficult to follow, and that it had little relevance to their interests. The instructor participants were in

Table 1: Course Topics by week, Flipped delivery teaching method schedule is aligned with Traditional and Lecture delivery. Duration in weeks is for classes held two versus three times per week; "0.5 week" is one of two or one-to-two of three classes.

Topic	Duration
Summaries, graphical and numerical	1.5 weeks
Normal distribution	1.5 weeks
Exam 1	1 day
Correlation and regression	2 weeks
Sampling and experiments	2 weeks
Exam 2	1 day
Probability	0.5 week
Sampling distributions	1 week
CIs and hypothesis test basics, known σ	2 weeks
Exam 3	1 day
1-sample t-test	0.5 week
2-sample t-test	1 week
1-sample p-test	1 week
Chi-square test	0.5 week
Review	1 week
Exam 4	1 day

two groups. Fifteen (15) course sections (classrooms of students) participated in the study coordinated by an experienced lecturer and were taught by experienced instructors under traditional pedagogy labeled "Traditional." Six (6) additional course sections were coordinated by the first author and were taught by experienced instructors as part of the experiment — four (4) new pedagogy sections were labeled "Flipped" (treatment) and two (2) control sections were taught in the traditional way and labeled "Lecture" (control). The "Lecture" control sections were in place to distinguish between the course coordinator and the teaching method achievement effects with the expectation that the course coordinator should not have an effect. Students enrolled without prior knowledge of the class format and randomized themselves into treatment groups. The study was pre-registered with an Institutional Review Board (IRB) and self-selected students provided prior consent to participate in an end-of-semester questionnaire. Questionnaire results show that the three groups were balanced with respect to age, gender, number of credits earned toward their degree, number of math/stat courses completed, and GPA.

This study compares our existing delivery methods (Traditional and Lecture) with a proposed alternative following the GAISE guidelines in the flipped classrooms (Flipped). Both methods cover the same topics with common student learning outcomes and use the same exams.

3.2. The course with a GAISE-based approach

3.2.1. Course overview

The introductory statistics course covers descriptive statistics, data collection, inferential statistics, and hypothesis testing (Table 1). The course is a three-credit-hour class meeting either two or three times per week over a 16-week academic semester. Both teaching methods also used the same interactive e-book called LaunchPad which included multimedia content, test banks, and adaptive quizzing (Moore *et al.*, 2015).

A chalkboard is available in each classroom in order to supplement classroom discussions. Each classroom is equipped with a computer projection system; however, no desktop or laptop computers is accessible in the classroom. Historically and currently, the introductory statistics course was taught

by lecturers, part-time instructors, and graduate student teaching assistants from the Mathematics and Statistics Department.

3.2.2. The sections with a traditional approach: Traditional/Lecture delivery

This teaching method consists of lecture-based instructor presentations, handouts, and a textbook as a reference for homework (not graded).

(a) Pre-class: No pre-topic preparation.

(b) In-class: Instructors have the discretion to lecture as they like, whether using chalkboard or slide presentation lectures or incorporating additional in-class quizzes.

(c) Post-class: Homework problems are optional and not graded, the e-book assignments (graded) were due the Saturday after the previous chapter lecture, and the online chapter Quizzes (graded) and online chapter Learning Curve (graded) were due the Sunday after each previous chapter lecture.

3.2.3. The sections with a GAISE-based approach: Flipped delivery

Each flipped delivery section had an instructor who was a graduate student or part-time instructor and a peer learning facilitator (PLF) who was an undergraduate in-class helper with a higher-level statistics background and an interest in teaching. Instructors had weekly coordination meetings and provided consistent instruction. The flipped delivery course had the following weekly features:

(a) Pre-class: Reading, Video lectures (graded), and reading Quiz (graded) for the student accountability of exposure to the material.

(b) In-class: Coordinator-written worksheet activities (graded for effort) to engage teams of students with the ideas preceded by brief topic summary discussions occasionally using statistical software with simulations or real data to illustrate and visualize concepts at the beginning of class, such as sampling distribution, confidence interval, or outliers. Students engaged in active struggle, but with PLFs, student teammates, and the instructor available to provide support before they felt frustrated.

(c) Post-class: Homework, online chapter Quizzes (graded), and online chapter Learning Curve (graded) concept questions for student accountability for mastery of the material on the same schedule as the Traditional delivery method.

The two delivery methods can be compared using the GAISE recommendations for how to teach introduction to statistics. The Traditional/Lecture delivery meets GAISE recommendations (1) Teach statistical thinking and (2) Focus on conceptual understanding. In addition, the Flipped delivery included in-class worksheets to (3) integrate real data with a context and purpose and (4) foster active learning, and used computer simulations to (5) use technology to explore concepts and analyze data. Finally, the flipped delivery coordinator does (6) use assessments to improve and evaluate student learning.

3.3. Data collection and analysis

This study pseudo-randomized all 20+ course sections between coordinators, the six experimental sections were pseudo-randomized to Flipped delivery treatment and Lecture control teaching methods,

and students registered for their course sections without knowledge of the different teaching methods. The primary common metric for comparison is the set of four course exams. This randomized design with imposed teaching methods provides a basis to infer causality on the exam scores due to teaching delivery methods (Flipped vs Lecture/Traditional) or due to coordinators (Lecture vs Traditional) rather than unlikely *a priori* differences between the students registered for different sections. Note that a difference between Lecture and Traditional delivery would indicate an effect due to course coordinator, which is not expected since the coordinator had almost no input into the Lecture sections since these were experienced Traditional delivery instructors.

The study also drew on multiple data sources to produce the understanding of quantitative data. Course instructors provided course materials including syllabus, distribution of course grades, and student evaluation data. Students provided student work and questionnaire responses. Some classroom practices differed; in particular, a few of the Flipped delivery sections unwittingly did not require attendance via a daily sign-in sheet, while all Traditional and Lecture delivery sections did (and those sections without a sign-in sheet did have slightly lower average scores).

The data were analyzed using quantitative and qualitative approaches and presented as summaries of pedagogy and student learning. We compared two teaching methods based on student performance through descriptive statistics and modeling — explaining the strengths and weaknesses of pertinent teaching methods through descriptive patterns, narratives of evidence and rationale — to categorize and summarize the data from these sources. First, we examined course materials such as syllabi, handouts, teacher notes, and student work samples. Second, a questionnaire was administered towards the end of the semester asking students to describe their perception of learning in the course and provided a qualitative indicator of course effectiveness. The questionnaire also included a Likert-type scale with a set of statements to measure student perceptions of success, adding context for the findings on course effectiveness. The questionnaire for this study was adapted from survey instruments designed to evaluate teaching innovation (Moore, 2003; Ramsden, 2003; Saunders, 2000). This survey had a demographic section and had statements regarding two main areas: categorical questions on course delivery and outcomes (40 items) and free response questions on learning experiences (3 items). Students indicated the extent to which they agreed for each statement using seven-, five-, or three-point Likert scales ranging from "strongly disagree" (score of 1) to "strongly agree" (score of 7) or other ordinal response categories. Categories were collapsed to "positive" vs "neutral and negative" then proportions were compared between Flipped and Traditional/Lecture delivery using permutation testing. Third, a record of student questionnaire responses and a record of course grades were analyzed to compare the two teaching methods. Numeric course grades were compared using ANOVA, while registrar letter grades used a Chi-square Test for Homogeneity.

Two quantitative measures of the course were assessed: (1) grades include the four exams and a weighted average of the exams, to exclude online material differences between teaching methods, and (2) the letter grade submitted to the Registrar. The Weighted exam score averages the exams with Exam 4 worth 50% more than each of the earlier exams, that is, Weighted = 2/9 (Exams 1 + 2 + 3) + 3/9 (Exam 4). Note that the content and style of the exams has not changed in several years, thus Flipped delivery Exams 1, 2, and 3 were modeled exactly on exams from previous years of the traditional delivery course, and Exam 4 was a common final designed by the Traditional delivery coordinator for all sections. The Final Exam 4 is the most comparable outcome. Exam 4 was administered simultaneously for all sections (in the earliest time slot on the first day of Finals week). Furthermore, it was graded by all course instructors in the same room: each instructor joined a team table to grade a single page, they coordinated the grading of that page following a rubric, and they would repeatedly choose a stack of exams to grade for students whom they did not teach. Note that the exam-weighted

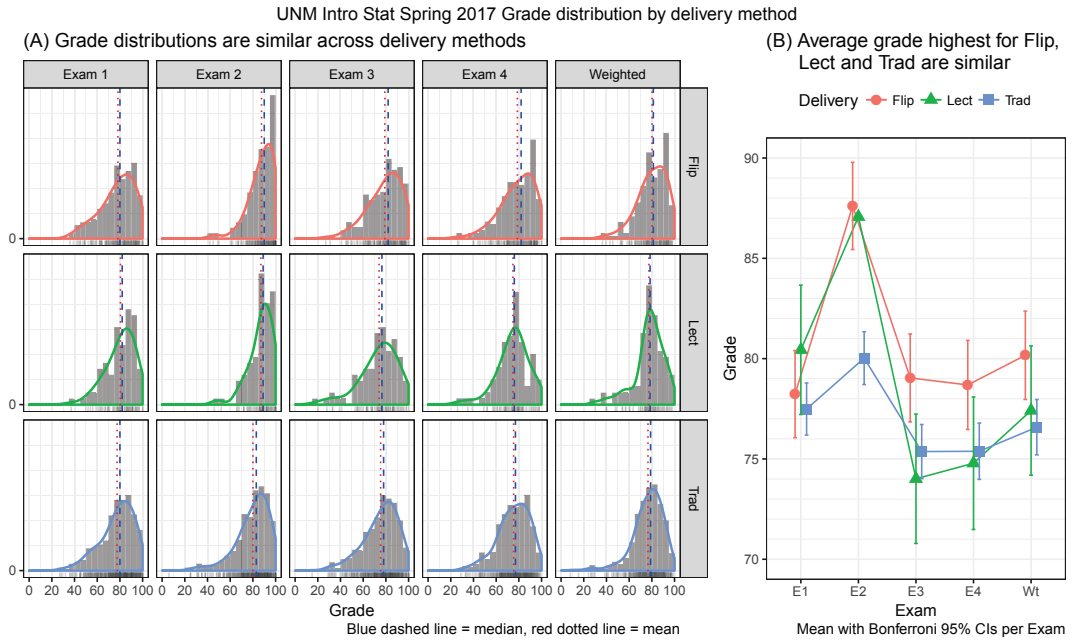

Figure 1: *(A) Grade distributions are similar across the flipped, lecture, and traditional delivery methods for Exams 1 through 4 and the weighted average of the exam scores, though average scores are slightly higher for flipped delivery. Sample sizes are Flip = 180, Lect = 84, and Trad = 459. (B) The line plot shows the mean with exam-wise Bonferroni 95% confidence intervals for each exam and the weighted exam average; it illustrates that the flipped delivery method is on average 3 to 4 points higher than traditional or lecture delivery methods. Sample sizes are Flip = 180, Lect = 84, and Trad = 459.*

scores also serve as important overall performance metrics for students who never purchased the online access to the e-book learning system — losing 10% to 15% of their grade — as there is no used textbook option for our students. The authors find this situation problematic since the online resources are expensive (roughly $100) and there are cases of disadvantaged students in our student population forgoing the online resources because the cost was prohibitively high.

4. Results

4.1. Summative outcomes

Notable results are shown in the following two figures. Figure 1(A) illustrates that grade distributions were similar across teaching methods (similar shape and spread) for the four exams and their weighted average, which was expected, though the means and medians differ slightly.

To focus on the differences between averages, Figure 1(B) shows that the Flipped delivery method was 3 to 4 points higher than Traditional and Lecture delivery methods on average for the Weighted average of Exams. This represents a 15% achievable gain (for the discussion of achievable gain in the physics classroom see Hake (2002)) from Traditional to Flipped delivery, where the achievable gain is defined as the proportion of possible additional improvement: $(80.2 - 76.6)/(100 - 76.6) = 15.4\%$.

The letter grades for Exam 4 and those submitted to the registrar are summarized in Figure 2. In these stacked barplots, cell areas are proportional to the numbers of respondents within each delivery

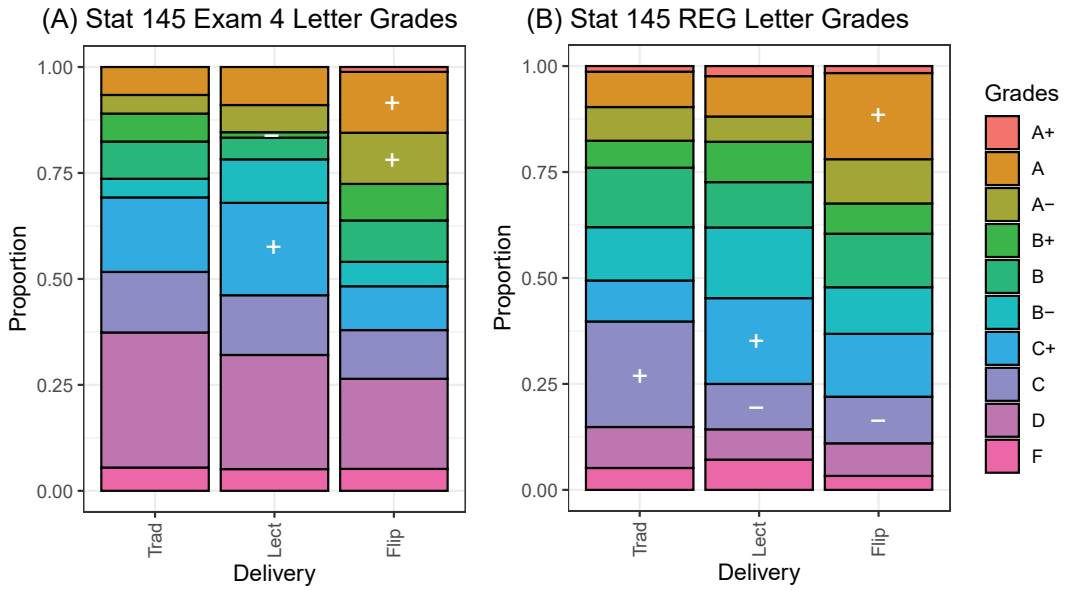

Figure 2: *The stacked barplots shows the proportions of students in each teaching delivery method receiving letter grades. (A) For Exam 4 letter grades, the best comparison between delivery methods, in the flipped delivery column, we observe more A− and above than B and above for traditional delivery or lecture delivery. Sample sizes are Flip = 177, Lect = 80, and Trad = 443. (B) For registrar letter grades, there are roughly twice as many As and half as many Cs in the flipped delivery compared to traditional delivery. The effective impact of the flipped delivery is that everyone improves by one-third of a letter grade. Sample sizes are Flip = 182, Lect = 84, and Trad = 755 (larger because these include all grades submitted to the registrar, not only shared instructor grade sheet data).*

method and white + and − signs indicate large Pearson residual from a chi-squared analysis contributing largely to differences between delivery methods. For Exam 4 letter grades (Figure 2(A)), the best comparison between delivery methods, in the flipped column, we observe more (A)− and above than (B) and above for traditional or lecture delivery methods. For registrar letter grades (Figure 2(B)), in the flipped column, we observe twice as many As and half as many Cs as the traditional delivery course, suggesting that the entire distribution of students in the flipped delivery classrooms achieved one-third of a letter grade higher than traditional delivery classrooms.

4.2. Questionnaire outcomes

We reviewed the student questionnaire responses and organized questions by category, seven categories of which are given in Table 2 (Table A.1 for all remaining questions). Statistical comparisons test whether the proportion of positive responses from flipped delivery students differs from traditional delivery students for each question. For Communication, all four were rated more highly by flipped delivery students; however, no individual measure was statistically significant. For Collaboration 1, significantly more flipped delivery students responded that they developed team work skills, had more discussion with others, and had more accessible time with the professor. For Collaboration 2, significantly more flipped delivery students experienced time in teams and more student/teacher interaction. Both methods have about a third of students wanting more professor accessibility, more

Table 2: Questionnaire outcomes with categorical responses collapsed to "positive" vs "neutral and negative," and the proportions were compared between flipped and traditional/lecture delivery using permutation testing. Questions were grouped by theme. Flipped delivery had 49 total respondents while traditional and lecture delivery had 152, though some nonresponses make the total frequencies slightly smaller for individual questions.

Question	Response category	Flip % (n)	Trad % (n)	p-value
Communication				
15. I feel I am better able to communicate with others.	Agree (5–7 of 7)	42% (21)	29% (43)	0.115
23. I feel I am better able to present my findings.	Agree (5–7 of 7)	56% (28)	40% (59)	0.068
25. I was able to see good ways of presenting information.	Agree (5–7 of 7)	58% (29)	48% (71)	0.270
33. The class improved my skills in written communication.	Agree (5–7 of 7)	39% (19)	23% (35)	0.054
Collaboration 1				
03. This class helped me to develop my team working skills.	Agree (5–7 of 7)	38% (19)	22% (33)	0.036
How often activity experienced: Discussion with others	once/week or more	85% (39)	37% (51)	< 0.001
How often activity experienced: Time for professor accessible (questions/answers)	once/week or more	94% (44)	78% (107)	0.030
What is your activity preference: Discussion with others	More	17% (8)	29% (40)	0.155
What is your activity preference: Professor accessibility (questions/answers)	More	36% (17)	32% (44)	0.719
Collaboration 2				
How often collaboration experienced: In teams of 2-4 students	once/week or more	51% (24)	19% (26)	< 0.001
How often collaboration experienced: Whole-class discussion	once/week or more	53% (25)	45% (61)	0.391
How often collaboration experienced: Student/teacher interaction	once/week or more	77% (36)	58% (80)	0.035
What is your collaboration preference: In teams of 2-4 students	More	32% (15)	39% (54)	0.478
What is your collaboration preference: Whole-class discussion	More	36% (17)	28% (39)	0.403
What is your collaboration preference: Student/teacher interaction	More	43% (20)	38% (52)	0.676
Feedback				
How often feedback experienced: Verbal feedback from instructor	once/week or more	85% (39)	56% (76)	0.001
How often feedback experienced: Verbal feedback from peer learning facilitator	once/week or more	63% (29)	12% (16)	< 0.001
How often feedback experienced: Verbal feedback from peers	once/week or more	48% (22)	30% (40)	0.039
What is your feedback preference: Verbal feedback from instructor	More	43% (20)	40% (54)	0.838
What is your feedback preference: Verbal feedback from peer learning facilitator	More	30% (14)	25% (33)	0.580
What is your feedback preference: Verbal feedback from peers	More	22% (10)	28% (37)	0.557
Course structure				
06. The instructor focused more on encouraging me to find information than on giving me facts.	Agree (5-7 of 7)	59% (29)	37% (56)	0.012
30. I usually had a clear idea of where I was going and what was expected of me in this class.	Agree (5-7 of 7)	76% (38)	65% (98)	0.220
Activities				
How often activity experienced: Assignment completion	once/week or more	94% (44)	76% (103)	0.017
What is your activity preference: Assignment completion	More	51% (24)	31% (43)	0.023
How often activity experienced: Lectured more than one-third of class time	once/week or more	77% (36)	92% (126)	0.011
How often activity experienced: Lectured less than one-third of class time	once/week or more	36% (17)	12% (17)	< 0.001
What is your activity preference: Lecture	More	34% (16)	14% (20)	0.007
Technology				
How often activity experienced: Utilizing technology	once/week or more	62% (29)	58% (79)	0.793
What is your activity preference: Utilizing technology	More	19% (9)	23% (32)	0.709
How often technology experienced: Personal laptops/pads	once/week or more	35% (16)	27% (37)	0.446
How often technology experienced: Instructor uses statistical software	once/week or more	61% (28)	33% (45)	0.002
How often technology experienced: Students use statistical software	once/week or more	30% (14)	27% (36)	0.762

Continued

Continued

Question	Response category	Flip % (n)	Trad % (n)	p-value
How often technology experienced: Presentation software (PowerPoint)	once/week or more	20% (9)	42% (57)	0.010
How often technology experienced: Paper handouts	once/week or more	98% (45)	34% (46)	< 0.001
What is your technology preference: Personal laptops/pads	More	13% (6)	32% (43)	0.022
What is your technology preference: Instructor uses statistical software	More	26% (12)	36% (49)	0.292
What is your technology preference: Students use statistical software	More	17% (8)	44% (60)	0.002
What is your technology preference: Presentation software (PowerPoint)	More	24% (11)	30% (40)	0.579
What is your technology preference: Paper handouts	More	30% (14)	43% (57)	0.203

teamwork, and more whole-class discussion. For Feedback, significantly more flipped delivery students responded that they had feedback from the instructor, from the peer-learning facilitator (PLF), and from peers. Both methods have 40% of students wanting more instructor feedback, and 25% wanting more PLF and peer feedback. For Course structure, significantly more flipped delivery students were encouraged to find information, while both methods could improve on setting student class expectations. For Activities, significantly more flipped delivery students experienced assignment completion, preferred assignment completion, and had lectures lasting less than one-third of the class time. A third of flipped delivery students would prefer more time for lectures. For Technology, significantly more flipped delivery students experienced the instructor using statistical software, the use of paper handouts; however, fewer flipped delivery students experienced presentation software and wanted to use statistical software.

We also analyzed students' verbatim feedback by asking three free response questions and identifying themes of major appreciation or concern throughout the course by teaching method (Table 3). Regarding the benefits of the course (Prompt 1), students in traditional delivery commented on the kind of knowledge (knowing) gains followed by skills, while students in flipped delivery addressed skills (doing) gained in the course followed by knowledge, learning process, and their views of statistics. In response to the question about the effective features of each teaching method (Prompt 2), students in traditional delivery picked lecture and homework as important aspects of their learning, while students in flipped delivery mentioned doing statistics through assignments and identified learning independently as effective aspects of the course. In addition to the difficulty of statistics as content, students in traditional delivery identified quizzes and tests as well as the difficulty of using the e-book as challenges in the course (Prompt 3). On one hand (Prompt 3), students in flipped delivery similarly mentioned assessment and content as the difficult aspect of learning elementary statistics; on the other hand (Prompt 2), unlike the students in traditional delivery, students in flipped delivery gave little mention about the difficulties experienced during the lectures.

5. Discussion

5.1. Reflecting on the comparison of two teaching methods

The entire distribution of students in flipped delivery performed better than students in traditional delivery. Though there are not differences on the first exam, flipped delivery did better on the next three exams including the final exam and the overall course grade. Regarding the outcome of Exam 1, we interpret the little difference for Exam 1 as evidence that both groups of students started similarly and the warm-up period accounts for students adjusting to the flipped delivery teaching method. It could

Table 3: Percent of student responses to a prompt on course experiences and representative comments

Themes	Flipped (n = 49)	Traditional (n = 136)	p-value
(Prompt 1: The *most useful thing/skill* I learned was . . .)			
Knowledge	30% (n = 15) "The various types of statistical tests and when to use them."	40% (n = 54) "The chi-square test" "The different formulas in each chapter."	0.191
Skills	43% (n = 21) "Being able to look at data and analyze/understand"	30% (n = 41) "Finding the P-values was statistically significant."	0.152
Learning process	14% (n = 7) "Work in groups and communicate with people, problem-solving skills."	8% (n = 11) "How to find the answers to my questions on my own"	0.323
Views towards statistics	11% (n = 5) "How to solve statistical problems that will later come up in my profession."	2% (n = 3) "How to view statistics as a whole"	0.062
(Prompt 2: What made learning *most effective* for me was. . .)			
Lecture	17% (n = 8) "In class examples" "Good explanations"	60% (n = 82) "The breakdown by the professor and the before exam reviews"	< 0.001
Doing statistics in class	19% (n = 9) "Asking questions and having the professor and the TA walking around while we did our assignments so that they could answer any questions that we had and correct our mistakes"	13% (n = 18) "Doing problems in class with the professor"	0.517
Homework	18% (n = 9) "Mandatory stats tutor videos, all of LaunchPad, flipped in-class assignments and lecture. Also the in-class worksheets" "LaunchPad Homework assignments"	18% (n = 24)	0.935
Learning independently	11% (n = 5) "The class set up. How we did stat tutors, then lectures, then applied both of those for worksheets and was able to ask questions if needed."	4% (n = 5) "Learning the information on my own using resources from LaunchPad"	0.183
Group work	6% (n = 3) "Handouts, peer work, and feedback"	5% (n = 7) "Definitely working in groups was the most helpful thing when it came to understanding the material."	0.955
(Prompt 3: The thing I found *most difficult* was. . .)			
Assessment	32% (n = 16) "The weekly quizzes."	41% (n = 56) "Some of the quizzes on LaunchPad. The online quizzes. They are much harder than the problems you go over in class so sometimes you had no idea how to do the problems"	0.217
Lecture	4% (n = 2) "Sitting through class"	24% (n = 33) "The class itself the lectures were a bit too much." "The teacher moved at a really fast pace and I found myself lost most of the time."	< 0.001
Using the e-book	19% (n = 9) "Trying to use [Launchpad] and the homesoftware. [Having to use] very specific terminology for answers, [and] listening to the Stat Tutors"	4% (n = 5) "The Launchpad system was not good. The teaching techniques were not the same as in the class and the homework/quiz questions were difficult to understand"	0.0049

Continued

Continued

Themes	Flipped (n = 49)	Traditional (n = 136)	p-value
Content	12% (n = 6) "Chapter 17"	17% (n = 23) "Chi squared"	0.317
Learning process	7% (n = 3) "Trying to learn something on my own when I couldn't really understand the instructor"	4% (n = 5) "The vocabulary and concepts and where to apply them for which situation"	0.713

be that students in the flipped delivery classrooms needed time to adjust to an unfamiliar teaching method where students are expected to put in more effort on previewing materials before the lecture so that they are ready to engage with class materials. Rather than simply taking notes and absorbing delivered knowledge, the flipped delivery classroom required students to ask questions for clarification and share ideas to extend their knowledge during the lecture, which was a different mindset of learning for students. Regarding the outcome of Exam 2, we note that scores for the lecture group's Exam 2 were high possibly due to an easier exam than for the traditional delivery sections. Eventually, the higher performance by the flipped delivery group on the final exam is quite telling and lends credence to the view that our students can and do perform better when they learn in a classroom that puts greater emphasis on using real data with a meaningful context, using active learning, using conceptual understanding, and using problem-solving as integrated practice.

Student feedback on the questionnaire is useful for understanding the specific nature of student learning in flipped delivery in contrast to learning through traditional delivery. Out of 40 items exploring students' learning experiences, we identified nearly half (19 items) to which students in these two teaching methods responded differently (at a 0.05 level and uncorrected for multiple comparisons). In particular, student responses suggest that the flipped delivery course fostered more collaboration, analytical, communicative, and problem-solving skills than the traditional delivery course. These skills are closely aligned to the six pedagogical GAISE recommendations set forth by the American Statistical Association (Carver *et al.*, 2016). We do not mean to argue that traditional delivery approaches inhibit the learning of statistics, rather, we feel strongly that flipped delivery courses built on the GAISE recommendations can impact students' summative performance as well as the traditional delivery course design. A flipped delivery approach, when implemented with careful consideration of the appropriate content of the course, can afford more meaningful learning opportunities that foster related statistical literacy skills and nurture a productive mindset towards statistics.

Student responses to free response prompts suggest that both traditional and flipped delivery methods can increase student knowledge and skills if implemented appropriately. However, we found nuance in the student responses in that students in flipped delivery thought their statistical skills were slightly more useful than just functioning as knowledge. These students demonstrated a higher awareness of the process of learning statistics and had different views towards statistics than the students in traditional delivery. Of interest to note is that students in flipped delivery perceived the class as more interesting and perceived that it made time pass faster than traditional delivery ($p = 0.0002$) since flipped delivery focused on activities. Additionally, students in flipped delivery expressed more difficulty in using an e-book than students in traditional delivery ($p = 0.0049$). We suspect that this struggle was due to the fact that flipped delivery used the e-book prior to class meetings. Flipped delivery emphasized important points and resolved misunderstandings in class after the first contact with the e-material; however, the initial struggle might have left an impression of difficulty in using the e-book for the course.

With the premise that introductory statistics in college serves as an entry (or exit) point for students considering their academic potential in pursuing STEM fields, it is significant to report that flipped

delivery, when aligned to GAISE, may better support our students in developing statistical literacy and productive learner mindsets for more advanced statistics than traditional delivery.

Acknowledgements

EBE supported by UNM Center for Teaching and Learning while a UNM Teaching Fellow and ASA Statistics Education Mentee.

Appendix:

The remaining questionnaire responses are given in Table A.1.

Table A.1: The remaining Questionnaire Outcomes not discussed in the text to complement Table 2

Question	Response category	Flip % (n)	Trad % (n)	p-value
01. I learned about how to present my findings to an audience.	Agree (5–7 of 7)	35% (17)	34% (52)	1
02. I found this class difficult.	Agree (5–7 of 7)	37% (18)	46% (69)	0.351
07. The class was more about analyzing and evaluating information than it was about memorizing it.	Agree (5–7 of 7)	74% (37)	63% (96)	0.219
09. This class helped me to discover what was expected of me as a learner.	Agree (5–7 of 7)	58% (29)	46% (70)	0.206
10. I enjoyed working in this class.	Agree (5–7 of 7)	60% (30)	50% (76)	0.306
13. I didn't need to apply anything I learned.	Agree (5–7 of 7)	14% (7)	11% (16)	0.690
14. There was a lot to learn.	Agree (5–7 of 7)	74% (37)	78% (116)	0.715
16. I felt I had to work hard to complete this class.	Agree (5–7 of 7)	70% (35)	66% (99)	0.728
17. I felt I was able to take more responsibility for my own learning.	Agree (5–7 of 7)	72% (36)	70% (105)	0.879
19. I found the class challenging.	Agree (5–7 of 7)	56% (28)	60% (91)	0.714
20. I feel more confident in my ability to solve problems.	Agree (5–7 of 7)	74% (37)	63% (95)	0.228
21. I felt I could get through the class simply by memorizing things.	Agree (5–7 of 7)	24% (12)	29% (44)	0.585
22. I felt a sense of control over my learning.	Agree (5–7 of 7)	62% (31)	53% (79)	0.325
26. It was always easy to know the standard of work expected.	Agree (5–7 of 7)	70% (35)	68% (100)	0.886
27. The class developed my problem-solving skills.	Agree (5–7 of 7)	62% (31)	61% (91)	1
28. The workload was too heavy.	Agree (5–7 of 7)	42% (21)	26% (39)	0.053
29. The class sharpened my analytic skills.	Agree (5–7 of 7)	60% (30)	55% (83)	0.681
31. To do well in this class all you really needed was a good memory.	Agree (5–7 of 7)	24% (12)	27% (41)	0.781
32. As a result of my class, I feel confident about tackling unfamiliar problems.	Agree (5–7 of 7)	46% (23)	44% (66)	0.935
36. I was generally given enough time to understand the things I had to learn.	Agree (5–7 of 7)	58% (29)	54% (81)	0.743
37. The assessment in this class required an in-depth understanding of the course content.	Agree (5–7 of 7)	68% (34)	68% (102)	1
39. The sheer volume of work to be got through in this class meant it couldn't all be thoroughly comprehended.	Agree (5–7 of 7)	38% (19)	30% (45)	0.381
40. I could recommend this class to a fellow student.	Agree (5–7 of 7)	66% (33)	57% (85)	0.319

References

Aliaga M, Cobb G, Cuff C, *et al.* (2005). Guidelines for assessment and instruction in statistics education (GAISE): College report. Available from: http://www.amstat.org/education/gaise/

Alldredge JR and Brown GR (2006). Association of course performance with student beliefs: An analysis by gender and instructional software environment, *Statistics Education Research Journal*, **5**, 64–77.

Bidgood P, Hunt N, and Jolliffe F (2008). ViSA-variety in statistics assessment, *MSOR Connections*, **8**, 25–26.

Blumenfeld PC, Soloway E, Marx RW, Krajcik JS, Guzdial M, and Palincsar A (1991). Motivating project-based learning: Sustaining the doing, supporting the learning, *Educational Psychologist*, **26**, 369–398.

Budgett S and Pfannkuch M (2007). Assessing students statistical literacy. *International Association for Statistical Education (IASE/ISI Satellite, 2007 Portugal)*.

Carver R, Everson M, Gabrosek J, *et al.* (2016). *Guidelines for assessment and instruction in statistics education (gaise) college report 2016*. VA: American Statistical Association, Alexandria. Available from: http://www.amstat.org/education/gaise/

Chelimsky E and Shadish WR (1997). *Evaluation for the 21st Century: A Handbook*, Sage Publications, Thousand Oaks.

Dierker L, Alexander J, Cooper JL, *et al.* (2016). Engaging diverse students in statistical inquiry: A comparison of learning experiences and outcomes of under-represented and non-underrepresented students enrolled in a multidisciplinary project-based statistics course, *International Journal for the Scholarship of Teaching and Learning*, **10**, 1–9.

Everson M, Zieffler A, and Garfield J (2008). Implementing new reform guidelines in teaching introductory college statistics course, *Teaching Statistics*, **30**, 66–70.

Forte JA (1995). Teaching statistics without Sadistics, *Journal of Social Work Education*, **31**, 204–218.

Franklin C, Kader G, Mewborn D, Moreno J, Peck R, Perry M, and Scheaffer R (2007). Guidelines for assessment and instruction in statistics education (gaise) report, American Statistical Association. Available from: http://www.amstat.org/education/gaise/GaiseCollege_Full.pdf

Gal I and Ginsburg L (1994). The role of beliefs and attitudes in learning statistics: Towards an assessment framework, *Journal of Statistics Education*, **2**, 1–16.

Garfield J and Ben-Zvi D (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics, *International Statistical Review*, **75**, 372–396.

Garfield J, delMas R, and Chance B (1999). Using technology to improve statistical reasoning. In *Proceedings of the JMM. Joint Mathematics Meetings*. Available from: http://www.tc.umn.edu/~delma001/stat_tools/

Garfield J, delMas R, and Zieffler A (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course, *The International Journal on Mathematics Education*, **44**, 883–898.

Garfield JB (1994). Beyond testing and grading: Using assessment to improve student learning, *Journal of Statistics Education*, **2**, 1–11.

Garfield JB, Ben-Zvi D, Chance B, Medina E, Roseth C, and Zieffler A (2008). Creating a statistical reasoning learning environment. In *Developing Students Statistical Reasoning*, Springer, 45–63.

Gelman A and Carlin J (2017). Some natural solutions to the p-value communication problem — and why they wont work, *Journal of the American Statistical Association*, **112**, 899–901.

Gelman A and Nolan D (2002). *Teaching Statistics: A Bag of Tricks*, Oxford University Press, Oxford.

Giraud G (1997). Cooperative learning and statistics instruction, *Journal of Statistics Education*, **5**, 1–12.

Gould R (2017). Data literacy is statistical literacy, *Statistics Education Research Journal*, **16**, 22–25.

Haak DC, HilleRisLambers J, Pitre E, and Freeman S (2011). Increased structure and active learning reduce the achievement gap in introductory biology, *Science*, **332**, 1213–1216.

Hake R (2002). Lessons from the physics education reform effort, *Conservation Ecology*, **5**, 1–61.

Hall MR and Rowell GH (2008). Introductory statistics education and the national science foundation, *Journal of Statistics Education*, **16**, 1–10.

Hassad RA (2011). Constructivist and behaviorist approaches: Development and initial evaluation of a teaching practice scale for introductory statistics at the college level, *Numeracy*, **4**, 7.

Hildreth LA, Robison-Cox J, and Schmidt J (2018). Comparing student success and understanding in introductory statistics under consensus and simulation-based curricula, *Statistics Education Research Journal*, **17**, 103–120.

Horton NJ (2015). Challenges and opportunities for statistics and statistical education: looking back, looking forward, *The American Statistician*, **69**, 138–145.

Johnson HD and Dasgupta N (2005). Traditional versus non-traditional teaching: Perspectives of students in introductory statistics classes, *Journal of Statistics Education*, **13**, 1–14.

Lane DM and Tang Z (2000). Effectiveness of simulation training on transfer of statistical concepts, *Journal of Educational Computing Research*, **22**, 383–396.

Macdonald D, Mitchell J, and Mayer D (2006). Professional standards for physical education teachers professional development: technologies for performance? *Physical Education and Sport Pedagogy*, **11**, 231–246.

Magel RC (1998). Using cooperative learning in a large introductory statistics class, *Journal of Statistics Education*, **6**, 1–9.

Mislevy RJ, Steinberg LS, and Almond RG (2003). Focus article: On the structure of educational assessments, *Measurement: Interdisciplinary Research and Perspectives*, **1**, 3–62.

Moore DS, Notz WI, and Fligner MA (2015). *The Basic Practice of Statistics* (7th ed), WH Freeman, New York.

Moore I (2003). *Evaluating a materials course* (Technical report), UK Centre for Materials Education. Available from: http://www.materials.ac.uk/guides/4-evaluating.pdf

Neumann DL, Hood M, and Neumann MM (2013). Using real-life data when teaching statistics: student perceptions of this strategy in an introductory statistics course, *Statistics Education Research Journal*, **12**, 1–12.

Nielsen PL, Bean NW, and Larsen RAA (2018). The impact of a flipped classroom model of learning on a large undergraduate statistics class, *Statistics Education Research Journal*, **17**, 121–140.

Paul W and Cunnington RC (2017). An exploration of student attitudes and satisfaction in a gaisein-fluenced introductory statistics course, *Statistics Education Research Journal*, **16**, 487–510.

Ramsden P (2003). *Learning to Teach in Higher Education*, Routledge, London.

Rinaman WC (1998). Revising a basic statistics course, *Journal of Statistics Education*, **6**, 1–14.

Rumsey DJ (1998). A cooperative teaching approach to introductory statistics, *Journal of Statistics Education*, **6**, 1–24.

Rumsey DJ (2002). Statistical literacy as a goal for introductory statistics courses, *Journal of Statistics Education*, **10**, 1–12.

Saunders M (2000). Beginning an evaluation with RUFDATA: theorizing a practical approach to evaluation planning, *Evaluation*, **6**, 7–21.

Utts J (2015). The many facets of statistics education: 175 years of common themes, *The American Statistician*, **69**, 100–107.

Winquist JR and Carlson KA (2014). Flipped statistics class results: Better performance than lecture over one year later, *Journal of Statistics Education*, **22**, 1–10.

Zieffler A, Garfield J, Alt S, Dupuis D, Holleque K, and Chang B (2008). What does research suggest about the teaching and learning of introductory statistics at the college level? a review of the literature, *Journal of Statistics Education*, **16**, 1–25.

Received August 26, 2019; Revised December 18, 2019; Accepted January 10, 2020