International Society of the Learning Sciences

Integrating Data Science Explorations in Science Classrooms

Bodong Chen, Rukmini Avadhanam
University of Minnesota
chenbd@umn.edu, avadh005@umn.edu

Abstract: We propose a technology-enriched learning intervention named DataX to engage students in analyzing authentic datasets about societal issues. In a pilot conducted in secondary science, students were first introduced to basic data concepts and then invited to analyze multivariate data about energy and chemical elements. Some students demonstrated sophisticated data skills and playfulness despite initial negative dispositions to data. This study showed early promise of DataX for engaging students in authentic data explorations.

Introduction

Data has been recognized as an important fabric of modern society. It is therefore crucial to seek means to cultivate data-literate citizens so that they are equipped with knowledge, skills, and dispositions important for disciplinary and civic engagement.

Data science as an emerging discipline has garnered significant attention across levels of education (Rosenberg & Chen, 2020; Wilkerson & Polman, 2020). In K-12, a key approach to supporting data science is to interleave it within existing subject areas including statistics, science, engineering, and social sciences. This paper reports on an ongoing design-based research project aiming to develop a technology-enriched learning intervention named *DataX* that positions students as "data analysts" who can build knowledge about important societal issues by collaboratively analyzing authentic datasets in specific disciplinary contexts. In the following sections, we describe the DataX project and report key findings from a pilot study.

The DataX Project

We conceptualize student experiences with data as an integral part of their disciplinary and civic engagement. To support this conception, DataX situates data in a larger inquiry process of value to students, aiming to develop "students' identities as agentive data practitioners" who recognize the roles of data in various social contexts and are both capable and willing to use data skills to drive social change (Wilkerson & Polman, 2020, p. 4).

The DataX online environment builds on CODAP (Common Online Data Analysis Platform), an open-source tool designed for Grades 6-14 students to view, transform, analyze, visualize, and interpret data. DataX extends CODAP from an individual-oriented tool to a community-centric, knowledge-building environment that allows learners to collaborate on data investigations. In particular, DataX allows learners to publish data analysis notebooks for others to comment on or remix.

With the designed technology, we conducted a pilot study to examine how students may use support provided by the DataX intervention. We asked: (a) What data experiences and dispositions did students bring into this intervention? (b) To what extent did DataX facilitate students' data explorations and data dispositions?

Methods

This study involved two public schools in a metropolitan area in the midwestern United States. Participants were two science teachers and 81 students from their 11th grade classes.

Before the classroom intervention, we worked closely with the teachers to co-design pedagogical supports responsive to their classes. Through multiple design meetings, we created an overarching plan for their classes to work on the NGSS crosscutting concept of "energy" from different angels.

During the pilot, a pre-survey on students' prior data experiences and data dispositions was first administered in both classes. After the pre-survey, the researchers conducted a mini lesson on several basic ideas about data, discussing "what is data" and exploring real-world examples about data. In the same session, students were introduced to the DataX environment, followed by a warm-up data activity on visualizing a public dataset. After the introductory session, two classes diverged: Class A did a data investigation on energy consumption in the US, whereas Class B investigated the properties of elements in the periodic table. Researchers and teachers provided continuous scaffolding in both classes. Although more DataX sessions were planned, COVID-19 and the emergency transition to remote learning disrupted the pilot, before we could engage students in further investigations.

Multiple sources of data were collected, including the pre-survey, student-generated data artifacts in DataX, videos of classroom activities, researcher field notes, and teacher interviews. To answer research

questions, we first analyzed student responses to the pre-survey and identified different data dispositions profiles. Using the profiles, we sampled several students and examined their data artifacts on DataX. Researchers' field notes, video recordings, and teacher interviews were incorporated as secondary data.

Results

Overall, students brought a wide range of data dispositions into the study. Descriptive analysis of the pre-survey showed that nearly 80% of the students found data to be interesting or very interesting; about 78% of students indicated that they like or love data-related activities. Students were approximately normally distributed in their interest in reading graphs, thinking about data, seeking data behind news stories, and exploring complex problems. In terms of experiences with data, about 45% of students indicated they were comfortable working with large amount of data, while the remaining 52% said no.

Starting from students' pre-survey responses, we purposefully sampled four students based on their data dispositions and inspected their DataX notebooks. Overall, we found some students with little data experiences or less positive dispositions ended up doing quite well in their data explorations. For example, Student 1 entered the pilot with an interest in data and data activities. This student' DataX artifacts showed the student was able to read the data and create a simple visualization on the natural state and boiling point of elements in the periodic table, but could not "read beyond the data" to interpret what the graph means. In contrast, Student 2 did not like doing data activities but was able to read "between" and "beyond" the data during the intervention. This student visualized the natural states and boiling points of different elements and was able to interpret that boiling solids would take longer than liquids or gases (see Fig. 1, left). In other cases, students demonstrated a willingness to play or tinker with data. Student 3 from Class B independently created a line graph (see Fig. 1, right) using the energy dataset. Even though the student did not come up with a solid interpretation, this representation is peculiarly novel, and he was looking closely at the graph to make sense of the multivariate relationships.

Teacher interviews indicated that the pilot gave students new opportunities to work on and manipulate data and helped students think more about data in science contexts. One teacher said: "DataX allowed students to connect actual scientific information about the concepts we were studying around energy [and] energy usage and apply that to [we] know not only [about] the world now [but also] the world before." Teachers also highlighted that the activities offered authentic learning experiences for students to connect learning to real life situations. Some students expressed an interest in analyzing energy consumption of countries their families immigrated from.

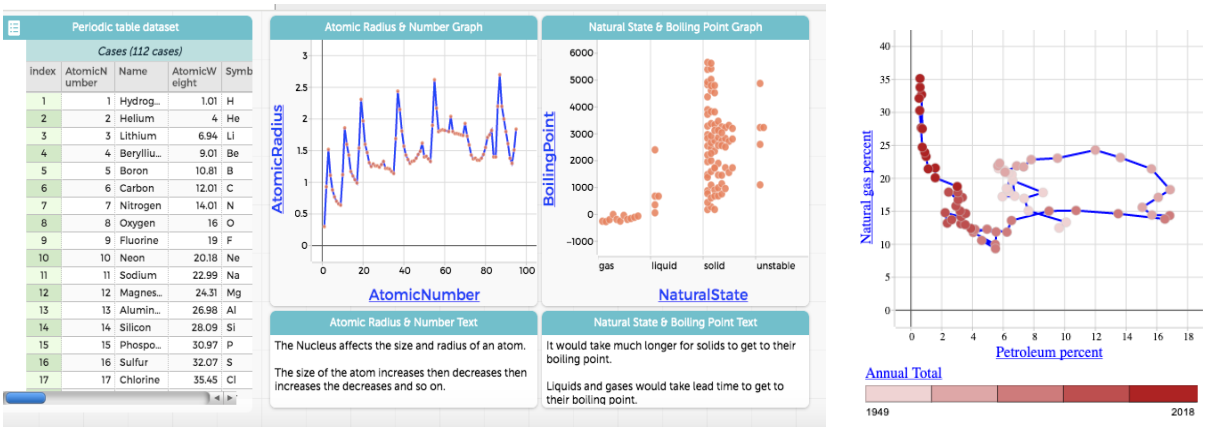

Figure 1. Screenshots of the DataX notebooks created by Student 2 (*left*) and Student 3 (*right*).

Conclusions

The DataX intervention showed early promise for engaging students in authentic data explorations relevant to important disciplinary contexts and societal issues. Future work will continue to revise the technological and pedagogical designs of DataX to facilitate student learning in data science.

References

Rosenberg, J., & Chen, B. (2020). *Exploring data science across the curriculum and across grade levels*. Symposium at the 2020 AERA Annual Meeting, San Francisco, California.

Wilkerson, M. H., & Polman, J. L. (2020). Situating data science: Exploring how relationships to data shape learning. *Journal of the Learning Sciences*, *29*(1), 1–10. https://doi.org/10.1080/10508406.2019.1705664