datAR: A Situated Learning Approach for Data Literacy Through Everyday Objects

Lilian Lopez
ETH Zurich
Zurich, Switzerland
llopez@ethz.ch

Zeyu Xiong
ETH Zurich
Zurich, Switzerland
zeyu.xiong@inf.ethz.ch

Kiara Chau
ETH Zurich
Zurich, Switzerland
kchaugarcia@student.ethz.ch

Gustavo Umbelino
University of Zurich
Zurich, Switzerland
gustavo.kreiaumbelino@zda.uzh.ch

Zihan Wu
University of Michigan
Ann Arbor, USA
ziwu@umich.edu

April Wang
ETH Zurich
Zurich, Switzerland
april.wang@inf.ethz.ch

Abstract

Data literacy, the ability to work with data, is essential for the younger generation. However, high school instructors often struggle to engage students from diverse backgrounds with abstract concepts that may not seem immediately tangible to them. This paper introduces datAR, an augmented reality application grounded in situated learning theory that integrates data into everyday objects. The tablet-based application uses tangible analysis blocks to help students explore and interact with data. We present a case study for the use of datAR in a local high school, where 15 students with little to no prior data science experience were introduced to data literacy through the analysis of nutrition information in familiar snacks. Our findings show that, despite students struggling to recognize the relevance of the learned concepts for their daily lives, datAR helped them develop data science skills through hands-on interaction with familiar objects.



Figure 1: datAR, a tablet-based application that uses augmented reality and tangible analysis blocks to help students explore and interact with data.

CCS Concepts

• **Human-centered computing** → **Interactive systems and tools**; • **Applied computing** → **Interactive learning environments**.

Keywords

Data Literacy, Situated Learning, Augmented Reality

ACM Reference Format:
Lilian Lopez, Zeyu Xiong, Kiara Chau, Gustavo Umbelino, Zihan Wu, and April Wang. 2025. datAR: A Situated Learning Approach for Data Literacy Through Everyday Objects. In *Proceedings of the 30th ACM Conference on Innovation and Technology in Computer Science Education V. 1 (ITiCSE 2025), June 27-July 2, 2025, Nijmegen, Netherlands.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3724363.3729047

This work is licensed under a Creative Commons Attribution 4.0 International License.
ITiCSE 2025, Nijmegen, Netherlands
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1567-9/2025/06
https://doi.org/10.1145/3724363.3729047

1 Introduction

Data literacy refers to the ability to collect, analyze, and interpret data, as well as understand its broader societal and ethical implications [12, 33]. Given the increasing role of data across industries and daily life, ensuring that youth develop data-related skills has become a global educational priority, as demonstrated by various national initiatives (e.g., US [7], Switzerland [27], and UK [40]).

However, teaching data science to non-technical students, who do not envision careers in data science or informatics, remains challenging for instructors. Common ways of learning data science require students to understand the syntax of coding in order to manipulate data [17, 36], which can be alleviated through low-code or no-code approaches [5, 35]. However, the fundamental challenge that instructors face when teaching students goes beyond programming; it relates to the lack of motivation, relevance, and engagement. We conducted formative interviews with six high school informatics instructors and found that one of the key challenges in teaching data literacy to non-technical students is maintaining students' engagement and interest. This is especially true for high school students, who are on the cusp of adulthood and independently embracing a digitally driven age [32]. Because students have diverse backgrounds in computational thinking and varied career plans, they demonstrate different levels of engagement with the course material. While using datasets relevant to students' learning

and lives (e.g., observing and measuring insect data in a biology context) can improve interest, data still requires interpretation through numbers on the computer, posing a cognitive barrier for many.

To help instructors engage non-technical students in learning data literacy, we propose a system design informed by *situated learning theory*. We refer to situated learning for data literacy as an instructional method that integrates data learning with physical objects, social contexts, and environments. This approach has the potential to bridge the gap between abstract data concepts and students' lived experiences, making learning more intuitive and relatable. In particular, tangible interactions, like linking data points to physical objects to make abstract concepts concrete, may enhance engagement and comprehension across learning styles. By making data literacy more tangible and experiential, instructors may effectively address issues of abstraction and engagement, making data literacy education more inclusive and effective. Building on this concept, Augmented Reality (AR) offers a compelling opportunity for implementing situated learning methods, as it allows students to interact with data in a physical space without the need for complex setups [16, 39]. By integrating digital information into the physical environment, AR may make abstract data concepts more accessible for non-technical students.

To this end, we designed *datAR*, a tablet-based AR application that allows students to scan everyday objects and attach data tags, facilitating data collection in a familiar context. We conducted a classroom deployment of datAR with 15 non-technical high school students, instructing them to complete data visualization and data analysis tasks using the datAR application during a 110-minute lecture. We reflected on the effectiveness of datAR in enhancing students' understanding and engagement with data literacy concepts through observations and discussions with students and instructors.

We found that students enjoyed learning with datAR in a collaborative setting. Preliminary evidence suggests they developed data literacy skills, although it is too early to confirm any learning gains. Some findings also indicated that students applied critical thinking while solving tasks with datAR. However, despite the learning situated around nutrition information on snack bars – a topic personally relevant to students – they struggled to see how the data concepts they learned could be applied to their everyday context. We discuss specific improvements for designing instructional activities that better connect data objects and concepts with meaningful, relatable, and situated activities for students.

This work makes two key contributions to the field of data literacy education: (1) The designed and developed datAR, a tablet-based AR application that allows students to learn data science by interacting with and analyzing data derived from everyday objects in their environment; (2) Preliminary insights on the benefits and challenges of this approach through a classroom deployment study in a high school informatics lecture.

2 Related Work

2.1 Fostering Data Literacy for Students

In our data-driven society, everyone must possess the knowledge and skills necessary to be data literate [14, 15]. *Data literacy* refers to not only the ability to collect, analyze, and visualize data but also the capacity to derive meaningful insights from it [10, 29]. As data becomes increasingly central to decision-making, research shows that teaching data literacy is most effective when students can connect it to their own experiences, especially using real-world examples and datasets that relate to students' daily lives, which help them better understand complex data concepts [21]. Students who work with familiar data are more likely to engage with the material and develop lasting analytical skills that they can apply in various situations [6].

Various educational tools have been developed to support different aspects of data literacy learning while emphasizing hands-on experience. Platforms like Callysto integrate hardware sensors with Python programming for data collection and analysis, allowing students to collect and analyze real-world environmental data [24]. Similarly, SensorBricks provides tangible components for gathering environmental data, offering an immersive learning experience [4]. Tools such as Orange employ block-based visual programming, allowing students to create interactive visualizations by connecting various analysis components [35]. This approach significantly reduces the technical barriers typically associated with data analysis tools. Furthermore, Co-ML offers a collaborative learning environment where students and their families can work together to build image classifiers, providing practical experience with data labeling and model training [34]. Many of these tools incorporate real-world datasets focusing on relevant societal issues such as environmental change [30], social media [14] and public health [25, 37], helping students understand how data literacy can address contemporary challenges while developing their technical competencies. Importantly, these educational approaches often integrate ethical considerations, teaching students to critically evaluate data privacy, bias, and responsible data use [2].

2.2 Integrating AR with Situated Learning Theory for STEM

Situated learning introduces the idea that knowledge occurs in the "real world" and is highly linked to specific content and situations [13, 19]. Furthermore, knowledge is integrated into the activity and environment, as well as social interactions. That is not to say that learning in an abstract classroom setting is inefficient. As John R. et al. show, abstract teaching instructions can be effective, and the degree to which knowledge is transferred depends on the presentation and the experimental situation [1]. Situated learning theories (SLT) imply immersive, hands-on education tools that contextualize the learning content and allow room for collaborative learning. datAR integrates physical objects from everyday context into classroom activities, creating a memorable link to student's lived experiences. Furthermore, datAR encourages peer collaboration, tying back to the notion that learning should happen in a social environment in situated learning theory. Collaborative environments have been linked to improved learning outcomes [9, 23], as students benefit from sharing diverse perspectives and problem-solving strategies, fostering better retention and understanding. Using augmented reality (AR) technology is a natural choice. AR allows us to contextualize information by overlaying real-world objects with the learning content. Smørdal et al. demonstrated this concept by designing a mobile AR application that allows students to explore a speculative scenario that takes place in their own physical location

but in an alternative timeline, a so-called situated simulation [31]. AR can also increase students' motivation [8, 24], particularly for students from non-technical backgrounds [22]. For instance, novice computer science students found AR visualizations more engaging and effective for learning data structures than paper-based or desktop applications [26].

While education tools grounded in SLT and supported by AR already exist for language learning [20], there is still limited work on creating AR educational tools grounded in SLT for *science learning* [18, 38]. datAR aims to contribute towards filling this gap by applying situated learning principles and contextualizing data science-related activities through familiar physical objects.

3 Formative Study

To inform our system design, we interviewed six high school CS instructors to understand the current practices and challenges of teaching data-related lectures, and identified the following themes with regard to current practices of teaching data literacy. Their classroom sizes ranged from 20 to 25 students.

Tools. The tools employed by instructors primarily centered around programming languages such as Python and SQL. Instructors reported using a variety of web-based IDEs to minimize the setup costs.Three instructors also mentioned using features such as automatic viewing of student solutions through platforms like JupyterHub, compatible with both laptops and tablets.

Competencies. Instructors emphasized several key areas in a data literacy curriculum to ensure a comprehensive understanding and practical application of data skills, including data collection, management, security, analysis, and interpretation:

(1) Interpreting from Data (n=6). Students need to be adept at interpreting data from everyday contexts such as newspapers, statistical figures, visual charts, and understanding the relationship between data.
(2) Statistical Analysis (n=6). Students should be able to import CSV files and perform basic statistical analyses using Excel or Python, including calculating the median, mean, and standard deviations.
(3) Data Collection (n=3). It is important to "discuss the differences in data collection for scientific [versus] non-scientific purposes" (P1). Students should engage in personal data projects where they collect their own data to make the learning process more engaging.
(4) Proper Data Management (n=3). Students need to be able to handle large data sets, and understand how to organize, clean, and classify data.
(5) Security (n=3). Participants reported teaching students about cybersecurity, data protection, and information security.

Challenges. Most (5/6) teachers found that teaching students with different prior knowledge in programming made it difficult for a one-size-fits-all model. Many (5/6) also worry about student engagement when the content was not graded or the grades did not matter to them, or because the concept was too difficult. Some (3/6) found creating course materials challenging, because it was difficult to balance creating original content and using existing resources, and to keep up with emerging technologies.



Figure 2: A student using datAR. An information layer is attached to a snack bar, and students can create pie charts visualizing the nutrition data of the snack.

4 System Design

Building on the formative study, we introduce datAR a tablet-based data analysis application that uses augmented reality and tangible analysis blocks to help learners explore real-world data embedded in everyday objects.

4.1 Gathering and Mapping Data

datAR provides three different ways for data collection. First, the student can simply import any `csv` file provided by their teacher or prepared by themselves. Second, the student can manually enter a new data point to an existing data table, providing the information for each data attribute. Lastly, the student can choose to scan any product bar code to automatically extract data from a physical product. The last two options for creating new data points are possible by directly tapping on a physical object through the camera view, scanning the bar code for automatic data extraction, or manually entering all the data attributes for the corresponding object. Once the student has successfully mapped a new data point to its respective physical object, a semi-transparent window overlays the AR view, displaying the data in a standard plain text format (see Figure 2). For automatic data extraction datAR uses the crowd-sourced Open Food Facts database[1] to find information based on the bar code of any common product found at a super market.

4.2 Visualization and Basic Analysis Blocks

A crucial element of the data analysis pipeline is the ability to graphically visualize entire data sets, helping students understand the operations performed on their data [28]. The physical blocks provided together with our tool enable students to perform data table operations as well as visualize the modified table.

Based on widely used data visualization types, we implemented visualization blocks for: Pie charts, bar charts, and scatter plots. The pie chart block serves to visualize an individual data point and its various data attributes, for example, the percentage of protein, carbohydrates, fats and salt in an energy bar; see Figure 2. On the other hand, the bar chart and scatter plot blocks are used when visualizing an entire data table, such as the price for various electronic

[1] https://world.openfoodfacts.org/

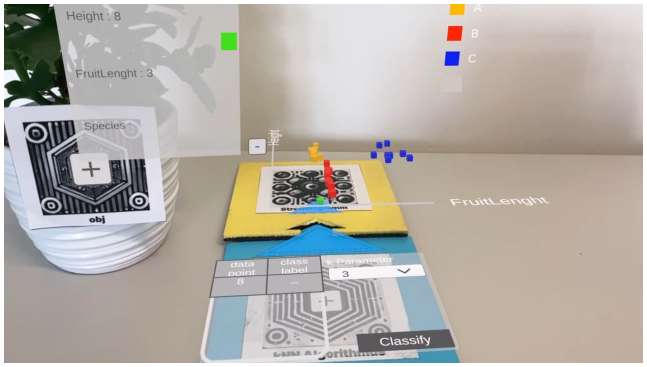


Figure 3: A screenshot from the datAR application. In the left corner Alice's plant with the attached data label. The blue block is a custom analysis block, which executes the k-nearest neighbor algorithm. The yellow block visualizes plant height and leave length from a plant collection. The species of each data point is color coded.

devices. In order to modify a data table students can use different basic analysis blocks. We implemented the ORDER BY block and WHERE block, inspired by the SQL language. In Figure 1 the students used the ORDER BY block to sort the data table containing various food items by the amount of carbohydrates. Furthermore, they used a scatter plot block to visualize the sorted values and color-coded the data points based on the nutri-score of each item.

4.3 Supporting Custom Analytical Blocks

To help instructors integrate our tool into their curricula, we provide them with the capability to design their own custom operation blocks. This allows them to introduce more advanced topics like machine learning to their students. We created a web platform that allows them to browse through a variety of blocks created by other educators, select the ones that best fit their needs, and seamlessly integrate them into their own classrooms. On the desktop interface, instructors can create custom analysis pieces by providing a Python code snippet and generating an AR marker for printing.

4.4 Example Use Case

Imagine Alice who wants to find out to which species her new houseplant belongs to. For that, she first measures the plants height and leave length, she can then attach the measured data attributes to her plant in the AR view; see Figure 3. As a next step, she loads a data file into the application, provided by her biology teacher and containing measurements from different known plant species. In addition, her computer science teacher also provided her with a custom analysis block, which executes the k-nearest neighbor algorithm. Activating the block in the datAR application and choosing the correct training data as well as the label to classify by, allows Alice to determine the species for her own plant.

4.5 Implementation

We used the Unity game engine with the AR foundation plugin[2] to render 3D models and animations. We used iPadPro, as it also comes with a `LiDAR` scanner – a technology developed specifically for mixed-realty applications. Our backend connects to the instructor's interface through the `FastAPI` service. We adopted a `RESTful API` to communicate between the front-end and the `MySQL` database.

5 Case Study: Classroom Deployment

To explore how AR education tools grounded in situated learning theory help with teaching non-technical students' data literacy, we conducted a classroom deployment of datAR with a local high school in Europe. The school places a strong emphasis on the liberal arts. We deployed the study during an informatics lecture at a local high school, which typically covers topics such as ethical implications of digital technologies, creative media usage, foundational programming, and computer technologies. We designed a data science lecture that used datAR to cover data concepts like sorting, visualization, and k-nearest neighbors (`kNN`) classification. The case study was approved by the ethics board of our institution.

5.1 Study Protocol

The 110-minute lecture, delivered by the first author, includes:

Part 1: A 10-minute introduction explained data concepts and basic visualization methods (e.g., pie and bar charts), followed by a 10-minute warm-up exercise introducing datAR's functionalities. The students were divided into six groups (G1-G6) to solve exercises on **sheet #1** and **sheet #2**.

Part 2: Each group received physical visualization and analysis blocks, an iPad, and an energy bar snack. Tasks on **sheet #1** included **(a)** extracting data from the snack to an existing table, **(b)** filtering data table entries, **(c)** adding a new data attribute, **(d)** sorting the data, and **(e)** visualizing the sorted data.

Part 3: After the break, students learned the kNN algorithm through a 10-minute explanation and an example scenario and received **sheet #2**. Tasks on **sheet #2** required students to: **(a)** train kNN to classify different energy bars as "good" or "bad" based on nutrition data, **(b)** visualize the dataset in a scatter plot, and finally **(c)** classify their own snack using different k values.

Part 4: To wrap up, students had 20 minutes to individually complete a quiz assessing their understanding of data science concepts and a questionnaire on their subjective feelings during the lecture.

5.2 Data Collection and Analysis

We collected screen recordings with audio of the iPads per group, the quiz answers and questionnaire responses for each individual student. We also took observational notes. In addition, we invited the original instructor of the course to observe the session and conducted a 20-minute post-session interview with them to collect their feedback. The screen recordings were analyzed by two authors fluent in the local language to collect students' quotes and track task completion of the worksheets per group. We employed a deductive data analysis approach to identify themes [3].

[2]https://docs.unity3d.com/Packages/com.unity.xr.arfoundation@6.0

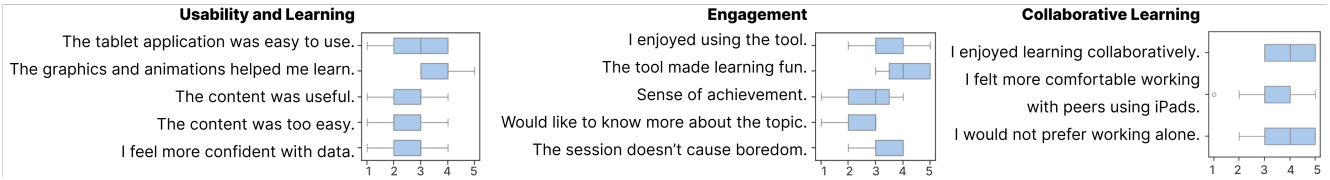


Figure 4: 5-Point Likert Scale Survey Results, regarding *usability and learning, engagement,* and *collaborative learning.*

5.3 Findings

Most groups successfully completed **sheet #1** (a) (b) (c) (e) (data extraction, selection, addition, and visualization) and **sheet #2** (a) (training kNN). However, for tasks **#1** (d) (sorting data) and **#2** (b) (c) (creating scatter plot and using different k values in kNN), groups skipped certain sub-tasks on the worksheets, and as a result needed increased assistance and provided incomplete or incorrect answers. The quiz, conducted on an individual basis, consisted of five questions asking for different skills, with seven out of 15 (47%) students getting four or more correct answers. In the following section, we further break down the learning process of students.

5.3.1 Learning data science skills with datAR. All six groups (100%) were able to collect data from their energy bars and answered the control question from **#1** (a) correctly.

For the data analysis task **#1** (b), four out of six groups (67%) successfully executed filtering or sorting operations to identify snacks with a fat content greater than 35g. Later on, in the quiz, 11 out of 15 (73%) students chose the correct option regarding sorting and filtering operations, demonstrating a strong grasp of these concepts. When asked to add a new data attribute showing the price per gram in sheet **#1** (c), four out of six groups (67%) succeeded. G4 and G6 did not provide any formula for calculating the new attribute, resulting in the addition of a new table column with only zero value entries.

All groups successfully created at least one meaningful data visualization using the provided blocks (i.e., pie chart, bar chart, scatter plot) required for three different sub-tasks in both sheets. While examining the visualization blocks, one student made thoughtful considerations when choosing a data visualization: "*An appropriate diagram? Yeah, I would say bars.*" (G2). Furthermore, G5 was able to discuss potential trends in the visualization "*so the most expensive ones are better than the cheapest ones*". However, G2 and G4 struggled to understand the bar graph settings needed for task **#1** (e) and did not select the correct data attributes to visualize.

Half of the groups (50%) selected the correct settings for training the kNN algorithm in task **#2** (a). However, due to time constraints and fatigue, none of the groups were able to solve task **#2** (c), which focused on the k parameter. This is also reflected in the individual assessment. In the quiz, 12 out of 15 students (80%) correctly classified a new data point presented in a scatter plot using the kNN algorithm. In contrast, only eight students (53%) correctly answered the question, "How does changing the k-value (the number of nearest neighbors) affect the prediction in the kNN algorithm?"

5.3.2 datAR Made Learning Fun. Eleven students (73%) rated the survey question "The tool made learning fun" with a grade 5 (strongly agree) or 4 (agree). Three students (20%) self-reported that the session was fun because it was different from the regular class: "*It was great. Something different than usual in class.*" Only two (13%) students agreed with a rating of four to the survey question "*This session caused me to become bored at times.*" From the screen recording, researchers also found that students would sound proud when the output of the application would change based on their actions. It appeared that bigger visual changes from the application elicited stronger reactions from the students. When successfully loading the AR table, students would make exclamations like, "*Guys, I have it! The table.*" (G6) or when adding a new data attribute to the table, "*What the heck! M and M cost only 0.6 per gram!*" (G5). These reactions showed that students were engaged during the session.

5.3.3 Learning with Peers. Most students benefited from collaborating with their peers for solving the tasks. We found that 10 out of 15 students (66%) rated the survey question "I enjoyed using the learning object collaboratively with my peers" as a 4 (agree) or a 5 (strongly agree). These findings serve as evidence that most of the students saw value in interacting with their peers to complete the assigned task with datAR. The discussions recorded while groups were solving the task with datAR were often meaningful and contributed towards the group completing a certain task. For example, while coming up with a suitable filtering condition, a group member questioned "*Why are you doing smaller and not greater?*" (G4) or while looking for the information to enter: A-"*Where can I read this*" B-"*Ah here. The weight is 20*" (G1).

5.3.4 Challenges from the study: What did not work. Although students successfully solved tasks using datAR, we noticed some things that did not work out well.

First, it was difficult for students to recognize the relevance of the learned data science skills. For the survey question, "The content of this session was useful to me" only 3 of the students (20%) agreed with a rating of 4 (as shown in Figure 4). Two students (13%) explicitly stated that even though they enjoyed the session, they were not interested in the topic of data science"*I think the project itself is good, the explanations were good but it is not something for me, personally.*" This strongly suggests that, although the students were familiar with the snack dataset, they were unable to translate the data science skills to future career scenarios or everyday contexts, as quoted from G3 while solving the tasks with the tool "*Who would use that in their free time? Yes, go look at a diagram like this.*"

In addition, while solving the tasks in groups using datAR, some students frequently expressed confusion about the task at hand. The question "*What do we have to do?*" could be heard in all the screen recordings. Moreover, when the AR overlay disappeared, students often asked "*Where did the table go?*" (G6), regularly spending quite

a bit of time moving the iPad around to find missing elements. We conclude that these observations, along with the lack of status information in the application, contributed to students experiencing frustration. This explains conversations captured from the screen recordings, such as "*We are fed up*" (G3), which was uttered shortly after "*I don't get the task*" (G3).

6 Discussion

In this section, we reflect on what aspects of the situated learning activity were successful and which were less effective. To summarize our findings, students enjoyed using datAR in a collaborative setting, with early indications of developing data literacy skills. We hope this discussion inspires further exploration of situated data literacy learning within the Computer Science Education (CSEd) and Human-Computer Interaction (HCI) communities, focusing on both education tools and the design of learning activities.

6.1 Social and Physical Context

Our classroom deployment demonstrated that datAR can effectively promote peer collaboration, which plays an important role in improving students' engagement with the class. In our case study, we observed that students were more engaged when each played a distinct role in the analysis process. Using tablets, which one student needed to hold, along with physical objects and tangible analytical blocks that another student manipulated, naturally fostered distributed collaboration, allowing team members to take on different responsibilities. Compared to laptops, iPad devices are easy to carry around, allowing peers to share their screens through physical interactions. This setup led to meaningful discussions as students worked together on data science tasks. Schools can benefit from this model as it fosters an inclusive learning environment in which students actively collaborate, reinforcing the social aspect of learning. Future work can further explore other collaborative learning activities. For instance, an "escape room" themed around data literacy and datAR could require students to solve data-related challenges to "escape" the room, thereby deepening their understanding through an immersive, problem-solving experience.

6.2 Making Programming Accessible

As data science skills rely on the ability of abstract thinking, it is crucial to provide novice learners with an appropriate translation into more accessible actions. The visual language and tangible analytical blocks in our system have demonstrated potential in making complex data explorations more approachable and comprehensible. However, datAR still requires evaluation through a comparative study to draw more definitive conclusions about its effectiveness. In addition, datAR allows instructors to scaffold the learning process by customizing the granularity of the analytical blocks to suit their students' needs. This design provides instructors the autonomy to tailor the tool for their classroom, such as adding distracting blocks to design a tangible Parsons problems [11]. However, this level of autonomy requires instructors to have comprehensive knowledge of data concepts and the ability to program the functionality of the analytical blocks – a challenge highlighted in our formative study. Our future work includes expanding the block-sharing platform, enabling instructors to create customized blocks quickly, and fostering a community where instructors can share and reuse materials.

6.3 Recognizing the Relevance

In developing our prototype, we aimed to harness data from everyday objects, making the interaction with data more memorable for students. During our user study, snacks were used as a primary example, demonstrating the prototype's ability to adapt to everyday items. However, while food nutrition is a relevant subject in everyday life, students did not see how the data concepts taught can benefit their future careers or daily lives. This could be explained by the fact that the classroom session failed to clearly state the benefits of data science skills during the lecturing portion and explicitly connect them to real-life scenarios. Furthermore, the hands-on learning activities did not reflect data literacy benefits sufficiently. As a result, students did not connect with the real-world application of visualizing and analyzing nutrition values. Future work should investigate whether this gap can be addressed by offering different learning scenarios, such as having students imagine themselves as school administrators selecting snacks for vending machines while balancing nutrition, taste, and budget. Alternatively, this gap might be bridged by contextualizing data within different domains, such as biology, personal health, or liberal arts.

6.4 Limitations

As our implementation of datAR is still in the early stages, it featured only a preliminary user interfaces. It lacked debugging information and detailed feedback, such as hints for creating visualizations or analysis. This limitation may have contributed to students quickly losing motivation and being more receptive to distractions.

Our evaluation of datAR was mainly focused on the students' perspectives and experiences, and did not take into account the interface design for instructors, such as creating new custom analysis blocks. Furthermore, as confirmed by their teacher, the participating students may not be fully representative as the class was quite heterogeneous, mixing students with different high school levels into a single one. The study also lacked a direct comparison with existing approaches, limiting the scope of the findings.

7 Conclusion

In this paper, we explored situated learning as a method for engaging non-technical students in data literacy. Through the design and deployment of datAR, a tablet-based AR application, we demonstrated how integrating physical objects with digital data representations can make abstract data concepts more tangible and accessible. Our classroom study with high school students revealed that situated learning approaches hold promise for enhancing student engagement and collaboration. Although we observed early indications of critical thinking and data literacy skills development, challenges remain in helping students connect the application of data concepts to their daily lives.

Acknowledgments

We thank the Dieter Schwarz Stiftung Foundation for generously funding this research.

References

[1] John Anderson, Lynne Reder, and Herbert Simon. 1996. Situated Learning and Education. *Educational Researcher* 25 (05 1996), 5–11. https://doi.org/10.2307/1176775

[2] Javiera Atenas, Leo Havemann, Caroline Kuhn, and Cristian Timmermann. 2023. Critical data literacy in higher education: teaching and research for data ethics and justice. In *Data Cultures in Higher Education: Emergent Practices and the Challenge Ahead*. Springer, 293–311.

[3] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis.* American Psychological Association.

[4] Hans Brombacher, Rosa Van Koningsbruggen, Steven Vos, and Steven Houben. 2024. SensorBricks: a Collaborative Tangible Sensor Toolkit to Support the Development of Data Literacy. In *ACM International Conference Proceeding Series*. Association for Computing Machinery. https://doi.org/10.1145/3623509.3633378

[5] Ruijia Cheng, Aayushi Dangol, Frances Marie Tabio Ello, Lingyu Wang, and Sayamindu Dasgupta. 2023. Concepts, practices, and perspectives for developing computational data literacy: Insights from workshops with a new data programming system. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (Chicago, IL, USA) *(IDC '23)*. Association for Computing Machinery, New York, NY, USA, 100–111. https://doi.org/10.1145/3585088.3589364

[6] Aayushi Dangol and Sayamindu Dasgupta. 2023. Constructionist approaches to critical data literacy: A review. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (Chicago, IL, USA) *(IDC '23)*. Association for Computing Machinery, New York, NY, USA, 112–123. https://doi.org/10.1145/3585088.3589367

[7] Erica Deahl. 2014. Better the data you know: Developing youth data literacy in schools and informal learning environments. *Available at SSRN 2445621* (2014).

[8] Chris Dede. 2014. Augmented Reality Teaching and Learning. *Handbook of Research on Educational Communications and Technology* (01 2014), 735–745. https://doi.org/10.1007/978-1-4614-3185-5_59

[9] Elise Deitrick, R. Benjamin Shapiro, Matthew P. Ahrens, Rebecca Fiebrink, Paul D. Lehrman, and Saad Farooq. 2015. Using Distributed Cognition Theory to Analyze Collaborative Computer Science Learning. In *Proceedings of the Eleventh Annual International Conference on International Computing Education Research* (Omaha, Nebraska, USA) *(ICER '15)*. Association for Computing Machinery, New York, NY, USA, 51–60. https://doi.org/10.1145/2787622.2787715

[10] Catherine D'Ignazio and Rahul Bhargava. 2016. DataBasic: Design Principles, Tools and Activities for Data Literacy Learners. *The Journal of Community Informatics* 12, 3 (2016), 83–107. Special Issue on Data Literacy.

[11] Barbara Ericson, Janice L. Pearce, Susan H. Rodger, Andrew Csizmadia, Rita Garcia, Francisco J. Gutierrez, Konstantinos Liaskos, Aadarsh Padiyath, Michael James Scott, David H. Smith, Jayakrishnan Madathil Warriem, and Angela Zavaleta Bernuy. 2023. Multi-Institutional Multi-National Studies of Parsons Problems. In *Proceedings of the 2023 Working Group Reports on Innovation and Technology in Computer Science Education*. Association for Computing Machinery, United States, 57–107. https://doi.org/10.1145/3623762.3633498

[12] Mark Frank, Johanna Walker, Judie Attard, and Alan Tygel. 2016. Data Literacy-What is it and how can we make it happen? *The Journal of Community Informatics* 12, 3 (2016).

[13] James G. Greeno. 2011. *Theories of Learning and Studies of Instructional Practice*. Springer New York. 41–71 pages. https://doi.org/10.1007/978-1-4419-7582-9_3

[14] Anna Hartl, Elena Starke, Angelina Voggenreiter, Doris Holzberger, Tilman Michaeli, and Jürgen Pfeffer. 2024. Empowering Digital Natives: InstaClone - A Novel Approach to Data Literacy Education in the Age of Social Media. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1* (Portland, OR, USA) *(SIGCSE 2024)*. Association for Computing Machinery, New York, NY, USA, 484–490. https://doi.org/10.1145/3626252.3630839

[15] Fiona M. Hollands, Daniella DiPaola, Cynthia Breazeal, and Safinah Ali. 2024. AI Mastery May Not Be For Everyone, But AI Literacy Should Be. In *Proceedings of the 2024 on ACM Virtual Global Computing Education Conference V. 1* (Virtual Event, NC, USA) *(SIGCSE Virtual 2024)*. Association for Computing Machinery, New York, NY, USA, 60–66. https://doi.org/10.1145/3649165.3690117

[16] Sara Invitto, Italo Spada, and Lucio Tommaso De Paolis. 2015. Augmented reality, embodied cognition and learning. In *Augmented and Virtual Reality: Second International Conference, AVR 2015, Lecce, Italy, August 31-September 3, 2015, Proceedings 2*. Springer, 125–134.

[17] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, et al. 2016. Jupyter Notebooks–a publishing format for reproducible computational workflows. In *Positioning and power in academic publishing: Players, agents and agendas*. IOS press, 87–90.

[18] Abhishek Kulkarni, Cecelia Albright, and Sharon Lynn Chu. 2024. Investigating Situated Learning Theory Through an Augmented Reality Mobile Assistant for Everyday STEM Learning. 955–956. https://doi.org/10.1109/VRW62533.2024.00274

[19] Jean Lave and Etienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation.* Cambridge University Press.

[20] Jaewook Lee, Sieun Kim, Minji Park, Catherine L Rasgaitis, and Jon E. Froehlich. 2024. Embodied AR Language Learning Through Everyday Object Interactions: A Demonstration of EARLL. In *Adjunct Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST Adjunct '24)*. Association for Computing Machinery, New York, NY, USA, Article 52, 3 pages. https://doi.org/10.1145/3672539.3686746

[21] Zipora Libman. 2010. Integrating real-life data analysis in teaching descriptive statistics: A constructivist approach. *Journal of Statistics Education* 18, 1 (2010).

[22] Shi Liu, Thimo Schulz, Peyman Toreini, and Alexander Maedche. 2023. An Immersive Learning Factory for AI & Data Literacy: An Exploratory Study in the Wild. In *ACM International Conference Proceeding Series*. Association for Computing Machinery, 349–353. https://doi.org/10.1145/3603555.3608564

[23] Owen HT Lu, Jeff CH Huang, Anna YQ Huang, and Stephen JH Yang. 2018. Applying learning analytics for improving students engagement and learning outcomes in an MOOCs enabled collaborative programming course. In *Learning analytics*. Routledge, 78–92.

[24] Aditya Mehrotra, Christian Giang, Noé Duruz, Julien Dedelley, Andrea Mussati, Melissa Skweres, and Francesco Mondada. 2020. Introducing a Paper-Based Programming Language for Computing Education in Classrooms. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education* (Trondheim, Norway) *(ITiCSE '20)*. Association for Computing Machinery, New York, NY, USA, 180–186. https://doi.org/10.1145/3341525.3387402

[25] Stephen J Mooney and Vikas Pejaver. 2018. Big data in public health: terminology, machine learning, and privacy. *Annual review of public health* 39, 1 (2018), 95–112.

[26] Husnu S. Narman, Cameron Berry, Alex Canfield, Logan Carpenter, Jeremy Giese, Neil Loftus, and Isabella Schrader. 2020. Augmented Reality for Teaching Data Structures in Computer Science. In *2020 IEEE Global Humanitarian Technology Conference, GHTC 2020*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/GHTC46280.2020.9342932

[27] Swiss Academies of Arts and Sciences. 2024. Swiss Data Literacy Charter. https://doi.org/10.5281/zenodo.11146023

[28] Pareena Padwal, Yashpreet Singh, Jeetesh Singh, and Suvarna Pansambal. 2021. DVAR: Data Visualization using Augmented Reality. In *2021 2nd Global Conference for Advancement in Technology (GCAT)*. 1–6. https://doi.org/10.1109/GCAT52182.2021.9587831

[29] Chantel Ridsdale, James Rothwell, Mike Smit, Hossam Ali-Hassan, Michael Bliemel, Dean Irvine, Daniel Kelley, Stan Matwin, and Brad Wuetherick. 2015. *Strategies and Best Practices for Data Literacy Education: Knowledge Synthesis Report.* Technical Report. Dalhousie University.

[30] Sheikh Ahmad Shah, Jaai Uday Phatak, Avneet Hira, Helen Zhang, and Mike Barnet. 2024. Smart Greenhouse: Bridging Physical Computing, Plant Science, and Data Literacy. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 2* (Portland, OR, USA) *(SIGCSE 2024)*. Association for Computing Machinery, New York, NY, USA, 1928. https://doi.org/10.1145/3626253.3635433

[31] Ole Smørdal, Gunnar Liestøl, and Ola Erstad. 2016. Exploring situated knowledge building using mobile augmented reality. *QWERTY–Open and Interdisciplinary Journal of Technology, Culture and Education* 11, 1 (2016), 26–43.

[32] Shashank Srikant and Varun Aggarwal. 2017. Introducing data science to school kids. In *Proceedings of the 2017 ACM SIGCSE technical symposium on computer science education*. 561–566.

[33] Matti Tedre and Henriikka Vartiainen. 2023. K-12 Computing Education for the AI Era: From Data Literacy to Data Agency. In *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1* (Turku, Finland) *(ITiCSE 2023)*. Association for Computing Machinery, New York, NY, USA, 1–2. https://doi.org/10.1145/3587102.3593796

[34] Tiffany Tseng, Jennifer King Chen, Mona Abdelrahman, Mary Beth Kery, Fred Hohman, Adriana Hilliard, and R. Benjamin Shapiro. 2023. Collaborative Machine Learning Model Building with Families Using Co-ML. In *Proceedings of IDC 2023 - 22nd Annual ACM Interaction Design and Children Conference: Rediscovering Childhood*. Association for Computing Machinery, Inc, 40–51. https://doi.org/10.1145/3585088.3589356

[35] University of Ljubljana, Bioinformatics Laboratory. 2024. *Orange: Data Mining Fruitful and Fun.* https://orangedatamining.com Accessed: 2024-07-31.

[36] Gregory Vial and Bogdan Negoita. 2018. Teaching programming to non-programmers: the case of Python and jupyter notebooks. (2018).

[37] Lance A Waller and Carol A Gotway. 2004. *Applied spatial statistics for public health data.* John Wiley & Sons.

[38] Yun Wen and Chee-Kit Looi. 2019. *Review of Augmented Reality in Education: Situated Learning with Digital and Non-digital Resources*. 179–193. https://doi.org/10.1007/978-981-13-8265-9_9

[39] Margaret Wilson. 2002. Six views of embodied cognition. *Psychonomic bulletin & review* 9 (2002), 625–636.

[40] Annika Wolff, Jose J Cavero Montaner, and Gerd Kortuem. 2016. Urban Data in the primary classroom: bringing data literacy to the UK curriculum. *The Journal of Community Informatics* 12, 3 (2016).