Accepted manuscript, to appear in: Proceedings of the ACM Global Computing Education Conference 2025 Vol 1 (CompEd 2025)

# The Data Case Study – A Teaching and Learning Method for Computer Science Education in Schools

Viktoriya Olari
viktoriya.olari@fu-berlin.de
Freie Universität Berlin
Berlin, Germany

Ralf Romeike
ralf.romeike@fu-berlin.de
Freie Universität Berlin
Berlin, Germany

## Abstract

With artificial intelligence (AI) topics entering computer science school curricula, there is a growing need for instructional approaches that support the teaching of data-related concepts and practices. In academic AI and data science programs, an established teaching and learning method is the data case study. However, despite its widespread adoption, reports on its didactic development in school settings are lacking. This article presents findings from a research project that theoretically and empirically identified the challenges of implementing the data case study in schools. Based on these findings, the method was further developed for secondary education. The result is a specification of the data case study as an action-oriented, learner-centered approach aligned with the goals of AI education in schools and designed to promote data-based judgment and problem-solving skills.

## CCS Concepts

• **Computing methodologies → Artificial intelligence**; • **Social and professional topics → K-12 education**; *Computing education*; • **General and reference** → *Empirical studies.*

## Keywords

Case-Based Teaching; Case Study Approach; Pedagogy; K–12; Secondary School Education; Computer Science Education; Data Science Education; AI Literacy; Data Literacy; Design-Based Research

**ACM Reference Format:**
Viktoriya Olari and Ralf Romeike. 2025. The Data Case Study – A Teaching and Learning Method for Computer Science Education in Schools. In *Proceedings of the ACM Global Computing Education Conference 2025 Vol 1 (CompEd 2025), October 21–25, 2025, Gaborone, Botswana.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3736181.3747164

## 1 Introduction

The data case study is a learner-centered, action-oriented instructional method widely used in academic education on artificial intelligence (AI) and data science [14, 30, 57].[1] University students work on data cases by applying theoretical knowledge of data concepts and practices to real-world datasets in authentic scenarios, thereby developing data-based judgment and problem-solving skills [19].

[1]Wright et al. provide illustrative examples (e.g., "Predicting Annual Air Pollution") in [57].


*CompEd 2025, Gaborone, Botswana*

ACM ISBN 979-8-4007-1929-5/2025/10
https://doi.org/10.1145/3736181.3747164

This process helps students internalize the fundamental role of data in AI systems, especially those that employ machine learning (ML) methods. Similarly, school education on AI prioritizes cultivating a fundamental understanding of data and strengthening students' ability to make data-based decisions and solve problems. According to the authors' national educational standards for secondary school education for computer science (CS), students should, for example, be able to describe the influence of the training data used when applying a ML method [11]. Given these learning objectives, the question arises whether the data case study can be fruitfully adapted as an instructional method for teaching AI in secondary school CS education.

The case study is already established as a teaching and learning method for fostering problem-solving skills in CS lessons [15]. However, the data case study represents a more complex form of learning, and its use in schools is likely to pose challenges that require careful consideration. For example, academic data cases demand extensive programming skills [30] and are time-consuming to prepare [19]. In secondary schools, however, many students lack the programming skills needed to work with data, and teachers have limited time for lesson planning.

To balance its potential with school-specific challenges, we adapted the data case study for use in secondary school CS education following the design-based research approach. In the process, we addressed two research questions:

**RQ1:** What challenges should be considered to fruitfully implement the data case study in CS lessons on the topic of AI?

**RQ2:** How can the data case study be adapted to address the identified challenges and meet the objectives of CS education on the topic of AI?

In the next sections, we outline the theoretical background, the methodology and the results of the research process. This work makes a twofold contribution: In addition to the theory-based and empirical adaptation of the data case study as a teaching and learning method, we provide examples of data cases for CS education that were tested under real school conditions.

## 2 Theoretical Background

Data is a central topic in CS education in schools (see, for instance, [4]) and is becoming even more relevant with the integration of AI content into CS curricula. Data is a fundamental component of AI systems, especially those built with ML methods [59]. To design and utilize such systems responsibly—competencies that are considered essential for K-12 education on AI [22, 35, 48]—an advanced understanding of data practices such as data collection, cleaning, feature engineering, and data splitting [32] and fundamental data concepts such as data-based task, data noise, or data bias [31] is necessary.

The most common approach to teaching ML topics is reportedly data-driven [25], i.e., it glassboxes the content related to creating and labeling datasets to train ML models while blackboxing the details of the ML algorithms; nevertheless, a recent study found that the role of data in current teaching approaches has largely been underestimated [34]. For instance, teaching approaches rarely address data cleaning, optimization, or visualization practices. At the same time, scholars emphasize that engaging with data can create substantial learning opportunities for students to understand AI technologies and develop a greater ownership for learning [41, 47, 60]. Studies that teach data concepts and practices in the context of AI education report on successful experiments with unplugged approaches [54] and project-based learning in textual [2] and graphical environments [9, 49].

To inform the design of learning arrangements for AI education on data concepts and practices, prior research on data literacy offers valuable insights [20, p. 5]. Besides providing strong examples for instructional materials (see, for instance, *Bootstrap World* platform [38]), these studies inform about the difficulties of teaching data-related topics. For instance, it is reported that generating and maintaining motivation and engagement is challenging when teaching data literacy and data science topics to school students [21, 43]. The reason is debated—besides the inappropriately selected context [26], researchers argue that the topic is too complex, students lack programming knowledge, and time constraints in schools make mastering the subject difficult [21]. To overcome these challenges, researchers recommend grounding learning arrangements in situated learning theory, an instructional method that integrates learning with physical objects, social contexts, and environments [21]. Others suggest employing an active learning approach based on the constructivist learning theory [10]. Positive effects have also been reported from teaching with real-life data following the constructionist approach [40], using low-code or no-code environments [21] to enable low-floor access for all students, and providing hands-on and module-based approaches [43].

In academic data science education, where students' backgrounds and prior knowledge are also heterogeneous [45, p. 44], an established teaching and learning method grounded in the tradition of situated, active, and constructivist learning is the data case study (also known as a "case study" or "lab") [3, 8, 12, 16, 30, 57]. At its core, it presents students with a specific *data case*—a description of a realistic, problematic situation accompanied by a dataset and an example of how to approach it [3, 30, 57]. Students' work on a data case is then structured in various ways. Some educators use step-by-step instructions [57], while others prefer open tasks and exercises [16, 19].

Preparing a data case requires significant initial investment in time and effort, and working with it in academic contexts requires prior knowledge of statistics and computing [14, p. 11,]. Because data cases are realistic situations without ideal, clear-cut solutions [3] and because the students need to absorb subject matter knowledge, statistical concepts, and computing skills simultaneously, teaching with data case study is challenging [14, p. 6]. Nevertheless, learning with data cases reportedly has positive effects on students' interest, motivation, engagement, knowledge, and skill acquisition [12, 16, 19, 51].

Reports on the didactic development of the data case study for CS school education are lacking. However, teaching and learning with the case study is known [58], although not as widely as in other science subjects such as chemistry [1] or biology [13]. For instance, case study is used to teach databases [37]. Since the data case study represents a more complex form of learning, involving challenges evident from data science education, the question arises of how the data case study can be adapted to enrich school education on AI.

## 3 Methodology

Enriching CS school education on AI with the data case study requires balancing potential advantages and challenges. To support this aim, we formed an expert panel comprising two teachers, both of whom teach Mathematics and one of whom teaches CS, a domain expert, and two researchers in CS education. The team met regularly over a period of two years and followed the design-based research approach after Prediger as the methodological framework [39] to adapt the data case study. The research process consisted of four closely interlinked phases, which we performed in three cycles: (1) specifying and structuring the learning objective, (2) (further) developing the design, (3) conducting and evaluating design experiments, and (4) (further) developing local theories.

**(1) What should students learn?** The first phase involved specifying the subject matter to be learned. To identify data concepts and practices relevant to teaching AI, we analyzed the subject literature and AI curricula for schools [31, 32]. From the resulting collections of concepts and practices, we selected those for the first research cycle that were also covered in academic data cases (e.g., *preparing data* [57]), further narrowing the selection in the second cycle.

**(2) How should students learn?** In the second phase, we iteratively developed data cases, resulting in over 20 data cases after two years of work (see examples[2] in Table 1). While some requirements for the design—the importance of cognitive activation of students in the design of learning-effective activities [18] or the need to address motivational challenges (see Section 2)—could be derived from the research, we expected unknown difficulties during implementation. To identify and tackle these at an early stage, each data case was developed in four steps: (1) discussion of potential contexts and challenges among the team of experts, (2) identification of suitable contexts and datasets and development of a proposal for the data case by CS education researchers, (3) technical validation of the proposal by the domain expert, and (4) didactic validation of the proposal by the teachers.

Potential contexts were drawn from various sources: the domain expert in the expert group; academic data cases [57]; scientific publications about AI in the domain literature, such as the overview provided by Rolnick et al. [44]; popular science literature (e.g., Rex et al. [42]); teaching materials on data science school education, such as those provided by the *Day of AI* program [23]; online resources on AI for professionals, such as online courses [5], tutorials [50] or books for AI practitioners [24]. The datasets for data cases

[2]The data cases are being gradually published in a digital repository [33].

came from local institutions such as the local air quality monitoring network [46], databases for scientific data such as PANGAEA [36], National Oceanic and Atmospheric Administration [28], National Snow and Ice Data Center [29]; research institutions such as GEOMAR Helmholtz Centre for Ocean Research; and platforms specialized in datasets for ML, such as the UCI Machine Learning Repository [52]. In some data cases, the students collected their own data or worked with the datasets collected by us. The *architecture* for the data cases—which is the way of structuring the work on the data case—was initially derived from academic data cases. Before implementation in CS lessons, we transformed it based on the challenges and requirements, as explained in Section 4.1.

**(3) How successful is the implementation?** The third phase involved implementing the data case study under real school conditions and identifying further challenges. In this phase, two teachers from the expert group led a three-month enrichment course that integrated CS and mathematics, using data cases. The course took place weekly during regular school hours in a secondary school and covered topics related to ML with a strong focus on data practices and concepts. To evaluate the implementation, one CS researcher and the domain expert participated in the lessons. The study was coordinated with and approved by the Senate Department for Education, Youth, and Family Affairs Berlin and the ethics committee of the Freie Universität Berlin. Before starting the research, the school director, teachers, students, and their guardians were given comprehensive information about the study to obtain informed consent for participation.

To investigate challenges and conditions for success, we collected data on a wide range of direct and indirect indicators, as proposed by Van Den Akker [53] for research projects focused on development. First, we measured knowledge among students in weekly pre- and post-tests, which consisted of single, multiple-choice, and open-ended questions (e.g., "Briefly describe the creation and evaluation of a ML model in chronological order. Which steps are carried out in which order?"). Moreover, we collected data on students' intrinsic motivation after working on a data case with a questionnaire on perceived competency, pleasure, freedom of choice, and pressure, each consisting of four items [55]. For each item (e.g., "During class, I was able to approach the work in the way I wanted to."), the students selected an answer on a five-point Likert scale (1 – not at all true, 2 – slightly true, 3 – somewhat true, 4 – mostly true, 5 – completely true).

Second, CS education researchers observed and videotaped the lessons. After each lesson, we also conducted semi-structured interviews with teachers. For instance, we asked teachers "What difficulties did students encounter with teaching and learning arrangement, and how did they overcome them?" (question B5)[3]. To identify areas of difficulty for students, we collected student artifacts such as worksheets every week.

To understand and address the difficulties and success factors (RQ1), we qualitatively analyzed student artifacts and teacher interviews. The artifacts were coded inductively by a researcher in the MAXQDA software, focusing on mistakes, and subsequently summarized as lists of common difficulties. The interviews with teachers were transcribed. Then, a CS education researcher identified text passages in interview transcripts related to questions on difficulty and success factors and coded them inductively. The following categories of difficulties emerged while coding: subject-specific challenges (problems with understanding the *data flow* concept—i.e., the complete path of the data from the first to the last computing unit (called "Widget" in Orange3)—, debugging strategy, technical nature), pedagogical challenges (problems with learning speed, learning process, learning control), and domain-specific challenges (articulation and interpretation of the result). The CS researchers synthesized the results and discussed the findings weekly within the expert group to ensure validity. We also analyzed the data from the questionnaires on students' motivation and knowledge gains to determine the efficacy of learning with data cases.

**(4) Which factors are essential for learning?** In the fourth phase, at the end of each cycle, we retrospectively reflected within the expert panel on challenges and the effectiveness of the data case design in supporting student learning. Based on these reflections, we made decisions about how to revise the data cases for the next cycle. The reflections were audio-recorded, transcribed, and analyzed by a CS education researcher with a focus on joint decisions made regarding the design. Through reflection at the end of the second and particularly the third cycle, after experimenting with several approaches—and in accordance with the philosophy of the design-based research, which postulates that discovery occurs through change [7, p. 145]—, we were able to formulate the school-specific elements of the data-case architectures and their effects on learning (RQ2). The result is a specification of the data case study as a teaching and learning method that meets the challenges and goals of CS education on the topic of AI in secondary schools.

## 4 Results

This section provides an overview of the study participants and explains the key challenges and coping strategies in the research development process (Q1). Subsequently, we present insights on the effectiveness of the data case study and provide its specification as a teaching and learning method for CS school education (Q2).

*Participants.* Forty-four students in Grades 9 and 10 participated in the study: 15 students in the first course *(cycle 1: f = 5, m = 8; not specified = 2; average age = 14.46 years)*; 13 students in the second course *(cycle 2: f = 4, m = 7; not specified = 2; average age = 14.7 years)*; 16 students in the third course *(cycle 3: f = 4, m = 10, not specified = 2; average age = 14.72 years)*. Each course began with four introductory lessons on the context, continued with lessons guided by the data case study method, included a two-day field trip to the GEOMAR Helmholtz Centre for Ocean Research – where the domain expert works – and concluded with a project assignment. The first and second courses each lasted 10 weeks (48 lessons of 45 minutes per course). The third course lasted 9 weeks and comprised 46 lessons. In total, students completed 6, 9, and 11 data cases in the first, second, and third courses, respectively.

### 4.1 Known and New Challenges

Identifying and overcoming challenges was an iterative process based on theory and practice. Once known challenges were overcome, new, more detailed ones arose. The following outlines the

[3]The interview protocol is available upon request from the first author.

**Table 1: Overview of exemplary data cases in three different architectures. Each data case starts with a description of a problematic situation, providing context and a question. The data cases are available through the digital repository [33].**

| Context and Question | Dataset and Data Source | Architecture | Learning Artifact(s) | Data Concepts (C) and Practices (P) |
|---|---|---|---|---|
| **Biodiversity:** *What minimum legal size should be defined for harvesting abalones?* | Dataset on abalones with their physical measurements such as rings, length etc. [27] | Bottom-up: Follow the step-by-step instructions with inquiry tasks after each step. | Paper: Completed workbook with answered inquiry tasks; Digital: a data flow in Orange3 | C: Statistic measurements (mean, median, mode); P: identify and eliminate anomalies; interpret statistics; transform features |
| **Air Pollution:** *How has air pollution in Berlin developed over the past 15 years?* | Dataset on air pollution in Berlin (1993–2025) with measurements of particulate matter, ozone levels etc. [46] | Top-down: Start with a prepared data flow and work through interpretation and reflection tasks. | Paper: Written report on the results of the data exploration and data flow structure | C: Trend analysis, scatter plot, line plot, data flow stages; P: interpret plots; describe data flow phases |
| **Glacier Melt:** *When will the Schneeferner glacier disappear?* | Dataset on the total area and height of the Schneeferner glacier (1892–2023) [56] | Puzzle-like: Start with a set of puzzle pieces and work backwards to determine the underlying data flow and widget configurations. | Paper: Poster with the reconstructed and annotated data flow including widget configuration details | C: Data flow, data object, regression; P: reconstruct the data flow; examine and verbalize stages in the data flow |

synthesized results on difficulties and strategies for overcoming them when adapting the data case study for schools.

**1. Subject-specific challenges.** As an action-oriented method, the data case study requires extensive programming skills, particularly in data manipulation, visualization, and modeling. In CS school education, we expected students to have little to no prior competence in these areas. Our expectation was confirmed in first lessons. The students had heterogeneous programming skills and little experience in working with data, such as in spreadsheet programs. To overcome this challenge and address initially known difficulties such as frustration potential when working with data known from prior research, we developed highly guided bottom-up data cases for the intuitive, flow-based environment of Orange3 [6] in the first cycle. Each data case started with an empty workspace in Orange3 (Figure 1 left; data case on biodiversity in Table 1) and guided students step by step through data practices in context. The evaluation showed that the bottom-up architecture successfully addressed the initial challenges. All students were able to create data flows according to instructions. However, the necessary overemphasis on the operation of Orange3 in the first cycle caused more important skills to be neglected, such as reflection on the process and the transition to the abstract level.

Therefore, in the second cycle, we developed a set of data cases with a top-down architecture in close alignment with the methodical approach used in an academic context (Figure 1 center; data case on air pollution in Table 1). Each data case began with a prepared data flow, and included a series of interpretation and reflection tasks. With this, we replaced the step-by-step instructions with a criteria-based reflection on the process. The weekly work results and final projects revealed that all students were able to understand, interpret, and modify prepared data flows and successfully create their own. However, two new challenges arose: (1) Some students had difficulty understanding the data flow concept. After the fourth data case in the second course, the teacher noted: *"They also actively used the help cards for the widgets. [...] But, for example, one group did not understand how the connection [...] between the widgets works, that the data flows from one widget to the next" (T2W5* [4]). (2) Additionally, students lacked debugging strategies: *"They also had problems with doing this data discovery [...]. They lack debugging strategies" (T2W5)* .

Both challenges were successfully addressed by explicitly teaching the data flow concept and debugging strategies in the third course using puzzle-like data cases. In this architecture, students started with several puzzle pieces (comprising widgets, tables, and templates for configuring the widgets) for a given context and worked backwards to determine the underlying data flow and reconstruct widget configurations (Figure 1 on the right, data case on glacier melt in Table 1). During the reconstruction process, students intuitively discovered elements that are essential for debugging data flows (such as the amount of data entering and leaving a widget).

**2. Subject-specific pedagogical challenges.** In addition to different prerequisites and subject-specific difficulties, the data case study must address various educational goals. Teacher interviews conducted at the beginning of the first cycle showed that meaningful communication among students is a central educational means and goal. Hence, the data case study as a teaching and learning method must offer appropriate opportunities. In academic education, however, data cases are often designed for individual work, allowing little exchange [57]. To address this challenge, various partner and group activities were integrated into the work on data cases in the first and second cycles (e.g., jigsaw teaching technique). These measures were partially successful. In partner work, we often observed exchange between students on the subject matter; group work, however, was less effective. Teachers criticized that data cases offer few opportunities for discussion: *"Where I really need to say*

[4] Read T2W5 as Cycle 2, Week 5.

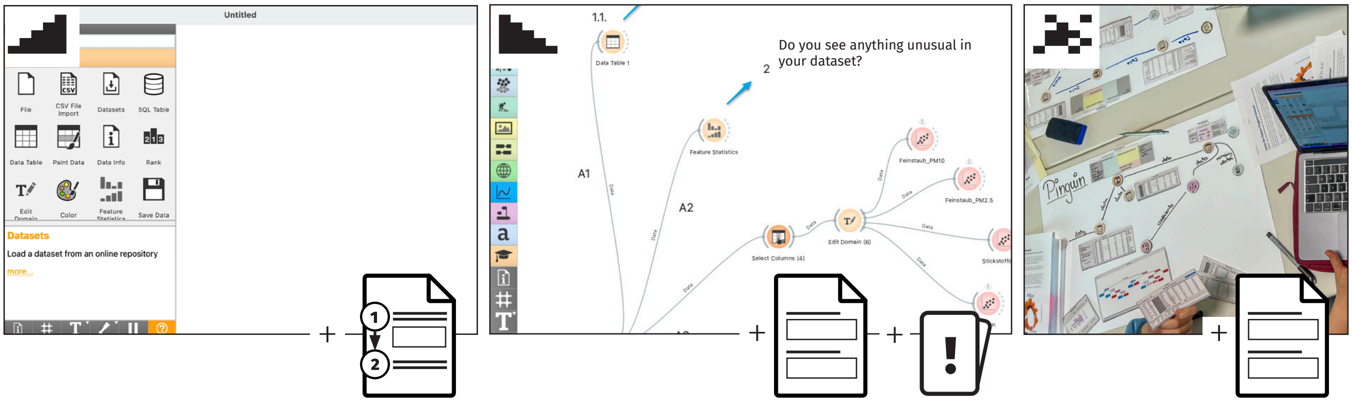


**Figure 1: Architecture of a data case: bottom-up ( ), top-down ( ), puzzle-like ( ).**

*something to someone else, where I really (...) get into a real conversation" (T1W3)*. In addition, students worked at different speeds, and some often had to wait for their team members. In-depth tasks proved unsuitable as a solution, as it was not the higher-performing students who were faster, but often the lower-performing peers who worked superficially. Through the development of a puzzle-like architecture in which students had to solve the task together, meaningful communication was successfully implemented in the third cycle.

**3. Domain-specific challenges.** To assess the plausibility of a dataset (e.g., in the context of data bias) or to interpret the results in context, domain knowledge is required. In academic education, students receive information material to familiarize themselves with the domain [30]. However, teachers and students lack the time to do so. One solution is to choose contexts in which teachers and students have prior knowledge. We chose the topic "environment," which is covered in earlier grades and is personally and socially relevant. However, it was observed that students' prior knowledge was often insufficient to evaluate results in context. Although this challenge was anticipated and addressed through informational material, its use in the first cycle proved only moderately successful: *"This idea that I have to incorporate knowledge from previous tasks into the new task is difficult for me as a student" (T1W3)*.

In the second cycle, the interpretation tasks were supplemented with explicit references to the information material (e.g., *"Compare the result with your intuition. If you don't have any intuition, read the information sheet on air pollution. State whether you think the prediction is plausible."*). Additionally, we created interpretation cards with criteria for describing the results. These measures were helpful, with one teacher summarizing: *"They [the students] were able to acquire the prior knowledge via the information cards and then apply it" (T2W5)*.

**4. Design challenges.** Some teaching approaches can be developed by teachers, while others require specialized expertise. One example is *Leitprogramm* — a workbook with 50–100 pages that requires professional expertise, didactic imagination, linguistic talent, and time to create [17]. Data cases are similar.

The central element of a data case is a prepared scenario with an associated dataset. Research on data science education has shown that searching for a dataset and preparing it is a time-consuming task [14]. We confirm this for secondary level education. Although we limited the context to a subject area for which many datasets are freely available and for which domain experts could assess the plausibility of the scenarios and the suitability of the datasets, the development effort per data case still ranged from 40 to 120 hours. The search for suitable datasets—well documented, comprehensible to secondary school students, and meaningful and technically viable for the data concept or data practice to be taught—was the most demanding task. Close cooperation between subject didactics, domain experts, and teachers is therefore essential for the development of new data cases. The actual implementation of finished data cases in the classroom was found to be manageable by the teachers.

## 4.2 Specification of the Data Case Study as a Teaching and Learning Method

In addition to the difficulties described above, other school-related challenges were successfully addressed during the process, including hardware limitations for complex data analysis and limited homework time. In this section, we present insights into the evaluation of the effectiveness of the data case study as a teaching and learning method, followed by its school-specific characterization resulting from the research process.

*4.2.1 Effectiveness.* In all three cycles, we observed that the students were dedicated to work on the data cases. According to the teachers, the students showed endurance and engagement: "*Well, once again, I had the impression that everyone was actually working and participating" (T3W2int2)*. The observations from the lessons and the findings from the teacher interviews are reflected in the weekly quantitative data on motivation. For instance, in T3W2, students (n=12) found the learning arrangement mostly interesting (*mean perceived pleasure = 3.73, SD = 0.59*). They were fairly satisfied with their performance (*mean perceived performance = 3.91, SD = 0.73*), somewhat agreed that they were able to choose from activities (*mean perceived freedom of choice = 3.24, SD = 0.86*), and mostly did not feel under pressure (*mean perceived pressure = 2.03, SD = 0.66*).

We noticed positive differences in the pre- and post-knowledge tests, indicating the knowledge gains among students for all three architectures. For example, in T2W2, where 11 students worked on a top-down data case, the mean score increased from *4.45 (SD = 1.57)* in the pre-test to *6.72 (SD = 1.48)* in the post-test, with *8* being a maximum score. The result of the paired *t*-test was statistically significant $t(10) = -4.65$ ($p < 0.0008$); *Cohen's d: -1.48*). By evaluating artifacts, we observed that students were able to examine data flow, identify and articulate its steps, and interpret the results (see Section 4.1). They could also modify the data flows and recognize errors in them. In the project assignments at the end of each course, most students could create their own data flows and answer questions on a given context, though with some errors and scaffolds.

*4.2.2 Specification of the data case study for school education.* The data case study is an action-oriented, learner-centered teaching and learning method. At its core is the work on a data case – a prepared scenario with an associated dataset. Key components of a school-specific data case are: (1) the description of a problematic situation from a context relevant to students; (2) a documented, real-life, tailored, and, if necessary, didactically enriched dataset to which selected data concepts and practices can be applied in a meaningful and technically correct manner; (3) a complete or partially developed data flow with tasks for describing and justifying the results and tasks for modifying and expanding the data flow, including ideas for further experiments; (4) a collection of domain information explicitly embedded in the tasks; (5) information material on the data mining tool and the interpretation of the results; (6) multi-level scaffolding material, learning controls, and exercises. A data case also includes (7) expected solutions for teachers with notes on problems in the dataset, prerequisites on the students' knowledge of the data mining tool, data concepts and practices; time requirements; and recommendations on social forms.

In schools, the data case is methodically implemented in three architectures, as presented in Section 4.1: top-down, bottom-up, and puzzle-like. The bottom-up and top-down architectures have some similarities to the methodical implementation of data cases in an academic context (e.g., [57]). A combination of top-down and puzzle-like architectures addresses many school-specific challenges.

In the bottom-up architecture, students build a data flow step by step and complete context-related tasks after each step. The learning artifacts are the data flow and the written analysis of the findings. This approach is particularly beneficial for high-achieving and interested students as well as teachers, as it offers a guided introduction to a specific topic (e.g., "exploring data") and is appropriate for individual work.

In the top-down architecture, students receive a complete data flow and are instructed to trace selected data practices and concepts in prepared scenarios and evaluate them according to criteria. Students work in three steps: (1) describe and interpret the results of the data flow and compare them with their intuition (in the context of Orange3, this is the work with leaf widgets); (2) examine intermediate steps and reflect on their significance in the data flow (in Orange3, this involves examining the preparation widgets that lead to leaf widgets); and (3) identify phases in the data flow. The learning artifact is a written report that presents findings from the data flow and describes its phases. The top-down architecture is particularly suitable when the goal is to critically reflect on the data flow and its phases.

In the puzzle-like data case, students work together without computers to reconstruct the data flow from given elements. This architecture is suitable for learning groups with heterogeneous prior knowledge and is particularly effective when teaching debugging skills required for the independent development of data flows or practicing meaningful communication and cooperative learning. The learning artifact is the data flow assembled from parts on a poster. This architecture can be expanded by using computers to check the results of the reconstruction.

Teaching and learning with data cases requires sufficient time. Depending on the architecture and complexity of a data case, the planned duration should be between one and four 45-minute lessons. Lessons in which data case study is used for learning are usually divided into the following phases: (1) preparation of the working environment; (2) familiarization with the data case and the objective; (3) meaningful individual/partner/group work on the data flow, supported by supplementary materials; (4) subject-specific consolidation of findings (ideally students independently compare their answers with solution cards, as this allows them to learn at their own pace); and (5) optional: reflection on the steps in the data flow.

## 5 Conclusion

In the research process, we investigated the difficulties of implementing the data case study in secondary school CS education on the topic of AI (RQ1). Known challenges—such as heterogeneous programming and data analysis skills, limited domain knowledge among students and teachers, and low motivation during data work—were addressed. However, when implemented in school practice, new difficulties emerged, including a lack of debugging skills, varying learning speeds, and limited opportunities for meaningful communication. In response to the difficulties, we developed school-specific architectures for the data cases (RQ2). The results indicate that the data case study is a suitable instructional method for secondary school CS education. Its three architectures (top-down, bottom-up, and puzzle-like) enable students to understand, evaluate, and apply AI-relevant data concepts and practices in real and authentic contexts. Thus, the use of this method in CS lessons on AI can enhance data-based judgment and problem-solving skills. Future research could broaden the focus beyond environmental contexts to other domains shaped by AI technologies and investigate the implementation of teaching and learning with data cases in larger student cohorts.

## Acknowledgments

We would like to thank the school students who participated in this study, as well as the teachers and the domain expert from the GEOMAR Helmholtz Centre for Ocean Research Kiel, who actively shaped the development of the data case study as a learning and teaching method. We are also grateful to the anonymous reviewers for their helpful comments. Parts of this paper were published in German at a local practitioner-oriented conference aimed at sharing approaches with CS teachers. We used DeepL Translate, DeepL Write, and ChatGPT to support the translation of German text and language refinement in this paper. We reviewed and revised all content thoroughly and take full responsibility for its final form.

## References


[1] Flávia M. Bernardi and Maurícius S. Pazinato. 2022. The Case Study Method in Chemistry Teaching: A Systematic Review. *Journal of Chemical Education* 99, 3 (March 2022), 1211–1219. doi:10.1021/acs.jchemed.1c00733

[2] Rolf Biehler and Yannik Fleischer. 2021. Introducing Students to Machine Learning with Decision Trees Using CODAP and Jupyter Notebooks. *Teaching Statistics* 43 (2021), S133–S142.

[3] Valentina Chkoniya. 2021. Success Factors for Using Case Method in Teaching Applied Data Science Education. *European Journal of Education* 4, 1 (April 2021), 77–86. doi:10.26417/236hbm84v

[4] Computer Science Teachers Association. 2017. CSTA K–12 CS Standards. (2017).

[5] DeepLearning.AI (instructor: Robert Monarch). 2025. AI for Good Specialization | DeepLearning.AI on Coursera. https://www.coursera.org/specializations/ai-for-good.

[6] Janez Demšar, Tomaž Curk, Aleš Erjavec, Črt Gorup, Tomaž Hočevar, Mitar Milutinovič, Martin Možina, Matija Polajnar, Marko Toplak, Anže Starič, Miha Štajdohar, Lan Umek, Lan Žagar, Jure Žbontar, Marinka Žitnik, and Blaž Zupan. 2013. Orange: Data Mining Toolbox in Python. *Journal of Machine Learning Research* 14, 71 (2013), 2349–2353.

[7] Matthew W. Easterday, Daniel G. Rees Lewis, and Elizabeth M. Gerber. 2018. The Logic of Design Research. *Learning: Research and Practice* 4, 2 (July 2018), 131–160. doi:10.1080/23735082.2017.1286367

[8] T. Grandon Gill. 2011. *Informing with the Case Method: A Guide to Case Method Research, Writing, & Facilitation.* Informing Science Press, Santa Rosa.

[9] Andreas Grillenberger and Ralf Romeike. 2019. About Classes and Trees: Introducing Secondary School Students to Aspects of Data Mining. In *Informatics in Schools. New Ideas in School Informatics*, Sergei N. Pozdniakov and Valentina Dagien\.e (Eds.). Springer International Publishing, Cham, 147–158.

[10] Orit Hazzan and Koby Mike. 2024. *Guide to Teaching Data Science: An Interdisciplinary Approach.* Springer, Cham.

[11] Lutz Hellmig, Steffen Burk, André Greubel, Martin Hennecke, Henry Herper, André Hilbig, Tilman Michaeli, Alexander Mittag, Arno Pasternak, Hermann Puhlmann, Gerhard Röhner, Michael Rücker, Thomas Schmalfeldt, Wolf Spalteholz, and Peer Stechert. 2025. Bildungsstandards Informatik für die Sekundarstufe I – Empfehlungen der Gesellschaft für Informatik e. V. (GI). (2025). doi:10.18420/REC2025_052

[12] Clyde Freeman Herreid. 2011. Case Study Teaching. *New Directions for Teaching and Learning* 2011, 128 (Dec. 2011), 31–40. doi:10.1002/tl.466

[13] Clyde F. Herreid, Nancy A. Schiller, and Ky F. Herreid (Eds.). 2014. *Science Stories You Can Count on: 51 Case Studies with Quantitative Reasoning in Biology.* NSTA Press, National Science Teachers Association, Arlington, Virginia.

[14] Stephanie C. Hicks and Rafael A. Irizarry. 2018. A Guide to Teaching Data Science. *The American Statistician* 72, 4 (Oct. 2018), 382–391. doi:10.1080/00031305.2017.1356747

[15] Thomas B. Hilburn and Massood Towhidnejad. 2007. A Case for Software Engineering. In *20th Conference on Software Engineering Education & Training (CSEET'07)*. IEEE, Dublin, Ireland, 107–114. doi:10.1109/CSEET.2007.5

[16] Erin Rae Hoffer. 2020. Case-Based Teaching: Using Stories for Engagement and Inclusion. 2, 2 (2020), 75–80.

[17] Urs Kirchgraber, Werner Hartmann, M. Bettinaglio, M. Adelmeyer, J.P. David, and K. Frey. 1997. Und Dann Und Wann Ein Leitprogramm! *Zentralblatt für Didaktik der Mathematik* 29, 6 (1997), 199–206.

[18] Eckhard Klieme. 2020. Unterrichtsqualität. In *Handbuch Schulpädagogik*. Waxmann, 394–409.

[19] Jana Lasser, Debsankha Manik, Alexander Silbersdorff, Benjamin Säfken, and Thomas Kneib. 2021. Introductory Data Science across Disciplines, Using Python, Case Studies, and Industry Consulting Projects. *Teaching Statistics* 43, S1 (July 2021). doi:10.1111/test.12243

[20] Duri Long and Brian Magerko. 2020. What Is AI Literacy? Competencies and Design Considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3313831.3376727

[21] 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)*. Association for Computing Machinery, Nijmegen, Netherlands and New York, NY, USA, 152–158. doi:10.1145/3724363.3729047

[22] Tilman Michaeli, Ralf Romeike, and Stefan Seegerer. 2022. What Students Can Learn About Artificial Intelligence – Recommendations for K-12 Computing Education. In *Towards a Collaborative Society Through Creative Learning*, Keane, Therese, Lewin, Cathy, Brinda, Torsten, and Bottino, Rosa (Eds.). Springer Nature Switzerland, Cham, 196–208. doi:10.1007/978-3-031-43393-1_19

[23] MIT App Inventor Team. 2025. IceMelt: Modeling and Making Predictions with Environmental Data – App Inventor. https://appinventor.mit.edu/explore/ai2/icemelt.

[24] Christoph Molnar. 2025. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable (3rd Ed.) – About the Book. https://christophm.github.io/interpretable-ml-book/#about-the-book.

[25] Luis Morales-Navarro and Yasmin B Kafai. 2024. Unpacking Approaches to Learning and Teaching Machine Learning in K-12 Education: Transparency, Ethics, and Design Activities. In *Proceedings of the 19th WiPSCE Conference on Primary and Secondary Computing Education Research.* ACM, Munich Germany, 1–10. doi:10.1145/3677619.3678117

[26] Stephan Napierala. 2020. The Road to Finding Interesting Contexts for Teaching Data Literacy at School. In *Proceedings of the 15th Workshop on Primary and Secondary Computing Education.* ACM, Virtual Event Germany, 1–2. doi:10.1145/3421590.3421620

[27] Sellers Nash, Warwick and Wes Ford. 1994. Abalone. doi:10.24432/C55C7W

[28] National Centers for Environmental Information (NCEI). 2025. Datasets Search | National Centers for Environmental Information (NCEI). https://www.ncei.noaa.gov/access/search/dataset-search.

[29] National Snow and Ice Data Center. 2025. Explore Data | National Snow and Ice Data Center. https://nsidc.org/data/explore-data.

[30] Deborah Ann Nolan and Terry Speed. 2001. *Stat Labs: Mathematical Statistics through Applications* (corr. 2. print ed.). Springer, New York Berlin Heidelberg.

[31] Viktoriya Olari and Ralf Romeike. 2024. Data-Related Concepts for Artificial Intelligence Education in K-12. *Computers and Education Open* 7 (Dec. 2024), 100196. doi:10.1016/j.caeo.2024.100196

[32] Viktoriya Olari and Ralf Romeike. 2024. Data-Related Practices for Creating Artificial Intelligence Systems in K-12. In *Proceedings of the 19th WiPSCE Conference on Primary and Secondary Computing Education Research.* Association for Computing Machinery, Munich, Germany, Article 5.

[33] Viktoriya Olari and Ralf Romeike (Eds.). 2025. *Data Cases: Data case study for computer science education on the topic of data and artificial intelligence.* Freie Universität Berlin, Computing Education Research Group. https://computingeducation.de/proj-datacases

[34] Viktoriya Olari, Kamilla Tenório, and Ralf Romeike. 2023. Introducing Artificial Intelligence Literacy in Schools: A Review of Competence Areas, Pedagogical Approaches, Contexts and Formats. In *Towards a Collaborative Society Through Creative Learning*, Therese Keane, Cathy Lewin, Torsten Brinda, and Rosa Bottino (Eds.). Vol. 685. Springer Nature Switzerland, Cham, 221–232. doi:10.1007/978-3-031-43393-1_21

[35] Organization for Economic Co-operation and Development (OECD). 2025. *Empowering Learners for the Age of AI: An AI Literacy Framework for Primary and Secondary Education (Review Draft).* OECD, Paris.

[36] PANGAEA. 2025. PANGAEA - Data Publisher for Earth & Environmental Science. https://www.pangaea.de.

[37] Penon, Johann and Spolwig, Siegfried. 1994. Video-Center – Eine Fallstudie Zur Einführung in Relationale Datenbanksysteme. LOG IN, 2, 32–35. *LOG IN* 2 (1994), 32–35.

[38] Daniel R Pimentel, Nicholas J Horton, and Michelle Hoda Wilkerson. 2022. Tools to Support Data Analysis and Data Science in K-12 Education. *National Academies* (2022).

[39] Susanne Prediger, Michael Link, Renate Hinz, Stephan Hussmann, Bernd Ralle, and Jörg Thiele. 2012. Lehr-Lernprozesse initiieren und erforschen. Fachdidaktische Entwicklungsforschung im Dortmunder Modell. *Der mathematische und naturwissenschaftliche Unterricht* 65, 8 (2012), 452–457.

[40] Prerna Ravi, Robert Parks, John Masla, Hal Abelson, and Cynthia Breazeal. 2024. "Data Comes from the Real World": A Constructionist Approach to Mainstreaming K12 Data Science Education. In *Proceedings of the 2024 on ACM Virtual Global Computing Education Conference V. 1.* ACM, Virtual Event NC USA, 271–274. doi:10.1145/3649165.3703623

[41] Yim Register and Amy J. Ko. 2020. Learning Machine Learning with Personal Data Helps Stakeholders Ground Advocacy Arguments in Model Mechanics. In *Proceedings of the 2020 ACM Conference on International Computing Education Research.* ACM, Virtual Event New Zealand, 67–78. doi:10.1145/3372782.3406252

[42] Markus Rex, Marlene Göring, and Sarah Pybus. 2022. *The Greatest Polar Expedition of All Time: The Arctic Mission to the Epicenter of Climate Change.* David Suzuki Institute : Greystone Books, Vancouver ; Berkeley ; London.

[43] Chantel Ridsdale, James Rothwell, Mike Smit, Michael Bliemel, Dean Irvine, Daniel Kelley, Stan Matwin, Brad Wuetherick, and Hossam Ali-Hassan. 2015. Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report. (2015). doi:10.13140/RG.2.1.1922.5044

[44] David Rolnick, Priya L. Donti, Lynn H. Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, Alexandra Sasha Luccioni, Tegan Maharaj, Evan D. Sherwin, S. Karthik Mukkavilli, Konrad P. Kording, Carla P. Gomes, Andrew Y. Ng, Demis Hassabis, John C. Platt, Felix Creutzig, Jennifer Chayes, and Yoshua Bengio. 2023. Tackling Climate Change with Machine Learning. *Comput. Surveys* 55, 2 (Feb. 2023), 1–96. doi:10.1145/3485128

[45] Aimee Schwab-McCoy, Catherine M. Baker, and Rebecca E. Gasper. 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:10.1080/10691898.2020.1851159

[46] Verkehr Senatsverwaltung für Mobilität, Klimaschutz und Umwelt (Land Berlin). 2025. Aktueller Luftqualitätsindex (LQI) | Berliner Luftgütemessnetz. https://luftdaten.berlin.de/lqi.

[47] 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*. ACM, Seattle Washington USA, 561–566. doi:10.1145/3017680.3017717

[48] Matti Tedre, Peter Denning, and Tapani Toivonen. 2021. CT 2.0. In *Proceedings of the 21st Koli Calling International Conference on Computing Education Research (Koli Calling '21)*. Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/3488042.3488053

[49] Matti Tedre, Henriikka Vartiainen, Juho Kahila, Tapani Toivonen, Ilkka Jormanainen, and Teemu Valtonen. 2020. Machine Learning Introduces New Perspectives to Data Agency in K—12 Computing Education. In *2020 IEEE Frontiers in Education Conference (FIE)*. 1–8. doi:10.1109/FIE44824.2020.9274138

[50] TensorFlow Team / Google. 2025. Time Series Forecasting | TensorFlow. https://www.tensorflow.org/tutorials/structured_data/time_series.

[51] Claire Toogood. 2023. Supporting Students to Engage with Case Studies: A Model of Engagement Principles. *Educational Review* (Nov. 2023), 1–15. doi:10.1080/00131911.2023.2281227

[52] Irvine University of California. 2025. UCI Machine Learning Repository – UC Irvine. https://archive.ics.uci.edu/.

[53] Jan Van Den Akker. 1999. Principles and Methods of Development Research. In *Design Approaches and Tools in Education and Training*, Jan Van Den Akker, Robert Maribe Branch, Kent Gustafson, Nienke Nieveen, and Tjeerd Plomp (Eds.). Springer Netherlands, Dordrecht, 1–14. doi:10.1007/978-94-011-4255-7_1

[54] Xiaoyu Wan, Xiaofei Zhou, Zaiqiao Ye, Chase K. Mortensen, and Zhen Bai. 2020. SmileyCluster: Supporting Accessible Machine Learning in K-12 Scientific Discovery. In *Proceedings of the Interaction Design and Children Conference*. ACM, London United Kingdom, 23–35. doi:10.1145/3392063.3394440

[55] Matthias Wilde, Katrin Bätz, Anastassiya Kovaleva, and Detlef Urhahne. 2009. Überprüfung einer Kurzskala intrinsischer Motivation (KIM). *Zeitschrift für Didaktik der Naturwissenschaften : ZfDN* 15 (2009), 31–45. doi:10.25656/01:31663

[56] Wilfried Hagg, Christoph Mayer, Markus Weber, Heidi Escher-Vetter, and Hermann Rentsch. 2025. Bayerische Gletscher: Nördlicher Schneeferner, Topographie. http://bayerischegletscher.userweb.mwn.de/nsf/nsf_topo.htm.

[57] Carrie Wright, Qier Meng, Michael R. Breshock, Lyla Atta, Margaret A. Taub, Leah R. Jager, John Muschelli, and Stephanie C. Hicks. 2024. Open Case Studies: Statistics and Data Science Education through Real-World Applications. *Journal of Statistics and Data Science Education* (2024), 1–30. doi:10.1080/26939169.2024.2394541

[58] Andreas Zendler (Ed.). 2018. *Unterrichtsmethoden für den Informatikunterricht*. Springer Fachmedien Wiesbaden, Wiesbaden. doi:10.1007/978-3-658-20675-8

[59] Daochen Zha, Zaid Pervaiz Bhat, Kwei-Herng Lai, Fan Yang, Zhimeng Jiang, Shaochen Zhong, and Xia Hu. 2025. Data-Centric Artificial Intelligence: A Survey. *Comput. Surveys* 57, 5 (May 2025), 1–42. doi:10.1145/3711118

[60] Xiaofei Zhou, Jessica Van Brummelen, and Phoebe Lin. 2020. Designing AI Learning Experiences for K-12: Emerging Works, Future Opportunities and a Design Framework. doi:10.48550/arXiv.2009.10228