Extended Abstract for the
Symposium on Integrating AI and Data Science into School Education Across Disciplines 2025

EMPOWERING STUDENTS TO GAIN INSIGHTS WITHIN DATA EXPLORATION PROJECTS IN THE CLASSROOM - USING, MODIFYING, AND CREATING DATA MOVES THROUGH A SCAFFOLDED USE OF DIGITAL TOOLS

Sven Hüsing[1] & Susanne Podworny[2]
[1]Paderborn University, Germany, sven.huesing@upb.de
[2]Paderborn University, Germany

Focus Topics: Learning materials; Data and Problems; Tools

Introduction

Working with and exploring data is an essential scientific practice from which researchers want to gain insights into personally or socially meaningful issues. However, what researchers have long known and consider common practice can also be a powerful tool for all citizens. Citizenship can be promoted through data science education projects (Geiger et al., 2023; Makar et al., 2023). Here, acquiring the capacity to work with and make sense of data is a significant step towards data literacy for learners, as Ridsdale et al. (2015) point out. In our view, an essential goal in this respect is for learners to carry out their own data projects to explore personally meaningful issues. In doing so, learners can be actively involved in the knowledge construction process and shape their community's knowledge and methods. The literature refers to *epistemic agency* (Odden et al., 2023), which we will discuss here. Additionally, we present opportunities for learners to develop epistemic agency in the context of data exploration projects and propose a combined framework for such projects by building on the Use-Modify-Create-framework (Lee et al., 2011) and the PPDAC-cycle (Wild & Pfannkuch, 1990).

From our point of view, data exploration projects serve as an ideal connection between mathematics and computer science education, as they build on statistical skills and the interaction with a computer for data processing and visualization to make complex relationships understandable.

Background

Statistical projects are recognized as a powerful teaching method for fostering data science skills (Gómez-Blancarte & Ortega, 2018). Projects let learners engage with key statistical concepts, including data, representations, and variability (Burrill & Biehler, 2011). Additionally, studies show that project-based learning can boost students' motivation (Bilgin, Newbery, & Petocz, 2015).

Wild and Pfannkuch (1999) introduced the widely used PPDAC cycle to structure statistical thinking in empirical investigations. The five phases—problem, plan, data, analysis, and conclusion—guide the data analysis process. In "problem," statistical questions and hypotheses are defined. "Plan" involves designing the data collection conducted in "data." The "analysis" phase covers statistical evaluation, and the "conclusion" interprets the results, which eventually then drive another iteration. This framework is commonly found and adapted in various ways in curricula and teaching materials to guide learners through the process of statistical data analysis.

Data literacy—the ability to collect, manage, evaluate, and apply data critically (Ridsdale et al., 2015)—is essential in such a data investigation. School education provides a strategic starting point to embed data literacy, equipping students with the skills needed for evidence-based decision-making and workforce readiness. For the "analysis" phase in the PPDAC cycle (Wild & Pfannkuch, 1999), the competencies related to "data evaluation" outlined by Ridsdale et al. (2015) can be highly relevant.

A way to foster those competencies is by using data moves (Erickson et al., 2019). Data moves refer to the essential actions of manipulating and transforming data to make it useful for analysis, especially when working with large, complex datasets. These actions include filtering, merging datasets, creating new groupings, or constructing measures, which enable goal-driven exploration and analysis. Unlike traditional statistical instruction, which often uses pre-structured datasets and predefined tasks, data moves emphasize the dynamic and transformative nature of data. Teaching data moves helps students understand that datasets are not static but socially constructed and modifiable, fostering deeper engagement with data analysis.

Learning to apply data moves might therefore help students explore personal interests through data. This way, students might gain responsibility for their own learning and might influence the

In S. Podworny & S. Schönbrodt (Eds.), *Towards Fostering AI and Data Science Literacy in Schools Across Disciplines. Proceedings of the 1st Symposium on Integrating AI and Data Science into School Education Across Disciplines (AIDEA1 2025)*, Salzburg, Austria.

knowledge-shaping processes in the classroom or their community—thus fostering their epistemic agency (Odden et al., 2023). Epistemic agency refers to the idea of learners taking ownership of their own inquiry processes regarding their epistemic interest and the methods used (Miller et al., 2018). Opportunities to foster epistemic agency include especially (a) building on students' knowledge, (b) gaining knowledge, (c) creating a knowledge product, and (d) identifying actions for change in local or global structures (Miller et al., 2018). From this viewpoint, we see a direct connection to empirical investigations, as described by Wild and Pfannkuch (1999), to foster data literacy (Ridsdale et al., 2015). As tools for conducting such inquiry processes, we consider CODAP (https://codap.concord.org/) and Jupyter Notebooks with Python (Granger & Perez, 2021) to represent two suitable platforms. These have different characteristics and, therefore, offer different use cases regarding learning goals.

CODAP is a digital tool for learning data analysis (Haldar et al., 2018). Its user-friendly interface requires no coding skills, thereby addressing especially middle school students. CODAP lets beginners quickly explore data and pursue their own questions through direct manipulation. At the other end, a Jupyter Notebook is a more professional tool for data analysis and visualization. Its flexibility and integration with the programming language Python makes it a powerful platform for data science. While some coding skills might be helpful for working in Jupyter Notebooks, they also provide a rich environment for learning, enabling students to explore data, document their processes, and develop skills to engage in programming for pursuing individually meaningful projects (Granger & Perez, 2021).

While the CODAP environment directly enables engaging in inquiry processes through its "hands-on" character, using Jupyter Notebooks and the programming language Python requires an external scaffold to support students in their interwoven programming and knowledge processes. In this regard, we support students with worked examples (Atkinson et al., 2000; Muldner et al., 2023). These are complete solutions to similar (programming) endeavors that the students can use as guidelines or as sources for programming snippets.

In the following, we present two exemplary teaching modules for data exploration projects within the CODAP or the Jupyter Notebook environment.

Exploring data with CODAP: A teaching unit on "data and data detectives"

The "Data and Data Detectives" module introduces students to reasoning about data through a project-based approach using the PPDAC cycle (Wild & Pfannkuch, 1999) and the JIM-PB questionnaire dataset (https://www.prodabi.de/en/unterricht/toolkit/jim-pb-daten/). That dataset contains responses on media and leisure activities from more than 1200 German students (Podworny, et al., 2022). Young people have been asked about their leisure and media behavior, for example, about the frequency of reading books and magazines online or offline, playing computer games, using social media platforms, using YouTube, etc. An example question is: "How often do you watch LetsPlay videos on YouTube?" with possible answers 'daily,' 'several times a week,' 'once a week,' 'twice a month,' 'once a month,' 'less often,' and 'never'. Using the CODAP platform, students are introduced to several data moves and explore this multivariate dataset to pose and analyze meaningful statistical questions. The teaching unit consists of eight 45-minute lessons and concludes with students exploring the JIM-PB data in small groups based on their own interests, thereby fostering their agency. The teaching unit addresses fundamental ideas in statistics, such as data, representations, and variability (Burrill & Biehler, 2011). Accompanying research shows that students are challenged in the final projects by using percentages, but they discover and present interesting relationships in the data (Podworny, 2024; Podworny & Fleischer, 2022).

Exploring Data with Python and Jupyter Notebooks: A Teaching Unit on "Analyzing Environmental Data Through Epistemic Programming"

The other module focuses on inquiry processes through programming in the context of environmental data (Podworny et al., 2022). Students can come up with individual initial research questions at the beginning of the teaching module that they want to answer by conducting an exploration

of environmental data. Within the teaching module, the students retrieve or collect suitable data for their individual inquiry and afterwards analyze them in the Jupyter Notebook environment. Throughout this process, they might also change their epistemic interests, e.g., as a consequence of an insight or idea they get during their analysis process. To scaffold the students' data exploration and programming process, a worked example of a CO_2 analysis is provided, introducing the students to the PPDAC cycle, relevant data moves, and the corresponding programming code. The students can then use the worked example as guidance for their data exploration processes and as a kind of "cheat sheet" for program code. With this help, the students create a so-called computational essay as a reproducible documentation of their individual programming- and insight processes (Hüsing & Podworny, 2022), combining code cells and explaining markdown cells within a Jupyter Notebook.

In this way, the students engage with data within an insight-driven programming approach, which might enable them to view programming as a means for exploring personal or societal interests (Hüsing et al., 2024a). Currently, accompanying research focuses on the question of whether this programming approach fosters epistemic agency in students' inquiry processes (Hüsing & Schönbrodt, 2024) as well as on the evaluation of worked examples as a scaffold (Hüsing et al., 2024a; Hüsing et al., 2024b) and computational essays as a potential learners' product (Hüsing & Podworny, 2022).

Discussion

In both teaching modules, we use a scaffolded environment in which the students can conduct a data exploration regarding personally meaningful questions or topics. In the first teaching module, CODAP itself represents a scaffold for the students to apply data moves, insofar as students can directly interact with the data. In the second teaching module, worked examples support the students by providing an exemplary data exploration with already working program code, alongside explaining texts for the respective data moves.

Through these scaffolds, we hypothesize that the students will be able to engage in a tinkering or opportunistic (programming) process where they can pursue individual interests or research questions (Brandt et al., 2008). In both teaching modules, the students learn the data moves gradually, first using them before adapting them for their individual endeavors, and finally creating their own knowledge product. In this regard, we adhere to the Use-Modify-Create learning progression by Lee et al. (2011) and the PRIMM approach by Sentance et al. (2019), which is used to teach learners how to use, adapt, and extend code to create their own programming product.

Our goal is to explore ways to integrate data projects into the classroom, empowering learners to independently use data moves to explore personally meaningful contexts and take ownership of their learning as active, knowledge-driven participants who are starting to develop data literacy. Therefore, we want to transfer the principles of the Use-Modify-Create model to the use, adaptation, and extension of data moves (Erickson et al., 2019) for knowledge building with regard to learners' personal or societal interests.

To engage young learners in data exploration, we propose integrating approaches from mathematics/statistics and computer science education by combining the Use-Modify-Create framework from Lee et al. (2011) with the PPDAC cycle (Wild & Pfannkuch, 1999) as illustrated in Figure 1.

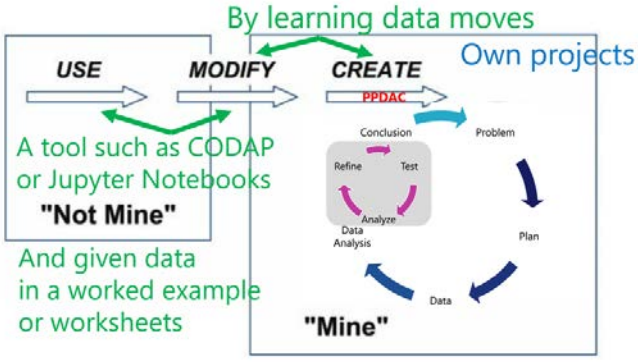

Figure 1. Engaging youth in data exploration, combining statistics and computer science education approaches (based on the Use-Modify-Create Learning Progression (Lee et al., 2011) and the PPDAC cycle (Pfannkuch & Wild, 1999))

Learners begin by using or exploring an existing scaffold—such as a CODAP activity or a worked example in a Jupyter Notebook—that is "not their own". After this "Use-Phase", they modify the existing scaffold in order to a) understand how and for which purpose to use the demonstrated data moves or programming steps (also see the concept of process-oriented worked examples in Van Gog et al., 2004) and to b) adapt it according to their own needs and interests. Through this guided experience of learning and practicing "data moves" (Erickson et al., 2019), they develop the skills needed to create their own data exploration projects. In the "Create-Phase", students then apply the PPDAC steps in several cycles, with particular emphasis on the analysis phase. Here, they engage in iterative cycles of analyzing, refining, and testing until they arrive at meaningful conclusions using digital tools.

In this way, we aim at enabling learners to devote themselves to personally meaningful questions in their data exploration endeavors, capable of using data moves as tools for gaining personally or socially meaningful knowledge in the context of data exploration projects, thereby fostering their epistemic agency.

References

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from Examples: Instructional Principles from the Worked Examples Research. *Review of Educational Research, 70*(2), 181–214. https://doi.org/10/csm67w

Bilgin, A. A., Newbery, G., & Petocz, P. (2015). Engaging and motivating students with authentic statistical projects in a capstone unit. In M. A. Sorto (Ed.), Advances in statistics education: developments, experiences and assessments. *Proceedings of the satellite conference of the International Association for Statistics Education (IASE)*, July 2015. Rio de Janeiro, Brazil: ISI.

Brandt, J., Guo, P. J., Lewenstein, J., & Klemmer, S. R. (2008). Opportunistic programming: How rapid ideation and prototyping occur in practice. *Proceedings of the 4th International Workshop on End-User Software Engineering*, 1–5. https://doi.org/10.1145/1370847.1370848

Burrill, G., & Biehler, R. (2011). Fundamental Statistical Ideas in the School Curriculum and in Training Teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics – Challenges for teaching and teacher education: A joint ICMI/IASE study* (pp. 57-69). Dordrecht: Springer Science+Business Media.

Erickson, T., Wilkerson, M., Finzer, W., & Reichsman, F. (2019). Data Moves. *Technology Innovations in Statistics Education, 12*(1). https://doi.org/10/gqv6c4

Geiger, V., Gal, I., & Graven, M. (2023). The connections between citizenship education and mathematics education. *ZDM – Mathematics Education, 55*(5), 923-940. https://doi.org/10.1007/s11858-023-01521-3

Gómez-Blancarte, A., & Ortega, A. S. (2018). Research on statistical projects: Looking for the development of statistical literacy, reasoning and thinking. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute

Granger, B. E., & Perez, F. (2021). Jupyter: Thinking and Storytelling With Code and Data. *Computing in Science & Engineering*, *23*(2), Article 2. https://doi.org/10.1109/MCSE.2021.3059263

Haldar, L. C., Wong, N., Heller, J. I., & Konold, C. (2018). Students making sense of multi-level data. *Technology Innovations in Statistics Education, 11*(1), 2–32. https://escholarship.org/uc/item/7x28z96b

Hüsing, S., & Podworny, S. (2022). Computational Essays as an Approach for Reproducible Data Analysis in lower Secondary School. *Proceedings of the IASE 2021 Satellite Conference*: *Statistics Education in the Era of Data Science*. https://doi.org/10.52041/iase.zwwoh

Hüsing, S., Schulte, C., Sparmann, S., & Bolte, M. (2024a). Using Worked Examples for Engaging in Epistemic Programming Projects. *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, 443–449. https://doi.org/10.1145/3626252.3630961

Hüsing, S., Sparmann, S., Schulte, C., & Bolte, M. (2024b). Identifying K-12 Students' Approaches to Using Worked Examples for Epistemic Programming. *Proceedings of the 2024 Symposium on Eye Tracking Research and Applications*, 1–7. https://doi.org/10.1145/3649902.3655094

Lee, I., Martin, F., Denner, J., Coulter, B., Allan, W., Erickson, J., Malyn-Smith, J., & Werner, L. (2011). Computational thinking for youth in practice. *ACM Inroads*, *2*(1), 32–37. https://doi.org/10/ggdnzr

Makar, K., Fry, K., & English, L. (2023). Primary students' learning about citizenship through data science. *ZDM – Mathematics Education, 55*(5), 967-979. https://doi.org/10.1007/s11858-022-01450-7

Miller, E., Manz, E., Russ, R., Stroupe, D., & Berland, L. (2018). Addressing the epistemic elephant in the room: Epistemic agency and the next generation science standards. *Journal of Research in Science Teaching*, *55*(7), 1053–1075. https://doi.org/10.1002/tea.21459

Muldner, K., Jennings, J., & Chiarelli, V. (2023). A Review of Worked Examples in Programming Activities. *ACM Transactions on Computing Education, 23*(1), 1–35. https://doi.org/10.1145/3560266

Odden, T. O. B., Silvia, D. W., & Malthe-Sørenssen, A. (2023). Using computational essays to foster disciplinary epistemic agency in undergraduate science. *Journal of Research in Science Teaching, 60*(5), 937–977. https://doi.org/10.1002/tea.21821

Podworny, S. (2024). Eine qualitative Studie zu Data Science Education: Schülerinnen und Schüler analysieren multivariate Daten. *Stochastik in der Schule 44*(1).

Podworny, S. & Fleischer, Y. (2022). An approach to teaching data science in middle school. In U.T. Jankvist, R. Elicer, A. Clark-Wilson, H.-G. Weigand, & M. Thomsen (Eds.), *Proceedings of the 15th international conference on technology in mathematics teaching (ICTMT 15)* (pp. 308–315). Danish School of Education, Aarhus University.

Podworny S., Fleischer, Y., Stroop, D. & Biehler R. (2022). An example of rich, real and multivariate survey data for use in school. *Twelfth Congress of the European Society for Research in Mathematics Education (CERME12)*, Bozen-Bolzano, Italy.

Podworny, S., Hüsing, S. & Schulte, C. (2022). A place for data science introduction in school: between statistics and programming. *Statistics Education Research Journal 21*(2), Special Issue on Data Science.

Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., & Wuetherick, B. (2015). *Strategies and best practices for data literacy education.* https://dalspace.library.dal.ca/items/ab6d9110-4739-4a74-8b27-4e03b19601e9

Sentance, S., Waite, J., & Kallia, M. (2019). Teachers' Experiences of using PRIMM to Teach Programming in School. *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, 476–482. https://doi.org/10/ggdpnn

Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2004). Process-Oriented Worked Examples: Improving Transfer Performance Through Enhanced Understanding. Instructional Science, 32(1/2), 83–98. https://doi.org/10.1023/B:TRUC.0000021810.70784.b0

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223–265. https://doi.org/10.1111/j.1751-5823.1999.tb00442.x