Adapting the Passion-Driven Statistics curriculum for transition to high school

Jennifer S. Rose, Ph.D.
Center for Pedagogical Innovation
Wesleyan University

Lisa L. Dierker, Ph.D.
Department of Psychology
Wesleyan University

Please address all correspondence to Jennifer Rose, Center for Pedagogical Innovation, jrose01@wesleyan.edu, 860-658-2406

Adapting the Passion-Driven Statistics curriculum for transition to high school

Summary

We adapted a multidisciplinary, project-based curriculum to introduce high school students to statistical inquiry. We examined student background characteristics, learning experiences and course outcomes for 30 students ages 12-14 entering 9th grade who were involved in a one-week, accelerated version of a Passion Driven Statistics curriculum. On average, students rated the experience as very rewarding and the majority (70.0%) felt they had accomplished more than they had expected. Over half found the course more useful compared to other courses they had taken and two-thirds reported receiving more individualized support. Over 70% reported interest in taking a future programming course, despite the fact that most of them entered the course with no previous programming experience. Results suggest that a multidisciplinary project-based curriculum could be adapted to an accelerated one-week summer experience for youth as young as 12 from under-represented communities, providing a model that introduces statistical inquiry in an engaging manner.

Keywords: Teaching Statistics, project-based curriculum, high school students, statistical inquiry

Introduction

In previous publications we describe the development of an introductory data analysis curriculum (Dierker et al., 2012; Dierker et al, 2015; Dierker et al, 2016; Cooper and Dierker, 2017) 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). Funded by the U.S. National Science Foundation, the project-based course (passiondrivenstatistics.com) engages students with real data and leading statistical software. To develop the curriculum, we followed 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) endorsed by the American Statistical Association. The GAISE report is the result of a collaboration among multiple statistical educators and researchers, whose recommendations were informed by outreach to the statistics education community and by reference to the statistics education literature (Wood et al., 2018).

The Passion Driven Statistics course seeks to spark interest and excitement in students about using data to answer research questions that interest them personally. It aims to train students to approach problems with a statistical eye, so their skills become a natural part of their intellectual toolkit and shape their perspective on the world. Students work on projects of their own design in class, allowing them to spend class time engaged in projects that pique their natural curiosity, learning the statistical methods necessary to answer questions specific to their interests. Students build on their statistical knowledge organically as they move through their project, forming more meaningful connections between concepts and statistical methods. By designing the curriculum around research projects of students' own choosing, the course offers individualized hands-on experience in data analysis and applied statistics. 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.

At the post-secondary educational level, this curriculum has attracted higher rates of under-represented minority (URM) students compared to 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 higher rates of both female and URM enrollment (Cooper & Dierker, 2017) in the project-based course. While URM students consider the material covered in the project-based course more difficult than non-URM students, URM students have 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 twice as likely as non-URM students to report increased interest in conducting research (Dierker et al, 2016). More recently, we presented data on an adaptation of the curriculum as a two week summer program for rising high school seniors as part of the federally funded educational initiative, Gaining Early Awareness & Readiness for Undergraduate Programs (GEAR UP). GEAR UP is sponsored by the U.S. Department of Education, and 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). This pilot program demonstrated both the feasibility of this model as well as its promotion of positive student experiences within the context of data analysis and applied statistics. Students for example, rated the experience as rewarding and the vast majority (78.1%) felt that they had accomplished more than they had expected (Dierker et al, 2017).

Based on these promising findings, we adapted the curriculum further in an effort to engage younger students in an accelerated and abridged version of the project-based curriculum, offered during a one-week summer program designed to expose adolescents from underrepresented communities to a college learning environment. Students enrolled in the GEAR UP program ranged in age from 12-14, and would be entering the 9th grade in the fall. The purpose of this abridged curriculum was to introduce these young adolescents to the process of statistical inquiry over the course of the week, from learning how to develop a testable research question through interpretation and presentation of findings. This report describes learning experiences and outcomes for those students participating in this project-based curriculum.

Methods

Participants

Students completing middle school were invited to attend an accelerated summer program offered on a college campus during June 2017 and were recruited through flyers, personal letters or conversations with teachers and/or GEAR UP Site Coordinators. The project-based curriculum was presented to students as “Summer Data Camp: Storytelling with Data”. Parents signed permission slips for students to attend the camp.

Students used the Wave 1 public data set from the National Longitudinal Study of Adolescent to Adult Health (Add Health; https://www.cpc.unc.edu/projects/addhealth/documentation/publicdata) in the course. They first explored the Add Health data codebook to develop a research question of interest to them, and then used the Add Health data set to conduct data management and analysis to answer their research question using basic descriptive statistics, graphing results, and interpreting and presenting findings. Students learned to write code-based data management and analysis programs by accessing the free SAS Studio cloud application through a web browser (https://www.sas.com/en_us/software/on-demand-for-academics.html). Classroom sessions included brief presentations and ample one-on-one mentoring for three hours on five

consecutive weekdays. Final projects were presented at a research poster session, during which students described their results.

Participating students represented 7 middle schools from 5 different school districts, 2 rural and 3 urban. Data were drawn from pre/post student surveys with the pre survey completed prior to the start of the summer program and the post course survey completed during the last class session. The sample included 30 students who completed a post survey, 24 of whom also completed a pre survey.

Measures

Background characteristics. Reported race/ethnicity were measured through non-mutually exclusive categories (Black, Hispanic, White, Asian, other). 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 and programming experience in the pre-survey. Students were asked whether they had previously “taken a course that taught how to analyze or look at data”. Prior experience with general programming and/or code-based statistical software was also evaluated by asking students to “describe any experience you have had with writing code or programming”.

Experiences with the project-based curriculum. Students’ perceptions of 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?”, 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.

Students were also asked if they would recommend this course to other students (Yes, probably; probably not; or no). Finally, in open-ended form, students were asked to describe what, if anything they had learned, noting in particular what each believed was most valuable.

Outcomes. To assess improvements in student’s ability to ask relevant and feasible questions with data, in both the pre and post surveys, students were asked to consider a hypothetical data set and to describe a possible research question that could be answered by analyzing the data. Responses were coded according to 1) whether or not answering the question would be feasible based on the description of the data set; and 2) the number of variables involved in the question (i.e. question complexity).

An additional outcome measure considered if students would like to take one or more courses as a follow-up to the project-based data analysis course. Options included a course in graphing and data visualization, constructing data sets, writing about data, computer programming, and/or other. Individual courses were examined separately in the pre and post survey and an

aggregate variable was constructed indicating whether or not a student endorsed an interest in taking at least one follow-up course.

Suggestions for future versions of the project-based course. In an open-ended format, students were also asked for their suggestions regarding future versions of the course.

Results

Student Characteristics

The mean age was 13.6 (s.d. 0.59; range 12-14) and 46.7% of the sample was female. A total of 30.8% of students were Hispanic, 53.8% White, and 15.4% Asian. Fifteen students (65.2%) reported that they would be the first generation in their family to attend college. For the majority of students (70.8%), this course was their first experience with programming and 45.8% reported that they had not learned how to analyze or look at data in a math class.

Based on responses from students completing the post-survey, self-reported experiences in the project-based course are presented in Table 1. Ratings for the question “how hard did you work in this class?” suggested that, on average, students believed that they worked very hard, but rated the experience as very rewarding, the majority (70%) felt that they had accomplished more than they had expected. All 30 students reported that they would (70%), or probably would (30%), recommend the course to other students.

Two-thirds of the sample reported receiving more individualized support compared to other courses they had taken and 53.3% felt that the individualized support was more useful than what they had received in other courses.

Table 1. Student experiences with the project-based curriculum (N=30)

Engagement	N=30
More effort put in than other courses *n* (%)	14 (46.7%)
How hard did you work in this class? (1 = not at all, 5 = extremely hard)	M 4.0 (IQR 1.0)
Rigor	
How challenging did you find this course? (1 = not at all to 5 = the most challenging)	M 3.0 (IQR 1.0)
Course was *more* challenging than other courses, *n* (%)	8 (26.7%)
Overall Impressions	
How rewarding did you find this course? (1 = not at all, 5 = extremely rewarding),	M 4.0 (IQR 1.0)
Accomplished *more* than expected *n* (%)	21 (70.0%)
Compared to other courses	
Course was *more* interesting *n* (%)	17 (56.7%)
Course was *more* useful *n* (%)	13 (43.3%)
More likely to use the skills and content that you learned *n* (%)	9 (30.0%)
Received *more* individualized support *n* (%)	20 (66.7%)
Individualized support was *more* useful *n* (%)	16 (53.3%)
Recommend	
Yes	21 (70.0%)
Probably	9 (30.0%)
Probably not	0 (0%)
No	0 (0%)

Note: M=median, IQR=interquartile range.

There was a wide variety of research questions presented at the poster session, for example "The association between paid labor and desire to go to college among adolescents"; "Gender differences in rates of depression"; and "The association between being lonely and too tired to do homework". Students enjoyed the opportunity to tell their data story at the poster session. A video showing students at the poster session can be found at https://passiondrivenstatistics.com/2018/07/30/11-passion-driven-statistics-utah-gear-up/.

Table 2 shows open-ended responses organized by theme to the post survey item asking students to describe what, if anything they had learned, noting in particular what each believed was most valuable. Although several students described steps needed to analyze data as most valuable, student comments most often reflected the value of learning to write SAS syntax (also referred to as "coding") and to run the syntax. Additional open-ended comments described understanding and communicating results and learning other forms of technology as most valuable. Further, individual support was the most valued by one student and another saw the unique value of personal engagement and effort.

Table 2. Open-ended responses to item asking students to describe most valuable thing they learned.

Theme	Responses
Steps needed to analyze data	"I feel like I learned another way to analyze data."; "I learned how to manage data. I believe that that was the most important part."; "The most valuable thing that I learned here was probably organizing the data."; "How to analyze data a lot better than in any other classes that I have had to analyze data and other things in"; "I have learned how to reconstruct data so I can learn and answer a question with it."
Learning to write and execute code	"I learned how to code, and how to make graphs, and learn more about data"; "The most valuable thing I have learned is coding and deciphering data."; "Coding was difficult but I learned a lot."; "I believe that the coding section was the best and most valuable."; "SAS coding"; "I learned how to code. I've never had this experience"; "I learned coding."; "I learned a lot of coding that will stick with me."; "Learning to code"; "I learned how to code."; "I learned more about coding itself and how it works."; "I learned how to use SAS coding and how to make good charts."; "Coding was very fun, also finding how to take data into coding and get output data to use was also very, very fun."; "SAS coding"; "The best thing that I learned was that I found out how to code."; "Learning to code, and understanding computer programing."; "I loved learning the SAS code. It was interesting and if I could learn more I definitely would."
Understanding and communicating results	"I found that the writing portion was valuable because it helped me expand my writing skills"; "I got more practice with interpreting graphs and charts"; "I learned more about how to properly present an idea using correct vocabulary and writing it in

	clear thoughts"; "how to put data on a graph"; "I learned how to find a question interesting to myself, learn and research about it, and present my findings"; "Finding ways to present my data set findings"
Learning other forms of technology	"I learned how to use google sheets and google slides for the first time."
Individual support	"The one-on-one time with the teachers really helped a ton."
Personal engagement and effort	"I feel like I learned that you don't find this class fun and engaging unless you make it so. You need to make it fun and engaging. What I believe was the most valuable was that this won't come easy at first."

In response to the question "How do you feel this experience could change or benefit your future?", one student responded, "It gives me experience on things that could be important in a future career or in college". Another student responded, "It really helps me learn more skills about statistics". Additional student interviews can be seen at (https://passiondrivenstatistics.com/2018/12/01/data-camp-at-american-preparatory-academy/).

Pre- and Post-Survey Outcomes

Of the 23 students who responded to the item requesting a possible research question based on a hypothetical data set in both the pre and post survey, fewer than half (n=10, 43.4%) wrote a question that could feasibly be asked (e.g. "Does gender and age affect the time you spend on the internet?"). Eight students (34.8%) wrote a research question in the pre-survey that could not be asked of the hypothetical data set (e.g. "How proficient are you doing with money?"), and five (21.7%) reported that they "Don't Know". In the post-survey, 18 students (78.3%) wrote a research question that could be feasibly asked of the data.

No clear improvements were found in question complexity among the 10 students who wrote a question that could be feasibly asked of the data in the pre-survey. However, among the eight students who did not articulate a question that could be asked by analyzing the data in the pre survey, but were able to in the post survey, all were able to articulate questions including more than one variable. These included "Does gender and age affect how much you are online?", "Does gender affect your personal income?", "What is the relation between age and income?", "Do males or females get paid more?", "When you are younger do you use more internet then older people?", "Does spending time online affect your income?", "Does gender manipulate how much income someone has?", and "Could the varying income of [an] individual affect the usage of online sources?"

Finally, when examining student interest in taking future courses in the area of data analysis, 29 of the 30 students (95.8%) reported interest in one or more follow-up courses including graphing and data visualization, constructing data sets, writing about data, computer programming/coding and/or other. However, no change was observed between the pre- and post-surveys. That is, students reported similar interest at both the start and end of the course. In both the pre- and post-surveys, students reported the highest rate of interest in computer/programming (i.e. 70.8% and 75.0% respectively).

Students' suggestions for future versions of the project-based course.

Students were also asked for their suggestions regarding future versions of the course, with their open-ended comments reflecting 1) *simplified and/or clearer content* (i.e. explaining things easier", "I suggest maybe try questions that aren't as confusing", "explain things easier." and "Teach at a 8th grade level not a college level"); 2) *greater support for presentations* (i.e. "I would probably make the presentations into power points, so that we can be confident in presenting inside and outside of our normal classes."); 3) *the need for additional personal support* (i.e. "For the future versions of this course, I would hire more mentors and teachers. I know that this version probably had as much mentors as possible, but with more, the students would be able to get help faster and finish the current project." and "have a student that somewhat knows what's going on so they can be leader"); and 4) *more fun* (i.e. "the topics are understandable, but try making more fun topics so it's not such a serious thing but you still learn in the class", "make it less boring" and "more specialized topics").

When asked for suggestions, two students suggested the need for more time on projects (i.e. "More time to complete it" and "maybe have more time"); while another felt we could "accomplish more in a shorter amount of time". Another two students suggested more opportunity for coding.

Finally, several others replied with their overall satisfaction with the course (i.e. "I thought that it was a good course", "I have no suggestions because this camp was actually better than I thought it would be.", "Nothing really it was great.", "I have absolutely no suggestions for a future version of this class because I thought it was an amazing class that all students should take" and "Just keep teaching the same thing.").

Discussion

In previous reports, we have described the development of an NSF funded project-based, data analysis and applied statistics course. We have also demonstrated 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, 2017). Despite the short duration of the adapted course described in the present study, we have demonstrated the feasibility of this project-based approach when meeting with students just three hours per day across a one-week period.

While, 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. Previous research has recognized the general challenge in teaching courses in data analysis and applied statistics. There is no typical student; instead, students come into courses with differing backgrounds, experiences, learning styles and levels of preparation. The present results suggest that our passion-driven, project-based approach to teaching data analysis and applied statistics can be effective in engaging youth as young as 12 in the statistical inquiry process. On the post-test, we saw an increase from 10 to 18 students who were able to develop a testable research question.

We were not able to measure the extent to which this exposure motivated students to take future courses in the area of data analysis, as the vast majority of the students (96%) reported

interest in taking future courses in the area of data analysis, and no change was observed between the pre- and post-survey responses. However, over 70% of the students reported that they were interested in taking a programming course in the future, despite the fact that most of them entered the course with no previous programming experience. In addition, statistical software programming turned about to be one of the skills they liked learning the most. As one student noted, "Whenever people talk about coding I always kind of think 'Like, that's a nerd', but I learned it in this course and I learned a lot, and I really liked it. It was fun."

Notably, this model provokes students to encounter (and struggle with) data. Although students believed that they worked very hard in the course, they rated the experience as very rewarding, and all 30 students reported that they would (70%), or probably would (30%), recommend the course to other students.

Although the 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 (Kubiatko & 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. All of our course materials, including the full set of videos used in the course taught to Wesleyan students, are freely available on the Passion Driven Statistics web site (https://passiondrivenstatistics.com/), and we encourage faculty to consider integrating the project-based course content. The Passion Driven Statistics curriculum could be adapted to provide additional mentoring and instructional support to younger students. Additional details about Passion-Driven Statistics, GEARUP partnership can be found at https://passiondrivenstatistics.com/2016/09/21/.

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 P334A150032 from the U.S. Department of Education. We appreciate the generous efforts of teachers and advisors who continue to contribute to the development and implementation of this program. The authors declare no conflict of interest.

References

Aliaga, M., Cuff, C., Garfield, J., Lock, R., Utts, J., and 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

Buck Institute for Education (BIE) and Boise State University, Department of Educational Technology. Retrieved April 1, 2012 from http://pbl-online.org/.

Carver, R., Everson, M., Gabrosek, J., Rowell, G. H., Norton, N., Lock, R., and 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.

Cooper, J. L., and Dierker, L. (2017) "Increasing Exposure to Programming: A Comparison of Demographic Characteristics of Students Enrolled in Introductory Computer Science Programming Courses vs. a Multidisciplinary Data Analysis Course," International Research in Higher Education, 2(1), 92-100.

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.

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.

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.

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), Article 2.

Dierker, L., Ward, N., Alexander, J., Donate, E. (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), Journal of Education and Training Studies. 5(4), 54-63.

Kubiatko, M., and 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.

Wood, B., Mocko, M., Everson, M., Horton, N.J., Velleman, P. (2018) Updated Guidelines, Updated Curriculum: The *GAISE College Report* and Introductory Statistics for the Modern Student, CHANCE, 31:2, 53-59.