Exploring Interest-Driven Data Science Through Participatory Design

Rotem Israel-Fishelson, Peter F. Moon, Daniel Pauw, David Weintrop
rotemisf@umd.edu, pmoon@umd.edu, dpauw@umd.edu, weintrop@umd.edu
University of Maryland

Abstract: Interest is a powerful catalyst for learning. Cultivating student interest is critical in data science education, where complex concepts and practical skills intersect. This paper explores the use of participatory design as a methodology for applying interest development theory to inform the design of an innovative data science curriculum for high school students. This study presents an analysis of participatory design activities using the Integrated Interest Development for Computing Education Framework (Michalis & Weintrop, 2022), highlighting various operationalizations of interest in data science contexts. By examining the multifaceted dimensions of interest, this work demonstrates how diverse activities provide a context for students to voice their interests. The paper's findings contribute to the discourse on innovative educational approaches that engage students in data science by aligning with their interests.

Introduction

Students need to develop data literacy to navigate our data-driven and information-rich world effectively (Martinez & LaLonde, 2020). The constant interaction of students with data presents a unique opportunity to situate data science in their lives (Wilkerson & Polman, 2020). By harnessing data that aligns with their interests, high school students can develop the skills and conceptual understanding to be informed citizens (Gould, 2021). Incorporating real-world cases that relate to their personal lives and interests can make the learning experience more relevant and meaningful (Lee et al., 2021). Students can learn the fundamental concepts of data science while developing a sense of ownership and empowerment. Such practices can foster a sense of responsibility and ethics in the use of data and help equip them to combat mis- and dis-information in our world (Carmi et al., 2020).

This paper presents efforts to use a participatory design (PD) and the Integrated Interest Development for Computing Education Framework (IIDCEF) (Michalis & Weintrop, 2022) to create an interest-driven high school data science curriculum. The PD sessions were designed to help us gain insights into the students' interests and their perceptions of data in their lives. We analyze and discuss the different dimensions of the framework across the PD and highlight how different operationalizations of interest are manifest in them. This work contributes to a growing body of literature on the role of interest-driven learning in data science. It shows how PD can serve as a mechanism to gain insight into learners' voices, values, and ideas to inform the design of learning experiences.

Literature review

Interest-driven learning

Interest Development Theory has emerged as a framework for understanding how individuals cultivate their interests and passions over time (Renninger et al., 2015). This theory states that interests develop when individuals explore, interact with, and derive meaning from a subject. This theory has gained considerable recognition because it provides a structured approach to enhancing learning experiences, particularly in science education (Harackiewicz et al., 2016). The IIDCEF builds on this work, incorporating three key dimensions:

- *Knowledge* – Knowledge acquisition in a specific area enhances interest through repeated experiences. Three key factors help develop interest and build knowledge: *challenge, relevance, and authenticity*. A *challenge* involves providing tasks supported by scaffolding adapted to the learners' abilities. *Relevance* means adapting learning contexts to the learner's personal experiences and cultural resources. *Authenticity* involves offering real-world scenarios for the knowledge and skills being developed.
- *Value* – Increased value generates interest, and heightened interest fosters a sense of value. Value has two parts: *Personal meaning* involves adapting the content to the learner's values and beliefs, while *personal usefulness* captures the practical benefits of the content for achieving specific goals.
- *Belonging* – The sense of belonging is linked to connecting one's background, experiences, and interests, learning new content, and feeling connected with peers and the community. Expanding notions of *what computing is* and *who can participate in computing* are essential to promote belonging.

Data science education

Data science is a multidisciplinary field related to perceiving, analyzing, and interacting with information. It involves practices for collecting, storing, extracting, and analyzing data to draw conclusions, predict, and make informed decisions (Schanzer et al., 2022). Data science education nurtures civic responsibility by informing students of their role as data producers and consumers and the potential dangers associated with irresponsible or biased uses of data (Carmi et al., 2020). To foster a genuine connection between students' identities and their social and cultural backgrounds, it is crucial to design learning experiences that include relevant datasets and activities (Lee et al., 2021; Wilkerson & Polman, 2020). Connecting the narratives of the educational activities and the studied datasets to the students' interests, views, and prior experiences can deepen engagement and increase the likelihood of knowledge acquisition (Brooks et al., 2021).

Participatory design

Participatory design (PD) is a research methodology that involves end-users in creating solutions, artifacts, and activities to ensure that their voices, values, and priorities are reflected in the process and the resulting final design (DiSalvo, 2016). This approach allows end-users to share unique insights about their needs, preferences, and concerns (Bødker et al., 2022). The methods used often involve a series of structured activities and interactions that facilitate collaboration and communication (Fails et al., 2013) and can help inform the design of culturally responsive curricula (Coenraad et al., 2022).

Methods

Participatory design sessions

We conducted PD sessions to inform the design of a new data science curriculum and ensure the resulting activities reflect the students' values, voices, and, most critically, interests. The central goal was to gain insights into students' interests and experiences with and perceptions of data. The sessions included discussions and seven hands-on design activities (e.g. Coenraad et al., 2022). The study had 28 high school students: 17 male and 11 female, 22 Black/African American, 2 American Indian or Alaska Native, 1 Hispanic, and 1 White student. Proper consent/assent forms were gathered. We describe two activities in the Findings section.

Data collection methods and analysis

We documented all design artifacts created during the PD sessions. Later, we analyzed them using an open coding approach to identify emergent themes (Saldaña, 2016). The analysis process resulted in mapping the design activities by the three dimensions and seven key factors of the IIDCEF. Given the context, we adjusted the last dimension to refer to data science factors rather than computing.

Findings

In this section, we detail two design activities carried out during the PD sessions and map them to the dimensions and key factors proposed by the IIDCEF, along with concrete examples. Figure 1 provides a comprehensive overview of how all PD activities are mapped according to the dimensions and the key factors. The inner ring includes the three dimensions of interest development: Knowledge, Value, and Belonging. The middle ring expands each dimension with the key factors representing unique aspects and approaches to foster interest development. The outer ring shows where the seven key factors can be found across our PD activities.

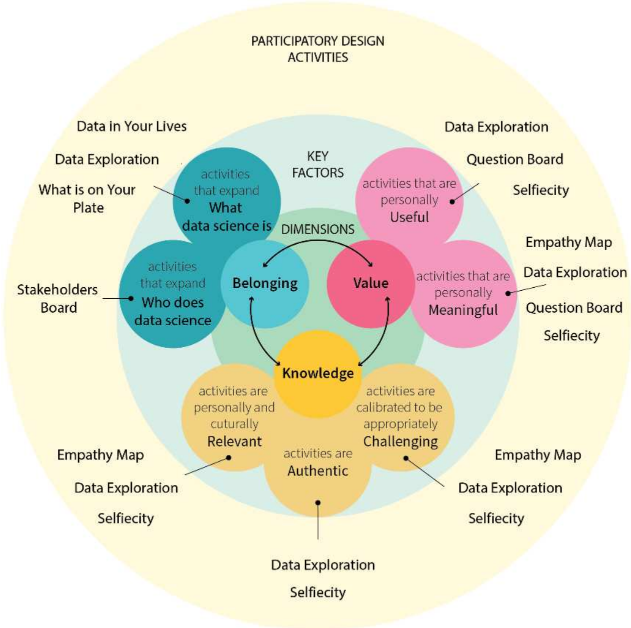
Empathy Map activity

This design activity is based on the User-Centered Design concept of a persona (Miaskiewicz & Kozar, 2011) and has students create personas to reflect on their identities and how data affects their lives. They were given posters divided into four quadrants labeled "Does", "Uses", "Concerns", and "Interests", and were asked to describe a made-up high school student's habits, interests, and concerns relating to data. For example, one group created Tiffany, a 17-year-old athletic student who uses the school's learning management system and her phone for social media apps. She enjoys traveling and sports and is concerned about failing her studies and managing her privacy online. This design activity resulted in nine Empathy Map artifacts describing different personas, including a graphic illustration of the persona at a varying level of detail.

We coded the Empathy Map activity as aligning with two dimensions of the framework: *Knowledge* and *Value*. The gradual construction of the persona and the collaborative work were important scaffoldings, suggesting it aligns with the key factor of *challenge*, especially when students encountered difficulty reflecting on their concerns and preferences. The activity also demonstrated *relevance* as the students reflected on personal data use.

Figure 1

Mapping the PD Activities to Dimensions and Key Factors of the Integrated Interest Development for Computing Education Framework (Michaelis & Weintrop, 2022).



This activity presented an opportunity to provide value to students and emphasized *personal meaning*. The personas revealed insights about students' extensive use of social networks and their interest in various topics, including sports, music, cooking, animals, going out, and video games. Further, the personas reflected students' concerns about privacy and security, data collection methods, and unauthorized personal data usage.

Data exploration activity

In this activity, students were asked to choose one question that interests them and then list the steps to answer it, detailing the practices involved, the prior knowledge used, and the resources needed. The students raised various questions, including '*How to make more money?'* and *'What is the best school to study law?'* In completing the activity, students revealed a breadth of data science practices they engage in. For example, the students suggested steps to obtain information on where to study law, such as asking a law-educated adult, looking at schools' websites, and visiting schools. This activity was coded as aligning with all three core dimensions of the framework: *Knowledge*, *Value*, and *Belonging*. During this activity, students were encouraged to reflect on their ability to answer questions using data. The task was *challenging*, requiring them to use mental processes uncommon in their daily lives. Moreover, it provided a learning context that was personally *relevant*, as the students were free to choose the question they were interested in. As for *authenticity*, the task was connected to a real scenario. This task was *valuable* as it helped the students meet a personal goal (*meaningful*) and understand how to answer questions that intrigue them (*useful*). As for *belonging*, this activity introduced topics and practices involved in data science and provided the students with a broad vision of what the field includes and how it can be applied in the context of their interests.

Discussion and conclusion

PD is a powerful approach for including students' voices and values in curriculum design (Coenraad et al., 2022). In this study, we combined PD activities to identify learners' needs, interests, and preferences. The activities enabled students to express themselves, reflect on their perceptions, and engage in foundational data science practices. These activities shed light on varied ways to spark students' interest in data science. The analysis of these PD activities revealed how each activity aligns with key aspects of interest development theory, including *knowledge*, *value*, and *belonging*. Each activity touched on one or more key factors while engaging students with a broad understanding of what data is, who participates in data science and what motivates them, what contribution and value the field has, or how it personally affects the students.

Some key factors could be further explored in future PD work. For example, the concept of *belonging* was partly addressed in the activities and could benefit from further attention. Specifically, the key factor of *who is doing data science* was expressed from the perspective of stakeholders who collect and analyze data, neglecting those who were historically excluded from computing. A broader understanding of who is involved in data science practices and who is represented or not in the data is also important for creating a sense of *belonging*. Some groups have been disproportionately impacted by biases and predatory applications of data-driven algorithms (Biehler et al., 2022), and data science education should tackle these issues and eradicate their future occurrences.

This work also demonstrates the application of the IIDCEF in a new context. The framework was originally proposed for computing education but can also speak to data science education. Thus, this work marks an initial step in a potentially generative line of work that draws on findings from computing education to inform data science instructional materials. This work has potential implications for educators and curriculum designers. Bringing a multifaceted conceptualization of interest development and design activities that tap into different aspects can spark interest in students, motivating them to explore and invest in new scholarly pursuits. The implications of this work extend beyond data science education and reveal a pathway for creating interest-driven curricula across the educational spectrum, emphasizing personal and cultural relevance in teaching and learning.

In data science education, tapping into students' interests can be a powerful catalyst for meaningful learning experiences. This work shows how PD can provide insight into ways to inform the design of interest-driven data science materials that foster a sense of belonging and inspire meaningful engagement. In doing so, we can seek to design effective, engaging, and equitable learning experiences that draw on the lived experiences of youth to prepare them to succeed in an increasingly data-driven world.

References

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.

Bødker, S., Dindler, C., Iversen, O. S., & Smith, R. C. (2022). What is participatory design? In S. Bødker, C. Dindler, O. S. Iversen, & R. C. Smith (Eds.), *Participatory Design* (pp. 5–13). Springer.

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.

Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review*, *9*(2), 1–22.

Coenraad, M., Palmer, J., Eatinger, D., Weintrop, D., & Franklin, D. (2022). Using participatory design to integrate stakeholder voices in the creation of a culturally relevant computing curriculum. *International Journal of Child-Computer Interaction*, *31*, 100353.

DiSalvo, B. (2016). Participatory design through a learning science lens. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4459–4463.

Fails, J. A., Guha, M. L., & Druin, A. (2013). Methods and techniques for involving children in the design of new technology for children. *Foundations and Trends® in Human–Computer Interaction*, *6*(2), 85–166.

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

Harackiewicz, J. M., Smith, J. L., & Priniski, S. J. (2016). Interest matters: The importance of promoting interest in education. *Policy Insights from the Behavioral and Brain Sciences*, *3*(2), 220–227.

Lee, V. R., Wilkerson, M. H., & Lanouette, K. (2021). A call for a humanistic stance toward K–12 data science education. *Educational Researcher*, *50*(9), 664–672.

Martinez, W., & LaLonde, D. (2020). Data science for everyone starts in kindergarten: Strategies and initiatives from the American Statistical Association. *Harvard Data Science Review*, *2*(3).

Miaskiewicz, T., & Kozar, K. A. (2011). Personas and user-centered design: How can personas benefit product design processes? *Design Studies*, *32*(5), 417–430.

Renninger, Martina Nieswandt, & Suzanne Hidi. (2015). *Interest in mathematics and science learning*. American Educational Research Association.

Saldaña, J. (2016). *The coding manual for qualitative researchers*. SAGE.

Schanzer, E., Pfenning, N., Denny, F., Dooman, S., Politz, J. G., Lerner, B. S., Fisler, K., & Krishnamurthi, S. (2022). Integrated data science for secondary schools: Design and assessment of a curriculum. *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education*, 22–28.

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.

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.