Integrating Real-World Authenticity into a High School Data Science Curriculum

Rotem Israel-Fishelson, David Weintrop
rotemisf@umd.edu, weintrop@umd.edu
University of Maryland

Abstract: In an increasingly data-driven society, it is essential that students understand and critically engage with the data that surrounds them. A key aspect of accomplishing this is helping students understand the importance of data and the impact it can have on their lives. This paper examines the role of real-world authenticity in a high school interest-driven data science curriculum. Through student reflections and project outcomes analysis, the study highlights how real-world data use fosters data practices by allowing students to see data science as relevant and applicable to real-life issues. Findings indicate that students perceived the data exploration activities as authentic and valued the meaningfulness of the data, recognizing its relevance to real-life contexts.

Introduction

Data influences our world and significantly affects our lives. Thus, there is a growing need to equip students with the skills to navigate and thrive in a data-driven society (Gould, 2021; LaMar & Boaler, 2021). Given the pervasiveness of data in students' lived experiences, there is tremendous potential and opportunity to situate data science instruction in ways that resonate with students' day-to-day lives (V. Lee & Wilkerson, 2018; Weintrop & Israel-Fishelson, 2024). This involves creating engaging and authentic learning experiences that promote the development of data science skills and practices such as data collection, cleaning, analysis, and visualization (Weiland & Engledowl, 2022). By situating data science education within authentic, real-world contexts, educators can create learning experiences that are both relevant and empowering for students (Wilkerson & Polman, 2020). However, a significant challenge in implementing such authentic experiences is the limited availability of curricular materials that provide students with meaningful opportunities to work with real-world data (Wright et al., 2024) and engender a sense of real-world authenticity in students (Shaffer & Resnick, 1999). In this paper, we investigate how real-world authenticity can be integrated into a high school data science curriculum and how students perceive it. This paper advances our understanding of authentic learning experiences and their effectiveness in data science education. It shows how real-time data exploration can be integrated into curricula to boost student perceptions of real-world authenticity and help them connect data science to the world beyond the classroom.

Literature review

The goal of authentic learning is to position learning experiences so that students grasp the value, usefulness, significance, and practicality of the knowledge they will gain (Hagenkötter et al., 2024). There are various forms of authenticity in learning, each emphasizing different aspects of how students connect their experiences to the knowledge they acquire. In this work, we are specifically interested in real-world authenticity, which seeks to reflect everyday life experiences (Shaffer & Resnick, 1999). Such authentic activities involve students tackling complex and ill-structured real-world problems through self-directed investigation using materials and tools commonly employed by professionals (Nachtigall et al., 2022). Data science education is a growing field that combines statistics, mathematics, computer science, and other application areas (Mike et al., 2023; Rosenberg & Jones, 2024). Data science curricula engage students with practices and fundamental skills to understand and interpret the data around them (Biehler et al., 2022; H. Lee et al., 2022). Robust data-gathering tools and real-world data sets can help create authentic learning experiences (Israel-Fishelson, Moon, Tabak, et al., 2024; V. R. Lee & Delaney, 2021; Moon et al., 2023; Pimentel et al., 2022). Incorporating students' interests and real-world data into the curriculum can enhance engagement, foster curiosity, improve the chances of knowledge acquisition, and create more inclusive learning experiences (Brooks et al., 2021; Wilkerson & Polman, 2020).

Methodology

The research was conducted in two computer science classes at a public charter high school in the Mid-Atlantic region of the United States during the 2023-24 academic year. In these classes, taught by a computer science teacher, students engaged with API Can Code, an interest-driven data science program that introduces students to

the computational foundations of data science through real-world data exploration (Weintrop & Israel-Fishelson, 2024). The curricular units center around data exploration activities in which students investigate various datasets that align with students' interests, as identified in our preliminary research (Israel-Fishelson, Moon, & Weintrop, 2024). The study included 25 high school students, 13 females and 12 males, with an average age of 17.2 years. Twenty-two students identified as Black/African American, one as American Indian or Alaska Native, and two chose not to disclose their race. Data was collected through surveys, interviews, and artifacts students produced during the lessons. The data was looking for instances that reflect real-world authenticity, which is the focal aspect of the research questions pursued in this work. To do so, we used Shaffer and Resnick's (1999) operationalization of real-world authenticity, which refers to learning activities that replicate tasks encountered in professional settings and connect to life beyond the classroom, highlighting the importance of practical knowledge application.

Findings

Real-world authenticity focuses on the connection between learning experiences and their applications in everyday life. Shaffer and Resnick (1999) argued that educational materials should *"reflect or recreate some aspect of the world outside of school…for example, authentic activities might ask students to investigate a 'real' problem"* (p. 198). Our analysis found that throughout the curriculum, students acknowledged the connection between the curriculum and real-world situations, showcasing a robust sense of real-world authenticity. The design of curricular activities helped students see the broader relevance of data science beyond the classroom, as one student noted during the post-unit 3 interview: *"The topics were real situations. We used a lot of real data from around the world. It was not made up"*.

The curriculum demonstrates how data science influences daily life and is crucial to various professions. As one student remarked, *"The class was relevant, especially with how he [the teacher] taught it, i.e., describing things and explaining that the whole world relates to data science."* Another student noted, *"We are doing things like the Top 100 movies; it is real stuff related to society. It is like we are doing real work but not like on a serious one like global warming"*. This showcases the ubiquity of data while also demonstrating the potential of situating data science in a context that resonates with learners, even if they acknowledge it is not the more pressing or important topic. Students also recognized the real-world applicability of data science. *"I could see how it will become useful. Even though it is not particularly hand-in-hand with what I'm doing, I'm going to handle a lot of data; it helps me learn how to analyze the data."*

To help make the curriculum authentic, the final project asked students to generate questions based on their interests and then go through a data science cycle to answer them, beginning with creating a program to retrieve and analyze API-based data and then visualize it to answer their questions. This project enabled the students to link and apply the data science practices they have learned to real-world scenarios grounded in their interests and experiences. This approach of anchoring learning activities in real-world data enabled students to link the curriculum to practical and meaningful uses, fostering an authentic learning experience. This aligns with research suggesting that students tend to be more engaged with material when they recognize its relevance to real-world contexts (Anker-Hansen & Andreé, 2019).

Discussion and conclusion

The learning sciences have long recognized the crucial role of authenticity in education, highlighting that learning becomes more effective when linked to real-life settings (Betz et al., 2016). Authentic learning experiences motivate students to engage deeply and purposefully with the educational content, making them essential for fostering meaningful and durable educational results. Data science is uniquely positioned to foster authentic learning experiences due to its inherent connection to real-world contexts and its reliance on diverse, complex datasets that reflect the kind of data encountered in everyday life. In high school data science curricula, the flexibility to integrate up-to-date datasets that span various domains—from social media trends and climate patterns to sports statistics and health data—enables students to engage with topics that can hold personal relevance or reflect real-world issues. Moreover, educators can incorporate datasets on topics that resonate with students' interests and experiences by utilizing and accessing data through services like publicly available APIs.

This work underscores the importance of designing educational experiences that connect classroom activities with real-world contexts, making data science both accessible and meaningful for students as they encounter real and live datasets. In doing so, this work shows the potential for data science as a domain that bridges the classroom to the real-world divide. Given the ubiquity of data in the world, foregrounding real-world authenticity in the design of a curriculum can produce learning experiences that will help learners make sense of the world around them.

References

Anker-Hansen, J., & Andreé, M. (2019). In pursuit of authenticity in science education. *Nordic Studies in Science Education*, *15*(1), Article 1. https://doi.org/10.5617/nordina.4723

Betz, A., Flake, S., Mierwald, M., & Vanderbeke, M. (2016). Modelling authenticity in teaching and learning contexts: A contribution to theory development and empirical investigation of the construct. In C. K. Looi, J. L. Polman, Cress, & P. Reimann (Eds.), *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences*. International Society of the Learning Sciences.

Biehler, R., Veaux, R. D., Engel, J., Kazak, S., & Frischemeier, D. (2022). Research on data science education. *Statistics Education Research Journal*, *21*(2), Article 2. https://doi.org/10.52041/serj.v21i2.606

Brooks, C., Quintana, R. M., Choi, H., Quintana, C., NeCamp, T., & Gardner, J. (2021). Towards culturally relevant personalization at scale: Experiments with data science learners. *International Journal of Artificial Intelligence in Education*, *31*(3), 516–537. https://doi.org/10.1007/s40593-021-00262-2

Gould, R. (2021). Toward data-scientific thinking. *Teaching Statistics*, *43*(S1). https://doi.org/10.1111/test.12267

Hagenkötter, R., Nachtigall, V., Rolka, K., & Rummel, N. (2024). Model authenticity in learning mathematical experimentation: How students perceive and learn from scientist and peer models. *European Journal of Psychology of Education*. https://doi.org/10.1007/s10212-024-00843-4

Israel-Fishelson, R., Moon, P. F., & Weintrop, D. (2024). Interest-Driven Data Science Curriculum for High School Students: Empirical Evidence from a Pilot Study. *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference*, 908–912. https://doi.org/10.1145/3628516.3659416

Israel-Fishelson, R., Moon, P., Tabak, R., & Weintrop, D. (2024). Understanding the Data in K-12 Data Science. *Harvard Data Science Review*, *6*(2). https://doi.org/10.1162/99608f92.4f3ac3da

LaMar, T., & Boaler, J. (2021). The importance and emergence of K-12 data science. *Phi Delta Kappan*, *103*(1).

Lee, H., Mojica, G., Thrasher, E., & Baumgartner, P. (2022). Investigating data like a data scientist: Key practices and processes. *Statistics Education Research Journal*, *21*(2), Article 2.

Lee, V. R., & Delaney, V. (2021). Identifying the content, lesson structure, and data use within pre-collegiate data science curricula. *Journal of Science Education and Technology*.

Lee, V., & Wilkerson, M. (2018). Data use by middle and secondary students in the digital age: A status report and future prospects. *National Academy of Sciences Engineering, and Medicine, Board on Science Education, Committee on Science Investigations and Engineering Design for Grades 6-12*, 1–43.

Mike, K., Kimelfeld, B., & Hazzan, O. (2023). The birth of a new discipline: Data science education. *Harvard Data Science Review*, *5*(4). https://doi.org/10.1162/99608f92.280afe66

Moon, P. F., Israel-Fishelson, R., Tabak, R., & Weintrop, D. (2023). The tools being used to introduce youth to data science. *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*, 150–159. https://doi.org/10.1145/3585088.3589363

Nachtigall, V., Shaffer, D. W., & Rummel, N. (2022). Stirring a secret sauce: A literature review on the conditions and effects of authentic learning. *Educational Psychology Review*, *34*(3), 1479–1516.

Pimentel, D. R., Horton, N., & Wilkerson, M. H. (2022). *Tools to Support Data Analysis and Data Science in K-12 Education* (p. 22). National Academy of Sciences Engineering, and Medicine, Board on Science Education, Workshop on Foundations of Data Science for Students in Grades K-12.

Rosenberg, J., & Jones, R. S. (2024). Data science learning in grades k–12: Synthesizing research across divides. *Harvard Data Science Review*, *6*(3). https://doi.org/10.1162/99608f92.b1233596

Shaffer, D. W., & Resnick, M. (1999). "Thick" authenticity: New media and authentic learning. *Journal of Interactive Learning Research*, *10*(2), 195–215.

Weiland, T., & Engledowl, C. (2022). Transforming curriculum and building capacity in K–12 data science education. *Harvard Data Science Review*, *4*(4). https://doi.org/10.1162/99608f92.7fea779a

Weintrop, D., & Israel-Fishelson, R. (2024). Bringing students' lives Into data science classrooms. *Harvard Data Science Review*, *6*(3). https://doi.org/10.1162/99608f92.6d2aec03

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.

Wright, C., Meng, Q., Breshock, M. R., Atta, L., Taub, M. A., Jager, L. R., Muschelli, J., & Hicks, S. C. (2024). Open case studies: Statistics and data science education through real-world applications. *Journal of Statistics and Data Science Education*.

Acknowledgments