

Project-Based Learning in Introductory Statistics: Comparing Course Experiences and Predicting Positive Outcomes for Students from Diverse Educational Settings

Lisa Dierker^{1*}

Jane Robertson Evia²

Karen Singer-Freeman³

Kristin Woods⁴

Janet Zupkus⁵

Alan Arnholt⁶

Elizabeth G Moliski⁷

Natalie Delia Deckard⁸

Kristel Gallagher⁹

Jennifer Rose¹⁰

¹Psychology, Department, Wesleyan University, Middletown, CT, US.

Email: ldierker@wesleyan.edu

² Department of Statistics, Virginia Tech University, Blacksburg, VA, US.

³ Psychology Department, SUNY-Purchase, Purchase, NY, US.

⁴ Department of Psychology, Southwestern Oklahoma State University, Weatherford, OK, US.

⁵ Mathematics Department, Naugatuck Valley Community College, Waterbury, CT, US.

⁶ Department of Mathematical Sciences, Appalachian State University, Boone, NC, US.

⁷ Department of Information, Risk, and Operations Management, University of Texas, Austin, TX, US.

⁸ Sociology Department, Davidson College, Davidson, NC, US.

⁹ Department of Psychology, Thiel College, Greenville, PA, US.

¹⁰ Center for Pedagogical Innovation, Wesleyan University, Middletown, CT, US.

Abstract

In order to evaluate the acceptability and potential impact of the Passion-Driven Statistics curriculum, this article describes background characteristics, and course experiences and outcomes of students enrolled in the multidisciplinary, introductory, project-based course in liberal arts colleges, large state universities, regional college/universities, and community colleges. We found that the course could be successfully delivered across these diverse educational settings. After controlling for educational setting and pre-survey responses to individual outcome measures, consistent predictors of positive course outcomes included student's initial interest in conducting research, their higher likelihood of enrolling in a statistics course if it were not required, finding the project-based course less challenging, and finding the research project more rewarding than other students. Regional college/university, and community college students reported working significantly harder in the course and finding the course more challenging than students taking the course at liberal arts colleges or state universities. Students from liberal arts colleges generally reported more positive course experiences than students from other educational settings. However, when compared to students from both liberal arts colleges and large state universities, those from regional colleges/universities reported being more likely to have learned more in the project-based course than in other college courses they had taken. Taken together, the project-based course was successfully delivered across diverse post-secondary educational settings and provides a promising model for getting students hooked on the power and excitement of applied statistics.

Keywords:

Teaching statistics
Passion-driven statistics
Project-based learning
Science
technology
engineering and mathematics
(STEM)
Flipped classrooms.

Licensed:

This work is licensed under a
Creative Commons Attribution
4.0 License.

Publisher:

Scientific Publishing Institute

1. Introduction

A central challenge of introductory statistics courses is the development of a curriculum that not only serves diverse students, but also sparks communication, reasoning, and collaboration that crosses traditional disciplinary boundaries. Many students express negative feelings about introductory statistics courses and emerge from these classes with few useful skills (Delia, 2017).

Petocz and Reid (2005) interviewed undergraduate students who had recently completed an introductory statistics class. They found that students who described the class as focusing on statistical techniques did not believe that statistical analyses were a useful way to evaluate information and planned to avoid the use of statistics in their professional lives. In contrast, students who described the class as focusing on analyzing and interpreting data were more likely to believe that statistical techniques could be used to understand real issues in a wide range of areas.

Bailey, Spence, and Simm (2013) suggest that developing analytic statistical thinking can best be achieved through inquiry-based projects. This type of project-based learning is most commonly defined as an instructional approach based on authentic, real-world activities that are aimed at engaging student interest and enthusiasm (Buck Institute for Education (BIE) and Boise State University, 2012; Krajcik & Blumenfeld, 2006).

Designed to answer a question or solve a problem, project-based learning allows students to face challenges that lead to answers, reflect on ideas, and make decisions that affect project outcomes (Aditomo, Goodyear, Bliuc, & Ellis, 2013). There is an emerging literature demonstrating that project-based learning is more effective in promoting deep thinking—the ability to apply knowledge, communication, and reasoning skills—when compared to traditional didactic approaches (Hickey, Kindfield, Horwitz, & Christie, 1999; Hickey, Wolfe, & Kindfield, 2000; Hmelo-Silver, Duncan, & Chinn, 2007; Lynch, Kuipers, Pyke, & Szesze, 2005; Walker & Leary, 2009).

In previous publications, we have described the development of a multidisciplinary, project-based introductory statistics course aimed at engaging students in applied statistical projects across both divisional and departmental boundaries (Dierker, Kaparakis, Rose, & Selya, 2012).

Funded by the U.S. National Science Foundation and first introduced into the curriculum at a selective liberal arts college, the project-based course follows each of the recommendations outlined in the Guidelines for Assessment and Instruction in Statistics Education (GAISE) report (Aliaga et al., 2005; American Statistical Association [ASA], 2014; Carver et al., 2016). Students choose their own research project, giving them time to think critically about statistical issues (Chance 2002), recognize the usefulness of data for answering questions of interest to them and to society (Neumann, Hood, & Neumann, 2013) tackle complicated real-world questions that involve more than one or two variables (De Veaux, 2015) and emphasize practical problem-solving skills that are necessary to answer statistical questions (Garfield, delMas, & Zieffler, 2012).

By supporting students as they make decisions about the most appropriate ways to visualize, explore, and analyze data, the course emphasizes conceptual understanding and application.

Rather than focusing on rules associated with traditional lists of statistical tools, we organized the course to focus on the decisions and skills involved in statistical inquiry. Basic themes such as measurement and descriptive and graphical representation are covered, as well as more specific inferential methods needed to test hypotheses and/or explore the empirical structure of data (Cobb, 2007; Gould, 2010; Horton, 2015). All, however, are introduced as the students' scientific questions dictate.

In this way, students are provided with opportunities to learn to select the tools that are most appropriate to address their research question(s) and to engage in statistical decision making. Using leading statistical software (e.g. R, SAS, Stata, SPSS, etc.) students are exposed to a variety of statistical methods (ANOVA, Chi Square Test of Independence, Pearson Correlation, etc.) and learn to choose and use them flexibly as they are needed.

Based on data collected at the liberal arts college where the curriculum was originally developed, we have demonstrated that the project-based course enrolls higher numbers of under-represented minority (URM) students compared to a traditional introductory statistics course (Dierker, Cooper, Selya, Alexander, & Rose, 2015). Further, while URM students considered the material in the project-based course more difficult than non-URM students, they demonstrated similar levels of increased confidence in applied skills and interest in follow up courses.

URM students were found to be twice as likely as non-URM students to report that their interest in conducting research increased after completing the project-based course (Dierker et al., 2016). Compared to students enrolled in a traditional introductory statistics course, those completing the project-based course have also been shown to be significantly more likely to report an increase in confidence between a pre and post course survey with regard to choosing the correct statistical test, managing data, and writing syntax or code to run statistical analyses. Project-based students were also more likely than students in the traditional course to show an increase in their interest in pursuing advanced course-work in statistics (Dierker et al., under review).

Because of a focus on programming in the context of data analysis, we have previously compared enrollment of the project-based course to traditional introductory programming experiences, revealing higher rates of female and URM enrollment compared to both a general introductory programming course and an introductory course representing a gate to the computer science major (Cooper & Dierker, 2017).

Overall, when compared to both traditional introductory statistics and introductory programming courses, the project-based course has been shown to attract students from a much wider range of academic backgrounds as measured by their Math SAT scores (Cooper & Dierker, 2017; Dierker et al., 2015).

In recent years, we have begun to disseminate the project-based curriculum across diverse post-secondary educational settings, providing workshops for faculty and making our project-based resources freely available (<http://passiondrivenstatistics.com>).

The present paper describes preliminary findings from these dissemination efforts by describing student background characteristics, as well as course experiences and outcomes, of those enrolling in the project-based introductory statistics course at liberal arts colleges, large state universities, regional state colleges/universities and community colleges.

2. Method

2.1. Participants

Pre and post course survey data from 1395 students enrolled in a project-based statistics course between fall of 2014 and spring of 2017 at one of 10 post-secondary institutions were examined. These participants included 446 students from one of 5 liberal arts colleges, 345 students from one of 2 large state universities, 525 students from one of 2 regional colleges/universities and 79 students from one of 2 community colleges. The average enrollment per class section across educational settings was 37.5 (mode=28, range: 10-130 students).

2.2. Project-Based Course Features

At each school, the multidisciplinary, project-based curriculum was delivered in the context of research projects of students' own choosing and individualized hands-on experience in data analysis and applied statistics was offered throughout the semester.

The curriculum focused on the decisions and skills involved in asking and answering questions with data with the goal of telling accurate and engaging stories. To accomplish this, students were provided with one or more large data sets and corresponding data dictionaries.

These data sets included Wave 1 of the U.S. National Epidemiologic Survey of Alcohol and Related Conditions, Wave 1 or Wave 4 from the U.S. National Longitudinal Study of Adolescent Health, the Outlook on Life Survey, the Youth Risk Behavior Survey, World Health/World Bank data, and a Mars Craters dataset.

Based on the student's choice of data, each student generated testable hypotheses, prepared data for analysis, selected and used descriptive and inferential statistical tools; and evaluated, interpreted, and presented research findings, orally, graphically, and/or as text. Final projects most often included a scientific poster presentation providing an empowering capstone experience allowing students to communicate methods, results, and insights not only to data experts but also to non-scientific audiences. In the absence of a poster presentation, final projects included electronic submission of scientific posters, final lab reports, and/or research proposals.

Each project-based course was considered introductory in that no prior training or coursework in statistics was required. In some cases, a quantitative reasoning course, high school algebra, pre-calculus, a passing grade on the university mathematics placement test, and/or a general psychology course was a required prerequisite for enrollment in the project-based course.

Course titles within the various educational settings included Introduction to Statistics, Psychological Statistics, Applied Data Analysis, Introduction to Data Analytics, Statistics for Social Sciences, Principles of Statistics, or Research Methods I: Design and Statistics.

Each course was semester-long, meeting 1, 2, or 3 times each week (i.e. 1 time per week for 3 hours, 2 times per week for 75, 80 or 85 minutes in each meeting, 3 times per week for 50 minutes or 1 hour in each meeting, or 1 time per week for 3 hours in addition to two 90-minute class sessions each week). Several instructors chose to implement a fully flipped-class model in which content and demonstrations were reviewed as text and/or videos outside of class and class time was used wholly to support students in work on their projects.

Others devoted approximately half of each class session to lectures and half to support of students' work. Each instructor chose the statistical software platform to be used in their course. These included R, SAS, Stata, JMP, SPSS, or StatCrunch.

At each school, the project-based course typically counted toward or was required by one or more majors including sociology, psychology, political science, human development, criminal justice, business, nursing, biology, and/or neuroscience and behavior.

The course also fulfilled a general education requirement in areas designated as natural science and mathematics, quantitative and symbolic reasoning, scientific and quantitative reasoning, quantitative literacy, or communication and symbols.

2.3. Measures

Data were drawn from student surveys completed before (pre) and after (post) each course. The pre course survey was completed prior to the end of the second week of classes and the post course survey was completed during the last weeks of the semester. Each survey took approximately 10-15 minutes to complete.

Background characteristics Age and gender were available from the course surveys. Reports of race/ethnicity, including White, Black, Hispanic, Asian and Other, were not mutually exclusive. Students also self-reported whether or not they were the first generation of their family to attend college and whether or not they were eligible for free or reduced lunch in high school.

Academic background was assessed by questions about previous statistics and programming courses in the pre-survey. Students were asked whether they had previously taken a statistics course (advanced placement or non-advanced placement statistics in high school or a college level statistics course). Any prior experience with general programming and/or code-based statistical software was also evaluated in the pre course survey (e.g., R, SAS, Stata, Matlab or Python, C++, Java, Java Script, HTML, etc.). Class year was dichotomized into 1st and 2nd year vs. upper-classmen.

A mutually exclusive hierarchy was created to categorize students by college major. A hierarchy of groups reflecting the following five mutually exclusive categories included: 1) science majors (e.g., biology, neuroscience, chemistry, physics, geology and environmental science, mathematics) with or without an additional major; 2) social science majors (e.g., psychology, economics, political science, sociology) without a science major; 3) art and humanities majors (e.g., history, music, ethnic studies, languages, anthropology, studio art, gender studies, English, journalism) without a science or social science major; 4) business majors (e.g. business administration, accounting, finance, marketing, advertising, management, international business, public relations) and 5) majors in the health professions (e.g., nursing, occupational therapy, physical therapy, dental hygiene, radiology).

Additional variables measuring how students were approaching the project-based course included whether or not the course was required ("Please indicate the primary reason that you enrolled in this course: required for my major or the major I intend to pursue", endorsed or unendorsed), the likelihood of taking a statistics class if it were not required ("How likely is it that you would have taken any course in statistics if such a course was not required for your studies", 1=very unlikely to 5=very likely) and interest in conducting research ("Are you interested in conducting research?", 1=not at all interested to 4=very interested).

Experiences with the project-based course. Based on the post course survey, students rated how well they understood the information presented in course resources (1=none of it to 5=all of it) and the usefulness of those resources (1=not at all useful to 5=extremely useful).

Course resources included instructor lectures, video presentations, and one-on-one support from instructor(s). Students were also asked about the amount of individualized support and how useful they perceived that individualized support to be relative to their experiences in other courses, dichotomized as more versus less or the same.

Students' perceptions of their engagement in the course were measured by questions about their preparation ("How frequently did you prepare by completing assigned materials before class?" on a scale from 1 = never to 5 = always) and their effort ("How hard did you work in this course?" on a scale from 1 = not at all hard to 5 = extremely hard). Students' perceptions of rigor were measured with the questions "How challenging did you find this course?" (from 1 = not at all to 5 = the most challenging) and "Was this course more challenging, less challenging or similarly challenging compared to other college courses you have taken?"

Overall impressions of the course were measured with the questions "How rewarding did you find the research project?" (1=not at all rewarding to 5=the most rewarding project I ever completed), "Did you accomplish more than you expected, less than you expected, or about the same as you expected?", "Did you find the course more interesting, less interesting, or the same compared to other college courses you have taken?", "Did you find this course more useful, less useful, or similarly useful compared to other college courses you have taken?", and "In terms of how much you learned, compared to other courses that you have taken at your school, in this course, did you learn more, less or the same amount?". All questions comparing experiences in the statistics course to other college courses were dichotomized as more versus less or the same.

Confidence and interest in future experiences. Increases in perceived confidence on specific data analysis and statistical skills were evaluated based on changes in student ratings from the pre to post course survey, using a scale of 1 (not at all confident) to 4 (very confident).

These skills were developing a research question, managing data (e.g. setting aside missing data, creating scales, and/or dichotomizing variables), choosing the correct test, checking assumptions, calculating a test statistic by hand, writing syntax to run a statistical analysis, graphing, effectively presenting results, and the more general category of conducting a statistical analysis of data. For students not rating themselves at the

maximum confidence level in the pre-survey, individual dichotomous variables were created for each skill indicating whether or not the student's confidence increased between the pre and post course surveys. In addition, the total number of these skills on which students increased from pre to post was calculated. A total confidence score was also calculated separately based on pre and post survey responses to each confidence question (range 9 to 36).

Interest in related courses from the pre to post survey was also assessed. Options included a course in advanced statistical tools, constructing a data set, graphing and data visualization, computer programming, science writing, or other.

Individual courses were examined separately and an aggregate variable based on pre and post survey data was constructed indicating whether or not there was an interest in a larger number of courses between the pre and post survey. A total course interest score was also calculated separately based on pre and post survey responses (range 0 to 6).

Increases in interest in using statistics were also measured by two questions on student expectations about using statistics in future employment and interest in pursuing advanced statistics coursework. Each of these questions was rated on a 5-point Likert scale in the pre and post surveys with higher values indicating increased interest in using statistics.

For students not rating themselves at the maximum interest level in the pre-survey, dichotomous variables were created for each question indicating whether or not the students' interest increased between the pre and post surveys. These items were also examined based on pre and post course responses for the entire sample.

2.4. Analyses

Chi-square Tests of Independence and ANOVA were used to evaluate differences by educational setting (liberal arts college, large state university, regional college/university, and community college) for categorical and continuous variables, respectively. The Duncan test was used to evaluate *post hoc* paired comparisons for ANOVA and a Bonferroni adjustment of .0001 was used to evaluate *post hoc* paired comparisons for Chi Square Tests of Independence.

Next, least absolute shrinkage and selection operator (LASSO) regression analysis (Tibshirani, 1996) was used to identify an interpretable subset of background and course experience variables to be included in multivariate regression models predicting course outcomes (i.e., post course confidence in statistical skills, interest in follow-up courses, interest in advanced statistical course work, and expectation of using statistics in future employment).

SAS PROC GLMSELECT, LASSO regression with the least angle regression selection algorithm (Efron, Hastie, Johnstone, & Tibshirani, 2004) was used to identify the interpretable subset of variables that were most strongly associated with the course outcomes. The LASSO regression model was estimated on a training data set consisting of a random sample of 70% of the students, and a test data set included the other 30% of the student. All predictor variables were centered and scaled prior to conducting the LASSO regression analysis. Cross validation was performed using *k*-fold cross validation specifying 10 folds.

The change in the cross validation mean squared error rate at each step was used to identify the best subset of predictor variables. After identifying the best LASSO model, the selected variables were included in final multiple regression models, controlling for educational setting and pre-survey responses to each course outcome.

3. Results

3.1. Student Background Characteristics by Educational Setting

A comparison of demographic and other background characteristics for students enrolled in the multidisciplinary, project-based course by educational setting is presented in Table 1.

Community college students were significantly older than liberal arts, large state university, and regional college/university students, and along with regional college/university students were somewhat more likely to report being the first in their family to attend college and having been eligible for free school lunches during high school. Students enrolled in the project-based course at a community college or regional college/university also reported the highest rates of Hispanic ethnicity (24.1% and 17.0% respectively).

Nearly 10% of liberal arts, regional college/university, and community college students reported Black race/ethnicity and the large state university and regional college/university settings showed the highest rates of females enrolled in the project-based course. Students from each educational setting showed statistically similar rates of previous experience with programming, having taken a statistics course in high school or college, and being a first or second year college student.

When considering the mutually exclusive categories for college major, students from liberal arts colleges were more likely to be science majors than students from the other educational settings. Students from the large state university were more likely to be a social science major than students from other educational settings. Community college students were more likely to be business majors compared to students from other educational settings, and students from both community colleges and regional colleges/universities were more

likely to major in a health profession than students from liberal arts colleges or the large state university. Students enrolled in the project-based course at a liberal arts college were significantly more likely to have taken the course even if not required and reported more interest in conducting research than those from the large state university, regional college/university, and community college settings.

Table-1. Student background characteristics by educational setting.

	Liberal Arts College	State University	Regional College/ University	Community College	
Demographics	n=446	n=345	n=525	n=79	Statistics
Age	M=20.4 ^{b,c} (s.d. 1.94)	M=19.9 ^c (s.d. 1.93)	M=20.6 ^b (s.d. 3.38)	M=23.5 ^a (s.d. 7.45)	$F(3,1291)=26.0,$ $p<.0001$
Gender (% female)	249 (52.2%) ^b	271 (79.0%) ^a	414 (80.4%) ^a	49 (64.5%) ^{a,b}	$X^2(3)=81.9,$ $p<.0001$
White	292 (65.8%) ^b	303 (87.8%) ^a	369 (70.3%) ^b	55 (69.6%) ^b	$X^2(3)=52.9,$ $p<.0001$
Asian	85 (19.1%) ^a	26 (7.5%) ^b	32 (6.1%) ^b	0 (0%) ^b	$X^2(3)=59.7,$ $p<.0001$
Black	47 (10.6%) ^a	10 (2.9%) ^b	49 (9.3%) ^{a,b}	7 (8.9%) ^{a,b}	$X^2(3)=17.3,$ $p<.0011$
Hispanic	43 (9.7%) ^{a,b}	14 (4.1%) ^b	96 (18.3%) ^a	19 (24.1%) ^a	$X^2(3)=51.9,$ $p<.0001$
First generation student	69 (15.7%) ^b	56 (16.3%) ^b	126 (24.4%) ^{a,b}	28 (35.4%) ^a	$X^2(3)=25.8,$ $p<.0001$
Free/reduced lunch eligibility in HS	66 (16.1%) ^{a,b}	32 (9.7%) ^b	140 (28.8%) ^a	21 (29.2%) ^a	$X^2(3)=53.1,$ $p<.0001$
Academic Background					
First or second year student	164 (38.9%)	163 (50.2%)	102 (42.0%)	41 (57.8%)	n.s.
Previous statistics course	177 (39.8%)	133 (38.6%)	161 (30.7%)	25 (31.7%)	n.s.
Programming experience	162 (47.9%)	145 (54.9%)	237 (56.4%)	42 (71.2%)	n.s.
Major					
Science	95 (33.5%) ^a	44 (13.5%) ^b	80 (17.1%) ^b	6 (12.2%) ^b	$X^2(3)=45.0,$ $p<.0001$
Social Science	143 (50.4%) ^{b,c}	260 (79.8%) ^a	296 (63.4%) ^b	13 (26.5%) ^c	$X^2(3)=86.9,$ $p<.0001$
Art and Humanities	40 (14.1%) ^a	6 (1.8%) ^{b,c}	2 (0.4%) ^c	5 (10.2%) ^{a,b}	$X^2(3)=84.0,$ $p<.0001$
Business	3 (1.1%) ^b	14 (4.3%) ^b	1 (0.2%) ^b	12 (24.5%) ^a	$X^2(3)=107.0,$ $p<.0001$
Health Profession	2 (0.7%) ^b	3 (0.9%) ^b	88 (18.8%) ^a	12 (24.5%) ^a	$X^2(3)=115.6,$ $p<.0001$
Approach to the course					
Required for major or intended major	159 (36.6%) ^c	329 (95.4%) ^a	486 (93.1%) ^a	62 (78.5%) ^b	$X^2(3)=509.7,$ $p<.0001$
Likelihood of taking statistics if not required (1=very unlikely to 5=very likely)	M=3.4 ^a (s.d. 1.17)	M=2.0 ^b (s.d. 1.02)	M=2.0 ^b (s.d. 1.02)	M=2.1 ^b (s.d. 0.91)	$F(3,916)=109.5,$ $p<.0001$
Interest in conducting research (1=not at all to 4=very)	M=2.9 ^a (s.d. 0.85)	M=2.1 ^b (s.d. 0.89)	M=2.5 ^c (s.d. 0.93)	M=1.9 ^d (s.d. 0.76)	$F(3,916)=52.4,$ $p<.0001$

Mean (M) and standard deviation (s.d.) presented for quantitative variables; n and % presented for binary variables.

Percentages are based on the number of respondents completing each item.

Results of post hoc paired comparisons are denoted with superscripts (a,b,c). Statistics sharing the same superscript within row are statistically similar. The Duncan test was used following ANOVA and a p<.0001 Bonferroni adjustment was used for Chi Square Tests of Independence.

3.2. Project-Based Course Experiences by Educational Setting

A comparison of project-based course experiences by educational setting is presented in Table 2. Instructor lectures were rated as *more useful* to students enrolled in the project-based course within community college and regional college/university settings, though no significant differences were found between students from different educational settings in ratings of how much of instructors' lectures were *understood*. In contrast, more of the video lectures were *understood* by liberal arts and state university students compared to students from a regional/college university setting, though there were no differences in how *useful* students found the videos across educational settings. Compared to other educational settings, liberal arts students enrolled in the project-based course were more likely to report receiving more individualized support than in other college courses they had taken. They also reported finding the one-on-one support more useful than in other college courses and understanding more of the one-on-one support from instructors compared to students at the state university, and regional college/university settings.

When considering rigor of the project-based course, regional college/university students reported working significantly harder in the course and finding the course more challenging than students taking the course in liberal arts, large state university, and community college settings. Community college students also reported working significantly harder in the course and found the course more challenging than liberal arts or state university students. Both regional college/university and community college students were more likely to report that the project-based course was more challenging and required more effort than other college courses compared to students from liberal arts and large state college settings. Project-based students from a regional college/university reported preparing for class sessions less frequently than students from the other educational setting.

Although students from liberal arts colleges compared to other educational settings found the project more rewarding and were more likely to find the course more useful than other college courses, students from the regional colleges/universities reported being more likely to have learned more in the project-based course than in other college courses they had taken compared to students from both liberal arts colleges and large state universities. Students from a regional college/university also showed the highest rate of having accomplished more than they had expected (42.4%).

Table-2. Project-based course experiences by educational setting.

	Liberal Arts College	State University	Regional College/ University	Community College	
Understanding Resources	n=446	n=345	n=525	n=79	Statistics
Instructor lectures (1=none of it to 5 = all of it)	M=3.7 (s.d. 0.97)	M=3.7 (s.d. 1.03)	M=3.6 (s.d. 1.00)	M=4.0 (s.d. 0.91)	n.s.
Video presentations (1=none of it to 5 = all of it)	M=3.8 ^a (s.d. 0.98)	M=3.9 ^a (s.d. 1.23)	M=3.4 ^b (s.d. 1.18)	M=3.6 ^{a,b} (s.d. 1.19)	$F(3,989)=9.8,$ $p<.0001$
One-on-one (instructor) (1=none of it to 5 = all of it)	M=4.4 ^a (s.d. 0.78)	M=4.0 ^b (s.d. 1.18)	M=3.8 ^c (s.d. 1.19)	M=4.2 ^{a,b} (s.d. 0.92)	$F(3,1272)=21.4,$ $p<.0001$
Usefulness of Resources					
Instructor lectures (1=not at all to 5=extremely)	M=3.2 ^c (s.d. 1.13)	M=3.6 ^b (s.d. 1.12)	M=3.8 ^a (s.d. 1.10)	M=4.0 ^a (s.d. 1.06)	$F(3,1367)=31.4,$ $p<.0001$
Video presentations (1=not at all to 5=extremely)	M=3.2 (s.d. 1.21)	M=3.4 (s.d. 1.40)	M=3.2 (s.d. 1.32)	M=3.5 (s.d. 1.38)	n.s.
Received <i>more</i> individualized support than other courses <i>n</i> (%)	305 (69.3%) ^a	91 (26.5%) ^b	186 (35.7%) ^b	23 (29.1%) ^b	$X^2(3)=179.3,$ $p<.0001$
Individualized support was <i>more</i> useful than other courses <i>n</i> (%)	227 (51.6%) ^a	84 (24.4%) ^b	173 (33.2%) ^b	23 (29.1%) ^{a,b}	$X^2(3)=69.1,$ $p<.0001$
Engagement					
How frequently did you	M=4.0 ^a	M=4.1 ^a	M=3.6 ^b	M=3.9 ^a	

prepare for class? (1 = never, 5 = always)	(s.d. 0.93)	(s.d. 0.96)	(s.d. 0.92)	(s.d. 0.89)	$F(3,1292)=18.7,$ $p<.0001$
How hard did you work in this class? (1 = not at all, 5 = extremely hard)	M=2.9 ^c (s.d. 0.96)	M=2.8 ^c (s.d. 0.95)	M=3.6 ^a (s.d. 0.97)	M=3.4 ^b (s.d. 0.89)	$F(3,909)=36.9,$ $p<.0001$
Rigor					
Course was challenging (1=not at all to 5=the most challenging)	M=2.8 ^c (s.d. 0.84)	M=2.8 ^c (s.d. 0.90)	M=3.5 ^a (s.d. 0.94)	M=3.2 ^b (s.d. 0.92)	$F(3,1380)=71.7,$ $p<.0001$
Course was <i>more</i> challenging than other courses, <i>n</i> (%)	61 (13.9%) ^b	73 (21.5%) ^b	309 (59.9%) ^a	41 (51.9%) ^a	$X^2(3)=263.4,$ $p<.0001$
Required more effort than other courses	51 (18.8%) ^b	63 (19.4%) ^b	133 (54.5%) ^a	34 (46.6%) ^a	$X^2(3)=111.0,$ $p<.0001$
Overall Impressions					
Course was <i>more</i> interesting than other courses	113 (25.7%) ^a	44 (12.9%) ^b	115 (22.2%) ^{a,b}	18 (23.4%) ^{a,b}	$X^2(3)=20.1,$ $p<.0001$
Course was <i>more</i> useful than other courses	223 (50.7%) ^a	75 (21.9%) ^c	185 (35.9%) ^b	23 (29.9%) ^{b,c}	$X^2(3)=70.8,$ $p<.0001$
Learned <i>more</i> than in other courses	67 (24.7%) ^b	59 (18.2%) ^b	100 (41.0%) ^a	27 (37.0%) ^{a,b}	$X^2(3)=40.6,$ $p<.0001$
Course was rewarding (1=not at all, 5=extremely rewarding)	M=3.6 ^a (s.d. 0.95)	M=2.6 ^c (s.d. 1.06)	M=3.1 ^b (s.d. 1.19)	M=3.2 ^b (s.d. 1.12)	$F(3,1380)=47.5,$ $p<.0001$
Accomplished <i>more</i> than expected	154 (35.0%) ^{a,b}	83 (24.1%) ^b	221 (42.4%) ^a	28 (35.4%) ^{a,b}	$X^2(3)=30.4,$ $p<.0001$

Mean (M) and standard deviation (s.d.) presented for quantitative variables; n and % presented for binary variables.

Percentages are based on the number of respondents completing each item.

Results of post hoc paired comparisons are denoted with superscripts (a,b,c). Statistics sharing the same superscript within row are statistically similar. The Duncan test was used following ANOVA and a $p<.0001$ Bonferroni adjustment was used for Chi Square Tests of Independence.

3.3. Increased Confidence and Interest by Educational Setting

A comparison of increased confidence and interest in future experiences by educational setting is presented in Table 3. Increased confidence in statistical skills was seen among students in each educational setting and ranged from approximately one-third to more than 80% of students not reporting maximum confidence in the pre-survey. Similar rates of students with increased confidence were seen in developing a research question, calculating a test statistic by hand, graphing, effectively presenting results, and conducting statistical analyses. Students enrolled in the project-based course at a liberal arts college showed higher rates of increased confidence in managing data, choosing the correct statistical test, checking whether the assumptions of a statistical test are met, writing code to run statistical analyses and total number of skills compared to other educational settings.

An interest in an increased number of courses related to statistics and quantitative research between the pre and post course surveys was also seen more often among students enrolled in the project-based course at liberal arts colleges compared to the other educational settings, though statistically similar rates of new interest were seen across educational settings for graphing/data visualization, computer programming and scientific writing. Rates of increased interest in pursuing advanced coursework in statistics and expecting to use statistics in future employment were also similar across educational settings.

Table-3. Increased confidence and interest in future experiences by educational setting.

	Liberal Arts College	State University	Regional College/ University	Community College	
Increased confidence in statistical skills	n=263	n=345	n=525	n=70	Statistics
Developing a research question	136 (67.0%)	138 (53.3%)	135 (62.8%)	27 (42.9%)	n.s.
Managing data	167 (68.2%) ^a	138 (47.3%) ^b	130 (56.5%) ^{a,b}	27 (42.9%) ^b	$X^2(3)=28.1$, $p<.0001$
Choosing the correct statistical test	194 (77.0%) ^a	174 (55.6%) ^b	139 (57.9%) ^b	36 (56.3%) ^b	$X^2(3)=32.1$, $p<.0001$
Checking whether the assumptions for a statistical test are met	168 (67.5%) ^a	156 (50.7%) ^b	129 (54.4%) ^b	30 (44.8%) ^b	$X^2(3)=20.5$, $p<.0001$
Calculating a test statistic such as <i>r</i> (the correlation coefficient) or <i>F</i> (found in ANOVA) by hand	123 (48.6%)	186 (59.1%)	109 (45.4%)	44 (66.7%)	n.s.
Writing syntax or code to run statistical analysis	208 (80.3%) ^a	137 (42.6%) ^b	138 (58.0%) ^b	33 (47.8%) ^b	$X^2(3)=87.4$, $p<.0001$
Graphing	124 (54.6%)	162 (57.5%)	100 (45.7%)	27 (46.6%)	n.s.
Effectively presenting research results	87 (46.3%)	132 (49.1%)	105 (47.5%)	30 (46.2%)	n.s.
Conducting a statistical analysis of data	109 (44.1%)	115 (38.3%)	70 (29.8%)	25 (37.9%)	n.s.
Number of skills with increased confidence	M=3.9 ^a (s.d. 1.92)	M=3.1 ^b (s.d. 2.13)	M=3.4 ^b (s.d.=2.04)	M=2.9 ^b s.d. (2.16)	$F(3,893)=9.0$, $p<.0001$
New interest in related courses	n=271	n=325	n=244	n=73	
advanced statistical tools	67 (24.7%) ^a	26 (8.0%) ^b	27 (11.1%) ^b	4 (5.5%) ^b	$X^2(3)=42.7$ $p<.0001$
constructing data sets	51 (18.8%) ^a	24 (7.4%) ^b	20 (8.2%) ^b	5 (6.9%) ^b	$X^2(3)=24.6$ $p<.0001$
graphing and data visualization	56 (20.7%)	55 (16.9%)	20 (8.2%)	8 (11.0%)	n.s.
computer programming	39 (14.4%)	23 (7.1%)	11 (4.5%)	5 (6.9%)	n.s.
scientific writing	33 (12.2%)	26 (8.0%)	15 (6.2%)	5 (6.9%)	n.s.
Interest in increased number of courses between pre and post surveys	177 (43.2%) ^a	84 (25.9%) ^b	47 (19.3%) ^b	12 (16.4%) ^b	$X^2(3)=45.2$ $p<.0001$
Increased interest in using statistics	n=438	n=344	n=521	n=79	
Increased interest in pursuing advanced course work in statistics	175 (40.1%)	105 (30.6%)	181 (34.8%)	31 (39.2%)	n.s.
Increased expectation of using statistics in future employment	122 (27.9%)	91 (26.5%)	115 (22.1%)	17 (21.5%)	n.s.

Mean (M) and standard deviation (s.d.) presented for quantitative variables; n and % presented for binary variables.

Percentages are based on the number of respondents completing each item.

Results of post hoc paired comparisons are denoted with superscripts (a,b,c). Statistics sharing the same superscript within row are statistically similar. The Duncan test was used following ANOVA and a $p<.0001$ Bonferroni adjustment was used for Chi Square Tests of Independence.

3.4. Predicting Course Outcomes

Table 4 presents multiple regression results predicting project-based course outcomes (i.e. confidence in statistical skills, interest in number of follow-up courses, interest in pursuing advanced course work in statistics and expectation of using statistics in future employment) following variable selection through LASSO regression analysis. After controlling for educational setting (liberal arts versus other) and pre-survey responses to confidence in statistical skills, a greater understanding of lecture material and of material

presented one-on-one by the instructor was significantly associated with greater confidence statistical skills following the project-based course. Finding the course less challenging and the research project more rewarding were also associated with higher greater confidence in statistical skills in the post survey. Interest in a greater number of follow-up courses was associated with a student's initial interest in conducting research, their reported likelihood of taking a statistics course if it were not required, being male, finding the course more interesting than other college courses, and finding the course less challenging. Similarly, specific interest in pursuing advanced coursework in statistics was associated with a student's initial interest in conducting research, their reported likelihood of taking a statistics course if it were not required, finding the course less challenging, and the project more rewarding. Finally, a greater expectation of using statistics in future employment was associated with a greater initial interest in conducting research, having found the project-based course more useful than other college courses, viewing the course as less challenging, and the research project more rewarding than previous project experiences. Reported effects were significant after controlling for educational setting (liberal arts vs. other) and pre-survey responses to each outcome measure.

Table-4. Regression models predicting positive course outcomes based on the post survey – Beta weight (p value).

Background	Confidence in statistical skills (post)	Interest in number of follow-up courses (post)	Interest in pursuing advanced course work in statistics (post)	Expectation of using statistics in future employment (post)
Gender (female vs. male)		-0.16 (.0399)		
Interest in conducting research		0.16 (.0001)	0.21 (.0001)	0.34 (.0001)
Likelihood of taking a statistics course if not required		0.12 (.0007)	0.12 (.0029)	
Experience with the course				
How much of the material presented during lecture do you feel you understood?	1.67 (.0001)			
How much of the material presented during one-on-one help from instructor do you feel you understood?	0.66 (.0003)			
How challenging did you find the course?	-0.79 (.0001)	-0.13 (.0005)	-0.15 (.0001)	-0.13 (.0001)
Found the course more interesting than other college courses		0.29 (.0005)		
Found the course more useful than other college courses				0.28 (.0001)
How rewarding did you find the research project?	1.29 (.0001)		0.21 (.0001)	0.15 (.0001)
Covariates (pre-survey)				
Liberal arts college vs. other	1.07 (.0074)	0.52 (.0001)	0.03 (n.s.)	-0.06 (n.s.)
Confidence in statistical skills	0.37 (.0001)			
Interest in number of follow-up courses		0.42 (.0001)		
Interest in pursuing advanced coursework in statistics			0.39 (.0001)	
Expectation of using statistics in future employment				0.44 (.0001)

4. Discussion

In previous reports we have described the development of a U.S. National Science Foundation funded project-based, data analysis and applied statistics course and its success in attracting students from a wide range of racial and academic backgrounds compared to both a traditional statistics course offered through a math department and introductory programming courses offered through computer science (Cooper & Dierker, 2017; Dierker et al., 2015). Further, we have found that students completing the project-based course were more likely to report an increase in confidence between a pre and post course survey and also more likely to show an increase in their interest in pursuing advanced course-work in statistics compared to those enrolled in a traditional statistics course (Dierker et al., under review). The present paper extends our previous work by describing recent dissemination efforts beyond the liberal arts college where the course was initially

developed. We compare student background characteristics, course experiences, and outcomes of students at liberal arts colleges, large state universities, regional state colleges/universities, and community colleges.

Overall, student experiences with the course were quite promising within each educational setting. For example, when compared to students taking the course in liberal arts colleges and large state university settings, regional college/university and community college students reported working significantly harder in the course and finding the course more challenging, as well as requiring more effort than other college courses they had taken.

Along with this increased sense of difficulty and effort, students from regional colleges/universities were more likely than students from liberal arts colleges and large state university settings to report learning more in the project-based course than in other college courses they had taken. Students from regional college/universities also showed the highest rate of having accomplished more than they had expected (42.4%), a statistically similar rate to students from community college settings (35.4%). Though students from the liberal arts colleges were more likely to consider the project-based course more useful than other college courses they had taken (50.7%), nearly one-third of students from both the regional colleges/universities and community colleges felt that the course was more useful. [Garfield and Ben-Zvi \(2007\)](#) stress the benefit to students from struggling to solve conceptually difficult problems and associating these problems with other areas of relevant knowledge. The productive resolution of temporary intellectual roadblocks is a common occurrence in this project-based course. We hypothesize that the successful resolution of these moments of intellectual struggle may have helped students to experience a stronger sense of accomplishment about their work in the class despite the difficulty.

An important feature of the project-based course, aimed at supporting the kind of productive struggle that promotes a sense of accomplishment, is individualized support ([Pasquale, 2015](#)). Based on the present analyses, between 26.5% and 69.3% of students across the four educational settings reported receiving more individualized support than in other college courses they had taken and many (range 24.4% to 51.6%) felt that the individualized support was more useful than what they had received in other college courses.

[Williams \(2010\)](#) found that the physical and psychological availability of an instructor significantly improves students' feelings about learning. In our model, which included a flipped classroom framework for at least half of in-class time, allowing class sessions to be largely devoted to one-on-one conversations, students are encouraged to "keep the conversation going" in ways that allow them to test what they are learning and use language that best translates their data into an accurate and compelling story.

Despite differences by educational setting in terms of student background characteristics and course experiences, large proportions of students from each educational setting reported increased confidence in individual statistical skills (approximately one-third to more than 80%).

Further, increased interest in pursuing advanced statistics course work (range 30.6% to 40.1%) and a greater expectation of using statistics in future employment (21.5% to 27.9%) were seen at statistically similar rates across academic settings. In many cases, students from the liberal arts colleges showed more improvement across course outcomes. However, after controlling for enrollment in a liberal arts college versus other educational setting, the most consistent predictors of course outcomes included prior interest in conducting research, a higher likelihood of taking the statistics course if not required, and finding the course both less challenging and more rewarding.

These findings demonstrate the potential importance of attitudes rather than a students' demographic characteristics. In fact, previous research has shown that students' interest in research is an important factor in retaining students in STEM ([Hunter, Laursen, & Seymour, 2007](#)) and in increasing skills and a sense of identity as a researcher ([Graham, Frederick, Byars-Winston, Hunter, & Handelsman, 2013](#); [Hurtado, Cabrera, Lin, Arellano, & Espinosa, 2009](#)). Students who express negative attitudes about their experience studying statistics are less likely to gravitate towards work in fields that require statistical literacy and are less likely to make use of statistical information to support evidence-based practice ([Petocz & Reid, 2005](#)). Future research is needed to better understand how to foster positive course outcomes among students who enter the introductory statistics course with negative attitudes toward the disciplinary content of the course.

The present results should be interpreted in the context of study limitations. First, identifying comparison groups in this type of research can be challenging given vast differences across classroom contexts and the fact that random assignment of students to different curricular experiences is often infeasible. Because of this, the present sample is a self-selected, incidental sample of students from four different educational settings. Thus, reported differences may have been influenced by one or more of the many unmeasured differences between courses or instructors. Further, while outcomes in self-reported confidence and interest in future data analysis and applied statistics experiences were examined, no outcome measure of achievement was available. Finally, the measures used in the present analyses represented single items rather than scales or subscales. Given that single items are often less reliable, these findings will need to be replicated within the context of expanded data collection efforts.

Despite these limitations, we have demonstrated the feasibility of this project-based approach across diverse educational settings. No single course will adequately prepare students for either the amount or complexity of data they will encounter as professionals and as citizens (Collins & Halverson, 2010). Thus, modern courses need to focus on imparting a deep interest and confidence among students, belief in importance of the discipline, and a desire to continue learning statistics and disciplines focused on data and computation. In this context, the present findings are promising. While we are not yet able to evaluate the potential impact of the project-based course in terms of actual future academic decision-making, the present findings suggest that a course in which students have the opportunity to answer questions that [they] feel passionately about through independent research based on existing data, may represent a promising option for engaging diverse students in the process of statistical inquiry. We believe that this course can benefit other schools not only through dissemination of our model and experiences, but also by making our resources widely available. We are happy to share our course materials with others and encourage faculty to consider integrating project-based course content into their curriculum (<http://passiondrivenstatistics.com>).

References

- Aditomo, A., Goodyear, P., Bliuc, A. M., & Ellis, R. A. (2013). Inquiry-based learning in higher education: Principal forms, educational objectives, and disciplinary variations. *Studies in Higher Education, 38*(9), 1239-1258.
- Aliaga, M., Cuff, C., Garfield, J., Lock, R., Utts, J., & Witmer, J. (2005). Guidelines for assessment and instruction in statistics education (GAISE): College report. *American Statistical Association*.
- American Statistical Association [ASA]. (2014). Curriculum guidelines for undergraduate programs in statistical science. Retrieved from <http://www.amstat.org/education/curriculumguidelines.cfm>.
- Bailey, B., Spence, D. J., & Sinn, R. (2013). Implementation of discovery projects in statistics. *Journal of Statistics Education, 21*(3), 1-24.
- Buck Institute for Education (BIE) and Boise State University. (2012). Department of Educational Technology. Retrieved from <http://pbl-online.org/>. [Accessed April 1, 2012].
- Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., . . . Wood, B. (2016). Guidelines for assessment and instruction in statistics education (GAISE) college report. Retrieved from <http://www.amstat.org/asa/education/Guidelines-for-Assessment-and-Instruction-in-Statistics-Education-Reports.aspx>.
- Cobb, G. W. (2007). The introductory statistics course: A ptolemaic curriculum? *Technology Innovations in Statistics Education, 1*(1), 1-16.
- Collins, A., & Halverson, R. (2010). The second educational revolution: Rethinking education in the age of technology. *Journal of Computer Assisted Learning, 26*(1), 18-27.
- Cooper, J., & Dierker, L. (2017). Increasing exposure to programming: A comparison of demographic characteristics of students enrolled in introductory computer science programming. *International Research in Higher Education, 2*(1), 92.
- De Veaux, R. (2015). What's wrong with Stat 101? Presentation given at the United States Conference on Teaching Statistics (USCOTS).
- Delia, D. N. M. (2017). Statistics education for undergraduate sociology majors: Survey findings across institutions. *Numeracy, 10*(2), 8.
- Dierker, L., Alexander, J., Cooper, J., Selya, A., Rose, J., & Dasgupta, N. (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), 1-11.
- Dierker, L., Cooper, J., Selya, A., Alexander, J., & Rose, J. (2015). Evaluating access: Comparing enrollment patterns in traditional versus multidisciplinary, project-based introductory statistics courses. *Journal of Interdisciplinary Studies in Education, 4*(1), 22-37.
- Dierker, L., Kaparakis, E., Rose, J., & Selya, A. (2012). Strength in numbers: A multidisciplinary, project-based course in introductory statistics. *Journal of Effective Teaching, 12*(2), 4-14.
- Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics, 32*(2), 407-499.
- Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistics Review, 75*(3), 372-396.
- Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM: The International Journal on Mathematics Education, 44*(7), 883-898.
- Gould, R. (2010). Statistics and the modern student. *International Statistical Review, 78*(2), 297-315.
- Graham, M. J., Frederick, J., Byars-Winston, A., Hunter, A., & Handelsman, J. (2013). Increasing persistence of college students in STEM. *Science, 341*(6153), 1455-1456.
- Hickey, D. T., Kindfield, A. C., Horwitz, P., & Christie, M. A. (1999). Advancing educational theory by enhancing practice in a technology-supported genetics learning environment. *Journal of Education, 181*(2), 25-55.
- Hickey, D. T., Wolfe, E. W., & Kindfield, A. C. (2000). Assessing learning in a technology-supported genetics environment: Evidential and systemic validity issues. *Educational Assessment, 6*(3), 155-196.
- Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and Achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*, 99-107. Available at: 10.1080/00461520701263368.
- Horton, N. J. (2015). Challenges and opportunities for statistics and statistics education: Looking back, looking forward. *American Statistician, 69*(2), 138-145.

- Hunter, A. B., Laursen, S. L., & Seymour, E. (2007). Becoming a scientist: The role of undergraduate research in students' cognitive, personal, and professional development. *Science Education*, 91(1), 36-74.
- Hurtado, S., Cabrera, N. L., Lin, M. H., Arellano, L., & Espinosa, L. L. (2009). Diversifying science: Underrepresented student experiences in structured research programs. *Research in Higher Education*, 50(2), 189-214.
- Krajcik, J. S., & Blumenfeld, P. (2006). Project-based learning. In Sawyer, R. K. (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 317-333). New York: Cambridge University Press.
- Lynch, S., Kuipers, J., Pyke, C., & Szesze, M. (2005). Examining the effects of a highly rated science curriculum unit on diverse students: Results from a planning grant. *Journal of Research in Science Teaching*, 42(8), 912-946.
- Neumann, D., Hood, M., & Neumann, M. (2013). Using real-life data when teaching statistics: Student perceptions of this strategy in an introductory statistics course. *Statistics Education Research Journal*, 12(2), 59-70.
- Pasquale, M. (2015). Productive struggle in mathematics. Interactive Technologies in STEM Teaching and Learning: Research brief. Retrieved from <http://interactivestem.org/wp-content/uploads/2015/08/EDC-RPC-Brief-Productive-Struggle.pdf>.
- Petocz, P., & Reid, A. (2005). Something strange and useless: Service students' conceptions of statistics, learning statistics and using statistics in their future profession. *International Journal of Mathematical Education in Science and Technology*, 36(7), 789-800.
- Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society, Series B*, 58(1), 267-288.
- Walker, A., & Leary, H. (2009). A problem based learning meta analysis: Differences across problem types, implementation types, disciplines, and assessment levels. *Interdisciplinary Journal of Problem-Based Learning*, 3, 6-28. Available at: 10.7771/1541-5015.106.
- Williams, A. (2010). Statistics anxiety and instructor immediacy. *Journal of Statistics Education*, 18(2), 1-18.