Participatory Design as a Mechanism for Informing an Interest-Driven Data Science Curriculum

ROTEM ISRAEL-FISHELSON and DAVID WEINTROP, University of Maryland, College Park, Maryland, USA

Objectives: Interest plays a central role in learning by shaping what, how, when, where, and why learning occurs. In data science education, where complex concepts, lived experiences, and practical skills intersect, capturing and cultivating student interest can be especially generative. This work explores approaches for designing and evaluating interest-driven data science instructional materials. *Methods*: This article presents a participatory design study that informs the development of a data science curriculum for high school students. To assess how well learner interests and values are reflected in the resulting curriculum, we used the Integrated Interest Development for Computing Education framework [56], which provides a concrete operationalization of interest that captures its multifaceted nature. *Findings*: The article demonstrates and discusses how participatory design can be used to identify students' interests and how those interests can be used to inform the creation of an interest-driven curriculum. Further, it highlights how different types of participatory design activities yield insight into different facets of students' interests and identities, which can then be used to design learning experiences. This work shows how the resulting Participatory Design (PD) reflects and harnesses the multifaceted nature of student interest and how it can be leveraged to design learning experiences that connect with learners' lived digital experiences. *Conclusions*: Participatory design is an effective student-centered approach for tailoring computational learning experiences aligned to students' voices, values, and interests. The use of various participatory design activities revealed different facets of students' interests that informed the creation of an interest-driven curriculum that could not have been created without the input of the students themselves.

CCS Concepts: • **Social and professional topics → K-12 education**;

Additional Key Words and Phrases: Data science education, Interest development theory, Participatory design, Curriculum design

ACM Reference format:
Rotem Israel-Fishelson and David Weintrop. 2025. Participatory Design as a Mechanism for Informing an Interest-Driven Data Science Curriculum. *ACM Trans. Comput. Educ.* 26, 1, Article 11 (December 2025), 34 pages.
https://doi.org/10.1145/3772279

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.
Authors' Contact Information: Rotem Israel-Fishelson (corresponding author), University of Maryland, College Park, Maryland, USA; e-mail: rotemisf@umd.edu; David Weintrop, University of Maryland, College Park, Maryland, USA; e-mail: weintrop@umd.edu.

ACM 1946-6226/2025/12-ART11
https://doi.org/10.1145/3772279

1 Introduction

In our increasingly interconnected and technologically driven world, data are ubiquitous. High volumes of data flow through our lives, shaping our experiences, informing our decisions, and molding our understanding of the world [8]. This pervasive data-driven reality presents opportunities and challenges to educators and the computer science education community. It is more important than ever that students develop computational and data literacy to navigate this information-rich landscape effectively and be prepared to make informed decisions about the role of data in their lives [54, 83, 90].

High school students are at the forefront of living in a data-driven world. Their participation extends beyond being data consumers as they are also active creators through sharing their thoughts and experiences on social media, studying and completing coursework in learning management systems, using sensors in their science classroom, and tracking their daily activities with wearable devices [36]. Simultaneously, students are exposed to a wealth of information, news, and online content, much of which is driven by data analytics, algorithms, and artificial intelligence [91]. This continuous interaction with data often occurs passively, without structured educational experiences that teach students how to critically engage with data. Thus, developing a robust data science curriculum at the high school level is crucial to transforming these passive interactions into meaningful learning opportunities that introduce and situate data science in students' lived experiences, making data science an engaging and relevant subject of study [49, 90].

By harnessing data that aligns with their interests, high school data science instruction can engage students in compelling learning experiences while fostering critical skills and conceptual understanding of the data-rich world around them [32]. Students can learn the fundamental concepts and computational practices involved in data collection, analysis, and visualization while utilizing programming languages or computer-based tools [83]. These skills, which draw on concepts, practices, and technologies from computer science, can foster a sense of responsibility and ethical awareness regarding data privacy and societal impact [2]. They can help students explore cultural trends, track environmental changes, and make informed decisions as civic actors entering adulthood. Further, they can provide a new way to explore their interests, helping them track their favorite sports stars or discover their next favorite band. Incorporating real-world data that relate to and draw from learners' personal lives, values, and interests can make the learning experience more relevant and meaningful [53]. Moreover, they can empower students to discern reliable information from misleading narratives and make informed decisions [12].

Despite the recognized importance of foundational data science skills, research on understanding the best way to introduce high school students to data science is only in its infancy. Further, there is a lack of established methods and frameworks for designing curricula that connect this important topic with students' interests and foster conceptual understanding [49, 87]. This is particularly critical given the direct and continued engagement students have with systems, platforms, and algorithms that use the data they consume and create. Helping students understand the role of data in their lives and empowering them to be informed data consumers and creators, and doing so in a way that resonates with their lived experiences, interests, and identities, is the central goal of this work. The overarching goal of this work is to better understand *how can we design a high school data science curriculum that situates the discipline within students' lived digital experiences.* In other words, how can we create a data science curriculum that draws on students' experiences and interactions with data? To pursue this goal, we employed **Participatory Design (PD)** as a design strategy for learning about students' interests and experiences with data for informing the development of an interest-driven high school data science curriculum. The decision to use PD as a means to ensure the contextualization of the content reflects the voices, values, and interests of the students it is being designed for led us to investigate the following three research questions:

(1) *What student interests emerge through PD that could inform the design of an interest-driven curriculum?*
(2) *How can these insights into student interest be integrated into a high school data science curriculum?*
(3) *In what ways are aspects of interest-driven learning reflected in the resulting curriculum?*

To answer the first research question, we present a qualitative analysis of the PD conducted, seeking to identify various interests expressed by students throughout the PD. An initial analysis of a subset of PD activities was previously published, demonstrating how the activities provide the opportunity for students to express their interests [43]. This work extends that initial analysis to include the full slate of PD activities and then presents the design process of how those insights were integrated into a data science curriculum. After presenting the interests identified in the PD data, we answer the second research question by providing a detailed description of the design process for three lessons in the curriculum, connecting data from the PD with specific data science content. For each of these lessons, we also discuss how the resulting curricular materials align with various facets of interest-driven learning. Finally, to answer the third question, we analyzed the complete curriculum using the **Integrated Interest Development for Computing Education (IIDCE)** framework [56]. The IIDCE framework serves as a way to operationalize interest-driven learning across three distinct dimensions and provides a way of understanding whether and how the chosen design approach resulted in a curriculum that reflects the multifaceted nature of interest-driven learning.

This work makes several contributions. First, it advances the nascent field of K-12 data science education by exploring the potentially generative approach of situating data science in students' lived experiences through foreground interest-driven data science learning opportunities. Given the impact data has on students' lives and the vast array of datasets across innumerable topics available to students and teachers, there is tremendous potential to bring together data science and interest-driven learning. This work presents a methodological approach and documentation of the results of that approach toward the goal of creating an interest-driven data science curriculum. Second, this work contributes to the literature on PD as a means to gain insight into students' voices, values, and interests. In particular, it employs well-established methods but focuses them on a novel goal: creating a data science curriculum. While previous work has used PD to inform the design of curricula (e.g., [15, 30]) and educational technologies (e.g., [57, 70]), its use toward interest-driven learning and data science is novel. A third contribution is in how this work operationalizes an interest-driven framework through its application of the IIDCE framework. While there is a deep and rich literature related to interest-driven learning, the application of a multifaceted conceptualization of interest-driven learning in a computing discipline is novel. Finally, the overarching approach of blending interest-driven learning, PD, and K-12 data science education is innovative. While previous research and design efforts have employed two of these areas, this work breaks new ground at the intersection of all three. Collectively, this research shows how these three areas can be combined to productively create instructional materials that introduce computationally rich disciplinary ideas in ways that leverage learners' interests, identities, and lived experiences. In doing so, the work presents a design process for creating data science education experiences that draw on learners' interests and ideas. Given the importance of all students developing foundational computational and data science skills, this work's contribution of a novel strategy for creating interest-driven data science learning materials adds to the methodological tool belt another means of bringing important computational and data science concepts into K-12 classrooms.

This article presents the PD sessions conducted, the data collected throughout them, and our analytic strategies for turning insights gleaned through PD into an interest-driven data science curriculum. We begin with a review of the three bodies of literature from which it draws: data science education, interest-driven learning, and PD. We then present our methodological approach, including a detailed discussion of the PD workshops carried out and the data science learning trajectory that served as the content backbone of the curriculum being designed. Next, we present our findings and the answers to the three sub-research questions. This begins with the results of our analysis of interests that emerged from the PD, followed by three examples of how those interest insights were integrated into curricular lessons, and finally, an analysis of how, where, and when interest was present in the resulting curriculum using the IIDCE framework as our analytic lens. That article concludes with a discussion of this work and its implications and contributions.

2 Background

2.1 Data Science Education

Data science is a multidisciplinary field related to how we perceive, analyze, and interact with information. Data science involves practices for collecting, storing, extracting, and analyzing data to draw conclusions, predict, classify knowledge structures, gain insights, and make informed decisions [32, 78]. Data science education plays a vital role in preparing students to navigate an increasingly data-driven world. Over the past 50 years, there has been an ongoing academic discourse on the definition of the field [24]. However, today, there is a consensus that it is a discipline situated at the intersection of statistics, computer science, mathematics, and application fields [5, 49].

The foundational role of computing and technology in data science is increasingly recognized [11, 61, 79]. This computational foundation is evident across various data science practices, including data collection/production [36], cleaning [89], analysis [64], and communication [81]. Both the Computer Science Teachers Association standards [80] and the Computational Thinking in Math and Science Taxonomy [88] emphasize connections between data science and computer science/computational thinking, underscoring the significance of integrating computing into data science education. This integration is particularly pronounced in instruction using authentic, real-world data sources [e.g., 89].

Rubin [76] highlights key aspects of data reasoning, advocating for learning experiences that underscore the importance of real-world data engagement while understanding the context in which it was collected. Rubin argues that the extensive work in statistics education conducted over the past three decades provides a solid foundation for advancing data science education. Recognizing the growing importance of data science has led to various educational initiatives worldwide promoting it as a topic of instruction in primary and secondary schools [73, 82]. The International Data Science in Schools Project is one such initiative that is cross-disciplinary and multinational. It provides training programs and frameworks for developing introductory data science courses for high school students [41].

In recent years, several year-long curricula have been developed for high school settings [e.g., 33, 86], including some with roots in computer science curricula [e.g., 78]. These curricula are crucial in shaping students' understanding of data science and determining how they are introduced to the discipline. Such curricula promote critical thinking and teach students to question assumptions, search for reliable evidence, and draw conclusions based on data [87]. Moreover, data science curricula can nurture civic responsibility by informing students of their role as data producers and consumers and the potential dangers associated with irresponsible [78] or biased uses of data [12, 78]. It is critical that students understand that data science technologies and the algorithms that drive artificial intelligence systems are not immune to bias and can introduce or perpetuate inequalities in society [3]. Therefore, data science curricula should refer to equality and social justice issues. By engaging with issues such as algorithmic bias or biased data collection, fairness, and ethical use of data, students can learn how to reduce the adverse effects of technology on marginalized communities and create a more just society [4, 10, 62].

Data science curricula should be tailored to the student populations and diverse institutional contexts [48]. To foster a genuine connection between students' identities and their social and cultural backgrounds, it is crucial to design data science educational experiences that include relevant datasets and activities. This approach can deepen students' affiliation with data science, promote interest, and improve their self-efficacy [53, 90]. 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 [9]. Additionally, it can inspire students who may lose interest in the field due to excessive emphasis on the technical or computational aspects of data analysis [23]. Therefore, it is crucial to examine what motivates and interests the students and design the curricula accordingly.

2.2 Interest-Driven Learning

Fostering students' interest in learning is a key objective of education, as it significantly impacts motivation, curiosity, engagement, and academic success [38, 74]. A growing body of research highlights the importance of interest in knowledge acquisition and promoting lifelong learning. Interest Development Theory, a concept rooted in educational psychology, has emerged as a framework for understanding how individuals cultivate and evolve their interests and passions over time [69]. This theory states that interests develop and flourish when individuals can explore, interact with, and derive meaning from a subject they are passionate about. Moreover, cultivating interest is a complex and multifaceted process influenced by various environmental factors and individual experiences [68]. Cultural context and social feedback are socio-cultural factors that also influence interest development. When the learning experiences are tailored to the learners' background, it allows them to leverage their previous experience and knowledge, express their perspectives, and feel empowered to engage with the learning tasks [1]. In recent years, this theory has gained considerable recognition due to its ability to provide a structured approach to enhancing learning experiences, particularly in science education [35].

Based on this work, Michaelis and Weintrop [56] developed the IIDCE framework. The framework seeks to bring together various operationalizations of interest and situate them in computing education contexts. It is built around three central dimensions of interest development: knowledge, value, and belonging. Here, we will briefly review these dimensions as they inform the design of our PD sessions.

(1) *Knowledge*—The theory suggests that knowledge acquisition in a specific area enhances interest through repeated experiences. A lack of knowledge encourages acquiring new knowledge by highlighting the gap between existing knowledge and what is needed to solve a problem. In addition, a personal interest in a particular field can also motivate a search for new knowledge and prompt learners' exploratory activities. Three key factors help develop interest and build knowledge: *challenge*, *relevance*, and *authenticity*. Challenge involves providing tasks that are appropriately calibrated to the learner's level of knowledge and skills. These tasks should be supported by scaffolding adapted to the tasks' complexity and the learners' abilities while also allowing for the scaffolding to be removed when appropriate. Relevance means adapting learning contexts to the learner's personal experiences and cultural resources, i.e., ensuring the topic is meaningful to the learner. Authenticity involves offering

real-world scenarios and purposes for the knowledge and skills being developed, making the learning meaningful.

(2) *Value*—The perception of value also plays a crucial role in developing interest. The relationship between interest and value is reciprocal—increased value generates interest and heightened interest fosters a sense of value. Value is subjective and varies from learner to learner, shaped by their beliefs, goals, and community values. *Personal meaning* and *usefulness* are two main factors in the perception of value. Personal meaning involves adapting the content to the learner's values and beliefs or connecting it to their identity. In contrast, personal usefulness refers to the practical benefits of engaging with the content to achieve specific goals.

(3) *Belonging*—The third dimension of interest development involves a learner's sense of belonging and connection to their learning content. Developing a sense of belonging is linked to connecting one's background, experiences, and interests, learning new content, and feeling connected with peers, teachers, and the content community. Fostering a sense of belonging is crucial to supporting the development of interest in computing. Belonging introduces a social aspect to interest development, considering how learners perceive their identity, abilities, and how others view them concerning their engagement in an activity. As articulated in the IIDCE framework, overcoming stereotypes in computing is essential to promote belonging, and two key factors in achieving this are expanding notions of *what computing is* and broadening the understanding of *who can participate in computing*.

In this work, we applied the IIDCE framework in evaluating an interest-driven data science curriculum. Details of this application are provided in Section 3.3.

2.3 PD

Participatory design (PD) is a research methodology that involves end-users in the creation of solutions, artifacts, and activities to ensure that their voices, values, and priorities are reflected in the process and the resulting final design [20, 27]. This approach acknowledges the power differential between designers and end-users by providing opportunities for end-users to share their unique insights about their needs, preferences, and concerns [7]. Moreover, as opposed to other user-centric design methodologies, it emphasizes agency and active collaboration, granting participants significant influence throughout the design process by enabling them to take on various roles. Participants can serve as users, testers, informants, design partners, co-researchers, or protagonists [25, 27, 47]. Each role reflects different levels of involvement and responsibility. For instance, users provide feedback based on their experiences, while design partners actively contribute to shaping the design process [27]. Co-researchers exchange and analyze contextual data from their surroundings, producing insights and situated knowledge that inform the creation of new and more relevant artifacts [25]. The role of the protagonist goes further, positioning participants as central agents whose perspectives and ideas drive the design outcomes, emphasizing their ownership and creative input throughout the process [47].

The PD approach involves methods and techniques for co-creating goals with end-users while simultaneously identifying constraints and opportunities within a particular context [22]. The methods used can vary, but they often involve a series of structured activities and interactions that facilitate collaboration and communication [29, 92]. Cooperative inquiry, for example, is a method that involves children as equal design partners throughout the entire design process [26]. Children contribute ideas, feedback, and evaluations through collaborative activities like brainstorming, iterative low-tech and high-tech prototyping, and reflective discussions, ensuring shared influence on the final product. Bonded Design focuses on long-term partnerships between designers and children. Through sustained collaboration, trust, and deeper engagement, children participate as co-creators, contributing to iterative development through recurring design sessions that allow for continuous refinement of ideas and mutual learning [50]. These approaches ensure user-centered solutions by incorporating multiple perspectives and enabling iterative development, making them valuable in various fields. However, these methods have their limitation. They are often considered resource-intensive, requiring considerable time and sustained engagement from both researchers and participants [34]. Power imbalances between researchers and participants can persist, even with intentional efforts to create an equitable environment [37]. Assertive participants may dominate discussions, while quieter individuals may be marginalized, leading to uneven contributions. Additionally, some participants may lack the subject-domain expertise needed to fully engage in the design process [92]. These limitations highlight the need for scaffolding and preparatory activities that build foundational knowledge to support meaningful participation.

Numerous studies have documented how PD can draw on the knowledge and values of teachers and students to create technologically rich learning materials [21, 58, 67, 72]. In addition, other studies have shown that PD is effective in iterative curriculum development, emphasizing the importance of continuous collaboration with end-users, including teachers and students, and incorporating feedback to refine and improve learning materials [15, 19]. Famaye et al. [30] provided a recent example of engaging students in PD sessions, holding different roles in the design process, and embedding their lived experiences in a machine learning program design. Polman et al. [66] have shown that involving teachers and students in the curriculum development process through PD sessions can yield valuable insights and improve teaching strategies. Instead of relying on top-down decision-making by curriculum developers, by actively involving students and soliciting their feedback, teachers can tailor curricular designs to better resonate with their diverse learning preferences and backgrounds. Student involvement in PD activities can give them the self-efficacy and vocabulary to help identify necessary and valuable topics and contribute to constructing interest-driven learning experiences [21]. The use of PD as an educational design methodology has yielded profound insights into how educational settings can be structured to foster meaningful learning experiences that better cater to learners' needs and preferences [39]. In this article, we elucidate the pivotal role PD plays in constructing an interest-driven data science curriculum that prioritizes learner agency and engagement.

3 Methods

This research aims to introduce high-school students to the computational foundations of data science in a way that resonates with and draws from their lived experiences. This work is part of a nascent research-practice partnership [13] with a public charter high school focused on technology and computer science education. Toward our goal of developing a novel introductory interest-driven data science curriculum that situates data science in the lived experiences of today's students [46], we employed a three-phase study design (Figure 1). First, we conducted a series of PD workshops to understand students' perceptions of data in their lives and identify their areas of interest. In the second phase, we used the data from the PD to develop a high school data science curriculum. Finally, in the third phase, we evaluated the curriculum to determine how and where it aligns with the IIDCE framework. This evaluation helped in revising the curriculum. Below is a breakdown of these phases.

3.1 Phase 1: PD Sessions

3.1.1 Research Settings. To inform the design of the curriculum and ensure the curricular activities and structures reflect the values, voices, and, most critically, interests of the students in our partnering school, we conducted three PD sessions grounded in the cooperative inquiry method [26]. In these sessions, we positioned participants as co-researchers, enabling them to

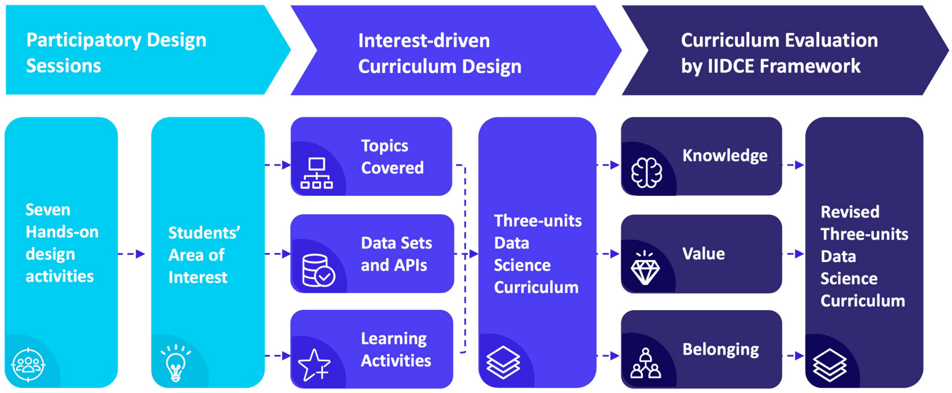

Fig. 1. The three-phase study design.

share, gather, and analyze contextual data from their own practices as we collaboratively sought to understand potential intersections of data science and interest [25]. The central goal of these PD sessions was to gain insights into students' interests, experiences with, and perceptions of data and data science. Two PD sessions lasted 90 minutes and were held after school, while a third extended PD session was carried out over one whole school day. The 90-minute sessions allowed for focused, intensive engagement with specific design activities while minimizing disruption to students' regular after-school schedules. The full-day session provided an opportunity for more in-depth exploration, enabling students to participate in a comprehensive design cycle that included generating, refining, and expanding their ideas. Each session involved a different set of students, reflecting a broad range of perspectives.

The sessions were conducted in classrooms at the partnering school, where students worked through a series of discussions and seven hands-on, distinct design activities [e.g., 16] to elicit their ideas and interests related to data science and inform the design of the curriculum. Two researchers facilitated the sessions, introducing each activity and circulating between groups to support and guide the design activities. This interactive structure underscored students' roles as co-researchers actively shaping the research process through collaborative exploration. Research protocols and data collection methods were approved by the IRB board. The PD consisted of the following activities (Table 1): Data in Your Lives, What is on your Plate? Empathy Map, Stakeholders Board, Data Exploration, Questions Board, and Selfiecity. Each activity was designed to stimulate reflective thinking and creative problem-solving, encouraging students to share experiences, generate insights, and propose ideas relevant to their lived experiences with data. Below is a detailed description of the activities.

Ice Breakers: Data in Your Lives Activities. In this opening design activity, students were asked: "*What do you think when you hear the word data?*". Students then scanned a QR code and typed their responses into an online data collection tool that provides real-time results that can be shared with participants and used to lead a class-wide discussion. Next, the facilitator expanded the discussion while presenting the infographic "Data Never Sleeps" (Figure 2), which shows the data being created each minute online (e.g., 231.4M e-mails sent a minute, 500 hours of video uploaded to YouTube, 1.7M pieces of content users share on Facebook). This activity aimed to help students understand that activities they engage in their daily lives (e.g., social media interactions) produce data and help them make connections between their personal lives and data science. Further, it highlights the scope and scale of data generation on the Internet.

Table 1. PD Activities

	Total Duration	Session One (N = 12) and Session Two (N = 9)	Session Three (N = 7)
Data in Your Lives	20 minutes	X	X
What's on Your Plate?	25 minutes	-	X
Empathy Map	20 minutes	X	-
Stakeholders Board	20 minutes	-	X
Data Exploration	25 minutes	X	-
Questions Board	15 minutes	X	-
Selfiecity Data Exploration	40 minutes	-	X
Summative Questionnaire	10 minutes	X	X

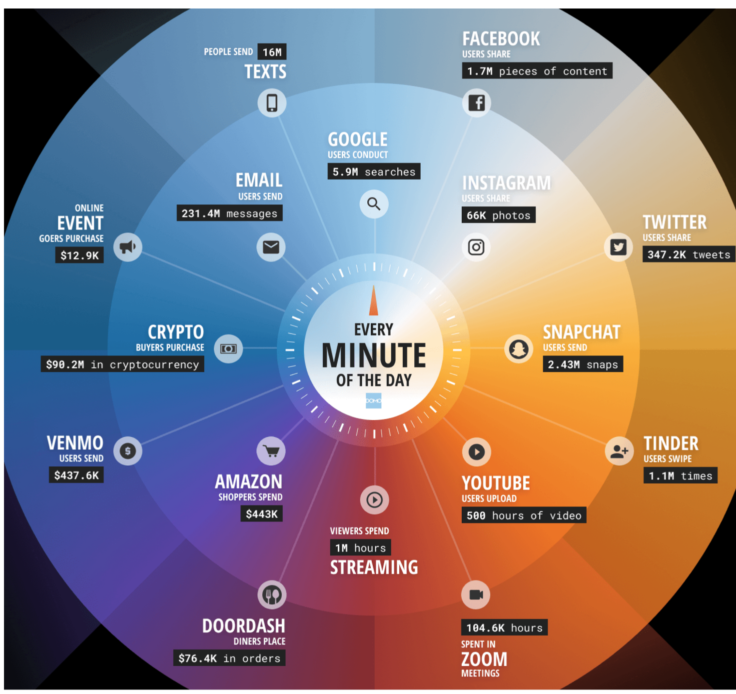

Fig. 2. The "Data Never Sleeps" infographic was introduced to the students.

What Is on Your Plate Activity. In this activity, inspired by a design thinking technique [63], students were given a paper plate and markers and asked to reflect on "*What data do you create and consume during your day?*". They were then instructed to illustrate it on the plate. Students were free to express themselves creatively on the plate (three examples are shown in Figure 3). Students first worked on their plates alone before sharing them with their peers. Similar to the previous one, this activity aimed to help students draw connections between their personal lives and data science.

Empathy Map Activity. This activity was based on the user-centered design concept of a persona [55]. It was intended to help students reflect on their identities and present how data affects their lives. Working in groups of three, students were given large posters that included a circle at the center with four quadrants around it labeled as follows: "Does," "Uses," "Concerns," and "Interests" (Figure 4, left). The groups were then asked to create a persona of an imaginary student who might

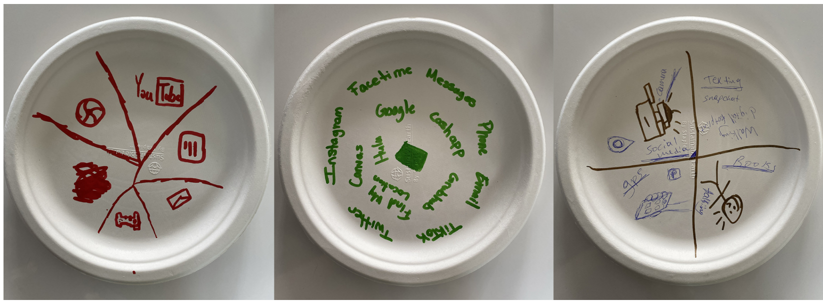

Fig. 3. Examples of the plates created by the students.

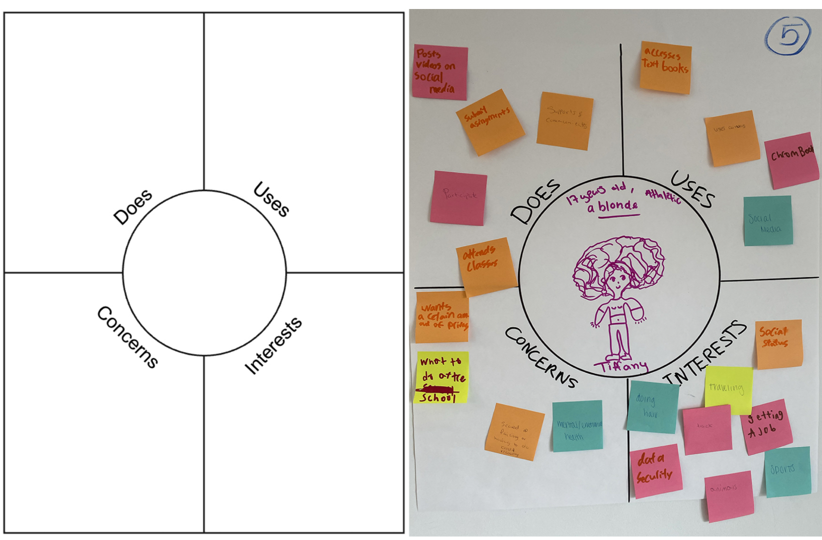

Fig. 4. Example of the empathy map template (left) and one created by the participants (right).

attend their school and to describe this person's habits, interests, and concerns on sticky notes. The center circle was to be filled in with a drawing of their imaginary student, and the responses were to be placed on sticky notes in the corresponding quadrant. The groups were provided with guiding questions to help them reify their student: *What do they do that creates data? What do they do that uses/consumes data? What concerns do they have about data? What are they interested in?* In completing the activity, students reflected on various activities, practices, and platforms where they regularly create and consume data. This activity aims to allow students to use the imaginary student persona as a proxy to express their habits, interests, and concerns without having to claim them personally, allowing for honest expression behind an imagined, anonymous student.

Stakeholders Board Activity. In this activity, students received a poster divided into three parts (Figure 5, left). In the center of the poster, they were asked to write all the stakeholders interested in the data they create in their daily lives. On the upper part of the page, they were asked to add sticky notes describing why companies collect data; on the lower part, they were requested to add sticky notes to name the different uses of the collected data. The goal was to gain insights into how students think about and understand the larger network of stakeholders involved in data collection

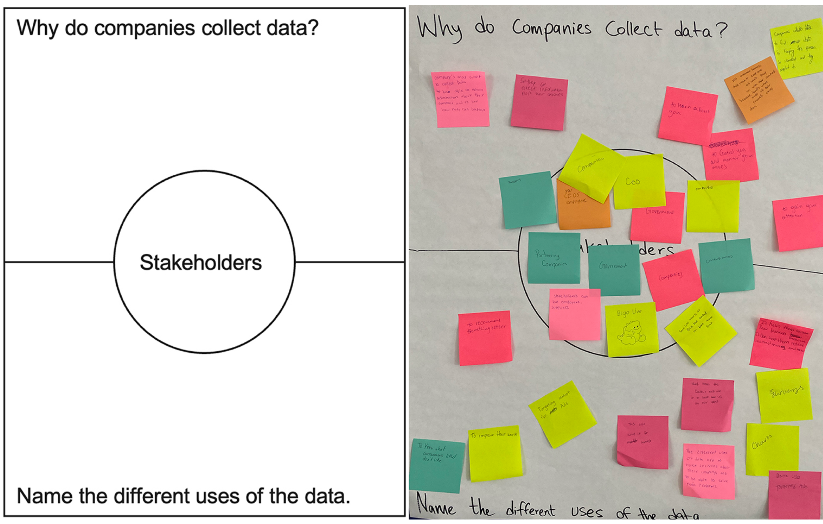

Fig. 5. Example of the stakeholders board template (left) and the board created by the participants (right).

and usage. We also hoped to learn who students think of when asked to articulate a list of people and groups participating in data science.

Data Exploration Activity. This data-driven inquiry activity was based on "Scenario Maps," a technique from design thinking aimed at creating a visual narrative of the user's experience [40]. Specifically, we used this exercise to uncover students' current data science practices. The students were asked to choose one question that interests them and then list the steps to answer it, detailing the data science practices involved, the prior knowledge, and the resources needed (Figure 6). The students came up with a variety of questions, including "*How do you make more money?*"; "*What is the best school to study law?*"; "*How do you raise your grades?*" and "*What are the ways to overcome social anxiety?*". To assist them in articulating the steps, they respond to a series of prompts, which serve as scaffolds designed to help them break the task down, attending to the data-related practices, knowledge, and resources associated with each step. In completing the scaffold, students revealed a breadth of data science practices they engage in, focusing on data collection, analysis, and interpretation.

Question Board Activity. In this activity, students were asked to generate questions related to specific topics using sticky notes and post them on a large poster board on the classroom wall. The list of topics was based on the interests expressed by the students during the Empathy Map activity. Afterward, each participant received a pack of blue and green stickers. The students used blue stickers to denote the topics or questions they found most exciting and green stickers for the most valuable questions (Figure 7).

"Selfiecity" Data Exploration. The final activity introduced students to the "Selfiecity" project (selfiecity.net). Selfiecity is a research project that examines the trends of selfies (self-portraits typically taken with smartphones) shared on social media platforms in five cities worldwide [84]. The project presents interactive infographics and findings about the demographics of people taking selfies, their poses, and expressions, each associated with a specific geographic region [42]. The Web site also enables users to interact with the data using a graphical user interface to apply various

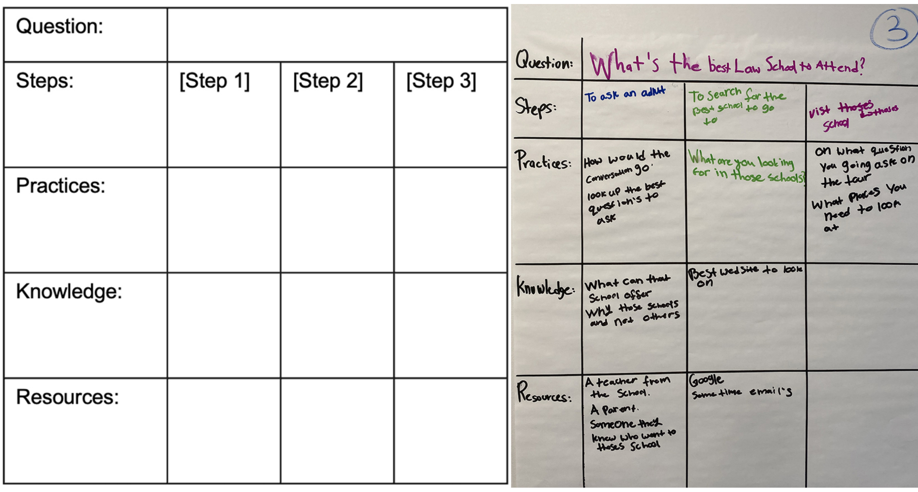

Fig. 6. The template provided for the data exploration activity (left) and the students' artifact for that task (right).

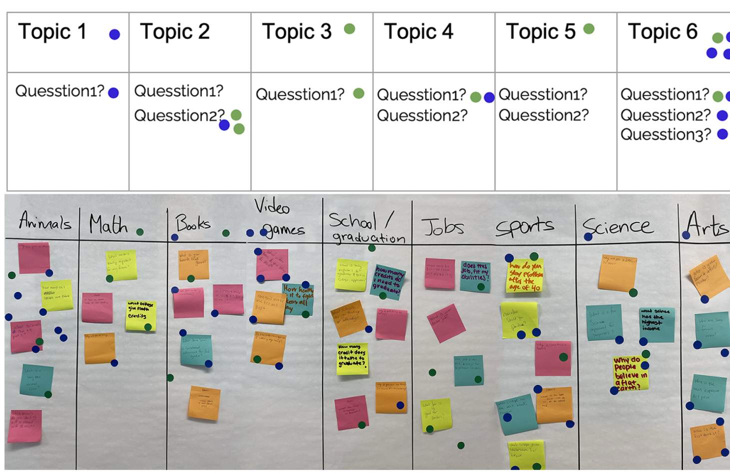

Fig. 7. Example of the questions board activity template (top) and the students' artifact for that task (bottom).

filters, such as city, demographics, pose, facial features, and the person's estimated mood, to explore the data.

During this design activity, students were asked: How could you use data like this to examine people's lives? What data do they create like this in their own life? and What (or who) do they think is missing from this project? This activity aimed to encourage the students to be critical and reflect on the potential biases inherent in the data collection and presentation methods, such as which cities were chosen and how the data was categorized. This activity culminated in having students design a potential research study with data from the Web site.

3.1.2 Participants. Twenty-eight students participated in the PD sessions. These students were recruited through in-person visits and a flyer posted in the partnering charter high school. The researchers coordinated with the school to find a time to visit classrooms and recruit students

Table 2. Participant Demographics

	Total (N)	Session One (After-School)	Session Two (After-School)	Session Three (During-School)
Gender				
Male	17	6	8	3
Female	11	6	1	4
Race				
Black or African American	22	11	6	5
Native American	2	0	2	1
Hispanic	1	0	0	1
White	1	1	0	0
Age				
14	1	0	1	0
15	5	3	2	0
16	7	3	3	1
17	13	6	3	4
18	2	0	0	2

that would not interfere with the course of studies. During this visit, a general explanation of the study was given to students, and consent forms were handed over, which included information about the sessions and an e-mail address for contact. Parent/guardian consent and child assent were received from all participants before the PD sessions. The gender, race, and age of participants are shown in Table 2. The students provided this demographic information by filling out a survey at the beginning of the sessions. As illustrated, most students (17 out of 28) identified as males, and most (22 out of 28) were Black/African American. Participants spanned all four years of high school. All participants had taken a computer science course before our sessions, and only one had taken a data science course.

3.1.3 Data Collection and Analysis. During the PD sessions, the researchers collected the artifacts created by the students through the design activities (e.g., sticky note-covered posters and decorated paper plates). These artifacts served as the primary data source and were supplemented by audio and video recordings, which were later transcribed. The researchers also observed students and took field notes on the physical conditions, participants' interactions and engagement, direct quotes, and insights, all while maintaining a factual and objective approach. Finally, all students completed a short demographics survey, which included questions about their race, gender, age, and prior experiences with computer science and data science. The data collected included 20 hours of video, 18 hours of audio, six paper plates, and 13 posters.

All the data collected during the sessions, including the discussions and design artifacts, were analyzed using inductive thematic analysis to identify emergent themes [77]. First, the researchers reviewed each text segment, identifying and labeling prominent ideas or themes without using preconceived categories. These initial descriptive themes reflected the variety of ways students described data. Upon further review, the researchers synthesized the themes into six higher-level categories, serving as high-level orienting codes to structure the analysis. Table 3 presents the coding scheme, including the definition and an example for each code from the data. Two researchers systematically and independently coded the data using this coding scheme [71]. Each text segment was coded according to one of the six codes, with researchers also adding notes or memos about their thoughts, questions, and insights. The resulting codes were compared, and an inter-rater reliability assessment using Cohen's kappa [17] yielded a coefficient of 0.8. The researchers discussed and resolved the remaining discrepancies between their analyses until complete agreement was achieved.

Table 3. Coding Manual for the PD Sessions Analysis

Code	Definition	Examples from Participants' Data
What	Student-provided definitions of data	"A set of gathered information"; "Various amounts of information being stored into a specific location (a computer or other virtual forms)"
Who	Student-provided entities and organizations interested in data	"Government"; "Foreign Agencies"; "Schools"; "Banks"; "Hospitals"; "Social Media Platforms"; "Hackers and Phishers"
Why	Student-provided reasons for entities and organizations to collect data	"To recommend something better"; "To improve their work"; "To make decisions"; "To control you and monitor your moves"
Where	Student-articulated sources of data (i.e., defining where data comes from)	"Social Media (e.g., Instagram, Twitter)"; "Netflix"; "Hulu"; "Canvas"; "Google"; "Online shopping"; "iMessage"
Concerns	Student-articulated issues regarding the collection, storage, sharing, and use of data	"Being tracked by the apps used"; "Being recorded without his consent"; "Someone wants to use your data against you"
Area of Interest	Student-articulated topics they are interested in	"Sports"; "Social Media"; "Video Games"; "Music"; "Movies and TV Shows"; "Animals"

3.2 Phase 2: Interest-Driven Curriculum Design

Drawing on insights from the PD sessions, we began developing API Can Code, an interest-driven introductory data science curriculum that introduces students to the computational foundations of data science through authentic, meaningful data exploration. We first determined the general structure and learning trajectory of the curriculum based on existing data science education literature [41, 51]. The result was a curriculum comprised of three units and a final project. Table 4 provides an overview of the units and the topics covered in each lesson. The first unit, "Data in Learners' Lives," is designed to help students understand what data is, where it comes from, and its impact on their lives. The lessons in this unit address data collection entities and their motivations, types of data, various data sources and their effects, and methods for evaluating datasets. The second unit, "Computational Foundation of Data Science," aims to help students gain the fundamental computational skills for programmatically retrieving, processing, and analyzing publicly available datasets that align with students' interests. Lessons in this unit focus on querying data from various APIs and then organizing and manipulating the results returned. This process involves writing programs in EduBlocks, a block-based programming tool that presents text-based programming languages like Python in a user-friendly way. The third unit, "Data Science Practices," introduces students to data science practices, focusing on data analysis and visualization to extract meaningful insights. The lessons focus on different types of visualizations and their applications, exploratory analysis, detecting misinformation and outliers, and utilizing statistical methods on data. These processes involve using Common Online Data Analysis Platform (CODAP), a free, user-friendly data visualization and analysis tool that allows users to analyze data, create various summary plots, and conduct basic statistical tests [14]. The final project asks students to pose a question, then identify and validate a data source that can shed light on the question, gather and organize the identified data, and then analyze it to answer the state research question. The last day of the curriculum has students present their questions, processes, and answers to their peers. After establishing the conceptual structure, we proceeded with figuring out how to situate the data science concepts and activities, including what datasets to use and ways to situate the content within students' lived experiences, and how to align it with their interests based on what was learned during the PD sessions. The details of this process are presented in Section 4.2.

3.3 Phase 3: Evaluating the Curriculum Using the IIDCE Framework

After the initial version of the curriculum was complete and had been taught once in a classroom setting, we began to reflect on our design process and conceptualize this manuscript. We decided to conduct an evaluative analysis to determine if and how the curriculum was interest-driven as intended. We sought frameworks to help operationalize the concept of interest-driven in the

Table 4. Topics Covered in Each Lesson

Unit	Lesson	Topics Covered [51]
Unit 1: Data in Learners' Lives	1.1 Introduction to Data	Nature of data > Numerical/Quantitative Data, Categorical Data; Data Sources > Public Access Datasets; Time Series Data > Date-and-time Variable Formats
	1.2 Data Collection and its Purpose	Ethics > Privacy, Bias
	1.3 Using Data	Time Series Data > Forecasting
	1.4 Sources of Data	Data Sources > Public Access Datasets; Ethics > Bias
	1.5 Evaluating Datasets	Nature of data > Tabular structure Ethics > Bias
	1.6 Data Collection: Impact and Equity	Data Sources > Surveys; Inquiry with Data > Type of Inquiry; Sampling and Simulating > Random Sampling; Generalizability
Unit 2: Computational Foundations of Data Science	2.1 What is Data Science?	Inquiry with Data > Data Cycle
	2.2 Manual Data Processing	Inquiry with Data > Data Cycle; Data Manipulation > Filtering; Mutating, Reshaping, Transforming
	2.3 Intro to Programming with EduBlocks: Filtering & Data Transformation	Nature of data > Tabular Structure; Data Manipulation > Filtering;
	2.4 Accessing Data with APIs	Data Sources > Public Access Datasets; Programming > Logic; Data Manipulation > Cleaning; Filtering; Mutating, Reshaping, Transforming
	2.5 Preparing Data for Analysis	Data Sources > Public Access Datasets; Programming > Logic; Data Manipulation > Filtering; Mutating, Reshaping, Transforming
	2.6 Data Analysis in Practice	Data Sources > Public Access Datasets; Programming > Logic; Data Manipulation > Cleaning; Filtering; Mutating, Reshaping, Transforming
Unit 3: Data Science Practices	3.1 Intro to Data Visualization	Data Visualization > Comparing Plots; Ethics > Misleading Information
	3.2 Exploratory Analysis with CODAP	Map Data > Dimensions of Map Data; Location and Region Data; Map-Plotting
	3.3 Graphs and Figures: One Variable	Data Visualization > Dotplot; Histogram; Bar Chart; Measures of Center > Mean; Median; Distributions and Variability > Distributional Shape
	3.4 Graphs and Figures: Two Variables	Nature of Data > Frequency Tables; Variable Association > Correlation; Third Variable; Data Visualization > Scatterplot
	3.5 Statistical Testing	Distributions and Variability > SD; Interquartile Range (IQR); Significance > Tests; Confidence Intervals
	3.6 Linear Models	Variable Association > Linear Models; Time Series Data > Forecasting

context of a computing curriculum, identifying the IIDCE framework [56] as a useful tool for the proposed analysis. The IIDCE framework was only consulted after the curriculum had been designed and piloted. Two researchers independently coded the activities in the lessons according to the three dimensions and seven key factors of the IIDCE framework. The goal of this process was to assess the extent to which the lessons aligned with the various dimensions of interest development theory. The researchers reviewed each lesson and coded them according to the coding scheme presented in Table 5. The resulting codes were compared, and an inter-rater reliability assessment using Cohen's kappa yielded a coefficient of 0.79, indicating substantial agreement. The researchers resolved discrepancies in the coding through discussions until complete agreement was achieved. It is important to note that the researchers adjusted the Belonging dimension of the IIDCE framework, shifting it from "computing" to "data science" to align with the goals of this work.

Table 5. Coding Manual for Curriculum Analysis Using the IIDCE Framework

Dimension	Key Factor	Definition from Michaelis and Weintrop [56]	Codes Used for Data Analysis	Example from API Can Code
Knowledge	Challenge	The level of difficulty or complexity presented by computing tasks or problems.	Providing tasks that are appropriately calibrated to the learner's level of knowledge and skills. These tasks should be supported by scaffolding adapted to the tasks' complexity and the learners' abilities, while also allowing for the scaffolding to be removed when appropriate.	Lesson 2.5. Use->Modify->Create instructional strategy and providing scaffolded programs with subgoal labels.
	Relevance	The extent to which computing activities connect to learners' experiences, or real-world applications.	Adapting learning contexts to the learner's personal experiences, communities, and cultural resources.	Lessons 2.1 and 2.2. Investigating local data from OpenDataDC.
	Authenticity	The degree to which computing tasks mirror real-world situations or professional practices.	Providing real-world scenarios to work through while presenting authentic tools and resources to use.	Unit 2. Problem-based learning that incorporate professional tools, materials, and datasets via RapidAPI.
Value	Personal meaning	The sense of personal significance or importance derived from engaging in activities or concepts.	Adapting the content to the learner's values and beliefs or connecting it to their identity.	Lesson 2.6. Adapting the BillboardAPI code to investigate their favorite artist.
	Usefulness	The perceived utility or practical value of engaging in the activities.	The practical benefits of engaging the content/activity for achieving specific goals.	Lessons 2.2. Enabling students to investigate data about their neighborhood via parking ticket records.
Belonging	What data science is?	Understanding of fundamental principles, processes, and practices of data science* and expanding the notions of what data science is.*	Presenting expansive framings of data science and what it is (i.e., understanding the fundamental principles, processes, and practices and how and when to apply them).	Presenting the Data Science Cycle, 5V's, and so on.
	Who does data science?	Identification with individuals or groups engaged in data science*-related activities with an emphasis on helping learners see themselves as part of the field.	Expanding the notion of who participates in data science and combating stereotypes, particularly regarding women and racial minorities.	Discussions of the stakeholders who collect our data and providing examples of data scientists from various backgrounds.

*The Belonging dimension of the IIDCE framework was shifted from "computing" to "data science" to align with the goals of this work.

3.4 Positionality

This research project was primarily carried out by the two authors of this manuscript, with the first author leading all of the research and analysis activities. The first author is an international, White, cis-gendered female researcher who has lived and worked in the region of the United States where the research was conducted for 3 years. The second author, who was actively involved in all aspects of this work, is a cis-gendered, White male with a background in computer science education and learning science. He lives within the school district where the research was conducted. Both authors have worked on numerous projects that involved close collaboration with this and other school districts and previously conducted similar work on the use of PD as a means of informing culturally relevant instructional materials [15]. Along with the two lead authors, two cis-gendered, White male graduate students assisted in this work, one of whom assisted in designing, running, and collecting data on the PD sessions (Phase 1), and one of whom also contributed to the design of the curriculum (Phase 2) and served as the second coder of the curriculum using the IIDCE (Phase 3).

We recognize that our identities shaped the work in several ways. Given that most of the participants did not share our demographic backgrounds, we were particularly mindful of centering students' perspectives and creating structured opportunities for them to articulate their lived experiences with data. This awareness motivated our reliance on the PD as a way of foregrounding students' voices. In addition, our disciplinary background also influenced the study. In particular, the second author's expertise in computer science education and learning sciences guided decisions about curriculum structure, while the first author's experiences shaped the facilitation and activities of the PD.

4 Findings

Our findings illustrate how PD activities yielded insights into student interest that could then inform the design of an interest-driven data science curriculum through a systematic, student-centered approach. To address the first research question, we show how the varied PD activities afforded varying pathways to learn about the students' interests. We first present an overview of the curriculum, detailing how and where PD-identified interests are reflected throughout the curriculum (Section 4.1) and then present three lessons in greater detail, one from each of the curriculum's three units (Sections 4.2.1–4.2.3). To answer the second research question, we explain how insights from the PD informed the design process by translating identified interests into concrete lesson components, including datasets and learning activities. Using the three lessons in Section 4.2, we show how insights from the PD informed each lesson design, resulting in learning activities that resonate with student interest. To address the third question, we present the results of our IIDCE framework analysis, showing how these lessons correspond to various dimensions of interest-driven learning. Furthermore, Section 4.3 offers an overview of the curriculum and its representation of interest-driven learning dimensions.

4.1 Interests Identified across the PD

The analysis of the data students produced during the PD activities shows how they collectively provide essential insights into students' interests and help shape a curriculum that reflects these interests. Analysis of the artifacts created in the What's on Your Plate and Empathy Map activities revealed that students use social networks extensively and are interested in the various platforms. We learned this from the "uses" and "does" descriptions of the personas created and from the drawings and notes on the students' plate artifacts. A synthesis of the findings from these two activities, along with the Data in Your Lives and Questions Board, revealed that students are most interested in sports, video games, music, movies and TV shows, animals, making money, and academic success. This was evident from the interests and hobbies they noted, as well as the questions they generated. In addition, the Data Exploration and Selfiecity activities suggest that they are interested in investigating issues directly related to their community or living environment. These topics can be explored further in a data science class. This conclusion is a key takeaway from this research and has been used to inform the entire curriculum. It is evident in the chosen datasets and learning activities that tackle issues relevant to the students' geographic locations and communities.

The three lessons presented in the following section provide detailed examples of how ideas gathered in the PD can inform the creation of a curriculum that aligns with the various facets of interest-driven learning. Looking across the curriculum, Table 6 summarizes where the different topics of interest identified in the PD sessions have been incorporated into the curricular materials. As seen in the table, some topics were integrated multiple times throughout the curriculum. This is related to the availability of up-to-date datasets and the conceptual connections needed between the areas of interest and the educational topics covered in the lessons. It should also be noted that throughout the second unit, one or two areas of interest were included in each lesson, whereas in the first and third units, multiple areas of interest were incorporated into each lesson. This is due to the unique structure of the second unit, which emphasizes practical application more

Table 6. Students' Areas of Interest as Reflected throughout the Curriculum

	Unit 1: Data in Learners' Lives						Unit 2: Computational Foundations of Data Science						Unit 3: Data Science Practices					
	1.1	1.2	1.3	1.4	1.5	1.6	2.1	2.2	2.3	2.4	2.5	2.6	3.1	3.2	3.3	3.4	3.5	3.6
Social Media		×		×			×			×							×	
Music					×						×							
Sports				×								×	×				×	
Video Games									×		×						×	
Movies and TV Shows		×				×			×	×			×					
Animals					×									×	×			
School/Education				×		×		×					×		×	×		
Going Out														×		×		
Environment and Health	×		×										×					×
Food and Cooking				×	×		×							×	×			
Money and Jobs	×														×		×	
Community			×	×	×	×	×	×									×	

than the others. Here, students learn to access and modify datasets through code they create in a block-based programming environment. Consequently, each lesson typically focuses on a single topic. In contrast, the first and third units included more examples tied to students' interests to explore various theoretical concepts. Another important issue to note is that there are areas of interest, such as the environment, health, and politics, that the students did not mention during the PD. This may suggest they are less interested in these topics; however, the researchers concluded that their inclusion should still be considered.

4.2 How PD Informed an Interest-Driven Data Science Curriculum

Having identified areas of interest through PD and established a conceptual sequence for a data science curriculum, we now demonstrate how these components merged to form an interest-driven data science curriculum. We present three case studies. Each case study details a lesson from a different curricular unit and discusses how the PD activities informed their design. We also analyze these lessons and determine whether they correspond with the dimensions and key factors of the IIDCE framework. After presenting the three case studies, in the next section, we take a step back and look at interest-driven characteristics across the entire curriculum.

4.2.1 Lesson 1.2—Data Collection and Its Purpose and Impact. Goals and Description. In the second lesson of Unit 1, students learn about different stakeholders who collect data and their motivations. They also identify the data collectors in their own lives and discuss issues of privacy and bias. This lesson comprises five main activities. The first one is a warmup in which students brainstorm five data resources in their lives. The second activity introduces the students to the "Data Never Sleeps" infographic. They independently explore the infographic and answer various questions about it. For example, what kind of data are being consumed and collected? What are the differences between versions 1.0 and 10.0 of the infographics? Who might collect or care about these data streams? This activity ends with a class discussion around these questions. In the third activity, students revisit their list of data sources from the warmup and are asked to come up with more data collectors. The students also share their answers with the class and then discuss three cases of people worldwide (a Hollywood actor, a Kansas farmer, and a salesman from Thailand) while considering the data they may collect, create, and consume. In the fourth activity, students are presented with two fictitious cases; the first deals with the sharing of users' personal information by a social media platform to third-party companies, and the second deals with the use of a ring camera that records people who come up and ring a doorbell and then saves information on a cloud-based service. The students are encouraged to think about what kind of violation of privacy exists in these cases and also to provide examples of such violations from their own experiences and ways companies or other entities could better protect user privacy. In the last activity, the students are introduced to the concept of bias in the data. They discuss cases of representation and under-representation of various people and groups in political polls and facial recognition software. At the end of the lesson, the students are given an exit ticket asking them to consider a situation where a movie production company wants to survey which movie idea will generate more revenue. Students are asked to reflect on this situation and share their thoughts on privacy and bias.

How the PD Informed This Lesson. The data gathered from four PD activities informed the design of this lesson: Data in Your Lives, What is on Your Plate, Stakeholders Board, and Empathy Map. An analysis of students' responses to the "Data Never Sleeps" infographic in the PD sessions revealed that students recognized the various data sources but were overwhelmed by the vast amount of data produced and consumed daily. For instance, one student was amazed by the number of transactions on Venmo, which signifies increased online spending: "*People are spending more money online. There are 437 [transactions] in Venmo! More and more people use that.*" In addition, some students identified trends that have prevailed over the years, specifically between the first version of the infographic (2013) and the most recent (2023). For example, one of the students noted, "*It seems that as far as social media, back in the day, people used Facebook, but now everybody* is *on Instagram, Snapchat, YouTube, and Twitter.*" Another student added, "*People are using more social media; they really like to post things.*" These quotes, and others like them, highlight an essential aspect of data science related to the amount of data collected. Students need to understand not only what data is but also their role in producing and consuming data, especially given the massive increase in data creation. Given these findings, we decided to incorporate the "Data Never Sleeps" infographic into Lesson 1.2 to raise awareness of the scope of data consumption and a lever for discussion about data collectors.

The data drawn from the "What is on Your Plate" activity also helped to inform this lesson. Analysis of students' artifacts depicted their consumption and production of data in social media, including TikTok, X (formerly known as Twitter), and Instagram. Social media sites were included on all the plates and were mentioned 15 times. All students also referenced communication apps. Some students referred to specific applications such as FaceTime, e-mail, and iMessage, while others referred to the actions performed, i.e., texting and talking. Additionally, utility apps, like Camera, Calendar, and Chrome, were featured in seven instances across four plates, while navigation apps, e.g., Find My Location or Google Maps, and entertainment apps, e.g., Hulu, appeared three times. Video game apps were mentioned twice. Only one plate included a reference to e-commerce. However, during the discussion, two other students mentioned that they consume and produce a lot of data on Amazon's Web site, and, in retrospect, they would add it to their plate. When the students shared the plate they created with the other participants, they provided explanations for their different uses. For instance, one of the students noted, "*I have a lot on this plate. I use the Find My Friend app to keep track of people's locations; I use that every day to make sure they're home safe. I use iMessage, FaceTime, and Gmail. When School was in session, I used Canvas on my phone. I also use TikTok, Instagram to post stuff, and YouTube and Hulu to watch the shows I like.*" Additional student comments further revealed their awareness of how apps collect and use data. For example, one student explained how Hulu generates recommendations based on viewing history, saying, "*You watch one show and then Hulu comes down with this list, based on what you've been watching. If I watch, for example, a scary movie, it will have a whole list of other different movies that are similar or that I may be interested in.*" Another student mentioned ad targeting on social media, noting, "*I put Instagram and TikTok on my plate. Based on the stuff you like and watch, they will be able to target ads.*" A third student reflected on data storage when using utility apps, stating, "*When you're* *taking a picture, you're storing this scene*." These comments indicate that students were not merely listing apps but were aware of their dual role as producers and consumers of data. They recognized how their interactions with the apps result in data creation (e.g., posting or communicating) and how that data are used by the apps (e.g., for recommendations or targeted ads). However, while students demonstrated some understanding of these processes, they were less familiar with how external systems extract and use their data. This finding from the PD strengthened our decision to use the "Data Never Sleeps" infographic in the lesson as a way to increase their awareness of different sources and uses of data. Moreover, it helped inform the specific prompts we included in the lessons, asking students to think about the data they create and the apps and activities that result in them creating their data with others.

Analysis of the Stakeholders Board activity in the PD revealed that overall, students understand why entities collect data and who those entities are. Among these stakeholders, corporations were mentioned the greatest number of times (17), followed by government agencies (11), social media platforms (8), educational institutions (6), medical organizations (4), financial institutions (1), and even malicious agents such as hackers (4). This diversity of stakeholders was summed up by one participant who said, "*Pretty much everyone who can profit off your data wants your data*." The students also discussed the reasons these entities collect data. 15 out of 37 answers given by the students referred to improving the user experience. That is, to gather data to customize the services according to the users' needs or interests and to provide a personalized experience. For example, one participant noted, "*TikTok collects data based on what you like, and it will show you videos based on that*." The second most common reason given (6 out of 37) was to inform decision-making. For instance, one student stated: "*The different uses of data are to make decisions about their company and to be able to solve their problems*." Following that notion, two respondents mentioned colleges that use data to determine the acceptance of candidates and award scholarships in their answers. As high school students, college admission is a topic of great interest to them. Another reason raised in 6 out of 34 answers given by the students was profitability, i.e., the profit companies make from data collection. For example, one participant explicitly wrote, "*They [companies] also use it for money*." It is clear from these responses that students understand companies' use of data in pursuit of financial gain. Three out of 37 answers referred to hackers using data maliciously, such as stealing personal information. One student mentioned, "*With hackers and phishers, it's all a lot simpler. They just want your data so they can get all of your other stuff*." Lastly, another reason mentioned for using data was related to regulatory matters. It was apparent from the responses provided by the students that they have a general understanding of why entities collect data and who those entities are. Nevertheless, as reported in previous research [44], there are flaws in students' understanding, and thus, it is important to extend and develop their perspectives. These data led us to design and include the two case studies as part of Lesson 1.2; the first focused on social media companies and how they collect and use user data, mainly how they might sell it to third-party companies. The second was on the ring cameras collecting video on people without their consent and sharing them in cloud-based storage. These cases are meant to serve as ways to engage learners in considering central aspects of our data-driven economy and thinking about the various stakeholders involved, who are creating the data, who are profiting from it, and what key ethical considerations are.

Lastly, the Empathy Map activity analysis also influenced the lesson design. During the PD, students created nine personas aged 16–19, describing their actions, feelings, interests, and concerns. Analysis of the "Concerns" quadrant revealed that privacy and security were at the top of their mind, along with concerns about data collection practices and unauthorized use of personal data. For example, in a created persona named Zamaya, the post-its included concerns such as: "*How many people collect data?*"; "*What is done with our data?*"; and "*Who watches our data?*". Based on this data, we address these concerns in a class by introducing case studies where privacy or bias is violated, which is followed by a structured class discussion. In doing so, we hoped to expand the students' understanding of the subject while discussing cases close to their personal lives.

Alignment of the Lesson to the IIDCE Framework. In our IIDCE framework coding, this lesson aligned with all three core dimensions of the framework: *Knowledge*, *Value*, and *Belonging*. The various activities throughout the lesson emphasize the factor of *personal meaning*, first when the students are encouraged to think about the data sources in their lives, and later when they share personal cases of privacy violations. The fourth activity in the lesson is consistent with the factor of *Relevance* as students consider various aspects of privacy violation related to their lives, including implications of the use of social media and the use of cameras on doors and doorbells, which is increasingly common in the city where this research is taking place. The case study activity and the class discussion on bias align with *usefulness*. Engaging in these activities can help students better understand the consequences of sharing data and the meaning of bias and underrepresentation in data collection. Furthermore, this lesson addresses the factors of *what is data science?* and *who does data science?* This lesson emphasizes that data science relates directly to their activities and the community they participate in. In attending to who the various stakeholders are in data, students are reasoning about the various people and organizations connected with the field, including themselves and their families. This lesson may broaden students' ideas about how they may become involved in data science, recognizing that it is not only data scientists who participate in, contribute to, and are affected by data science.

4.2.2 Lesson 2.4—Accessing Data with APIs Using RapidAPI. Goals and Description. In the fourth lesson of Unit 2, students are introduced to the concept of APIs and learn how to create a block-based program in EduBlocks to access and manipulate API data from a Top 100 movie dataset. The goal of this lesson is for students to learn what an API is and how to programmatically query one to retrieve data. First, they learn what APIs are and their importance and see examples of APIs or services they may be familiar with. Later, the students are introduced to RapidAPI, a large repository that allows students to browse and search publicly available APIs. Then, they create their account and learn how to retrieve and manipulate data by writing short EduBlocks programs. For this lesson, students are to query an API that provides information on the Top 100 Movies on the **Internet Movie Database (IMDB)**. The lesson's structure employs a Use→Modify→Create structure [31]. First, they use a pre-written program and identify the dataset's variables. The students are given a program that retrieves all the information about the Top 100 movies and are asked to investigate the genres associated with the movies returned by the API call. Then, they are challenged to build upon and modify the provided API calls to expand the area of inquiry. More specifically, the students are challenged to extract data about the highest-rated movie of the 2000s, the highest-rated horror movie, and the Top 100 movies from the 1970s. Lastly, the students are asked to create programs using the IMDB Top 100 Movies API to pursue a question that interests them.

How the PD Informed This Lesson. Two PD activities, the Empathy Map and the Data Exploration activity, informed the design of this lesson. Analysis of the persona artifacts created by students in the Empathy Map activity revealed that TV shows and movies were among the topics that interested them. At the end of each PD session, students completed a short online questionnaire in which they were asked about their usage of social networks and mobile apps and their preferences for topics to study within data science courses. The survey responses confirmed the findings from the empathy map, showing that students' second-favorite interests were TV shows and movies, and music (both with an average of 3.86 out of 5), after video games (average of 3.9).

Based on this data from the students, we decided to use the IMDB Top 100 Movies API in this lesson (and data related to video games, i.e., Mario Kart, in the previous lesson). The Data Exploration activity was another source of inspiration for this lesson. It was a thought exercise that focused on performing a data cycle. Students were asked to reflect on their practices and answer questions using data. Along the way, they responded to a series of prompts, which serve as a scaffolding designed to help them break the task down into stages, attending to the practices, knowledge, and resources they use to answer questions in their daily lives. For example, the students who dealt with where to study law stated steps to obtain the information while searching for and preparing preliminary questions. These steps included locating an adult (parent or teacher) who studied law and conducting an interview with them to learn from their experience, looking at law schools' Web sites, and visiting the schools. In Lesson 2.4, the students investigate real data retrieved through the API call. Initially, they are given pre-written pieces of code that they are asked to run. After that, they modify the code according to the task requirements. Finally, they are expected to formulate a question about the data and write their program using the knowledge and practices they have gained. To complete this task, the students must locate the relevant values/parameters in the dataset and plan how to programmatically manipulate and organize the data to achieve the desired results to answer their questions. This could be done, for example, by creating an IF statement.

Alignment of the Lesson to the IIDCE Framework. Lesson 2.4 was analyzed using the IIDCE framework and, like Lesson 1.2, was found to align with all three core dimensions: Knowledge, Value, and Belonging. During this lesson, students are *challenged* to write their own EduBlocks program. This task follows the Use→Modify→Create structure designed to provide scaffolds, help the students complete it successfully, and ensure an appropriate level of challenge based on their prior programming experiences. Using up-to-date data from IMDB is consistent with the relevance factor, as this topic is in the students' interests, as demonstrated by data collected during the PD sessions. The final part of the lesson provides opportunities to make the activity take on *personal meaning* as the task invites them to formulate the questions and explore the data based on their personal movie tastes. Additionally, retrieving the data through an API call is an activity that is consistent with the factor of authenticity, as it is a common computing task that mirrors professional practices in the real world and stands in sharp contrast to working with a fixed dataset provided by the teacher or curriculum. In addition, the task given to the students aims to introduce some of the topics and practices involved in data science, such as data sources, data cleaning, and data manipulation. Moreover, it aims to provide the students with a broad vision of *what data science is*, showing how data and data science intersect with topics and industries of interest.

4.2.3 Lesson 3.4—Graphs and Figures for Two Variables. Goals and Description. In the fourth lesson of Unit 3, students are introduced to different multivariate graphs, namely scatterplots and binned dot plots. The goal of this lesson is for students to learn how to create scatterplots and binned dot plots in CODAP and be able to interpret them. The lesson starts with a warm-up activity where students are given a graph showing reasons for choosing a college. They are asked to identify the type of graph, the main variables and trends it represents, and the story the graph portrays. Afterward, the students are introduced to the Roller Coasters dataset integrated into CODAP. They are given prompt questions about the dataset (like those given in the warm-up activity), and then they are asked to create a scatterplot and describe the relationship between the variables. They are prompted to describe the data in the graph in terms of its trend, linear form, and correlation strength and examine if there are any outliers. Later, the students are asked to change the variables, create new scatterplots and another two-way binned dot plot, and compare the results to earlier data visualizations. In the exit ticket, the students are asked to describe the direction, form, strength, and outliers of a scatterplot presenting the correlation between test scores and final grades.

How the PD Informed This Lesson. The Question Board and Selfiecity Data Exploration activities influenced Lesson 3.4. The Question Board activity yielded 74 unique, student-generated questions about sports, school, video games, money, animals, art, science, books, cooking, social media, and going out (leisure). Sports was the topic with the highest number of questions (13 out of 74), followed by video games and school (each with 11 questions out of 74). After posting questions, students voted on the questions regarding excitement and importance. Of the question topics posted, animals were the topic that received the highest number of total votes, followed by video games, sports, going out, and school. Going out was ranked in second place in terms of excitement and in fourth place in terms of importance, while school was ranked in fourth place both in terms. The various questions raised by the students and the ranking of the various topics helped choose the examples and datasets that will be integrated into this lesson and throughout the entire curriculum. We decided to include research activities related to school/studies and leisure (roller coasters in this case) in this lesson to spark more interest among students.

The Selfiecity Data Exploration activity also inspired the design of this lesson. The researchers' observations during the PD revealed that the students were very involved throughout the activity, especially when asked to identify the trends and the story the graph portrays and when challenged to create research questions and think through how to answer them. They demonstrated critical thinking by questioning the methods used to collect and interpret data. For example, one of the groups suggested to examine the relationship between selfies and the real emotions of teenagers. They suggested gathering data through "*A secure online platform where participants could upload a selfie and verify their identity by linking it to their SSN. This platform would enable users to set their mood to ensure accurate data that is not subjected to the interpretation of an algorithm or a person.*" This task of working with an engaging dataset to think about where the data came from, what questions could be answered with it, and how that data could be displayed to convey information yielded positive feedback from students, so we sought to invite students to engage in similar data analysis practices in Lesson 3.4. By integrating such practices, we aimed to. Foster the kind of analytical thinking and curiosity observed during the PD.

Alignment of the Lesson to the IIDCE Framework. Similar to the previous lessons presented, Lesson 3.4 was found to align with the IIDCE framework's three core dimensions: Knowledge, Value, and Belonging. The activities in this lesson emphasize the *challenge* factor. First, the students were given guiding questions (scaffolding) to help interpret the presented graphs. Then, they created new graphs and identified emerging trends independently. Throughout the lesson, the difficulty level increased according to the construction of the student's knowledge. The activities emphasize the *relevance* aspect as the dataset and graphs articulate students' interests based around roller coasters, including data from roller coasters from amusement parks within driving distance of the school where the lessons are being taught. Allowing students to create their graphs using CODAP shows the *authenticity* of the assignments. Students can use their graphing skills to compare different variables (e.g., wooden versus steel roller coaster speeds), interpret the results, and draw conclusions. In this way, students are asked to carry out a data visualization process to produce a graph that conveys insight rather than just talking about a potential path of analysis or a hypothetical graph or visual. Choosing to start the lessons by having students analyze data visualizations related to reasons for choosing a college highlights the factor of *usefulness*, as engaging with this activity can benefit students in choosing a college in the future. The lesson activities focus on both aspects of Belonging. Students broaden their understanding of *what data science is* by experiencing data science practices, emphasizing data visualization and interpretation.

4.3 Expression of Interest throughout the Curriculum

In the case studies presented above, we described how the dimensions and key factors of the IIDCE framework manifested. As mentioned above, a significant emphasis was placed on integrating datasets and examples corresponding to the student's interests throughout the curriculum. This was a persistent focus across the entire curriculum. For example, during the first unit, students

Table 7. Alignment of the Curriculum to the IIDCE Framework

		Unit 1: Data in Learners' Lives						Unit 2: Computational Foundations of Data Science						Unit 3: Data Science Practices					
		1.1	1.2	1.3	1.4	1.5	1.6	2.1	2.2	2.3	2.4	2.5	2.6	3.1	3.2	3.3	3.4	3.5	3.6
Knowledge	Challenge			×	×	×	×	×	×	×	×	×	×		×	×	×	×	×
	Relevance		×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	
	Authenticity							×	×	×	×	×	×		×	×	×	×	×
Value	Personal meaning	×	×	×	×	×	×	×			×	×						×	
	Usefulness		×			×	×	×	×					×	×		×	×	
Belonging	What data science is?	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
	Who does data science?	×	×	×	×	×	×	×					×		×	×			

are given the opportunity to investigate a Heat Sensitivity Exposure dataset focused on the area surrounding their school. This dataset is used to teach them about the Data-Information-Knowledge-Wisdom architecture. This conceptual framework exemplifies the hierarchy of data processing and transformation from raw data to meaningful insights and actionable wisdom [75]. In a later part of the unit, the students are requested to assess the quality of a dataset about a well-known coffee brand, attending to key aspects of the dataset while discussing potential bias and equity issues. Once they have analyzed the dataset, they are instructed to find a dataset related to a topic they are interested in and perform a similar analysis. In the second unit, students analyze a dataset based on Mario Kart characters and their vehicles, a beloved and well-known video game. Later, they are introduced to RapidAPI and challenged to author programs to manipulate data from various APIs, including the IMDB Top 100 Movies, Billboard Top 100, and NBA APIs. In the third unit, the students experiment with visualization practices. They analyze datasets about roller coasters across the US, mammals, and seals' migratory patterns. Then, they learn to visualize the data they manipulated in EduBlocks using Zillow (a real estate Web site) and Earthquake APIs. These are just a few of the datasets and topics incorporated into the curriculum to try to connect the materials with students' interests. Collectively, the curriculum offers students opportunities to gain a fundamental understanding of what data science is, its practices, and how to apply them in different contexts. An analysis of the lessons and the datasets and activities students engage with shows how each of the three units implements the dimensions of the interest development theory: Knowledge, Value, and Belonging (Table 7).

Despite our efforts to ground the curriculum in students' lived experiences and interests, our analysis, shown in Table 7, reveals room for improvement. For example, Units 2 and 3 need more emphasis on personal meaning. This is crucial in engaging students and adding value to their learning experience. It can be fostered by tailoring the content to the student's cultural resources, beliefs, or identity or by allowing them to choose and work on topics they are passionate about. Related to this, we found a series of consecutive lessons where usefulness and consideration of "Who Does Data Science?" facets of interest were absent. These lessons place a greater emphasis on programming and show that, in our first iteration of this work, we fall victim to some of the same shortcomings related to representation and relevancy that often plague introductory computer science curricular materials [59]. After completing the above analysis, we revised the curriculum to address these oversights in an effort to provide more consistent and thorough coverage of aspects of interest-driven learning through all three units.

We have modified some of the lessons to better align with the dimension of Value and, specifically, the key factors of personal meaning and usefulness. In Lesson 2.5, for example, in which students investigate the Billboard Music API, we made adjustments to allow students to explore their favorite artists and songs. To emphasize the added value and usefulness of this data exploration, we added an activity encouraging them to reflect on their insights and think about additional analyses that could be performed using that data. An example of how we addressed the lack of personal meaning is that we modified the summative activity in Lesson 3.2, in which the students researched data on roller coasters across the USA. Students now have the option to choose which state to explore. This encourages them to reflect on their roller-coaster riding experiences, decide which variables to graph, and describe the characteristics of their favorite roller coaster.

In the second and third units, we also made changes to strengthen the dimension of Belonging, especially in expanding the concept of who participates in data science. For instance, in Lesson 2.6, while working with NBA API data, we noticed that this aspect was not explicitly addressed. After evaluating the lesson using the IIDCE framework, we added a class activity where students are given prompts on a dedicated slide to discuss how coaches and players can use data from this API to evaluate their performance from game to game or season to season and to analyze how to improve. These prompts also highlight how people indirectly involved, such as fans, can use this data to track the stats of their favorite players and teams. Similarly, in Lesson 3.6, in which students work with the EveryEarthquake API, we made changes and emphasized that geologists might perform such a data science activity to investigate earthquake trends thoroughly. These revisions show how the IIDCE framework can shed light on where interest is and is not incorporated into instructional materials and highlight the ease with which various facets of interest can be incorporated into data science courses. In doing so, it shows how insights related to learners' interests, identities, and lived experiences gathered through PD can be incorporated into a data science course.

5 Discussion

5.1 Influence of PD on Lesson Design

PD is a crucial methodology for designing learning environments because it actively involves students and teachers in the design process, ensuring that educational materials are relevant and engaging [39]. It is a powerful approach that provides ways for students' voices and values to be included in curriculum design [15, 19]. This approach is compelling for creating interest-driven computational learning experiences as PD allows educators to tailor the learning materials to students' intrinsic interests and learning needs, leading to enhanced motivation and improved educational outcomes. PD is also helpful in developing learner experiences for contemporary and societally relevant topics, like data science. Given the practical applications of data science and the various ways it interacts with learners' lived experiences, PD provides insight into how youth think about, experience, and use data, which can significantly impact student engagement with the resulting learning materials [53].

In this study, we employed seven different PD activities to gather insights to guide the creation of an interest-driven data science curriculum. The design activities enabled students to interact with data, reflect on its significance in their lives, and apply data science practices. Analyzing the data from these activities helped us understand how students perceive data in their daily lives and identify gaps in their understanding. Moreover, it helped us identify students' interests and preferences, which we then integrated into the curriculum through tailored assignments, datasets, and APIs. The PD also helped us discover ways to spark students' interest beyond topics they are passionate about. This included using warm-up activities to stimulate the learners' curiosity, creating personas to reveal habits, interests, and concerns, identifying data collectors and their motivations, and experimenting with exploratory data visualization activities. These activities also directly informed the structure of the curricular lessons, leading to the decision to integrate warm-up activities at the beginning of each lesson and to integrate two successful PD activities into the curriculum. The "Data Never Sleeps" activity revealed students' fascination with the vast amount of data collected by the various applications, sparking their eagerness to explore more statistics. Similarly, the "Selfiecity Data Exploration" activity resulted in high student engagement and fruitful discussion around the provided graphs.

The PD sessions also helped inform the specific datasets chosen and the type and level of scaffolds provided. For example, we worked to incorporate more real-life scenarios that require critical thinking and advanced data manipulation techniques that were relevant to students based on data from the PD. The lessons were designed to provide an appropriate level of challenge, balancing difficulty with their achieved competency, while the topics and datasets were carefully selected to match students' experiences, communities, and cultural resources. Some datasets were selected based on geography (e.g., data related to grocery stores near students' schools or housing data from different states), while others were chosen to reflect students' interests as identified during the PD, such as data on the best movies, NBA teams and players, or Billboard hits. We also sought to give learners agency when choosing the subjects or objects of the investigation (e.g., analyzing data about a specific artist they admire using a dataset provided in class or selecting a topic that interests them for their final project investigation). This personal connection is crucial for fostering a deeper engagement with the material [81], as students can see how data science could impact their interests and future aspirations [52].

The PD process was instrumental in shaping our interest-driven curriculum. By actively involving students in the design process, we gained deep insights into their interests, preferences, and motivations, as well as capabilities that informed the creation of engaging and relevant educational content. The PD process revealed the topics and activities that produced deep engagement while also revealing aspects of data science with which students had prior experiences or, alternatively, were less familiar. In this way, the PD enabled us to craft a curriculum that included both low-floor data analysis activities and discussion prompts for all students and more challenging assignments where more substantial scaffolds were needed (and provided). The structure of the PD activities and the involvement of the students helped to generate insights regarding the nature and frequency of scaffolds throughout the curriculum. These insights led us to design lessons that included various activities, such as interactive data analysis projects, class discussions, and individual research assignments. Additionally, we incorporated various strategies to assist students, such as dividing the tasks into smaller parts, providing worksheets and visual aids, including step-by-step scaffolded activities, and encouraging group work and class discussion.

This study not only demonstrates how PD can inform curriculum design but also offers insights into the challenges and benefits of using PD to co-design with students who lack deep subject-domain expertise in data science. One significant challenge we encountered was that, although students regularly interact with data through apps and social media, they initially had a narrow view of what data science entails, particularly its analytical processes. They were more familiar with data consumption (e.g., through recommendations or targeted ads) than with data analysis practices. To address this, we designed PD activities that encouraged students to reflect on how data is produced, collected, and used in their everyday lives, which helped broaden their understanding of data science. Moreover, we provided additional support in the form of guided exploration activities and structured prompts. Despite their limited domain expertise, students' perspectives were instrumental in shaping the curriculum. By involving students in tasks that required them to generate questions, explore datasets, and interpret data visualizations, we gained valuable insights into how novice learners approach data science. Furthermore, using multiple types of PD activities proved essential for understanding how students think about data and what they find engaging. Each activity contributed a different layer of insight, helping us design lessons that not only aligned with their interests but also challenged their thinking in meaningful ways. This process demonstrated that collaborating with students, even those with limited domain expertise can be meaningful in informing a curriculum design as long as adequate scaffolding is incorporated into the design process.

5.2 Combining PD, Interest-Driven Learning, and Data Science

While the specific goal of this larger research project was to create an effective, engaging data science curriculum, the approach taken has the potential to be broadly useful in computing education. Specifically, bringing together PD, interest-driven learning, and data science highlights synergies between the three central pillars of this work that can inform future work in the computing education space. A feature of data science, and computing more broadly, is its ubiquity and protean nature. By this, we mean it is possible to situate computing instruction in countless thematic contexts. Carefully selecting compelling contexts through which to teach can substantially impact learning outcomes and directly address longstanding issues of equity and belonging [6, 18, 60]. In the case of data science, the ubiquity of data and ease of accessing novel datasets make it particularly well suited for creating interest-driven learning. The existence of large repositories of publicly available data provides a means to empower learners to go out and find datasets that align with their interests. Given the importance of interest in learning [74] and the ability to situate computing in various contexts, bringing interest-driven learning to computing has tremendous potential, as outlined by the IIDCE framework. This work takes the next step beyond the IIDCE framework by leveraging PD as a methodological approach for realizing the goal of creating interest-driven computing learning experiences.

A central challenge in creating interest-driven learning is ensuring that the learning activities are, in fact, of interest to the students. A strength of PD is its ability to provide a voice to stakeholders, in this case, students, who otherwise would not have their ideas, values, or interests heard. As shown in this work, the various PD activities provided a means of sharing the students' interests with researchers in ways more nuanced and comprehensive than just asking them to complete a survey. Table 6 shows how the PD session revealed a large and diverse set of student interests that were then used to inform the resulting curriculum.

Finally, just as PD can be useful for shedding light on students' interests, so too can it be useful for helping us understand how to create effective and appropriate data science learning experiences. Given the relative recency of data science as a high school discipline, the nascent field of K-12 data science education is still working to understand what concepts and practices drawn from professional fields look like in a high school classroom. Using PD to learn what students know about data science, their experiences with data and data science, their current perceptions of the field, and their understanding of its impact on their lives serves as invaluable insights in creating a new curriculum. This feature of PD has direct applications for other emerging computing fields. For example, the rapid rise of fields like artificial intelligence and techniques like machine learning has led to educators and policymakers making calls to bring these cutting-edge technologies into classrooms [85]. Figuring out how best to do so, understanding what students do and do not know, and thinking through the design of tools for such educational experiences is a problem well suited for PD. In fact, some of this work is already underway, using PD to explore students' understanding of and methods for learning about machine learning [28].

Given computing's ever-evolving landscape and computing education's need to keep up, having strategies for bringing technologies and ideas from industry and society into the classroom in ways that reflect and respect what students know and care about is essential. With this work, we show

how PD can meet this challenge. Further, it shows how interest-driven learning can inform the design of these new, novel learning experiences. While in our case, the context was data science, it is easy to envision similar work focused on robotics, machine learning, artificial intelligence, or whatever the next wave of computing innovation is.

5.3 Operationalization of Interest through the IIDCE Framework

After creating and piloting the materials, we used the IIDCE framework [56] to assess the new interest-driven data science curriculum informed by students' insights from the PD sessions. The goal was to evaluate whether or not the PD-informed curricular materials achieved the goal of being interest-driven. This framework provided a multifaceted lens to evaluate if, how, and where the lessons operationalize different instantiations of interest. This analysis revealed how and where each curricular unit addressed the three dimensions of interest: knowledge, value, and belonging.

The idea that learning experiences need to be "real" is operationalized in a few ways in the IIDCE framework, including via authenticity and usefulness. Informed by these dimensions, the curriculum has students use authentic tools and practices. By combining real-time data from APIs with user-friendly interfaces (i.e., EduBlocks and CODAP), students can delve into trends and find answers to their questions. This approach provides an engaging introduction to data science and helps students grasp the value and usefulness of fundamental data science skills [45].

The IIDCE framework was useful in evaluating the curriculum we created and providing a roadmap for further improving it to better align with the full array of interest-driven learning dimensions. As shown in Table 7, our analysis revealed gaps in some dimensions of interest-driven learning. For example, the key factor of who is involved in data science was sometimes implied in the curriculum but not explicitly discussed. Moreover, it is often expressed from the perspective of stakeholders who collect and analyze data, neglecting those 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 crucial for creating a sense of belonging and directly confronting longstanding issues of a lack of representation in computing-related fields. Research shows that specific populations have been disproportionately impacted by biases and predatory applications of data-driven algorithms [e.g., 5, 62], and data science education must address these issues directly and intentionally to eradicate their perpetuation [65, 78]. These important issues are addressed in the new version of our curriculum. In this way, we can see how the IIDCE framework helps make concrete the, at times amorphous, idea of interest-driven learning.

5.4 Practical Implications for Designers, Educators, and Scholars

The findings from this research have several practical implications for educators and scholars in the computer science education community. In this work, we emphasize the utility of PD as a strategy to inform the creation of student-informed, interest-driven learning materials and demonstrate the utility of the IIDCE framework as an evaluative tool. Given the vision of Computer Science for All that our community is working toward, creating instructional experiences that resonate with learners of differing backgrounds and experiences is essential. This work shows the role PD can play in accomplishing this goal. Educators seeking to create interest-driven materials should consider incorporating PD methods to ensure that curricula are relevant and engaging for students. While this work shows how PD can be used for designing interest-driven curricula in data science, its applicability to other computer science topics may vary depending on the level of student familiarity with the subject and the flexibility of the educational context. A similar process has been used for the design of an introductory middle school CS curriculum [15], but how this approach can help for high school CS courses and beyond remains an open question. Our findings suggest that PD may be particularly effective for interdisciplinary computer science topics with clear real-world connections, where students' personal experiences can meaningfully inform the learning process. Designers and educators should consider using PD when developing curricula for socially embedded topics such as artificial intelligence, data privacy, cybersecurity, and machine learning, as these areas may benefit most from student insights and interest-driven learning.

In this work, the school environment also played a role in shaping the PD process. Working within a public charter school with a flexible curriculum allowed us to conduct extended PD sessions and gain deeper insights into students' perspectives. In more traditional educational settings, where time is more limited, it may be necessary to implement alternative PD methods or spread the activities over multiple sessions. We acknowledge that running extracurricular PD sessions is a significant undertaking, and there may also be more informal ways to incorporate PD ideas and activities into existing class time. Using class as a chance to teach and learn from students can potentially provide invaluable insights into ways to frame and teach topics. Doing so can deepen engagement and naturally promote inclusivity by valuing the voices of the students being taught. Incorporating students' voices and suggestions can ensure that the curriculum addresses their preferences and increases their engagement and motivation.

The use of the IIDCE framework presented above also has potential implications for designers, educators, and scholars. As shown above, the IIDCE framework can provide a structured and concrete operationalization of interest-driven learning that reflects the multifaceted nature of interest. Using the framework to evaluate a curriculum or learning experience can help ensure it attends to various forms of interest. Additionally, the framework can help identify potential pathways to deepen learner interest by identifying underdeveloped aspects of interest. Educators and designers can create a balanced curriculum that nurtures long-term interest and engagement in data science by focusing on all three dimensions of interest.

This work also provides a methodological contribution by demonstrating the applicability of the IIDCE framework to evaluate interest-driven learning in computer science-adjacent contexts such as data science. In doing so, it shows the potential for using the framework to inform other computer science and neighboring fields, such as robotics, engineering, artificial intelligence, and computational social sciences and arts. This work marks an initial step in a potentially generative line of work at the intersection of design and interest development. Bringing a multifaceted conceptualization of interest development and design activities that tap into different aspects that can spark interest in students, motivating them to explore, innovate, and get interested and invested in new scholarly pursuits.

5.5 Limitations

This study has several limitations that should be acknowledged. First, the findings are situated within a specific population of high school students in one geographic region and thus may not be generalized to students in other contexts, age groups, or educational settings. Second, the study was conducted at a particular moment in time where certain technologies, usages, and pop culture trends were salient to students. Given the rapid pace at which digital technologies and pop culture evolve, some of the examples and interests identified by the students may become irrelevant in the future. Finally, this article argues for the use of PD as a methodology for incorporating student interest into the design of instructional materials. To demonstrate this, we present the design of a data science curriculum but acknowledge that data science as a discipline provides more flexibility and more opportunities to integrate student interests than other disciplines. As such, conducting similar work in other domains or with different grades may look different.

6 Conclusion

Interest development theory posits that individuals are more likely to engage deeply with a subject and learn more effectively if the topic aligns with their interests, passions, goals, and sense of identity. Given the ubiquity of data today, there is tremendous potential for creating data science learning opportunities informed by interest development theory. This study demonstrates how PD, grounded in the cooperative inquiry method, can support this alignment by actively involving students in the design process and fostering a sense of belonging and cultural relevance to the field of data science. Our work broadens the literature on PD by highlighting the contribution of conducting distinct and complementary PD activities with high school students to gather insights into students' knowledge, interests, and perceptions. This approach not only informed the design of a more engaging and relevant data science curriculum but also shed light on the challenges and opportunities of co-designing with novice learners and ways to support their engagement with computational learning experiences. The implications of our study extend beyond data science education to guide educators and curriculum designers in the creation of interest-driven instructional materials across computer science and computer science-adjacent disciplines. The various PD activities and methods, along with the use of the IIDCE framework, can be applied to a wide array of computing fields, given the importance of computing in the lives of today's learners. Such a student-centered approach can cultivate a sense of belonging in the field of computing, which is essential for broadening participation in computing. By emphasizing engagement and personal and cultural relevance, we have the potential to transform how we introduce learners to the ways computing is impacting their lives and prepare them to succeed in an increasingly data-driven world.

References

[1] Azevedo, F. S. 2019. A pedagogy for interest development: The case of amateur astronomy practice. *Learning, Culture and Social Interaction* 23 (Dec. 2019), 100261. DOI: https://doi.org/10.1016/j.lcsi.2018.11.008

[2] Bargagliotti, A. Arnold, P, and Franklin, C. 2021. GAISE II: Bringing data into classrooms. *Mathematics Teacher: Learning and Teaching PK-12* 114, 6 (Jun. 2021), 424–435. DOI: https://doi.org/10.5951/MTLT.2020.0343

[3] Belitz, C. Ocumpaugh, J. Ritter, S. Baker, R. S. Fancsali, S. E, and Bosch, N. 2023. Constructing categories: Moving beyond protected classes in algorithmic fairness. *Journal of the Association for Information Science and Technology* 74, 6 (2023), 663–668. DOI: https://doi.org/10.1002/asi.24643

[4] R. Benjamin. 2019. *Race after Technology: Abolitionist Tools for the New Jim Code*. John Wiley & Sons.

[5] Biehler, R. Veaux, R. D. Engel, J. Kazak, S, and Frischemeier, D. 2022. Research on data science education. *Statistics Education Research Journal* 21, 2 (Jul. 2022), 1–1. DOI: https://doi.org/10.52041/serj.v21i2.606

[6] Boaler, J. 1993. The role of contexts in the mathematics classroom: Do they make mathematics more "real"? *For the Learning of Mathematics* 13, 2 (1993), 12–17.

[7] S. Bødker, C. Dindler, O. S. Iversen, and R. C. Smith. 2022. What is participatory design? *Participatory Design*. S. Bødker, C. Dindler, O.S. Iversen, and R.C. Smith (Eds.), Springer International Publishing, 5–13.

[8] Bowler, L. Acker, A. Jeng, W, and Chi, Y. 2017. "It lives all around us": Aspects of data literacy in teen's lives. *Proceedings of the Association for Information Science and Technology* 54, 1 (2017), 27–35. DOI: https://doi.org/10.1002/pra2.2017.14505401004

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

[10] J. Buolamwini and T. Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77–91.

[11] Cao, L. 2017. Data science: A comprehensive overview. *ACM Computing Surveys* 50, 3 (Oct. 2017), 1–42. DOI: https://doi.org/10.1145/3076253

[12] Carmi, E. Yates, S. J. Lockley, E, and Pawluczuk, A. 2020. Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review* 9, 2 (2020), 1–22. DOI: https://doi.org/10.14763/2020.2.1481

[13] Coburn, C. E and Penuel, W. R. 2016. Research–practice partnerships in education: Outcomes, dynamics, and open questions. *Educational Researcher* 45, 1 (2016), 48–54.

[14] CODAP. 2022. Common Online Data Analysis Platform. Retrieved December 19, 2022 from https://codap.concord.org/

[15] Coenraad, M. Palmer, J. Eatinger, D. Weintrop, D, and 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 (Mar. 2022), 100353. DOI: https://doi.org/10.1016/j.ijcci.2021.100353

[16] M. Coenraad, J. Palmer, D. Franklin, and D. Weintrop. 2019. Enacting identities: Participatory design as a context for youth to reflect, project, and apply their emerging identities. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, 185–196.

[17] Cohen, J. 1960. *A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement* 20, 1 (1960), 37–46.

[18] Cooper, S and Cunningham, S. 2010. Teaching computer science in context. *ACM Inroads* 1, 1 (Mar. 2010), 5–8. DOI: https://doi.org/10.1145/1721933.1721934

[19] D. Couso. 2016. Participatory approaches to curriculum design from a design research perspective. *Iterative Design of Teaching-Learning Sequences: Introducing the Science of Materials in European Schools*. D. Psillos and P. Kariotoglou (Eds.), Springer Netherlands, 47–71.

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

[21] B. DiSalvo and C. DiSalvo. 2014. Designing for democracy in education: Participatory design and the learning sciences. In *Learning and Becoming in Practice: The International Conference of the Learning Sciences (ICLS) 2014*. Joseph L. Polman, Eleni A. Kyza, D. Kevin O'Neill, Iris Tabak, William R. Penuel, A. Susan Jurow, Kevin O'Connor, Tiffany Lee, and Laura D'Amico (Eds.), Vol. 2, International Society of the Learning Sciences, Colorado, CO, 793–799.

[22] B. DiSalvo, J. Yip, E. Bonsignore, and C. DiSalvo. 2017. *Participatory Design for Learning: Perspectives from Practice and Research*. Taylor & Francis.

[23] Donoghue, T. Voytek, B, and Ellis, S. E. 2021. Teaching creative and practical data science at scale. *Journal of Statistics and Data Science Education* 29, sup1 (Jan. 2021), S27–S39. DOI: https://doi.org/10.1080/10691898.2020.1860725

[24] Donoho, D. 2017. 50 years of data science. *Journal of Computational and Graphical Statistics* 26, 4 (Oct. 2017), 745–766. DOI: https://doi.org/10.1080/10618600.2017.1384734

[25] F. van Doorn. 2016. *Children as Co-Researchers in Design: Enabling Users to Gather, Share and Enrich Contextual Data.*

[26] A. Druin. 1999. Cooperative inquiry: Developing new technologies for children with children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 592–599.

[27] Druin, A. 2002. The role of children in the design of new technology. *Behaviour and Information Technology* 21, 1 (Jan. 2002), 1–25. DOI: https://doi.org/10.1080/01449290110108659

[28] U. Dwivedi, S. Elsayed-Ali, E. Bonsignore, and H. Kacorri. 2024. Exploring AI problem formulation with children via teachable machines. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–18.

[29] Fails, J. A. Guha, M. L, and 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 (Dec. 2013), 85–166. DOI: https://doi.org/10.1561/1100000018

[30] Famaye, T. Irgens, G. A, and Adisa, I. 2025. Shifting roles and slow research: Children's roles in participatory co-design of critical machine learning activities and technologies. *Behaviour & Information Technology* 44, 5 (Mar. 2025), 912–933. DOI: https://doi.org/10.1080/0144929X.2024.2313147

[31] D. Franklin, M. Coenraad, J. Palmer, D. Eatinger, A. Zipp, M. Anaya, M. White, H. Pham, O. Gökdemir, and D. Weintrop. 2020. An analysis of use-modify-create pedagogical approach's success in balancing structure and student agency. In *Proceedings of the 2020 ACM Conference on International Computing Education Research*, 14–24.

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

[33] R. Gould, S. Machado, T. A. Johnson, and J. Molyneux. 2022. Introduction to data science curriculum. Retrieved from https://curriculum.idsucla.org/

[34] Guha, M. L. Druin, A, and Fails, J. A. 2013. Cooperative inquiry revisited: Reflections of the past and guidelines for the future of intergenerational co-design. *International Journal of Child-Computer Interaction* 1, 1 (Jan. 2013), 14–23. DOI: https://doi.org/10.1016/j.ijcci.2012.08.003

[35] Harackiewicz, J. M. Smith, J. L, and Priniski, S. J. 2016. Interest matters: The importance of promoting interest in education. *Policy Insights from the Behavioral and Brain Sciences* 3, 2 (Oct. 2016), 220–227. DOI: https://doi.org/10.1177/2372732216655542

[36] Hardy, L. Dixon, C, and Hsi, S. 2020. From data collectors to data producers: Shifting students' relationship to data. *Journal of the Learning Sciences* 29, 1 (Jan. 2020), 104–126. DOI: https://doi.org/10.1080/10508406.2019.1678164

[37] Harrington, C. Erete, S, and Piper, A. M. 2019. Deconstructing community-based collaborative design: Towards more equitable participatory design engagements. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 216:1–216:25. DOI: https://doi.org/10.1145/3359318

[38] Hidi, S and Harackiewicz, J. M. 2000. Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research* 70, 2 (Jun. 2000), 151–179. DOI: https://doi.org/10.3102/00346543070002151
[39] C. Hoadley. 2017. How participatory design has influenced the learning sciences. *Participatory Design for Learning: Perspectives from Practice and Research*. B. DiSalvo, J. Yip, E. Bonsignore, and C. DiSalvo (Eds.), Routledge, 22–27.
[40] IBM. 2018. Enterprise Design Thinking: Field Guide. Retrieved February 02, 2024 from https://www.ibm.com/design/thinking/
[41] IDSSP Curriculum Team. 2019. Curriculum Frameworks for Introductory Data Science. Retrieved August 11, 2023 from http://idssp.org/files/IDSSP_Frameworks_1.0.pdf
[42] Selfiecity. 2023. Investigating the style of self-portraits (selfies) in five cities across the world. Retrieved August 09, 2023 from https://selfiecity.net/
[43] R. Israel-Fishelson, P. F. Moon, D. Pauw, and D. Weintrop. 2024. Exploring interest-driven data science through participatory design. In *Proceedings of the 18th International Conference of the Learning Sciences (ICLS '24)*, 1159–1162.
[44] R. Israel-Fishelson, P. F. Moon, D. Pauw, and D. Weintrop. 2024. Using participatory design to gain insight into how students make sense of data in their lives. In *Proceedings of the 2024 Symposium on Learning, Design and Technology*, 85–94.
[45] Israel-Fishelson, R. Moon, P. F. Tabak, R, and Weintrop, D. 2025. Preparing students to meet their data: An evaluation of K-12 data science tools. *Behaviour & Information Technology* 44, 5 (Mar. 2025), 934–953. DOI: https://doi.org/10.1080/0144929X.2023.2295956
[46] R. Israel-Fishelson, P. F. Moon, and D. Weintrop. 2024. Interest-driven data science curriculum for high school students: Empirical evidence from a pilot study. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference*, 908–912.
[47] O. S. Iversen, R. C. Smith, and C. Dindler. 2017. Child as protagonist: Expanding the role of children in participatory design. In *Proceedings of the 2017 Conference on Interaction Design and Children*, 27–37.
[48] Janeja, V. P. Sanchez, M. Khoo, Y. X. Von Vacano, C, and Chen, L. K. 2024. Adopting foundational data science curriculum with diverse institutional contexts. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, 576–582.
[49] LaMar, T and Boaler, J. 2021. The importance and emergence of K-12 data science. *Phi Delta Kappan* 103, 1 (Sep. 2021), 49–53. DOI: https://doi.org/10.1177/00317217211043627
[50] Large, A. Nesset, V. Beheshti, J, and Bowler, L. 2006. "Bonded design": A novel approach to intergenerational information technology design. *Library & Information Science Research* 28, 1 (Mar. 2006), 64–82. DOI: https://doi.org/10.1016/j.lisr.2005.11.014
[51] Lee, V. R and Delaney, V. 2022. Identifying the content, lesson structure, and data use within pre-collegiate data science curricula. *Journal of Science Education and Technology* 31, 1 (Feb. 2022), 81–98. DOI: https://doi.org/10.1007/s10956-021-09932-1
[52] V. R. Lee, P. Sarin, I. Sieh, and A. Fuloria. 2023. Addressing the data set dilemma with personally relevant data generation and distributed labeling in the classroom. In *Proceedings of the 16th International Conference on Computer-Supported Collaborative Learning - CSCL 2023*. C. Damşa, M. Borge, E. Koh, and M. Worsley (Eds.), 420–423.
[53] Lee, V. R. Wilkerson, M. H, and Lanouette, K. 2021. A call for a humanistic stance toward K–12 data science education. *Educational Researcher* 50, 9 (Dec. 2021), 664–672. DOI: https://doi.org/10.3102/0013189X211048810
[54] Martinez, W and LaLonde, D. 2020. Data science for everyone starts in kindergarten: Strategies and initiatives from the American statistical association. *Harvard Data Science Review* 2, 3 (Sep. 2020). DOI: https://doi.org/10.1162/99608f92.7a9f2f4d
[55] Miaskiewicz, T and Kozar, K. A. 2011. Personas and user-centered design: How can personas benefit product design processes? *Design Studies* 32, 5 (Sep. 2011), 417–430. DOI: https://doi.org/10.1016/j.destud.2011.03.003
[56] Michaelis, J. E and Weintrop, D. 2022. Interest development theory in computing education: A framework and toolkit for researchers and designers. *ACM Transactions on Computing Education* 22, 4 (Dec. 2022), 43:1–43:27. DOI: https://doi.org/10.1145/3487054
[57] L. H. Musaeus, M. G. Petersen, and C. N. Klokmose. 2024. Bringing teachers and researchers together through participatory design and cooperative prototyping in computing education. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, 902–908.
[58] B. Naimipour, M. Guzdial, and T. Shreiner. 2020. Engaging pre-service teachers in front-end design: Developing technology for a social studies classroom. In *Proceedings of the 2020 IEEE Frontiers in Education Conference (FIE)*, 1–9.
[59] National Academies of Sciences, Engineering, and Medicine. 2021. *Cultivating Interest and Competencies in Computing: Authentic Experiences and Design Factors*. National Academies Press.
[60] National Academies of Sciences, Engineering, and Medicine. 2018. *How People Learn II: Learners, Contexts, and Cultures*. National Academies Press.
[61] Nolan, D. Temple Lang, and D. 2010. Computing in the statistics curricula. *The American Statistician* 64, 2 (May 2010), 97–107. DOI: https://doi.org/10.1198/tast.2010.09132

[62] C. O'Neil. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown.

[63] Panke. 2019. Design thinking in education: Perspectives, opportunities and challenges. *Open Education Studies* 1, 1 (2019), 281–306.

[64] Pfannkuch, M. Ben-Zvi, D, and Budgett, S. 2018. Innovations in statistical modeling to connect data. *Chance and Context. ZDM* 50, 7 (Dec. 2018), 1113–1123. DOI: https://doi.org/10.1007/s11858-018-0989-2

[65] Philip, T. M. Olivares-Pasillas, M. C, and Rocha, J. 2016. Becoming racially literate about data and data-literate about race: Data visualizations in the classroom as a site of racial-ideological micro-contestations. *Cognition and Instruction* 34, 4 (Oct. 2016), 361–388. DOI: https://doi.org/10.1080/07370008.2016.1210418

[66] J. L. Polman, A. M. Kohnen, M. P. Whitacre, R. M. Davidson, and E. H. Gebre. 2017. Evolving curricular designs through teacher adaptation and implementation with students over time. In *Participatory Design for Learning*. B. DiSalvo, J. Yip, E. Bonsignore, and C. DiSalvo (Eds.), Routledge, 59–70.

[67] Potvin, A. S. Teeters, L. P. Penuel, W. R, and Dimidjian, S. 2024. Humanizing co-design through attention to educators' affective and relational experiences. *Journal of the Learning Sciences* 33, 1 (Jan. 2024), 41–79. DOI: https://doi.org/10.1080/10508406.2024.2318557

[68] K. A. Renninger and S. Hidi. 2015. *The Power of Interest for Motivation and Engagement*. Routledge.

[69] Renninger, Martina Nieswandt, and Suzanne Hidi. 2015. *Interest in Mathematics and Science Learning*. American Educational Research Association.

[70] T. F. Revano and M. B. Garcia. 2021. Designing human-centered learning analytics dashboard for higher education using a participatory design approach. In *Proceedings of the 2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, 1–5.

[71] C. Rivas. 2012. *Coding qualitative data. In Researching Society and Culture*. C. Seale (Ed.), SAGE Publications, 367–392.

[72] J. Roschelle, W. R. Penuel, and N. Shechtman. 2006. Co-design of innovations with teachers: Definition and dynamics. In *Proceedings of the 7th International Conference on Learning Sciences*, 606–612.

[73] J. M. Rosenberg, M. Lawson, D. J. Anderson, R. S. Jones, and T. Rutherford. 2020. Making data science count in and for education. In *Research Methods in Learning Design and Technology*. Routledge, 94–110.

[74] J. I. Rotgans and H. G. Schmidt. 2017. The role of interest in learning: Knowledge acquisition at the intersection of situational and individual interest. In *The Science of Interest*. P.A. O'Keefe and J.M. Harackiewicz (Eds.), Springer International Publishing, 69–93.

[75] Rowley, J. 2007. The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science* 33, 2 (Apr. 2007), 163–180. DOI: https://doi.org/10.1177/0165551506070706

[76] Rubin, A. 2020. Learning to reason with data: How did we get here and what do we know? *Journal of the Learning Sciences* 29, 1 (2020), 154–164.

[77] J. Saldaña. 2016. *The Coding Manual for Qualitative Researchers*. SAGE.

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

[79] Schwab-McCoy, A. Baker, C. M, and Gasper, R. E. 2021. Data science in 2020: Computing, curricula, and challenges for the next 10 years. *Journal of Statistics and Data Science Education* 29, sup1 (Jan. 2021), S40–S50. DOI: https://doi.org/10.1080/10691898.2020.1851159

[80] Computer Science Teachers Association. 2017. *CSTA K-12 Computer Science Standards*. Retrieved January 02, 2025 from https://csteachers.org/k12standards/

[81] Stornaiuolo, A. 2020. Authoring data stories in a media makerspace: Adolescents developing critical data literacies. *Journal of the Learning Sciences* 29, 1 (Jan. 2020), 81–103. DOI: https://doi.org/10.1080/10508406.2019.1689365

[82] Sukol, S. 2024. *Beyond Borders 2024: Primary and Secondary Data Science Education around the World. Data Science 4 Everyone*

[83] Sukol, S. 2024. *State of the Field: Data Science and Data Literacy Education in US K-12*. *Data Science 4 Everyone*

[84] A. Tifentale and L. Manovich. 2015. Selfiecity: Exploring photography and self-fashioning in social media. In *Postdigital Aesthetics: Art, Computation and Design*. D.M. Berry and M. Dieter (Eds.), Palgrave Macmillan UK, 109–122.

[85] D. Touretzky, C. Gardner-McCune, F. Martin, and D. Seehorn. 2019. Envisioning AI for k-12: What should every child know about AI? In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, 9795–9799. DOI: https://doi.org/10.1609/aaai.v33i01.33019795

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

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

[88] Weintrop, D. Beheshti, E. Horn, M. Orton, K. Jona, K. Trouille, L, and Wilensky, U. 2016. Defining computational thinking for mathematics and science classrooms. *Journal of Science Education and Technology* 25, 1 (Feb. 2016), 127–147. DOI: https://doi.org/10.1007/s10956-015-9581-5

[89] Wilkerson, M. H. Lanouette, K, and Shareff, R. L. 2022. Exploring variability during data preparation: a way to connect data, chance, and context when working with complex public datasets. *Mathematical Thinking and Learning* 24, 4 (Oct. 2022), 312–330. DOI: https://doi.org/10.1080/10986065.2021.1922838

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

[91] Wise, A. F. 2020. Educating data scientists and data literate citizens for a new generation of data. *Journal of the Learning Sciences* 29, 1 (Jan. 2020), 165–181. DOI: https://doi.org/10.1080/10508406.2019.1705678

[92] J. Yip, T. Clegg, E. Bonsignore, H. Gelderblom, E. Rhodes, and A. Druin. 2013. Brownies or bags-of-stuff? Domain expertise in cooperative inquiry with children. In *Proceedings of the 12th International Conference on Interaction Design and Children*, 201–210.

Received 23 August 2024; revised 15 September 2025; accepted 9 October 2025