Branching out data science education: Developing task and computational environment design principles for teaching data science at the high school level through an international research collaboration

Anna Fergusson[1], Susanne Podworny[2], Yannik Fleischer[2], Sven Hüsing[2], Malia S. Puloka[1], Rolf Biehler[2], Maxine Pfannkuch[1], Stephanie Budgett[1] and Michelle Dalrymple[3]
[1]University of Auckland, New Zealand; [2]Paderborn University, Germany; [3]Cashmere High School, New Zealand
a.fergusson@auckland.ac.nz

Data science is rapidly emerging as a desired component of education, yet generalisable design principles for tasks and computational environments remain underdeveloped. This paper reports on a design-based research collaboration between Germany and New Zealand, which aims to develop and refine such design principles for high school data science education. Drawing on literature and prior research, we propose that the design of tasks and computational environments should integrate humanistic, algorithmic, and programming approaches. The three key design "branches" of our research are: (1) emphasising a humanistic focus through data construction and exploration; (2) using decision trees to develop algorithmic modelling concepts; (3) using computer programming for gaining insights, supporting creativity, and model tinkering. These branches are illustrated through examples from our collaborative work. Our research contributes to the development of effective educational strategies for data science education and could inform the development of professional development for high school teachers.

INTRODUCTION

As an emerging academic discipline, data science requires novel pedagogical approaches that adapt to the evolving nature of data, rapid technological development, and the need to prepare students for an ever-changing future (Burrill & Pfannkuch, 2024; Ridgway, 2015). Following the lead of universities, many countries around the world have introduced data science as a subject at the high school level, creating new curricula by drawing on, extending, and enhancing content from statistics and computer science (e.g., ProDaBi in Germany, IDS in the US). Despite extensive research efforts, mostly confined to specific national curricula, there remains a significant research knowledge gap pertaining to the development of generalised theoretical frameworks or pedagogical principles that can effectively guide data science education across different educational systems. Hence, the goal of our research project is to create new pedagogical materials and frameworks that can serve as guides for high school data science teachers, regardless of country. In this paper, we present our emergent findings from the first phase of our collaborative research, in which we focused on identifying three key design branches that could inform the development of tasks and computational environments for teaching data science at the high school level.

OUR COLLABORATIVE RESEARCH APPROACH

Our collaboration brings together researchers from Paderborn University, Germany and the University of Auckland, New Zealand. Collectively, our research team has explored the following aspects of data science education: task design principles for introducing computer programming as part of statistical and algorithmic modelling; the design of computational learning environments for statistics and data science; professional development of high school data science teachers; humanistic and computational approaches to constructing and exploring data; development of a programming approach for novices, called *Epistemic Programming*, that centres around gaining insights related to personal or societal interests through projects such as data exploration; introduction of machine learning methods, such as classification models and decision trees, to school-level students; and educationally designing and connecting different unplugged and computational tools.

Due to our shared focus on data science education at the school level, in particular with respect to design intentions and principles for tasks and tools, our main research question for this collaborative project is: *In what ways should tasks and computational environments be designed, so that students at the high school level are supported to engage with data science?*

In: Birk, L., Loth, G., Jotzo, L., Binder, K., & Frischemeier, D. (Eds.) (2026). *Statistics and Data Science Education in STEAM. Proceedings of the Satellite Conference of the International Association for Statistical Education (IASE), September/October 2025, Münster, Germany*.

To explore answers to this research question, we are using a Design-Based Research approach (DBR), as our goal is to develop new instructional theories and design principles that can influence teaching practice and education research (e.g., McKenney & Reeves, 2018). The DBR process used for this project will involve interconnected phases of grounding, embodying, iterating, and reflecting (Hoadley & Campos, 2022). In the grounding phase, we used relevant literature, our teaching experience, existing research tasks and curriculum materials we had developed, and our previous research findings to identify potential principles for designing high school data science tasks and computational learning environments that could be applied across different educational systems. In the embodying phase, we will use the design principles identified in the grounding phase to create new teaching activities and computational environments. In the iterating phase, these teaching activities will be implemented with students and teachers in New Zealand and Germany across different classroom settings, to allow us in the reflecting phase to develop more generalised design principles for data science education at the high school level that can be applied within different countries.

OUR THREE DESIGN BRANCHES

Based on analysis during the grounding phase of our DBR research, we identified three key design "branches" that connect across our collective research: (1) emphasising a humanistic focus through data construction and exploration; (2) using decision trees to develop algorithmic modelling concepts; (3) using computer programming for gaining insights, supporting creativity, and model tinkering. We will summarise the educational research basis for each branch and illustrate these design "branches" using specific features of tasks or computational environments developed by our collaborative research team.

Branch 1: Emphasising a humanistic focus through data construction and exploration

Data science students' first experiences with data often involve pre-existing datasets that are far removed from their personal lives and that do not support them to make connections with their culture or lived experiences. Furthermore, without students engaging with tasks that involve the construction of data and variables, they risk developing an under-appreciation of the origins of the data and the nature of the variables, which has serious implications for any data-based models created. High school students currently have limited experience with complex and messy data sets, and challenges occur when reasoning with data, for example, with statistical concepts such as variation, distribution, group comparison and association (Biehler et al., 2018). Students of all ages struggle to define categorical variables and reason with conditional proportions (Budgett & Puloka, 2019). However, these data reasoning skills are crucial for building students' critical thinking and data literacy. In our own teaching practice, we have found that students often struggle with reasoning with two-way tables, including confusion matrices, as these involve proportions.

In our previous research, we identified effective pedagogical strategies for teaching data construction and reasoning. For instance, modern sources of data such as social media usage (Fergusson & Bolton, 2018) or survey-based data collected from students themselves (Podworny, Fleischer, et al., 2022) can be authentic pedagogical contexts for learning. We have found that well-designed statistical software environments such as CODAP or Python in Jupyter Notebooks can support curiosity-driven data exploration (Hüsing, Schulte, et al., 2024; Podworny, Hüsing, & Schulte, 2022), and that constructing new variables from unstructured text can be an accessible entry point into cleaning complex and messy data sets using tools such as spreadsheets, data analysis software and computer programming (Fergusson et al., 2025). Tasks that allow students to use subjective measures, such as the religious origins of their names (Puloka & Pfannkuch, 2023) or the ethnic diversity of movie posters (Caetano et al., 2023), appear to be important for validating students' differing cultural backgrounds and perspectives.

We have also found that unplugged environments using data cards can be effective for initial experiences with constructing new variables (e.g., Fergusson & Pfannkuch, 2024; Fleischer et al., 2024). Linking physical sorting of data cards to two-way tables of counts and other data representations, such as human models, also appears to assist the development of proportional reasoning (Podworny, Fleischer, & Hüsing, 2022). Students can be supported to construct data from text and images by providing purpose-built computational environments that allow them to access data from online databases, create new variables based on features of images by moving photos around on a screen, and then "jump into" statistical software tools to explore the data and new variables they constructed (Fergusson & Wild, 2021). Figure 1 provides visual examples from our research of students constructing new variables by labelling within unplugged environments and purpose-built computational environments.

(a)

(b)

Figure 1. Visual examples from our research of students constructing new variables by labelling (a) within an unplugged environment, and (b) within a purpose-built computational environment.

Branch 2: Using decision trees to develop algorithmic modelling concepts

Algorithmic models influence nearly every aspect of students' digital technology interactions. Decision trees can provide an accessible entry point for data science students to learn about Machine Learning (ML) and algorithmic models. However, algorithmic models represent a substantially different way of thinking about data and models compared to what has been traditionally taught in high school mathematics and statistics classrooms. We have identified effective pedagogical strategies for teaching machine learning and algorithmic modelling concepts in our previous research. For example, we have found that exploring data to introduce decision trees can help students understand the nature of the model and the algorithm that created the model (Fleischer & Biehler, 2025).

Our research also indicates that informal approaches for introducing classification and prediction models can build awareness of ML ideas, such as bias related to unbalanced data sets, and awareness of concepts such as overfitting and generalisability (Fergusson & Pfannkuch, 2022a). Unplugged environments and data cards appear to work well to introduce decision trees to all students (Podworny, Fleischer, & Hüsing, 2022), particularly younger students (Fleischer et al., 2024). Involving students in the labelling of the variable that needs to be predicted can help develop statistical and computational thinking, and consideration of whether human subjective judgements can be replaced by computational model-based automations (Fergusson & Pfannkuch, 2024).

We have also found that carefully designed computational environments that use visualisations and animations can help in understanding decision trees (Biehler & Fleischer, 2021). These kinds of computational environments also allow for the use of digital data sources, providing an authentic context for introducing classification ideas (Fergusson & Pfannkuch, 2024). We have demonstrated that describing and explicitly connecting modelling actions for different tools (e.g., unplugged, GUI-driven, and code-driven) can support students to move between these tools when creating models (Fergusson & Pfannkuch, 2022b; Fleischer & Podworny, 2022). We hypothesise that tool-related conceptual progressions for ML involve gradually increasing the amount of automation used to develop the algorithmic model (Fleischer, Biehler, et al., 2022). Figure 2 provides examples from our research that demonstrate different tools for supporting student engagement with decision trees.

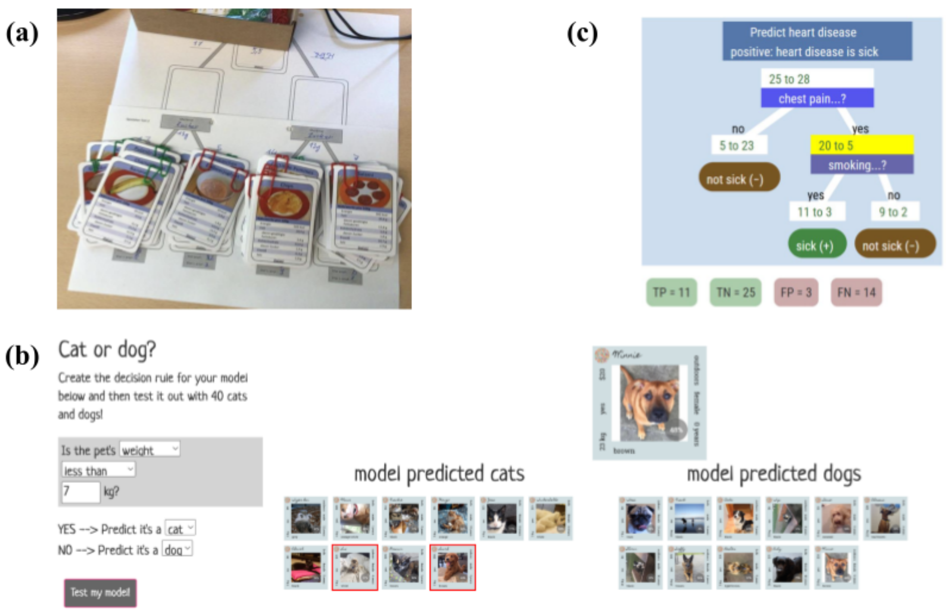

Figure 2. Examples from our research of students engaging with decision trees within (a) unplugged environments, (b) a purpose-built computational environment, and (c) the CODAP environment using the ARBOR decision tree plugin.

We have found that high school students and teachers can use code-based approaches to develop and evaluate classification and prediction models (Fergusson & Pfannkuch, 2022a; Fleischer, Biehler, et al., 2022), and that combining data cards and GUI software can be effective prior learning experiences before coding (Fergusson & Pfannkuch, 2022b; Fleischer & Biehler, 2025). Our research has identified that care needs to be taken about how to present metrics associated with evaluating decision trees, for example, by starting with the misclassification rate and later using different statistical measures like sensitivity and precision for a more sophisticated and context-related evaluation (Biehler & Fleischer, 2021).

Branch 3: Using computer programming for gaining insights, supporting creativity, and model tinkering

For effective implementation of data science education at the high school level, the tasks and computer programming environments for students have to be designed in a way so that students are not overwhelmed, but at the same time can utilise programming as a means for creativity and authentic scientific inquiry. Through our combined research, we have found that programming can be used as a means for exploring personally or socially relevant topics (Hüsing, 2021; Hüsing, Schulte, et al., 2024), for example, by providing access to authentic data (e.g., environmental data) from online databases (Fergusson & Pfannkuch, 2022a; Hüsing, Schulte, et al., 2024; Hüsing, Sparmann, et al., 2024) or creating visualisations based on student-selected Google search terms (Fergusson, 2023). Students can also use code-driven tools to “tinker” with model parameters (Fleischer, Biehler, et al., 2022) and explore creative solutions based on probability models (Fergusson, 2022).

Computer programming environments can be designed in ways that can help students to develop their epistemic agency, that is, empower them to take control of their own learning and knowledge-building processes (Odden et al., 2022). Connected with this idea is the development of a programming approach, called epistemic programming, focusing students’ learning on gaining insights, for example, through data exploration (Hüsing et al., 2023). We have found that worked examples can be useful tools for demonstrating and inspiring code-driven data exploration (Fergusson, 2023; Hüsing, Schulte, et al., 2024; Hüsing, Sparmann, et al., 2024) and that students can produce diverse responses when working within computer programming environments that support curiosity-driven decisions about data and modelling (Caetano et al., 2023). Specifically, programming environments that support the writing and execution of both markdown and code, such as Jupyter notebooks or RMarkdown documents, can be helpful for teaching reproducibility (Fergusson & Pfannkuch, 2022b) and support the creation of computational essays in the context of data exploration projects (Hüsing & Podworny, 2021).

It is also important to carefully sequence and scaffold students' initial experiences with code to provide an inclusive learning experience (Fergusson, 2023). The use of simple webpage-based interfaces that support students to interact with small "chunks" of code can be effective for introducing programming to novices (Fergusson & Pfannkuch, 2024), particularly when used within purpose-focused tasks that guide students to focus on "tinkering" with specific aspects of code to evaluate given statements (Fergusson & Pfannkuch, 2022a). Figure 3 provides examples from our research of computational environments we have designed to support student engagement with programming through Jupyter notebooks that hide code using interactive widgets, interactive online coding tasks explicitly connecting GUI-driven modelling actions with code-driven modelling actions, and the directed use of worked examples to guide the creation of computational essays also within a Juptyer notebook environment.

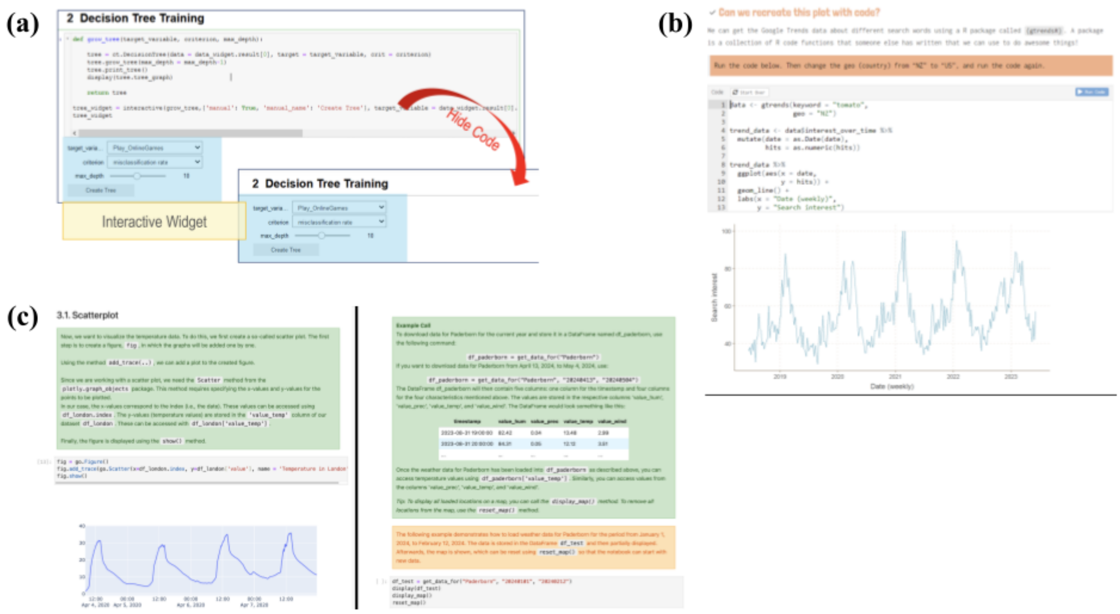

Figure 3. Examples from our research of computational environments to support student engagement with programming through (a) educationally-designed Jupyter notebooks, (b) interactive online coding tasks, and (c) using worked examples to guide the creation of computational essays.

Additionally, programming environments that progressively reveal components of worked examples, which students can then adapt to suit their personal interests, can support students new to programming to develop their statistical thinking (Fergusson, 2023; Fergusson & Pfannkuch, 2022a).

FUTURE WORK

These three design branches provide the grounding for the next phases of our research, where we aim to further extend and "strengthen" these branches by exploring the design features of tasks and computational environments that still require a greater understanding of how to engage students at the high school level with data science. Regarding the first design branch, *emphasising a humanistic focus through data construction and exploration*, our research has not yet established whether the specific data contexts we have used for constructing new variables from text and images will transfer across different countries and cultures. While we have found that using an unplugged approach first when constructing data can help students better understand the origins of the data and the nature of the variables, we have not explored if these understandings transfer to explorations with data they have not constructed. We hypothesise that if we can more explicitly connect physical experiences with sorting and cross-classifying data cards in unplugged environments, to reasoning with data representations created within purpose-built computational environments, students will be supported to explore and reason with data more effectively. We plan to develop new teaching activities that support students in personally constructing data sets by extracting features from their surroundings, text and images, first within unplugged environments and then within computational environments. This will involve designing a new computational environment that provides similar support for data construction and reasoning as that afforded by an unplugged environment using physical data cards. The design of the associated learning tasks will specifically target student understanding of categorical variables and proportional reasoning, including conditioning, to lay the foundations for learning about decision trees.

For the second design branch, *using decision trees to develop algorithmic modelling concepts,* our research has not yet established how important the data and model context are for developing students' understanding of the societal consequences of algorithmic models being used in "real life" practice. For instance, students may care more about a model "getting it wrong" if the consequence is of personal relevance. While students can build decision trees using large survey-based data (Fleischer, Biehler, et al., 2022), model contexts such as recommendation systems or predicting available car parks (Fleischer, Hüsing, et al., 2022) may help students better understand the utility and impact of ML. We plan to develop new teaching activities where students can explore ML and algorithmic concepts, such as overfitting and bias, within the modelling context of data-based decision trees. This will also involve developing computational environments that specifically target student engagement with key aspects of algorithmic models, using digital image data and model "tinkering".

With respect to our third design branch, *using computer programming for gaining insights, thereby supporting creativity and model tinkering,* we also plan to further explore the relationship between specific features and affordances of the computer programming environment and the degree of epistemic agency and creativity demonstrated by students when engaging in data or modelling-focused explorations. From our own teaching experiences, we know that particularly younger and more inexperienced programmers can make use of code-driven tools for conducting data science projects when there is well-thought-out scaffolding, for example through orientation or prepared worked examples. We also conjecture that the nature of the data context and the provided motivation for data exploration are key factors in fostering epistemic agency and curiosity-driven learning. Hence for this proposed project, we plan to develop a computational environment that will allow students with little or no programming experience to interact with worked examples for data exploration and modelling, by providing different features like a mechanism for setting aspects of the code as non-editable or highlighting aspects that students might want to change with regard to their own interests (cf. Fergusson, 2023).

CONCLUSION

Data science education is a rapidly evolving field that requires novel pedagogical approaches to keep pace with technological advancements and prepare students for future challenges (Gould, 2017; Burrill & Pfannkuch, 2024). Our paper has reported on an international design-based research collaboration between Germany and New Zealand aimed at developing generalisable principles for designing tasks and computational environments in high school data science education. Our work integrates humanistic, algorithmic, and programming approaches across three key design branches that emphasise: constructing and exploring personally relevant data; introducing algorithmic thinking through decision trees; and fostering creativity and agency via insight-driven computer programming. By exposing students to diverse data sources and technologies, and centring student-driven, curiosity-based learning (Ridgway, 2022), our research supports the development of essential data literacy and computational skills. Through the iterative design and refinement of tasks and environments, informed by implementation with both students and teachers in both countries, we aim to develop generalised theoretical frameworks and pedagogical principles for data science education. Our research can contribute to the development of effective educational strategies for this emerging discipline and can inform the development of professional development for high school teachers.

ACKNOWLEDGEMENTS

The project on which this paper is based was partially funded by the German Federal Ministry of Education and Research, the German Academic Exchange Service, and Education New Zealand (Manapou ki te Ao). The responsibility for the content of this paper lies with the authors.

REFERENCES

Biehler, R., & Fleischer, Y. (2021). Introducing students to machine learning with decision trees using CODAP and Jupyter Notebooks. *Teaching Statistics*, *43*, 133–142. https://doi.org/10.1111/test.12279

Biehler, R., Frischemeier, D., Reading, C., & Shaughnessy, J. M. (2018). Reasoning about data. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 139–192). Springer. https://doi.org/10.1007/978-3-319-66195-7_5

Budgett, S., & Puloka, M. (2019). Making sense of categorical data–question confusion. In S. Budgett (Eds.), *Decision making based on data. Proceedings of the Satellite Conference of the International Association for Statistical Education*, ISI/IASE.

Burrill, G., & Pfannkuch, M. (2024). Emerging trends in statistics education. *ZDM–Mathematics Education, 56*(1), 19–29. https://doi.org/10.1007/s11858-023-01501-7

Caetano, S-J., de Sousa, B., Fergusson, A., Le, L., Gibbs, A. L., White, B., & Damouras, S. (2023). Putting research into practice: Applying evidence-based principles to foster student learning in statistics and data science. In E. M. Jones (Ed.), *Fostering learning of statistics and data science. Proceedings of the Satellite conference of the International Association for Statistical Education*, ISI/IASE. https://doi.org/10.52041/iase2023.701

Fergusson, A. (2022). *Towards an integration of statistical and computational thinking: Development of a task design framework for introducing code-driven tools through statistical modelling* [Doctoral dissertation, ResearchSpace@ Auckland]. https://hdl.handle.net/2292/64664

Fergusson, A. (2023). Designing positive first experiences with coding for introductory-level data science students. In E. M. Jones (Ed.), *Fostering learning of statistics and data science. Proceedings of the Satellite Conference of the International Association for Statistical Education*, ISI/IASE. https://doi.org/10.52041/iase2023.503

Fergusson, A., & Bolton, L. (2018). Exploring modern data in a large introductory statistics course. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the 10th International Conference on Teaching Statistics*, ISI/IASE.

Fergusson, A., & Pfannkuch, M. (2022a). Introducing high school statistics teachers to predictive modelling and APIs using code-driven tools. *Statistics Education Research Journal*, *21*(2), 8. https://doi.org/10.52041/serj.v21i2.49

Fergusson, A., & Pfannkuch, M. (2022b). Introducing teachers who use GUI-driven tools for the randomization test to code-driven tools. *Mathematical Thinking and Learning*, *24*(4), 336–356. https://doi.org/10.1080/10986065.2021.1922856

Fergusson, A., & Pfannkuch, M. (2024). Using grayscale photos to introduce high school statistics teachers to reasoning with digital image data. *Journal of Statistics and Data Science Education, 32* (4), 345–360. https://doi.org/10.1080/26939169.2024.2351570

Fergusson, A., Pfannkuch, M., & Budgett, S. (2025). Data cleaning doesn't happen in a vacuum: An initial exploration of high school statistics teachers' data practices with messy data. In J. Kaplan & K. Luebke (Eds.), *Connecting data and people for inclusive statistics and data science education. Proceedings of the Roundtable conference of the International Association of Statistics Education*, ISI/IASE. https://doi.org/10.52041/iase24.301

Fergusson, A., & Wild, C. J. (2021). On traversing the data landscape: Introducing APIs to data-science students. *Teaching Statistics*, *43*, S71–S83. https://doi.org/10.1111/test.12266

Fleischer, Y., & Biehler, R. (2025). Exploring students' constructions of data-based decision trees after an introductory teaching unit on machine learning. *ZDM – Mathematics Education*, *57*(1), 153–173. https://doi.org/10.1007/s11858-025-01663-6

Fleischer, Y., Biehler, R., & Schulte, C. (2022). Teaching and learning data-driven machine learning with educationally designed Jupyter Notebooks. *Statistics Education Research Journal*, *21*(2), 7. https://doi.org/10.52041/serj.v21i2.61

Fleisher, Y., Hüsing, S., Biehler, R., Podworny, S., & Schulte, C. (2022). Jupyter notebooks for teaching, learning, and doing data science. In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering & educating today's learners in statistics. Proceedings of the 11th International Conference on Teaching Statistics*. ISI/IASE.

Fleischer, Y., Podworny, S., & Biehler, R. (2024). Teaching and learning to construct data-based decision trees using data cards as the first introduction to machine learning in middle school. *Statistics Education Research Journal*, *23*(1), Article 3. https://doi.org/10.52041/serj.v23i1.450

Fleischer, Y. & Podworny, S. (2022). Teaching machine learning with decision trees in middle school using CODAP. 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. 280–281). Aarhus University. https://doi.org/10.7146/aul.452

Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal*, *16*(1), 22–25. https://doi.org/10.52041/serj.v16i1.209

Hüsing, S. (2021). Epistemic Programming - An insight-driven programming concept for Data Science. In O. Sepällä & A. Petersen (Eds.), *Proceedings of the 21st Koli Calling International Conference on Computing Education Research* (pp. 1–3). ACM. https://doi.org/10.1145/3488042.3490510

Hüsing, S., & Podworny, S. (2021). Computational essays as an approach for reproducible data analysis in lower secondary school. In R. Helenius & E. Falck (Eds.), *Statistics education in the era of data science. Proceedings of the Satellite conference of the International Association for Statistical Education*, ISI/IASE.

Hüsing, S., Schulte, C., Sparmann, S., & Bolte, M. (2024). Using worked examples for engaging in epistemic programming projects. *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1* (pp. 443–449). ACM. https://doi.org/10.1145/3626252.3630961

Hüsing, S., Sparmann, S., Schulte, C., & Bolte, M. (2024). Identifying K-12 students' approaches to using worked examples for epistemic programming. In M. Khamis, Y. Sugano, & L. Sidenmark (Eds.), *Proceedings of the 2024 Symposium on Eye Tracking Research and Applications* (pp. 1–7). ACM. https://doi.org/10.1145/3649902.3655094

Hüsing, S., Schulte, C., & Winkelnkemper, F. (2023). Epistemic programming. In S. Sentence, E. Barendsen, N. R. Howard, & C. Schulte (Eds.), *Computer Science Education: Perspectives on teaching and learning in school.* (pp. 291–304) Bloomsbury Academic. https://doi.org/10.5040/9781350296947.ch-022

Mike, K., & Hazzan, O. (2022). Machine learning for non-majors: A white box approach. *Statistics Education Research Journal*, *21*(2), Article 10. https://doi.org/10.52041/serj.v21i2.45

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., Fleischer, Y., Hüsing, S. (2022). Grade 6 students' perception and use of data-based decision trees. In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering & educating today's learners in statistics. Proceedings of the 11th International Conference on Teaching Statistics*. ISI/IASE.

Podworny, S., Fleischer, Y., Hüsing, S., Biehler, R., Frischemeier, D., Höper, L. & Schulte, C. (2021). Using data cards for teaching data based decision trees in middle school. In O. Sepällä & Petersen (Eds.), *Proceedings of the 21st Koli Calling International Conference on Computing Education Research* (pp. 1–3). ACM. https://doi.org/10.1145/3488042.3489966

Podworny S., Fleischer, Y., Stroop, D. & Biehler R. (2022). An example of rich, real and multivariate survey data for use in school. In J. Hogden, E. Geraniou, G. Bolondi, & F. Ferretti (Eds.), *Proceedings of the Twelfth Congress of the European Society for Research in Mathematics Education (CERME12)*. ERME / Free University of Bozen-Bolzano. https://hal.science/hal-03751842v1

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), Article 6. https://doi.org/10.52041/serj.v21i2.46

Puloka, M. & Pfannkuch, M. (2023). What's in a Pasifika name? Constructing a name dataset. *Statistics and Data Science Educator*. https://sdse.online/lessons/SDSE23-001/

Ridgway, J. (2015). Implications of the data revolution for statistics education. *International Statistical Review, 84*(3), 528–549. https://doi.org/10.1111/insr.12110

Ridgway, J. (2022). *Statistics for empowerment and social engagement. Teaching statistics to develop informed citizens*. Springer. https://doi.org/10.1007/978-3-031-20748-8