

2017

Engaging Underrepresented High School students in Data Driven Storytelling: An Examination of Learning Experiences and Outcomes for a Cohort of Rising Seniors Enrolled in the Gaining Early Awareness and Readiness for Undergraduate Program (GEAR UP)

Lisa Dierker

Wesleyan University, ldierker@wesleyan.edu

Nadia Ward

Jalen Alexander

Emmanuel Donate

Follow this and additional works at: <http://wescholar.wesleyan.edu/psycfacpub>

Recommended Citation

Dierker, Lisa; Ward, Nadia; Alexander, Jalen; and Donate, Emmanuel, "Engaging Underrepresented High School students in Data Driven Storytelling: An Examination of Learning Experiences and Outcomes for a Cohort of Rising Seniors Enrolled in the Gaining Early Awareness and Readiness for Undergraduate Program (GEAR UP)" (2017). *Faculty Scholarship*. 5.
<http://wescholar.wesleyan.edu/psycfacpub/5>

This Article is brought to you for free and open access by the Psychology at WesScholar. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of WesScholar. For more information, please contact ljohnson@wesleyan.edu.

Engaging Underrepresented High School students in Data Driven Storytelling: An Examination of Learning Experiences and Outcomes for a Cohort of Rising Seniors Enrolled in the Gaining Early Awareness and Readiness for Undergraduate Program (GEAR UP)

Lisa Dierker¹, Nadia Ward², Jalen Alexander¹, Emmanuel Donate³

¹Psychology Department, Wesleyan University, Middletown, CT, USA

²Department of Psychiatry, Yale University, School of Medicine, New Haven, CT, USA

³Department of Physics and Astronomy, University of Georgia, Athens, GA, USA

Correspondence: Lisa Dierker, Psychology Department, Wesleyan University, 206 High Street, Middletown CT 06459, USA.

Received: February 3, 2017

Accepted: February 23, 2017

Online Published: March 1, 2017

doi:10.11114/jets.v5i4.2187

URL: <https://doi.org/10.11114/jets.v5i4.2187>

Abstract

Background: Upward trends in data-oriented careers threaten to further increase the underrepresentation of both females and individuals from racial minority groups in programs focused on data analysis and applied statistics. To begin to develop the necessary skills for a data-oriented career, project-based learning seems the most promising given its focus on real-world activities that are aimed at engaging student interest and enthusiasm. *Method:* Using pre and post survey data, the present study examines student background characteristics, learning experiences and course outcomes for a cohort of 33 rising high school seniors involved in a two-week, accelerated version of a project-based data analysis and applied statistics curriculum. *Results:* On average, students rated the experience as rewarding and the vast majority (78.1%) felt that they had accomplished more than they had expected. Based on responses to both the pre and post course surveys, roughly half of the students reported increases in confidence in applied skills (i.e. developing a research question, managing data, choosing the correct statistical test, effectively presenting research results, and conducting a statistical analysis of data), while more than 80% reported increased confidence in writing code to run statistical analyses. Fully 84.4% of students reported interest in one or more follow-up courses with interest in computer programming being endorsed by the largest number of students (53.1%). *Conclusions:* These findings support previous research showing that real-world, project-based experiences afford the best hope for achieving the kind of analytic and statistical literacy necessary for meaningful engagement in research, problem solving and professional development.

Keywords: passion driven statistics, underrepresented students, programming, STEM, project-based learning

1. Introduction

A large number of the top 25 highest paying in-demand jobs require specialized or general skills in data analysis and applied statistics, for example, data base administrator, data scientist, analytics manager, solutions architect, etc. (Glassdoor, 2015). While these and other technical occupations have been predicted to rise by 17.7% between 2012 and 2022 (Bureau of Labor Statistics, 2013), the relevance of data-oriented skills in many other growing careers that have previously been less directly focused on analytics is also expanding (U.S. White House, 2014).

These trends threaten to further increase current disparities given the continued underrepresentation of both females and individuals from racially underrepresented groups across all positions within the technology industry. Recent reports from large technology companies (e.g., Apple, Facebook, Google, Twitter, Microsoft, Cisco, Intel, etc.) have focused attention on this lack of diversity. Employee data from 2014 indicates that most large technology companies are more than 70% male and with few exceptions, little more than 5% Black or Hispanic (Mangalindan, 2014). These disparities are even more pronounced when considering diversity in both leadership and technical roles where, at Google for example, females hold as little as 17% of the technology positions, while Blacks and Hispanics occupy only 3% of positions in corporate leadership (Google, 2015).

The most often cited reason for early departure from technical and science oriented programs is uninspiring or unsuitable pedagogical practices (Handelsman, 2005; Seymour, Hunter, Laursen, & Deantoni, 2004). A central challenge to addressing disparities is the development of curricula that not only attracts diverse students, but also sparks communication, reasoning and collaboration that cross disciplinary boundaries. Previous authors have suggested that this can best be achieved through inquiry-based projects (Bailey, Spence, & Sinn, 2013) that allow students to “decompose their topic; identify key components; abstract and formulate different strategies for addressing it; connect the original question to the statistical framework; choose and apply methods; reconcile the limitations of the solution; and communicate findings” (Nolan & Temple Lang, 2009). This type of project-based learning is most commonly defined as an instructional approach based on real-world activities that are aimed at engaging student interest and enthusiasm (BIE, 2012; Krajcik & Blumenfeld, 2006). This approach allows students to face challenges that lead to answers, reflect on ideas and make decisions that affect project outcomes because it is designed to answer a question or solve a problem (Aditomo, Goodyear, Bliuc, & Ellis, 2013). There is an emerging literature showing that project-based learning in many contexts is more effective in promoting deep thinking, the ability to apply knowledge, communication and reasoning skills, when compared to traditional didactic approaches (e.g. Harada & Yoshina, 2004; Hickey, Wolfe, & Kindfield, 2000; Hickey, Kindfield, Horwitz, & Christie, 1999; Hmelo-Silver, Duncan, & Chinn, 2007; Langer, 2001; Lynch, Kuipers, Pyke, & Szesze, 2005; Walker & Leary, 2009; however, see Kirschner, Sweller, & Clark, 2006).

In previous publications we have described the development of an introductory data analysis curriculum (Dierker, Kaparakis, Rose, Selya, & Beveridge, 2012; Dierker, Cooper, Selya, Alexander, & Rose, 2015; Dierker et al., 2016) aimed at increasing access to applied statistics skills both generally and for students traditionally marginalized within many technical fields (e.g., women, underrepresented students, educationally disadvantaged students, etc.). Funded by the National Science Foundation, the project-based course engages students with real data and code-based statistical software. Closely following the recommendations outlined in the Guidelines for Assessment and Instruction in Statistics Education (GAISE) report (Aliaga et al., 2005; ASA, 2014; Carver et al., 2016), the curriculum is designed around research projects of students’ own choosing and offers individualized hands-on experience in data analysis. Work is organized to focus on the decisions and skills involved in statistical inquiry including the ability to tell accurate and engaging stories with data. Basic themes such as measurement, descriptives and graphical representation are covered, as well as more specific inferential methods needed to test hypotheses and/or explore the empirical structure of data (ASA, 2014). These skills are introduced as the students’ scientific questions dictate. Final projects are presented at a research poster session in which students have the opportunity to describe their results and process of inquiry, including the different decisions made along the way, their premises, and conclusions.

At the undergraduate level, this curriculum has been found to attract higher rates of under-represented minority (URM) students than a traditional introductory statistics curriculum (Dierker et al., 2015). We have also compared enrollment of the project-based course to traditional introductory programming experiences, a general introductory programming course and an introductory course representing a gateway to the computer science major, and found that there are now higher rates of female and URM enrollment (Cooper & Dierker, under review). More recently, we have shown that while URM students often considered the material covered in the project-based course more difficult than non-URM students, URM students’ demonstrated similar levels of increased confidence in applied skills and interest in follow-up courses as non-URM students. Following the course, URM students were also found to be twice as likely to report that their interest in conducting research had increased (Dierker et al., 2016)

Based on these promising findings, we engaged URM high school students in an accelerated version of this project-based curriculum to evaluate its feasibility, student experiences and outcomes related to improved confidence and attitudes, and interest in data analysis and applied statistics. Students identified in this project were enrolled in a federally funded educational initiative called Gaining Early Awareness & Readiness for Undergraduate Programs (GEAR UP) sponsored by the U.S. Department of Education. This program is designed to increase equity and access on the part of low-income and minority students to increase matriculation into- and out of- institutions of higher learning (Department of Education, 2008). The present paper describes background characteristics, learning experiences and course outcomes for students exposed to the project-based curriculum during a two-week college residential summer program.

2. Method

2.1 Participants

The sample included 33 rising high school seniors enrolled in a mid-sized, urban public school district in New England. Participants were invited to attend an accelerated summer residential program in July 2016 based on academic performance (e.g., GPA of 3.0 or higher) and teacher recommendation. Students participating in the summer program represented seven comprehensive high schools from the district. All students were identified GEAR UP students, some

of whom had been with the program since its inception in 2011. GEAR UP espouses a cohort model where, beginning in seventh grade, an entire cohort of students are followed through middle school, high school and into their first year of college (in this case, 1,633 students). Through their participation in GEAR UP, these students have access to an array of academic enrichment and support services that are implemented year round. These services include, a classroom-based social-emotional learning program, academic advising, tutoring mentoring, college tours, educational trips, after school programs, Saturday Academy programs (which includes SAT Prep), and summer residential programs.

As part of the two-week summer residential program, the project-based curriculum was presented to students as “Storytelling with Data” and included developing a research question, accessing and managing data, testing a hypothesis with basic inferential tools, graphing results, and interpreting and presenting findings (see <http://passiondrivenstatistics.com/>). Students learned to write code-based programs in SAS by accessing the free SAS Studio cloud application through a web browser (SAS Studio, Version 3.5). Classroom sessions included brief presentations and one-on-one mentoring for two hours on nine consecutive weekdays that concluded with a research poster session attended by mentors, GEAR UP staff, college instructors, parents, and local professionals.

Data for the present analyses were drawn from pre/post student surveys with the pre survey completed prior to the start of the residential summer program and the post course survey completed during the last class session. Each survey took approximately 10 minutes to complete.

2.2 Measures

2.2.1 Background Characteristics

Reports of race/ethnicity, including White, Black, Hispanic, Asian and Other, were measured through non-mutually exclusive categories. Students also self-reported their age, gender and whether they will be the first generation of their family to attend college.

Academic background was assessed by questions about previous statistics, mathematics, and programming experience in the pre survey. Students were asked whether they had previously “taken a course that taught about statistics or how to analyze and look at data?” Self-perceived skills in mathematics was measured by the questions, “How good at mathematics are you?” on a scale from 1 (very poor) to 5 (very good) and “How well did you do in mathematics courses you have taken in the past?” from 1 (very poorly) to 5 (very well). 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, HTML, etc.).

2.2.2 Experiences with the Course

Based on the post course survey, students rated the *usefulness* of course resources (i.e., in class presentations, video lectures and presentations) from 1 (none of it) to 5 (all of it). Students’ perceptions of their engagement in the course were measured by two questions about their *effort* (“How hard did you work in this course?” on a scale from 1 = not at all hard to 5 = extremely hard) and “Compared to other courses you have taken, did you put in more effort, less effort or a similar amount of effort?”. 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 courses you have taken?” Overall impressions of the success of the course were measured on a scale of 1=not at all to 5=extremely rewarding with the question “How *rewarding* did you find this course?”, and the questions “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 courses you have taken?”, “Did you find this course *more useful*, less useful or similarly useful compared to other courses you have taken?”, “Did you *learn* more, less or a similar amount compared to other courses” and “Are you more likely, less likely or similarly likely to *use the content and skills* that you learned in this course, compared to other courses?” 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. All questions comparing experiences in the project-based course to other courses were dichotomized as more vs. less or the same.

2.2.3 Increases in Confidence

Increases in perceived confidence on specific data analysis and statistical skills were evaluated based on changes in student ratings from the pre to post survey, using a scale of 1 (not at all confident) to 4 (very confident). These skills were developing a research question, managing data, choosing the correct test, writing syntax to run a statistical analysis, interpreting results, graphing, effectively presenting results, and the more general category of conducting a statistical analysis on 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. Post-test ratings of confidence, from 1 (not at all confident) to 5 (extremely confident), in abilities to learn more statistics, analyze data, answer questions with data and master introductory material were also examined.

2.2.4 Attitudes and Interest in Future Experiences

An increase in students' interest in conducting research was measured by comparing the pre survey and the post survey responses, each measured on a scale of 1 (not at all interested) to 5 (extremely interested).

Increases in *interest in using statistics* were measured by two questions on student expectations about using statistics in the future, 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.

A final outcome measure considered if students would like to take *one or more courses as a follow-up* to the project-based statistics course. Options included a course in advanced statistical tools, programming, data set construction, graphing and data visualization, science writing, or computer programming. Individual courses were examined separately and an aggregate variable based on post survey data was constructed indicating whether or not a student endorsed an interest in taking at least one follow-up course.

2.3 Analyses

Overall means and frequencies were first examined for continuous and categorical variables respectively. Next, Analysis of Variance (continuous outcomes) and Chi Square Tests of Independence (binary outcomes) were run to evaluate possible associations between student background characteristics and each measured experience and outcome of the project-based course. Findings were first examined in unadjusted models. Significant findings were then adjusted for multiple comparisons (Bonferroni adjustment, $p < .001$).

3. Results

3.1 Student Characteristics

Demographic and other student background characteristics are presented in Table 1. The mean age was 16.6 (s.d. 0.66 – range 15 to 18) and females accounted for 63.6% of the sample. A total of 48.5% of students were Black, 30.3% were Hispanic, 21.2% White, and 3.0% Asian. Nearly one-fifth of the sample (18.8%) endorsed 'Other' ethnicity. Nine students (28.1%) reported that they would be the first generation in their family to attend college.

A total of 15.1% of students reported that they were "very good" at math and 30.3% reported that they had done "very well" in previous math courses they had taken. Fully 2/3 of the students reported no previous experience with programming and only one student reported having ever used a code-based statistics program.

One-third of the sample reported having previously taken a course that was entirely focused on statistics or how to analyze and look at data. Another 36.3% reported that these topics had been covered as part of a math class.

Table 1. Student background characteristics

	<i>N</i> = 33
<i>Demographics</i>	<i>n</i> (%)
Gender (% female)	21 (63.6%)
White	7 (21.2%)
Asian	1 (3.0%)
Black	16 (48.5%)
Hispanic	10 (30.3%)
First generation college student	9 (28.1%)
Age	M 16.6 (SD 0.66)
<i>Academic Background</i>	
Taken a course on statistics or data analysis	11 (33.3%)
Learned about statistics or data analysis as part of a math course (but no statistics course)	12 (36.4%)
Any programming experience	11 (33.3%)
How good at mathematics are you? (1 = very poor to 5 = very good)	M 3.5 (SD 1.06)
How well did you do in mathematics courses you have taken in the past? (1 = very poorly to 5 = very well)	M 3.7 (SD 1.18)

3.2 Student Experiences with the Project-based Curriculum

Based on post survey responses, self-reported experiences in the project-based course are presented in Table 2. Video lectures and presentations were rated as being somewhat more useful than in-class presentations. Overall, students considered the course moderately hard and more than 1/3 reporting putting more effort into the project-based course than other courses they had taken. On average, students rated the experience as rewarding and the vast majority (78.1%)

felt that they had accomplished more than they had expected. Though on average the course was rated as moderately challenging, more than half of the students reported that they found it more challenging than other courses they had taken.

More than half of the sample reported receiving more individualized support compared to other courses they had taken and 43.8% felt that the *individualized support was more useful* than what they had received in other courses. While only 1/3 of students found the project-based course either more *useful* or more *interesting* than other courses that they had taken, more than half reported that they had *learned more* in the project-based course than in other courses. A total of 28.1% of student believed they were *more likely to use the skills* and content they had learned in the future.

Analyses examining the possible association between student background characteristics and experiences with the project-based course revealed relatively few differences. Females, however, (90.0%) were more likely than males (58.3%) to report having accomplished more than expected in the course $\chi^2(1) = 4.40, p < .04$, while Black students (60.0%) were less likely than non-Black students (94.1%) to report having accomplished more than they expected, $\chi^2(1) = 5.4, p < .02$. Females (75.0%) were also more likely than males (25.0%) to report that they had learned more in the project-based course than in other courses they had taken. $\chi^2(1) = 7.62, p = .006$. Black students were significantly less likely (6.7%) than non-Black students (58.8%) to find the course more useful than others they had taken, $\chi^2(1) = 9.6, p < .002$. Finally, those students without any prior experience with programming found the project-based course significantly more rewarding than those with prior programming experience, $F(1,31) = 7.1, p < .01$. When adjusting for multiple comparisons, none of these differences reached statistical significance (Bonferroni adjustment $p < .001$)

Table 2. Student experiences with the project-based curriculum

<i>Usefulness of resources</i> (1 = not available to 5 = extremely useful)	<i>N</i> = 33
In class presentations	M 2.8 (SD 1.07)
Video lectures and presentations	M 3.8 (SD 0.79)
<i>Engagement</i>	
More effort put in than other courses <i>n</i> (%)	12 (37.5%)
How hard did you work in this class? (1 = not at all, 5 = extremely hard)	M 3.69 (SD 0.93)
<i>Rigor</i>	
How challenging did you find this course? (1 = not at all to 5 = the most challenging)	M 2.69 (SD 0.90)
Course was <i>more challenging</i> than other courses <i>n</i> (%)	18 (56.3%)
<i>Overall Impressions</i>	
How rewarding did you find this course? (1 = not at all, 5 = extremely rewarding)	M 3.56 (SD 0.95)
Accomplished <i>more</i> than expected <i>n</i> (%)	25 (78.1%)
<i>Compared to other courses</i>	
Course was <i>more interesting</i> <i>n</i> (%)	11 (34.4%)
Course was <i>more useful</i> <i>n</i> (%)	11 (34.3%)
Learned <i>more</i> in this course <i>n</i> (%)	18 (56.3%)
<i>More likely to use the skills and content that you learned</i> <i>n</i> (%)	9 (28.1%)
Received <i>more individualized support</i> <i>n</i> (%)	18 (56.3%)
<i>Individualized support was more useful</i> <i>n</i> (%)	14 (43.8%)

Percentages based on valid data.

3.3 Increased Confidence in Basic and Applied Statistics Skills

Based on responses to both the pre and post course surveys ratings of increased confidence in applied statistical skills are presented in Table 3. Roughly half of the students reported increases in confidence in the majority of skills (e.g. developing a research question, managing data, choosing the correct statistical test, effectively presenting research results, conducting a statistical analysis of data), while more than 80% reported increased confidence in writing code to run statistical analysis. Students reported increases in their confidence on average for four of the eight applied skills. Students reported being moderately confident on average, following the project-based course in terms of their ability to learn more statistics, analyze data, answer questions with data and master introductory material.

On average, Black students reported increased confidence in a smaller number of applied skills following the project-based course (M 2.6, SD 0.70) than non-Black students (M 5.3, SD 2.26), $F(1,31) = 14.4, p = .0007$. Individual skills for which Black students were less likely to rate their confidence as having increased included choosing the correct statistical test, effectively presenting research results, and conducting a statistical analysis of data. Experiences in the project-based course were also found to differ according to whether students had prior experience with programming. Specifically, students without programming experience were more likely (90.5%) than students with

programming experience (54.6%) to report that they accomplished more than they expected, $\chi^2(1) = 5.4, p < .02$. Those without programming experience were also *more* likely to report an increase in confidence for data management (60.0% vs. 20.0%, $\chi^2(1) = 4.3, p < .04$), presenting results (58.8% vs. 14.3%, $\chi^2(1) = 4.0, p < .05$) and conducting a statistical analysis of data (70.0% vs. 22.2%, $\chi^2(1) = 5.7, p < .02$). When adjusting for multiple comparisons, increased confidence in a smaller number of applied skills following the project-based course for Black vs. non-Black students remained significant (Bonferroni adjustment $p < .001$).

Table 3. Increased confidence in applied skills.

	<i>N</i> = 33
<i>Increased Confidence on Skills</i>	<i>n</i> (%)
developing a research question	14 (53.9%)
managing data	14 (46.7%)
choosing the correct statistical test	16 (57.1%)
writing syntax or code to run a statistical analysis	25 (80.7%)
interpreting results	17 (63.0%)
graphing	14 (60.9%)
effectively presenting research results	11 (45.8%)
conducting a statistical analysis of data	16 (55.2%)
Number of skills with increases, <i>M</i> (<i>SD</i>)	M 4.1 (SD 2.44)
<i>Confidence Levels</i>	
(1 = not at all confident, 5 = extremely confident)	
How confident are you of your ability to learn more statistics?	M 2.5 (SD 0.95)
How confident are you of your ability to analyze data?	M 3.5 (SD 0.95)
How confident are you of your ability to answer questions with data?	M 2.9 (SD 1.09)
How confident are you that you mastered the introductory statistics material?	M 3.2 (SD 1.09)

3.4 Improved Attitudes and Interest in Future Exposure to Statistics

Improved attitudes and interest in future courses is presented in Table 4. Approximately 1/3 of students reported increases in conducting research and interest in pursuing advanced statistics coursework.

Fully 84.4% of students reported interest in one or more follow-up courses including (i.e., advanced statistical tools, constructing data sets, graphing and data visualization, scientific writing and computer programming) interest in computer programming being endorsed by more than half of the sample. Measures of attitudes and interest were found to be statistically similar by student background characteristics.

Table 4. Attitudes and interest in future exposure to statistics.

	<i>N</i> = 33
<i>Interest in using statistics</i>	<i>n</i> (%)
Increased interest in conducting research	7 (31.8%)
Increased interest in pursuing advanced statistics coursework	9 (36.0%)
<i>Interest in future courses</i>	
Advanced statistical tools	6 (18.8%)
Constructing data sets	9 (28.1%)
Graphing and data visualization	12 (37.5%)
Computer programming	17 (53.1%)
Scientific writing	9 (28.1%)
<i>Interest in any of the courses</i>	27 (84.4%)

4. Discussion

In previous reports we have described the development of an NSF funded project-based, data analysis and applied statistics course and its success in attracting higher rates of under-represented students and students from a wide range of academic backgrounds compared to both a traditional statistics course offered through a math department and introductory programming courses offered through computer science (Dierker et al., 2015; Cooper & Dierker, under review). The goal of the present study was to evaluate the success of a data analysis and applied statistics course with underrepresented high school students designed to increase confidence in applied skills and foster positive attitudes toward future learning. To that end, survey data from rising high school seniors participating in the project-based course as part of a two-week college residential GEAR UP program were examined.

Despite the short duration of the course, more than half of the students reported that they had learned more in the project-based course than in other courses they had taken in high school. At the same time the majority of students reported that they found the course more challenging than other courses they had taken. 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. This is supported by the fact that the vast majority of students (78.1%) felt that they had accomplished more than they had expected.

An important ingredient for the kind of productive struggle that promotes a sense of accomplishment is individualized support (Pasquale, 2015). Based on the present analyses, more than half of the students in the project-based course reported receiving more individualized support than in other high school courses they had taken and again, a majority felt that the individualized support was more useful than what they had previously received. Williams (2010) found that the physical and psychological availability of an instructor significantly improves students' feelings about learning. In our model, which includes a "flipped classroom" framework in which lectures are removed from class sessions and time is devoted to one-on-one conversations among instructors, students and peer mentors, 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.

Though some trends were seen suggesting that females, non-Black students or those without prior programming experience judged the course as more rewarding and/or more useful than other courses they had taken, and also reported accomplishing more than expected, after adjustment for multiple comparisons, none of these differences were significant. Furthermore, students with different background characteristics in terms of race, ethnicity, first generation college student status, perceived math ability, and prior experience with statistics showed no differences in their ratings of their own engagement in the course or its rigor.

Differences were observed in changes in confidence level in applied skills. On average, Black students reported increased confidence in a smaller number of applied skills following the project-based course than non-Black students. Individual skills for which Black students were less likely to rate their confidence as having increased included choosing the correct statistical test, effectively presenting research results, and conducting a statistical analysis of data. Notably, the project-based course described in this paper is designed to provide under-represented minority (URM) students with a culturally supportive learning environment. Fullilove and Treisman (1990) found that the addition of problem-solving sessions designed to build a culture of collaborative study among URM students to a calculus class increased the numbers of African American and Latino students majoring in math. Central to this model is the idea that to facilitate the success of URM students we must consider students' academic efforts in the context of their broader experiences and their previous education. While the collaborative nature of the project-based curriculum may provide this sort of context, it is likely that different students came to the experience with varying levels of background experience and confidence in their approach. Further, the course offered students a new way of working together that they do not experience in high school. That, coupled with students' different learning styles and personality characteristics (introversion/extroversion), may have influenced the few group differences that were detected.

Importantly, increases in confidence in writing code to run statistical analyses showed no differences by student racial or background characteristics and more than 80% of high school students in the project-based course reported increased confidence in this skill. Further, more than half of the students reported that they were interested in taking a programming course in the future, despite the fact that the vast majority of the students entered the course with absolutely no previous programming experience. This suggests that exposing high school students to project-based experiences in data analysis not only provides opportunities for students to gain basic programming skills, but also sparks an interest in future programming opportunities. Through the use of formal syntax to write complex programs aimed at managing and analyzing data, the project-based course introduces students to several of the same concepts and skills covered in traditional programming courses (Nolan and Temple Lang 2010; ACM, 2013). Importantly, our approach of using contextualized projects and research to motivate students is one that has also been undertaken in computer science courses (Jenkins 2001; Goldweber et al., 2012) and has been shown to be particularly motivating to female students (Rader et al., 2011). More research is needed to consider the future coursework and career choices of these students and how their experiences in the project-based data analysis course might offer another possible pathway into technical occupations.

Taken together, we have demonstrated the feasibility of this project-based approach when meeting with students just two hours per day across a two-week period. This work also confirms the importance of examining outcomes beyond the typical content knowledge exams. This is especially relevant as the most beneficial effects of project-based learning are on measures that consider the application, as opposed to the memorization, of knowledge (Dochy, Segers, Van den Bossche, & Gijbels, 2003) and it is these additional application outcomes and attitudes about future learning that are most important to continued persistence in statistics and data analysis. 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). Modern courses need to focus on imparting a deep interest among students, belief in importance of the discipline and a desire to continue learning statistics and disciplines focused on data and computation. Thus, the present findings

are promising in that more than 80% of students reported interest in one or more follow-up courses including scientific writing, constructing data sets, graphing and data visualization, advanced statistical tools and/or programming.

Despite these encouraging results, the findings should be considered within the context of study limitations. First, because of the small sample size, we were not able to examine how students' experiences with the course may predict increasing confidence and/or interest in future learning over and above their background characteristics when entering the course. Further, the sample included URM students achieving at least a 3.0 grade point average and thus does not fully represent the experiences of rising high school seniors. However, previous research has shown that African American and Hispanic college students with high grade point averages and SAT scores above 600 typically do not pursue STEM college majors for reasons including poor teaching in STEM courses, lack of encouragement from teachers or parents, and self-perception of their own inability to be successful in STEM majors (George, Neale, Van Horne, & Malcolm, 2001; Grandy, 1998). Thus, this is an important group to target with an innovative, welcoming curriculum that sparks interest in continued study. While we are not yet able to evaluate the potential impact of the project-based course in terms of actual future academic decisions-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.

Project-based learning, of course, presents its own challenges and customarily, courses offering opportunities for independent statistical inquiry have required a smaller group format compared to courses that rely on more traditional pedagogies. While this accelerated version of the course included class sizes of approximately 17 students each, at the college level, we are now teaching the course to sections of 50+ students. With appropriate supporting infrastructure (e.g. helpful materials for outside-of-class time that enhance engagement in active class sessions and a network of peer mentors), we hope to demonstrate that this project-based approach can be delivered to high school students by expending no more resources than required within a traditional classroom.

Previous research has recognized the general challenge in teaching courses in statistics. There is no typical statistics student; instead, students come into statistics courses with differing backgrounds, experiences, learning styles and levels of preparation. This project-based course provokes students to encounter (and struggle with) the central concepts and principles. Although our model focuses on applied statistics and programming, the emphasis on authentic real-world activities with the goal of sparking interest and enthusiasm (BIE, 2012) can be achieved in curricular content as diverse as science (Kubiatico & Vaculová, 2011) and foreign language (Danan, 2010) instruction. In addition to the specific skills most directly emphasized, project-based courses provide students with experience in communication, organization and time management (BIE, 2012). We believe that this course can benefit other universities and high schools not only through dissemination of our model and experiences, but by making our newly developed resources widely available. We are happy to share our course materials with others and encourage faculty to consider integrating project-based course content (<http://passiondrivenstatistics.com>). Additional details about Passion-Driven Statistics, GEARUP partnership can be found at <http://passiondrivenstatistics.com/2016/09/21/gear-up/>

Acknowledgements

This research was supported by grant 0942246 and 1323084 from the National Science Foundation, Transforming Undergraduate Education in Science, Technology, Engineering and Mathematics (TUES), and U.S. Department of Education P334A110229. We appreciate the generous efforts of our colleagues and advisors who contributed to the development and implementation of this course.

References

- ACM/IEEE-CS Joint Task Force on Computing Curricula. (2013). Computer Science Curricula 2013. *ACM Press and IEEE Computer Society Press*. <https://doi.org/10.1145/2534860>
- 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, 1239-1258. <https://doi.org/10.1080/03075079.2011.616584>
- 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*[online]. Available at <http://www.amstat.org/education/gaise/>
- American Statistical Association [ASA]. (2014). Curriculum Guidelines for Undergraduate Programs in Statistical Science" Retrieved November 1, 2016 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. [ww2.amstat.org/publications/jse/v21n3/bailey.pdf](http://www2.amstat.org/publications/jse/v21n3/bailey.pdf)

- Buck Institute for Education (BIE) and Boise State University, Department of Educational Technology. Retrieved April 1, 2012 from <http://pbl-online.org/>.
- Bureau of Labor Statistics, U.S. Department of Labor. (2013, December 19). Employment projections: Employment by detailed occupation. Retrieved March 20, 2015 from http://www.bls.gov/emp/ep_table_102.htm
- Carver, R., Everson, M., Gabrosek, J., Rowell, G. H., Norton, N., Lock, R., ... Wood, B. (2016, February). Draft: "Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report" [online]. Available at: http://www.amstat.org/education/gaise/collegeupdate/GAISE2016_DRAFT.pdf
- Collins, A., & Halverson, R. (2010). The second educational revolution: Rethinking education in the age of technology, *Journal of Computer Assisted Learning*, 26, 18-27. <https://doi.org/10.1111/j.1365-2729.2009.00339.x>
- Cooper, J. L., & Dierker, L. (2016, under review). Increasing Exposure to Programming: A Comparison of Demographic Characteristics of Students Enrolled in Introductory Computer Science Programming Courses vs. a Multidisciplinary Data Analysis Course, Manuscript submitted for publication.
- Danan, M. (2010). Dubbing Projects for the Language Learner: A Framework for Integrating Audiovisual Translation into Task-Based Instruction, *Computer Assisted Language Learning*, 23, 441-456. <https://doi.org/10.1080/09588221.2010.522528>
- Department of Education. (2008). Early Outcomes of the GEAR UP Program, Final Report, Retrieved August 11, 2016 from <https://www2.ed.gov/rschstat/eval/highered/gearup/early-outcomes.pdf>.
- Dierker L., Kaparakis E., Rose, J., Selya, A., & Beveridge, D. (2012). Strength in Numbers: A Multidisciplinary, Project-Based Approach to Introductory Statistics Education, *Journal of Effective Teaching*, 12, 4-14. <https://doi.org/10.20429/ijstl.2016.100102>
- 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). <https://doi.org/10.20429/ijstl.2016.100102>
- Dierker, L., Cooper, J., Selya, A., Alexander, J., & Rose, J. (2015). Evaluating Access: A Comparison of Demographic and Disciplinary Characteristics of Students Enrolled in a Traditional Introductory Statistics Course vs. a Multidisciplinary, Project-Based Course, *Interdisciplinary Studies of Education*, 4, 22-37. http://coejournal.astate.edu/index.php/jise/article/viewFile/Access/pdf_46
- Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: A meta-analysis. *Learning and instruction*, 13(5), 533-568. [https://doi.org/10.1016/S0959-4752\(02\)00025-7](https://doi.org/10.1016/S0959-4752(02)00025-7)
- Fullilove, R. E., & Triesman, P. U. (1990). Mathematics Achievement among African American Undergraduates at the University of California, Berkeley: An Evaluation of the Mathematics Workshop Program, *Journal of Negro Education*, 59, 463-478. <https://doi.org/10.2307/2295577>
- Garfield, J. and Ben-Zvi, D. (2005). A Framework for Teaching and Assessing Reasoning about Variability, *Statistics Education Research Journal*, 4(1), 92-99. [iase-web.org/documents/SERJ/SERJ4\(1\)_Garfield_BenZvi.pdf](http://iase-web.org/documents/SERJ/SERJ4(1)_Garfield_BenZvi.pdf)
- George, Y. S., Neale, D. S., Van Horne, V., & Malcolm, S. M. (2001). In pursuit of a diverse science, technology, engineering, and mathematics workforce. *American Association for the Advancement of Science*. ehweb.aaas.org/mge/Reports/Report1/AGEP/
- Glassdoor. (2015, February 17). 25 highest paying jobs in demand. Retrieved March 20, 2015 <http://www.glassdoor.com/blog/highest-paying-jobs-demand/>
- Goldweber, M, Barr, J., Clear, T., Davoli, R., Mann, S., Patitsas, E., & Portnoff, S. (2012). A framework for enhancing the social good in computing education: A values approach. In *Proceedings of the 17th Annual Conference reports on Innovation and Technology in Computer Science Education - working group reports, ITiCSE-WGR '12*, 2012. <https://doi.org/10.1145/2426636.2426639>
- Google. (2015). Diversity. Retrieved March 20, 2015 from <http://www.google.com/diversity/at-google.html>
- Grandy, J. (1998). Persistence in science of high-ability minority students: Results of a longitudinal study. *Journal of Higher Education*, 589-620. <https://doi.org/10.2307/2649210>
- Handelsman, J. (2005). More women in science. *Science*, 309, 1190-1191. <https://doi.org/10.1126/science.1113252>
- Harada, V. H., & Yoshina, J. M. (2004). Moving from Rote to Inquiry: Creating Learning that Counts, *Library Media Connection*, 23, 22. https://www.researchgate.net/publication/234671410_Moving_from_Rote_to_Inquiry_Creating_Learning_That_Counts

- 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*, 25-55. https://doi.org/10.1207/S15326977EA0603_1
- 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, 155-196. www.tandfonline.com/doi/pdf/10.1207/S15326977EA0603_1
- 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, <https://doi.org/10.1080/00461520701263368>
- Jenkins, T. (2001). The Motivation of Students of Programming. *ITiCSE* (2001). 53–56. <https://doi.org/10.1145/507758.377472>
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction does not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching, *Educational Psychologist*, 41, 75-86. https://doi.org/10.1207/s15326985ep4102_1
- Krajcik, J.S. & Blumenfeld, P. (2006). Project-Based Learning, In Sawyer, R. K. (Ed.), *The Cambridge Handbook of the Learning Sciences*, 317-333, New York: Cambridge University Press.
- Kubiatko, M., & Vaculová, I. (2011). Project-Based Learning: Characteristic and the Experiences with Application in the Science Subjects, *Energy Education Science and Technology – Social and Educational Studies*, 3, 65-74. http://www.kubiatko.eu/clanky_pdf/project_based_learning_%20characteristic_and_the_experiences_with_application_in_the_science_subjects.pdf
- Langer, J. A. (2001). Beating the Odds: Teaching Middle and High School Students to Read and Write Well, *American Educational Research Journal*, 38, 837–880. <https://doi.org/10.3102/00028312038004837>
- 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, 912-946. <https://doi.org/10.1002/tea.20080>
- Mangalindan, J. P. (2014, August 29). How tech companies compare in employee diversity. Retrieved August 10, 2016 from <http://fortune.com/2014/08/29/how-tech-companies-compare-in-employee-diversity/>
- Nolan, D., & Temple Lang, D. (2010). Computing in the Statistics Curricula. *Am. Stat.*, 64(2010), 97–107. <https://doi.org/10.1198/tast.2010.09132>
- Nolan, D., & Temple, L. D. (2009) Approaches to broadening the statistics curricula. In M. C. Shelley, L. D. Yore & B. Hand (Eds.), *Quality Research in Literacy and Science Education* (pp. 357-381). London, UK: Springer. https://doi.org/10.1007/978-1-4020-8427-0_18
- Pasquale, M. (2015). Productive Struggle in Mathematics. *Interactive Technologies in STEM Teaching and Learning: Research brief*. <http://interactivestem.org/wp-content/uploads/2015/08/EDC-RPC-Brief-Productive-Struggle.pdf>
- Rader, C., Hakkarinen, D., Moskal, B. M., & Hellman, K. (2011). Exploring the appeal of socially relevant computing: are students interested in socially relevant problems? In *Proceedings of the 42nd ACM technical symposium on Computer science education, SIGCSE '11*, 423–428, New York, NY. <https://doi.org/10.1145/1953163.1953288>
- Seymour, E., Hunter, A., Laursen, S., & Deantoni, T. (2004). Establishing the benefits of research experiences for undergraduates in the sciences: First findings from a three-year study. *Science Education*, 88, 493-534. <https://doi.org/10.1002/sci.10131>
- U.S. White House, *Ready to Work: Job-Driven Training and American Opportunity*. 2014; <http://ow.ly/VkuOJ>.
- 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, DOI: 10.7771/1541-5015.106. <http://dx.doi.org/10.7771/1541-5015.1061>
- Williams, A. (2010). Statistics Anxiety and Instructor Immediacy, *Journal of Statistics Education*, 18, 1-18. www2.amstat.org/publications/jse/v18n2/williams.pdf

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the [Creative Commons Attribution license](https://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.