Work In Progress: Improving student engagement in undergraduate bioinformatics through research contributions

Dr. Jessica Dare Kaufman, Endicott College

Jessica Kaufman began her engineering career as a chemical engineering major at The Cooper Union for the Advancement of Science and Art. After graduation, she worked as a process engineer, primarily in food and pharmaceuticals. Her work in biopharmaceuticals inspired her to earn a doctorate in Biomedical Engineering at Boston University. Since 2008, Jessica has worked at Endicott College and taught a wide range of biotechnology and bioengineering courses. Her primary research interests are bioinformatics and the mechanics of biomaterials.

Work in Progress: Improving student engagement in undergraduate bioinformatics through research contributions

Abstract

CUREs (Course-Based Undergraduate Research Experiences) are a model for changing laboratory design to include hallmarks of authentic research. CUREs have been shown to have similar improvements in student research skills to apprentice-based research experiences [1]. Unlike apprentice-based experiences, course-based research is available to all students in a program. At Endicott College, the undergraduate bioinformatics lecture course is required for two majors, Biology/Biotechnology and Bioengineering. The addition of an authentic in silico research experience as a project for this lecture course would thus expose all students in both majors to an independent research experience. A single post-project survey was used to measure differences in student perception of themselves as a researcher by asking students to rate how the research project increased their knowledge in the use of scientific practices, was an act of discovery, required collaboration, required iteration, and had relevance outside the course. These categories have been used to assess other CUREs and define the characteristics a CURE [2]. Students who participated in both the traditional lab to sequence DNA and the computational analysis were compared to students who only participated in the computational research project. The initial results indicate that there was no significant difference between the survey responses of the two groups and that a computational CURE may have similar impact without including a traditional lab component. Further study of the project design and impact on students is planned for future semesters.

Introduction

Most CUREs have been designed for laboratory courses or for joint lecture and laboratory courses. This model works well for investigations in molecular biology or chemistry, but many computational research tools are taught in a lecture course only. There is evidence that lecture courses can also be improved with the CURE model. The Genome Solver Project is a recent example of a similar bioinformatics project designed for a lecture course at five institutions. With the incorporation of a CURE, researchers were able to show increased student performance on a post-course quiz[3]. When the Small World Initiative CURE was taught at Florida State University, student lecture grades were improved [4]. In this proposed study, I will examine if student engagement in research can be similarly impacted for students in a lecture by comparing two groups of students within the same lecture course.

The inclusion of quantitative biology is an important goal in many of the proposals to improve undergraduate biology education [5]. In Krim et al, three types of research experience programs

are compared: course-based undergraduate research experiences, undergraduate research experiences and teacher research experiences [6]. One conclusion from this literature review is that persistence in science is influenced by the level of project ownership [6].

Course and Project Description

BIO340, Bioinformatics, is a required course for two majors at Endicott College. Students in the Biology/Biotechnology major take a two-course sequence junior year in the field of genomics. The fall course, Genes and Genomes, includes a laboratory focused on resequencing a microbial genome. The spring course is Bioinformatics. The first group of students participated in two CUREs, a laboratory CURE in Genes and Genomes lab and a computational CURE in Bioinformatics lecture. The second group of students only participated in a computational CURE in the Bioinformatics lecture course.

The project in Genes and Genomes incorporates many elements of an authentic research experience. Students isolate DNA from a commercial microorganism in a probiotic consumer product. In Fall 2019, the students isolated DNA from Life Extension brand Bifido GI Balance pills (Life Extension, Ft. Lauderdale, FL). The pills are labeled as containing 2 billion colony forming units of Bifidobacterium longum BB536. After fragmentation of genomic DNA with a Covaris ultrasonicator (Covaris, Woburn, MA), the students then use the NEBNext Ultra II library preparation kit (New England Biolabs, Ipswich, MA) to prepare libraries for Illumina Sequencing. The samples are sequenced by the class on an Illumina MiSeq using a 600-cycle v3 kit (Illumina, San Diego, CA).

The project in Bioinformatics also incorporates aspects of research. Students obtain the FASTQ files from Illumina's cloud portal, Basespace, and upload this raw data to usegalaxy.org, the public instance of Galaxy [7]. FASTQ files are trimmed, aligned to the Bifidobacterium reference genome from National Center for Biotechnology Information (NCBI), and variant called to produce a vcf file for each student's assigned gene using a workflow designed by each student. Students then use a Shiny website written for this project to convert their VCF file that lists variants to a FASTA sequence file [8]. The project files for the website are available on Github [9]. Finally, students submit their FASTA sequence file to NCBI to complete the assignment and gain project ownership from sharing their research results.

Methods and Timeline

The bioinformatics project will act as a CURE because it incorporates key factors including **outside stakeholders** by asking students to publish their FASTA sequence files to Genbank, **discovery** since there were previously no published sequence files for this strain, **collaboration** as students are encouraged to help each other choose methods and best settings for the variant

call workflow, and **iteration** as students must rerun tools to find the workflow that produces the most accurate list of variants in their assigned gene.

Preliminary Results

In Spring 2020, there are 22 students enrolled in the Bioinformatics course. The survey shown in Appendix A was submitted to the Endicott College Internal Review Board and was approved on February 21, 2020. The voluntary survey was given to students in class on March 11, 2020 and was completed by 18 of the 22 students. The survey results for all Likert scale questions are shown in Figure 1. Students differ in major, year, and prior experiences, but Figure 1 only considers if a student participated in a lab and a computational CURE or only the computational CURE. The small sample size of 18 students makes analyzing the role of confounding factors challenging. I plan to continue this study with future cohorts to look at these factors such as major.

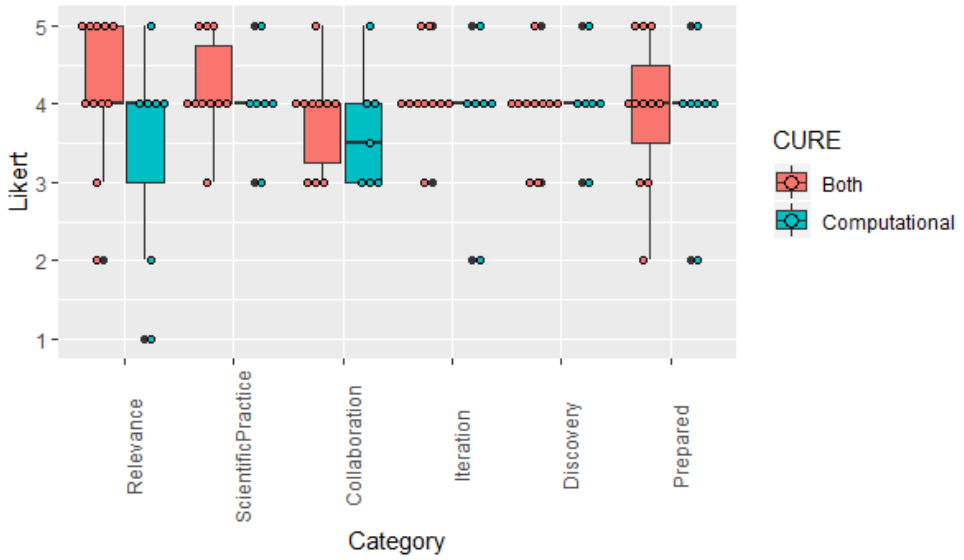

Figure 1. Box plot with individual survey results for students who participated in lab and computational CURE (Both) and students who only in computational CURE (Computational).

A Mann-Whitney-Wilcoxon test was used to compare the Likert scale results from each survey question by CURE group. None of the results were significantly different with the following p-values: Relevance (0.2085), Scientific Practice (0.5708), Collaboration (0.5611), Iteration (0.7405), Discovery (0.7909), and Feel Prepared for own Research Projects (0.9601). This preliminary result supports the hypothesis that there would be no significant difference between the groups. Further study of the impact of this computational CURE is needed to examine the role of project design, student major, year of study, and other confounding factors.

References

[1] C. Shapiro, et al., "Comparing the Impact of Course-Based and Apprentice-Based Research Experiences in a Life Science Laboratory Curriculum," *J. Microbiol. Biol. Educ.*, vol. 16, no. 2, pp.186-197. Dec. 2015.

[2] L. C. Auchincloss, et al., "Assessment of course-based undergraduate research experiences: a meeting report," *CBE Life Sci Educ*., vol. 13, no. 1, pp. 29–40. Oct. 2017.

[3] V. Mathur, et al., "The Genome Solver Project: Faculty Training and Student Performance Gains in Bioinformatics, " *J. Microbiol. Biol. Educ.*, vol. 20, no.1, pp. 4-14. Apr. 2019.

[4] J. P. Caruso, et al., "Citizen Science: The Small World Initiative Improved Lecture Grades and California Critical Thinking Skills Test Scores of Nonscience Major Students at Florida Atlantic University," *J. Microbiol. Biol. Educ.*, vol. 17, no. 1 pp. 156-62. Mar. 2016.

[5] A. M. Depelteau, et al., "SYMBIOSIS: development, implementation, and assessment of a model curriculum across biology and mathematics at the introductory level," *CBE Life Sci Educ*., vol. 9, no. 3, pp. 342-7. Fall 2010.

[6] J. S. Krim, et al., "Models and Impacts of Science Research Experiences: A Review of the Literature of CUREs, UREs, and TREs," *CBE Life Sci Educ*., vol. 18, no. 4, pp. ar65. Winter 2019.

[7] E. Afgan, et al., "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update," *Nucleic Acids Research*, vol. 46, no. W1, pp. W537–W544, July 2018.

[8] "Endicott College Bioinformatics: VCF2FA ," Endicott College November 26, 2019. [Online]. Available: https://ecbio340.shinyapps.io/R-VCF2FA-master/

[9] "Shiny App to convert VCF to FASTA sequence file," Class resources for BIO340 Bioinformatics at Endicott College, Department of Bioengineering, Beverly, MA. [Online]. Available: https://github.com/profbiot/R-VCF2FA

Appendix A

Proposed survey as submitted to Endicott College Institutional Review Board

BIO340 CURE Survey

Question Title
1. Have you taken BIO311 Genes & Genomes?
Yes in Fall 2019
Yes in Fall 2018
Yes in Fall 2017
No

Question Title
2. How would you rate how the Unit 1 Project: Variant Calling increased your knowledge for each category?

Greatly decreased
Decreased
Neutral
Increased
Greatly increased

Use of scientific practices
Use of scientific practices Greatly decreased
Use of scientific practices Decreased
Use of scientific practices Neutral
Use of scientific practices Increased
Use of scientific practices Greatly increased

Discovery
Discovery Greatly decreased
Discovery Decreased
Discovery Neutral
Discovery Increased
Discovery Greatly increased

Collaboration
Collaboration Greatly decreased
Collaboration Decreased
Collaboration Neutral
Collaboration Increased

Collaboration Greatly increased

Iteration
Iteration Greatly decreased
Iteration Decreased
Iteration Neutral
Iteration Increased
Iteration Greatly increased

Relevance outside this course
Relevance outside this course Greatly decreased
Relevance outside this course Decreased
Relevance outside this course Neutral
Relevance outside this course Increased
Relevance outside this course Greatly increased

Question Title
3. After this course, I feel more prepared for my own research projects.
Strongly Disagree
Disagree
Neutral
Agree
Strongly Agree

Question Title
4. After the semester is complete, your project will be assessed with a rubric for how well it met course objectives. Do you consent to this assessment data (not your report or your grade) being shared in a poster presentation at the American Society for Engineering Education?
Yes
No

Question Title
5. Was this project similar to research in laboratory classes? How was it similar? How could this project better connect to open-ended research questions?