DATA SCIENCE EDUCATION K-12

INDIVIDUAL PAPER: Iterating Toward a Student-Centric High School Data Science Curriculum

Peter F Moon, University of Maryland, pmoon@umd.edu
Rotem Israel-Fishelson, University of Maryland, rotemisf@umd.edu
Myles Proctor, Washington Leadership Academy, mproctor@wlapcs.org
David Weintrop, University of Maryland, weintrop@umd.edu

Session Description

Tailoring learning materials and activities to the learners is crucial for enhancing their engagement and interest. With a student-centered approach and iterative design, we developed a new interest-driven API-based data science curriculum for high school students. We revised our pilot curriculum based on feedback from our pilot teacher, student performance and course evaluations, and class observations. Key modifications included incorporating real-world examples of data science applications, expanding coding activities, and redefining class discussions to improve student involvement. Here, we summarize some of these changes made to support the development of data scientist identities and increased student engagement. This work highlights the significance of research-practice partnerships and recommends leveraging feedback from both educators and students to enhance curriculum delivery in K-12 settings. It contributes to the evolving field of data science education in K-12 classrooms and emphasizes the value of collaborative curriculum development based on practical classroom experience and feedback.

Submission [850 words]

Background

Researchers are currently developing an increasing number of curricula for data science education at the K-12 level (e.g., Gould et al., 2022; Bootstrap, 2022; YouCubed, 2022; CodeHS, 2022; Walker et al., 2023). While much has been published on the curricular design of these programs, there is less focus on their pedagogy (Mike, 2020), a vital component of helping students develop their knowledge of data science.

Research-practice partnerships have great potential to align educational research and curriculum development with school and classroom goals (Welsh, 2021). Such partnerships are particularly powerful because teachers may have valuable insights into implementing and improving the curriculum effectively to support the identified learning goals. In this work, we describe how the data collected from students, along with our partnership with our pilot teacher, helped us enhance our pedagogy in a high school data science curriculum that situates data science in the lived experiences of students.

Methods & Participants

From November 2023 to March 2024, we piloted our initial three-unit curriculum, API Can Code 1.0, with a pilot teacher teaching two computer science courses in Python at a public charter school in the US. Twenty-five students (15 male, 10 female) enrolled between the two courses participated in the study. All the students completed 24 lessons, with each lesson lasting 90 minutes. More details about the curricular units and activities can be found here[1].

We collected data from several sources. All class meetings were video recorded using a camera set up in the classroom. Students completed a survey before and after the course that included questions on their learning and enjoyment of the course, as well as questions about their identity as data scientists. In addition, at the end of each unit, we interviewed five students. We also received data on lesson effectiveness from our pilot teacher during our weekly Zoom meetings.

Results

A major goal of the course was to help foster a data scientist identity in students and expand the notion of who participates in data science. However, students responding to the question *"How much did you feel like a data scientist throughout this curriculum?"* averaged a 3.25 out of 5, less than we hoped. We built on this finding in two

[1] The full curriculum as well as accompanying materials can be accessed at https://apicancode.umd.edu

ways: first, we added additional references to what data science can look like throughout the course. For example, when discussing an NBA dataset, we added prompts to discuss how coaches and players can do data science using team and individual performance datasets to compare players, look for patterns, and strategize for future games. We made similar modifications in another lesson that uses an earthquake dataset, emphasizing the contribution of data science to the work of geologists. Second, based on responses to the question *"When did you feel like a data scientist in this curriculum?"* that focused on coding activities, we expanded on our existing EduBlocks programming activities. We added chunks of code that explored the structure of data (such as nested structures within JSON files) that students could swap in and out of their programs to explore what the chunks did (Figure 1).

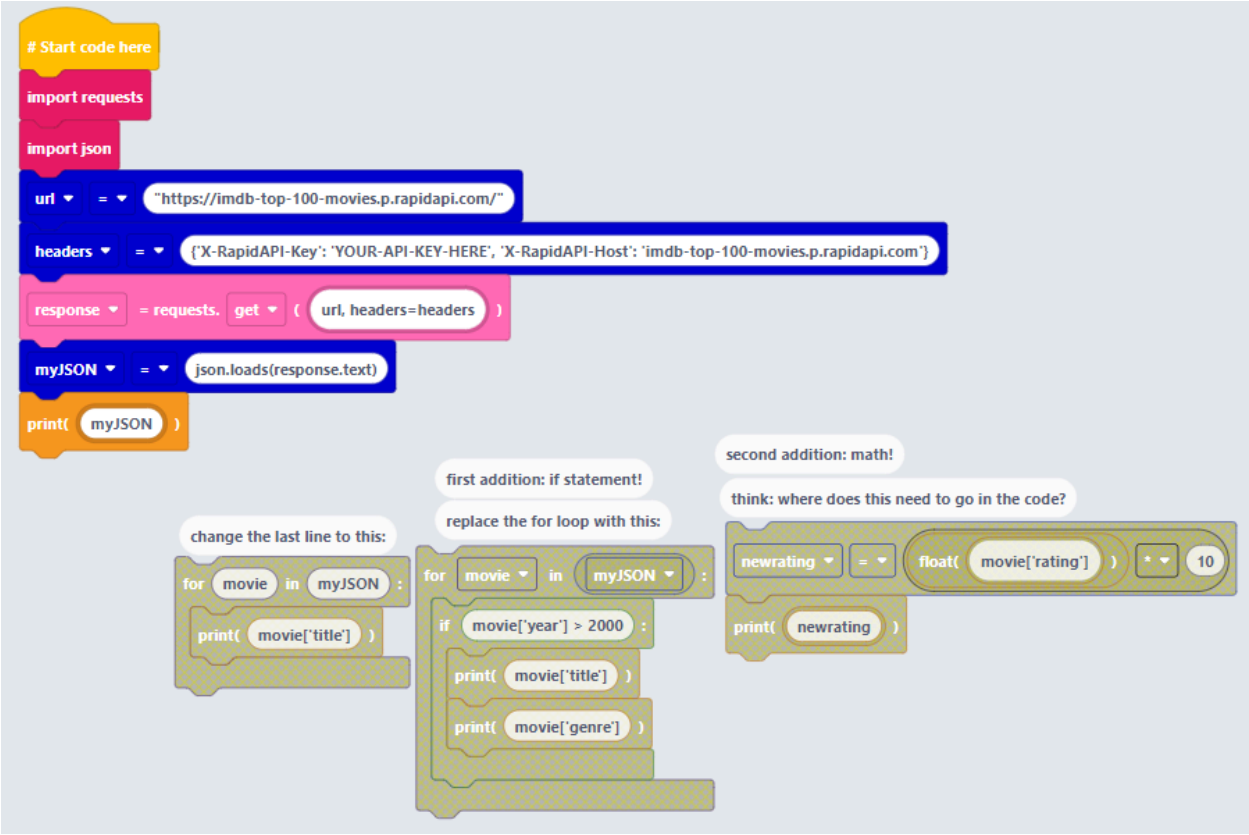

Figure 1. Starter code, with commented chunks that students can add and modify to explore.

Student engagement is a concern for any curriculum. Both classroom video and feedback from our pilot teacher suggested that students often lost focus and interest during "Turn-and-Talk" activities, where we asked them to discuss answers to questions with a partner, and in class discussions. Our pilot teacher suggested using a "Class Debate" prompt to reframe class discussions, which he had used successfully in our lesson on data privacy (Figure 2). We modified several lessons throughout the curriculum to include this and similar discussion starters that encouraged students to defend a data science viewpoint based on what they had learned from class.

Figure 2. A "class debate" prompt encouraging students to defend their opinions around data privacy.

Takeaways

Data science education in K-12 classrooms is a rapidly-developing field. Curriculum designers should consider ways to emphasize the development of data science identity through their lessons, as well as considering how to boost student engagement in discussions and other activities throughout the course. A major asset in this area may be evaluations from learners and feedback and especially suggestions from partner or focal instructors teaching the curriculum, who may have practical experience with grade-level pedagogy to enhance the delivery of the curriculum.

Acknowledgments

This work is supported by the National Science Foundation (Award # 2141655). Any opinions, conclusions, and/or recommendations are those of the investigators and do not necessarily reflect the views of the National Science Foundation.

References

Bootstrap. (2022). *Bootstrap:Data Science.* https://bootstrapworld.org/materials/data-science/

CodeHS. (2022). *CodeHS - Data Science Course.* https://codehs.com/course/data_science/overview

Gould, R., Machado, S., Johnson, T.A., & Molyneux, J. (2022). *Introduction to data science curriculum.*

Mike, K. (2020, August). Data science education: Curriculum and pedagogy. In *Proceedings of the 2020 ACM Conference on International Computing Education Research* (pp. 324-325).

Walker, J. T., Barany, A., Acquah, A., Reza, S. M., Barrera, A., Guzman, K. D. R., & Johnson, M. A. (2023). Coding like a data miner: A sandbox approach to computing-based data science for high school student learning. *2023 IEEE Frontiers in Education Conference (FIE)*, 1–5.

Welsh, R. O. (2021). Assessing the quality of education research through its relevance to practice: An integrative review of research-practice partnerships. *Review of Research in Education*, *45*(1), 170-194.

YouCubed. (2022). *Explorations in data science.* Youcubed High School Data Science Course. https://hsdatascience.youcubed.org/