A study of the role of data in statistical and mathematical modelling

CAMILLA HELLSTEN ØSTERGAARD AND BRITTA EYRICH JESSEN

This paper examines the role of data in mathematical and statistical modelling, addressing its importance for students' inquiry in mathematics education. The study focuses on how students' actions with data contribute to delimiting problems, constructing models, and validating results. We revisit two Danish case studies: grade five students investigating physical activity with TinkerPlots and upper secondary students designing a facial recognition system with GeoGebra. Using the anthropological theory of the didactic, we analyse students' question–answer processes, milieus, and data moves. Findings show that data served as a driver for inquiry, enabling autonomy and innovation, but also highlighted the need for orchestration and teacher support to develop conceptual understanding. The study underlines potentials and challenges of integrating data-driven modelling in school settings.

Modelling has emerged as a prominent topic in school mathematics over several decades, with a consensus around its pivotal role across primary, secondary, and tertiary levels (Houston et al., 2008). Data is recognised as crucial in modelling activities. Without going into further details we note that data is portrayed as a central component of the modelling process, serving as "the epistemological basis for the different sub-processes" (Blomhøj & Kjeldsen, 2006, p. 166) and playing an important role as a driver for the inquiry process in both statistics and mathematics (Wild & Pfannkuch, 1999; Chevallard, 2019). In this paper, our focus is on the role of data in two rather different case studies on mathematical and statistical modelling respectively. Below, we elaborate on the choice of cases.

Mapping the landscape of modelling, Groshong (2016) strives to delineate types of modelling and highlights their potential roles and functions in primary and secondary mathematics. In her analysis, Groshong (2016), refers to the four basic types of models outlined by Edwards & Hamson

Camilla Hellsten Østergaard, *University College Copenhagen*
Britta Eyrich Jessen, *University of Copenhagen*

Østergaard, C. H. & Jessen, B. E. (2025). A study of the role of data in statistical and mathematical modelling. *Nordic Studies in Mathematics Education*, 30 (4), 195–219.

(2007) arguing that both empirical and mechanistic models can manifest as either deterministic or stochastic. Mechanistic models will rarely draw on data where the empirical models will, where regression can be placed as both deterministic and stochastic models. Groshong explains that: "Deterministic models ignore random variation, so the model outputs are described by their input. Stochastic models [...] include randomness, so they produce different values with given initial conditions" (Groshong, 2016, p. 20). In addition to the categorisation, we see overlaps between mathematical and statistical modelling drawing on data in various forms when focusing on data moves as described by Fitzallen (2013) and McLean and colleagues (2023), in delimiting a modelling problem, or on the validation and analysis process of models. This is already reflected in existing research, though not emphasising the role of data (Groshong, 2016; Kawakami & Minero, 2021, Noll et al., 2022). In the following, we outline how data is addressed or omitted in modelling literature in mathematics and statistics education to point out how our analysis moves this research forward.

Several studies indicated differences between modelling in school contexts and in scholarly contexts, though the latter is assumed to be the reason for the former due to the didactic transposition (Jessen & Kjeldsen, 2021). In school modelling in Nordic countries, we see modelling often represented through modelling competency, modelling with an emphasis on the intra-mathematical components and in very fragmented way detached from scholarly or real-world modelling (e.g. Frejd, 2013; Frejd & Bergsten, 2016; Jessen & Kjeldsen, 2021; Berget, 2022). None of these studies explicitly mentions the role of data. In the modelling cycle referred to by Blomhøj and Kjeldsen (2006), data is considered as one of the elements placed in the middle and affecting all the phases in the cycle. More recent studies, such as Kawakami and Mineno (2021), emphasise the role of data who argue that there exist rich potentials in studying the role of data, and that a "data-based modelling approach can be used for constructing, validating, and revising various models while flexibly combing the mathematical, statistical and contextual approaches [...]" (Kawakami & Mineno, 2021, p. 398). Thus, instead of data being part of modelling in vague forms, they propose to let data be a driver for the modelling process. Despite potentials of letting data play a key role in e.g. validation, analyses indicate that "the validation does not amount to a huge part of the students' modelling activities in general." (Jablonski, 2023, p. 324). So, to harvest these potentials we might need to consider how this can be nurtured through task design and orchestration in classrooms when dealing with empirical deterministic models, but also within stochastic modelling where: "[...] little attention has been paid to empirical investigations of how and why students develop sampling models when investigating a categorical variable whose values are nominal." (Ärlebäck, Frejd & Doerr, 2021, p. 158). Thus, statistics education is of course dealing with data, but less with the role of data.

More generally, within statistics education, we see an aspiration to move in the direction of data modelling inspired by mathematical modelling, but this has other reasons as well: "Another influence was advances in technology, which resulted in innovative software being produced that enabled young students [...] new opportunities were arising to rethink and study the nature and role of statistical modeling." (Pfannkuch et al., 2018, pp. 113-114). This agenda has further been pushed as: "The advent of data science has provided new challenges and opportunities for statistics education researchers to re-think what it means to model. [...] How can new simulations and visualizations using technological tools increase students' access and agency for telling data stories, and enrich understanding in modeling processes?" (Noll et al., 2022, p. 331), and such data-driven understanding and storytelling already happens. Kawakami and Mineno (2021) assert that further research is needed to examine the data-based modelling processes undertaken by students, where the activity may go across the deterministic and stochastic domains. They conducted a study in a lower secondary school, where students were faced with a modelling activity regarding population estimate, where students to various degrees drew on mathematical, statistical or contextual approaches – or combined them. This underscores that certain data modelling endeavours may exhibit both deterministic and stochastic characteristics (cf. Groshong, 2016). However, Jessen and Kjeldsen (2021) note that the epistemological perspective is not always transposed into classroom practices, in particular losing inquiry traits and modelling being a "messy" process. Data modelling may revive these elements of school modelling, as data modelling has been regarded as the core process of inquiry, wherein data is utilized to address genuine questions (Hancock et al., 1992; English, 2012). Data serves not only as an element of mathematisation, employed to illustrate, or theorise the underlying statistical or mathematical structure of information to solve existing problems, but also as the source of statistical and mathematical knowledge, ideas and concepts (Lesh et al., 2008; Wild & Pfannkuch, 1999). Still, more research in this area is needed.

Realisations of data driven inquiries have been explored at university level (Serrano et al., 2010; Bosch et al., 2022), but at secondary school level, these studies do not explicitly link modelling and the teaching of statistics (e.g. Freixanet et al., 2023). Verbisck and colleagues conclude from a study drawing on Study and Research Paths (SRP) that "[...] among the contents activated by the SRP, aspects related to statistics and data processing appeared only tangentially." (Verbisck et al., 2022, p. 5), and the study aligns with earlier studies, which claim that mathematics and statistics diverge in their approaches to emphasising data (e.g. Cobb & Moore, 1997; Gattuso & Ottaviani, 2011). Where statistics always links data to a context and acknowledges the uncertainty inherent in interpreting data and solutions due to data variability, in contrast, mathematics primarily deals with logic and operations concerning numerical values within data, identifying patterns, abstracting generalisations based on data, and "generally adopts a deterministic view of the data and derived interpretations and conclusion" (Kawakami & Mineno, 2021, p. 390). Although certain teacher education contexts emphasize the learning of statistical knowledge, such as Bayesian statistics, over the exploration of data-driven inquiry processes (Hakamata et al., 2022), it is essential for statistical education to foster investigative thinking and the capacity to explore and reason with data, rather than merely teaching descriptive techniques. Therefore, we need to address the role of data in terms of actions and how it supports reasoning. Büscher (2019) emphasised data modelling as a means of structuring phenomena, formalizing communication, and creating evidence. McLean and colleagues (2023) introduced the concept of data moves encompassing six core actions: filtering, grouping, summarising, calculating, merging/joining, and making hierarchy. Data moves serve as precursors to visualisations, aligning with the dynamic process of changing representations to enhance comprehension, as discussed by Wild and Pfannkuch (1999). In both mathematical and statistical modelling, it is crucial to analyse data from diverse perspectives to develop new models for interpretation and predictions. We take the approach of focusing on students' actions with data and how this leads to reasoning when revisiting existing studies on modelling where data are key for the modelling process.

By analysing existing case studies, we seek to find ways to characterise the roles played by the data and linkages to the task design of each case. Barquero and Jessen (2020) argue how different theoretical frameworks impact the kind of modelling activities implemented in classrooms and the ways modelling is considered as content to be taught or if modelling is considered a driver for learning content knowledge while conducting modelling processes. In this paper, we analyse two existing case studies, 'health and physically activity' with grade five students (Østergaard & Larsen, submitted) and 'facial recognition system' with upper secondary school students (Jessen, 2024). The original papers are based on the anthropological theory of the didactic (ATD) and of course this impacts the conclusion we reach in this study as well, but the insights presented here will have general relevance to the modelling community. Below, we present the theoretical background of this paper, before we present our research question regarding the role of data in modelling processes.

The anthropological theory of didactic and data driven modelling

In this paper, we draw on ATD (Chevallard, 2019) and its perspective on modelling, conceptualised as SRP, where a "[...] mathematical activity is essentially a modelling activity in itself" (García et al., 2006, p. 232). Not in the sense that modelling is just another aspect of mathematics, but modelling being more than mathematisation and translations between extra- and intra-mathematical domains. We consider modelling to be the development of answers based on existing knowledge, data, new knowledge and more, pieced together through inquiry processes done by the learner. In this paper, both cases start from an initial question, Q_0 related to the real world, initiating a modelling activity, which leads to the development of answers. Both questions and answers can be either intra-mathematical or linked to the real world.

SRP as a design tool for modelling education, begins with an open generating question Q_0 that is accessible to students but requires their active engagement in inquiry processes to develop an answer; these processes lead to derived questions Q_{ij}, where student may question content from the Q_0. The derived questions lead to study processes where students study new knowledge from various resources or works, W_k. Works are often in school settings: the textbooks, online learning platforms, teacher presentations including lectures, videos (e.g. YouTube), podcasts, or newspapers. To grasp the new knowledge, students draw on their prior learning, which we consider existing answers, A_l^0. Thus, students decompose the new knowledge before piecing it together in terms of producing new answers to Q_{ij}, which is considered (re)construction of knowledge (Barquero et al., 2013; Jessen, 2017; Chevallard, 2019). The reconstruction of knowledge is what we call the research process and may require students to draw on some data, D_m. Data unfamiliar to the student will often require students to study what they are and what they can tell (Chevallard, 2019; Jessen, 2024). Other times, the data is ready to be put into use for the development of some answer, often requiring data moves as those listed by McLean and colleagues (2023). Data can be either qualitative or quantitative in nature.

We can link processes of posing questions and developing answers to the notion of praxeology. All human activities can be described in terms of praxeologies (Chevallard, 2006; Winsløw, 2011), which consist of a *praxis* and a *logos* part. The praxis for how you greet friends can have very different expressions as a nod, a handshake, or a hug in most Nordic countries. In southern Europe cheek kissing is common. Reasons for this lie within the logos, which for most people are rather implicitly given and bound in culture and tradition. Still, we consider the praxis the know-how and the logos the know why (Chevallard, 2006). In mathematics, we divide the "know-how" in type of task, T, and technique(s), τ, which solves the type of task. The type of task might be "find the solutions to a quadratic equation on the form $ax^2+bx+c=0$, where none of the coefficients are zero". One can apply the technique represented by the formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

The "know why" (the techniques solve the problem) is divided into the technology, θ, and the theory, Θ. The technology is defined as the discourse regarding how the technique solves the type of task, and theory as a higher level of justification for the arguments. In this case, being grounded in school algebra for upper secondary. We may add that this answer is an example of intra-mathematical modelling, where the formula above draws on techniques from school algebra pieced together to solve the type of task at hand. Often in schools, students are told both the "know how" and the "know why" by teachers who justify the presentation of this by inviting the students to apply the "know how" on certain problems (Artigue et al., 2020). Engaging students in SRP shifts the focus from transmitting established praxeologies to exploring open questions, fostering a collaborative process where students and teachers construct new praxeologies. The approach also highlights the importance of autonomous inquiry processes, enabling students to actively pursue the construct of logos. We can link questions, posed during an SRP, with tasks and the study of works and data, leading to the construction of answers, which develop students' praxeological equipment, and thereby learning mathematics and modelling (for an elaborate example, see Hakamata et al., 2022).

To analyse processes linked to data moves and the role of data, we draw on the notion of Herbartian Schema, where a group of learners, X, take on the study of a question, Q_0, under the guidance of the teacher(s) y (or Y), while interacting with the milieu, M, to develop an answer, $A^\heartsuit$. In short, the Herbartian schema can be expressed as:

$$[S(X;Y;Q_0) \Rightarrow M] \Rightarrow A^\heartsuit$$

This means that the didactic system, S, consisting of learners, teachers and question addressed, interact with the milieu, M, when developing an answer to the question, which is specific to the system's interaction with the milieu (and at the hearts of the actors). We denote this answer $A^{\heartsuit}$, as it is different from other answers, e.g. those of a textbook. The milieu in ATD is "a system of objects acting as a fragment of "nature" for Q, able to produce objective feedback about its possible answers" (Kidron et al., 2014). The milieu consists of students' prior knowledge, A_l^0, resources and data brought in by teachers or learners, this means the milieu can be described as the set:

$$M = \{A_i^0, \dots, W_j, \dots, D_k, \dots\}$$

Thus, the milieu can be considered elements representing the perceived reality the learners are striving to model and therefore, need to interact with. The entire process can be depicted in question-answer dialectics, mapped as below in a QA-map (see figures 1 and 7), where we in this paper draw on (Jessen, 2024 and include data and works, as they become central for the modelling process.

The indices of questions, answers, data and works in the figures refer to how they are connected, not representing the chronology of the modelling process. The final questions often become close to type of tasks, and therefore the map represents the praxeological organisation realised by the didactic system, *S*, where ideally elements of logos are developed.

Linking this with the meanings of data modelling presented by Büscher (2019), we see that studying the milieu in terms of works, W_j, and data, D_k, can be a way to describe and structure the modelling process. Questioning and formulating partial or whole answers are ways to formalise communication and develop new answers and questions in terms of praxeological organisations against the milieu containing data providing objective feedback (Kidron et al., 2014). This is a way to describe how data modelling is creating evidence (Büscher, 2019). Moreover, when identifying the praxis part of the praxeologies, we are identifying the data moves, and the core actions of filtering, grouping, summarising, calculating, merging/joining, and making hierarchy (McLean et al., 2023). We consider these actions core techniques drawn upon by students and drivers for building praxeologies in the same sense as moves lead to visualisations, and thereby the dynamic processes of changing representations which may entail comprehension (Wild & Pfannkuch,1999) and the development of more coherent praxeological organisations. Therefore, when analysing the role of data, we are interested in the dialectics between students studying data sets and the students' construction of new knowledge or answers based on the data. Thus, we are in the dialectics between questions and answers, study and research, as well as media and milieu and if this links to specific data moves or actions. This leads us to formulate the research questions of this paper: what roles played by data are identified in the question-answer and media-milieu dialectics? And what techniques are identified during the cases?

Methodology

In this paper, we revisit two previously conducted descriptive case studies (Yin, 2013; 2014), both based on in-depth classroom observations and analyses of two implemented Danish SRPs—one situated in a primary school context, the other in upper secondary school. These previous implementations aimed to explore, implement, and analyse real-world problems in mathematics education (Østergaard & Larsen, submitted; Østergaard & Larsen, 2023). The first case study, which centres on the theme of health and physical activity, is grounded in the theoretical principles of SRP and conducted within a design study framework (Bakker, 2018). This case draws on empirical data from the third iteration of a collaborative project involving three primary school teachers and a researcher with expertise in statistics education. The study follows the design study methodology incorporating preliminary, a priori, and in vivo analyses as described in the original publication (Østergaard & Larsen, submitted). The second case, focusing on a facial recognition system, is based on data from the realisation of a reform effort and examines the integration of technology and inquiry into teaching practices analysed through the lens of SRP. Unlike the first case, the design of this study was developed entirely by the teacher in collaboration with teacher colleagues. It includes *in vivo* and *a posteriori* analyses, as discussed in the original study (Jessen, 2024).

In our analysis, we revisited the a posteriori analysis drawing on SRP as an analytic framework (Hansen & Winsløw, 2011; Winsløw et al., 2013). This involved examining classroom activities, including whole-class discussions and group work, by identifying significant contributions from both students and the teacher all related to the overarching question, Q_0, to examine the role of data in the two cases.

The data are analysed initially by identifying the questions addressed in the teacher and student classroom contributions in the classroom, and whether they played a role in initiating or advancing the development of statistical or mathematical knowledge. Some of these contributions explicitly called for new techniques or concepts not yet developed in the classroom dialogue, such as proposals for further data collection, hypotheses requiring mathematical justification, or suggestions that implicitly pointed towards the need for specific statistical tools. Dialogue was identified as answers if it provided discourse that could support the resolution of the posed questions. Specifically, answers were defined as those contributions that incorporated aspects of statistical or mathematical techniques, technologies, or theories. The identified questions and answers were depicted in a QA map. In parallel, we identified media and milieus dialectic. This means we identified the media consulted by students and the milieu against which the preliminary answers or models were tested. We described this as specific branches of the QA map using the Herbartian schema (Chevallard, 2008, 2019). Here we emphasise the prior learning, works and data from different media drawn upon by students. The praxeological modelling process can be identified in the QA map. In the last part of the analysis, we identified data moves (McLean et al., 2023) linked to employed praxeologies linked and to the nature of the data drawn upon.

Analysis

Both cases are conducted on Danish school contexts, where modelling has played a dominant role for decades (Niss & Jensen, 2002; Niss, 2018; Niss & Højgaard, 2019). For grade 4-6 students, the curriculum states that students should be able to complete simple modelling processes and know these. The student should be able to apply and know about existing models. Statistics is not described in terms of modelling, but students should be able to complete statistical inquiries, find and compare simple descriptors. For both mathematical modelling and statistics, students should be able to handle real world situations and problems. The modelling process is formulated as a four-step cyclic activity starting from formulation of problem, translating it into a mathematical model, treatment of the model, and interpretation of the model and its results in relation to the initial problem (Danish Ministry of Education, 2019).

For upper secondary mathematics, the curriculum lists modelling as part of the identity of mathematics, disciplinary goals (linked to stochastics as well), and a core element of mathematics. The curriculum promotes a similar simplified modelling cycle to describe the process but also favours autonomous inquiry processes where students formulate and pursue their own questions (Danish Ministry of Education, 2017). In this sense, the upper secondary curriculum can favour SRP activities (Jessen & Kjeldsen, 2021). In addition to the upper secondary school, there exist guidelines for cross-educational competences and abilities, such as innovation. Thus, all disciplines must include creative and innovative activities (Danish Ministry of Education, 2022). The second case study was an attempt by teachers to capture this (Jessen, 2024). However, in this paper, we simply analyse the activity of developing a facial recognition system, as a modelling activity. When presenting each case, we start by presenting the local context of the in vivo analysis.

'Health and physical activity'

The case of 'health and physical activity' presents how Danish grade-five students dealt the generating question, Q_0: 'Are we physically active?' in a sequence of eight mathematics lessons. The students were novices to SRP based activities, making sense of data and analysing data using the dynamic visualisation software, TinkerPlots (Konold & Miller, 2021). The aim of the activity was to engage students in real-world questions, formulate problems, generate and analyse data, and interpret and validate the results. The students were used to explore small textbook tasks, without sharing in plenum discussions regarding broader statistical concepts. Previously, we have used data from the class discussing students' development of statistical literacy and statistical reasoning (Østergaard & Larsen, submitted). In this analysis, we focus on the crucial role of data in modelling using technology.

The QA-map in figure 1 visualises grade-five students' study and research process generated from Q_0, how students formulated a hypothesis, H_1, researched for established answers, $A_2^0 \dots A_{15}^0$, in provided media (newspaper, articles). E.g. A_2^0: 97% of girls in the world invest less than one hour a day in physical activities. The students posed derived questions $Q_{2,1} \dots Q_{15,1}$, where e.g. $Q_{3,1}$: What are the number of steps children walk a day? and $Q_{5,1}$: Are boys more active than girls? To answer these derived questions, the students explored self-generated data by structuring, filtering, grouping and connecting the data from different perspectives and developed new answers, $A_{2,1,1} \dots A_{2,1,x}$. These answers were discussed in plenum, where the teacher through new derived questions, Q', guided the students to visit and discuss works, W_j, which included the logos of statistical descriptors, e.g. mean and frequency and relative frequency.

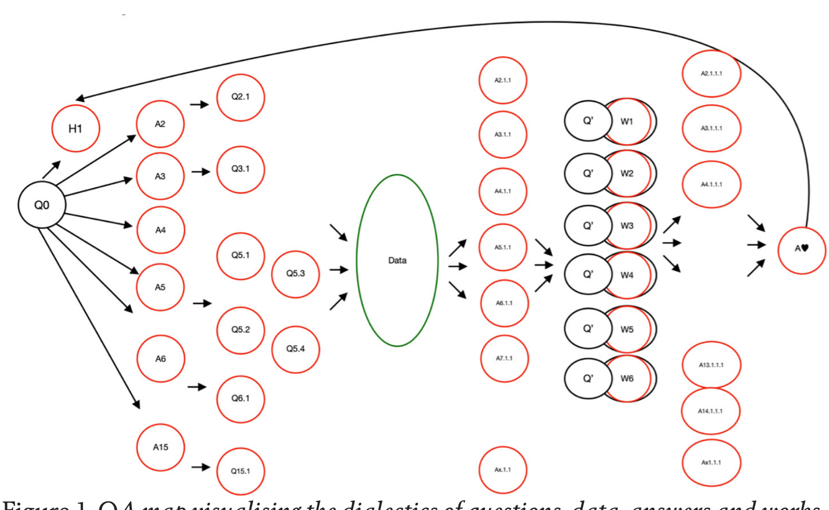

Figure 1. Q*A-map visualising the dialectics of questions, data, answers and works. Teacher contributions are in black, and students' contributions in red.

Thus, we see a lot of initial questions creating a need for data, which again leads to some first answers where the teacher invites students to revisit or study the statistical descriptors. An integral component of the milieu, M_1, was the data collected by students through a survey on their physical activities. The data, D_k, comprised 15 questions concerning various aspects such as gender, sleeping habits, time spent bike riding, daily step count, duration of digital device usage, and more. A total of 67 students responded to the survey. The data served as the focal point of students' inquiry, propelling their study and research endeavours forward. Another part of the milieu was the technology used, TinkerPlots, which proposed a new way for students to access, analyse, and visualise data, which scaffolded students' experiments with data (Konold & Higgins, 2003). In TinkerPlots students were encouraged to apply informal techniques (Pfannkuch et al., 2018), enabling them to actively model data by organising it into groups, stacking it to create new visualisations, and simultaneously exploring multiple observations. These techniques allowed students to construct their own intuitive models, fostering an understanding of trends and patterns within the data.

Some of the instrumented techniques offered by TinkerPlots support statistical creativity, particularly by enabling students to construct new graphical representations of data. These techniques were used by students to explore and compare different types of diagrams—analysing and interpreting them to assess whether, and how, a given diagram could help answer the derived question. In this process, TinkerPlots continuously shifted roles: from being part of the milieu in which students interact with data through actions such as clicking, dragging, and stacking, to becoming a tool for statistical inquiry and statistical interpretation. This exploration led to the construction of visualisations and the formulation of partial answers, which required further investigation and comparison with students' prior knowledge and their perception of the real-world context previously studied. In this sense, TinkerPlots not only functioned as a dynamic part of the exploratory milieu but also gradually became a media—or even a form of work—in itself. Thus, the initial milieu for data exploration can be described as:

$$M_1 = \{A_i^0, \dots, D_k\}$$

Which is elaborated when students and teachers collectively visited statistical work, W_l, and teachers raised questions regarding the statistical descriptors, Q_{ij}, to challenge students' newly developed answers, $A_{2,1,1}, \dots$:

$$M_2 = \{A_{2,1,1} \dots, D_k, Q', W_l \dots\}$$

In the subsequent analysis, we elaborate on four selected examples of students' data moves when modelling data.

Due to health recommends, students raise the derived question $Q_{3,2}$: What are the number of steps children walk a day? The first example of data modelling illustrates how students saw the data as pointers (Konold et al., 2015) and researched the data card (figure 2 to the left) to find the card which represent them and their friends. The students further saw data as a case value of an attribute (Konold et al., 2015), e.g. stating "Ahh, this student spends more than five hours a day using a digital device" (figure 2 to the right, discussing the yellow circle).

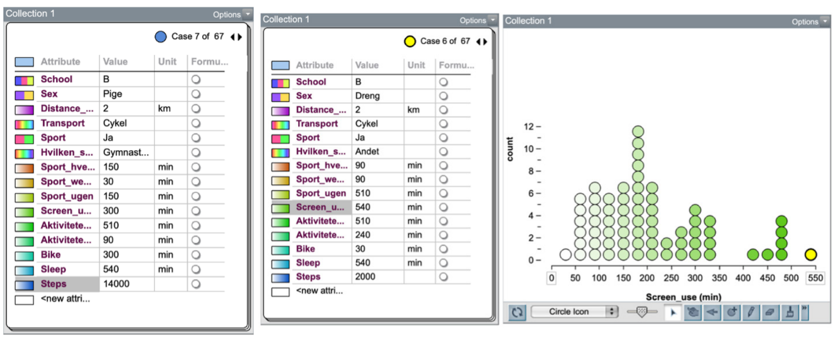

Figure 2. *Examples of focusing on data as a case value of attributes.*

The second example of data driven modelling illustrates how students structured the data of steps in less formal ways by separating and stacking data and by constructing pie charts, and how students modelled the data into visual representations with little connection to Q_0 and the derived questions, e.g. comparing steps with steps (figure 3 to the left) and stacking value bars, which illustrate the total sum of all steps of students (figure 3 to the right). The students did not know how to read or interpret their new graphical models, and therefore they were initially unable to study them.

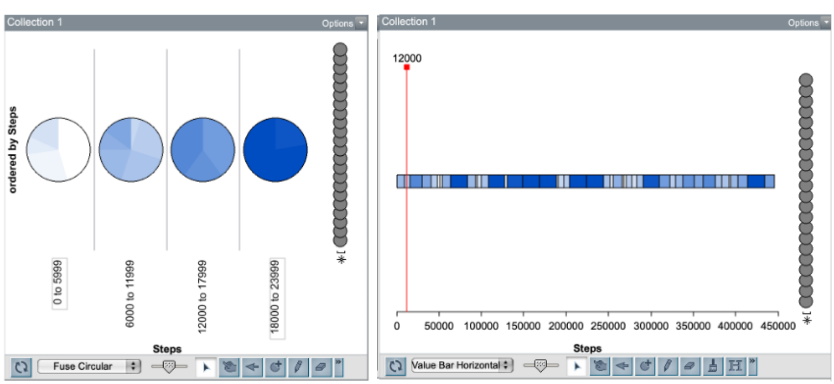

Figure 3. *Visual representation of informal methods for organising data.*

The third example of data driven modelling illustrates how the data were used as a classifier (Konold et al., 2015), finding the mean, and how students modelled data as an aggregate (Konold et al., 2015) when comparing students at different schools.

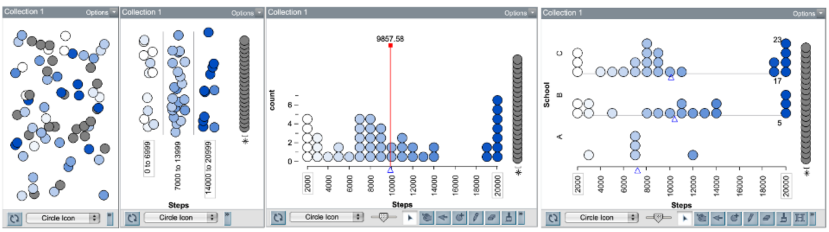

Figure 4. *Data as a classifier and data as an aggregate.*

In figure 4 we see how the students have used an instrumented technique in TinkerPlots to find the mean. The instrumented process is inquiry-based but does not display logos.

The fourth example of data driven modelling illustrates how students' models of data, in TinkerPlots, were used to visit different statistical works. In figure 4, we see two graphical representations which form the base to build students' understanding of the concept 'how many' (frequency and relative frequency) and of the 'mean' as a "statistical idea for describing and understanding data and as an algorithm for solving statistical problems" (Østergaard & Larsen, 2023). For students to understand the statistical concepts, the teacher and students visited statistical works, posed new derived questions, e.g. "What is a mean?" to bridge praxis and logos with a request for the students to deconstruct the algorithm of the arithmetic mean and describe and elaborate the (instrumented) technique, e.g. how the bar graph (figure 5 the right) can visualise the mean and the concept of add and divide. This might support the building of logos of a praxeology solved by instrumented techniques. It may lack theoretical grounding, but TinkerPlots makes it tangible.

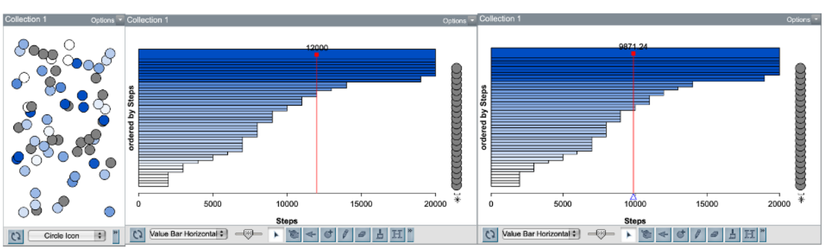

Figure 5. *Reference lines illustrating key questions: "How many?" (representing the recommended daily number of steps) and "What is the mean?" (representing the mean).*

In the above four examples of data driven modelling, the students studied different statistical descriptors in plenum discussions, where the teacher and other students functioned as media presenting works (notions) to each other to be studied further. Also, student researched and explored the descriptors in the digital milieu offered by TinkerPlots. Students' activities alternated throughout the SRP between students' research with data and the study process, where the students, together with the teacher, studied the statistical models, descriptors and answers found in media or developed in the groups. The dialectic of media and milieus became visible when the students developed new knowledge to improve their answers.

'Facial recognition system'

The second case is teaching mathematics applied to a real-world context, aiming to develop students' innovative competences in upper secondary school, grade 10. The activity was designed by the teacher, covered two hours of teaching in the late spring, where students have been taught vectors in 2D including a little geometry, notions of functions, descriptive statistics, and regression. The most recent topic addressed were polynomials with an emphasis on parabolas and second-degree equations. The design included an open question and a handout with a face printed on it, including a grid representing a coordinate system, with points and coordinates given. After each lesson, plenum sessions for sharing were planned. The focus was for students to experience through inquiry how mathematics can be used to produce facial recognition systems. Previously, the case has been analysed in Jessen (2024), discussing the potentials of the design in terms of engaging students autonomously in inquiry processes (Bosch & Winsløw, 2016, Jessen, 2017), carrying the potential of revisiting geometry learning and developing this further regarding properties of equilateral triangles (Jessen, 2024). Still, from the perspective of mathematical modelling as a data driven process, the case is relevant to revisit. The generating question of this SRP is, Q_0:

> "Q_0,: You are designers of digital solutions to security systems. You have a costumer who needs a facial recognition system. For efficiency the system should only require three measures in order to recognise people. What measures are needed and why?" (Jessen, 2024).

The entire class performed very differently during the two hours. In this analysis, we focus on one group being most elaborate in their use of data. Using the Herbartian schema, we see that the class is handed the problem, Q_0, the handout, and data, D_1, in the sense that the picture can be considered qualitative data for the design of the facial recognition system, which is prompted to draw on mathematics using the coordinates of the points to some calculations, constructions, etc. Thus, the teachers handed over a milieu consisting of $M = \{A_1^0, A_2^0, \dots, D_1\}$, where A_i^0 is students' existing knowledge. Initially, students discussed what counts as a measure, Q_1, and if one set of coordinates can count as such, $A_1^{\heartsuit}$, then, Q_2, which ones to choose? And Q_3, what kind of mathematical object can be constructed from three points in the plane? They considered polynomials, $Q_{3,1}$, and linear functions, $Q_{3,2}$ and if their existing knowledge regarding those notions could be used to model the system. The answers to these questions required the notion of functions and represent models of the facial recognition system. This process can be described as:

$$[S(X_1; y; Q_3, Q_{3,1}, Q_{3,2}) \rightarrow \{A_1^0, A_2^0, \dots, D_2\}] \rightarrow A_{3,1}^{\heartsuit} \cup A_{3,2}^{\heartsuit}$$

The answers were studied by students, who discovered that regardless of the points chosen for the regression, then $R^2 = 1$. From their existing knowledge, A_i^0, they knew this would not be true. So, they questioned, $Q_{3,3}$, if triangles would be a suitable basis for the facial recognition system.

First, students worked with pen and paper from D_1, and then moved their analyses to GeoGebra (GeoGebra Team, 2023), as they had the handout available electronically, which made the analyses etc., more swift procedures, see figure 6.

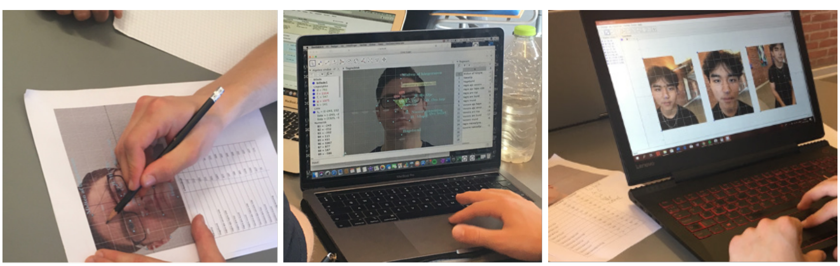

Figure 6. *The left picture shows students' initial work with the handout, in the middle, the initial work in GeoGebra. The right picture shows their own pictures when testing hypotheses (Jessen, 2024).*

We consider this part of the students' data moves. We see students filter data and calculate with data in order to structure the problem. The calculations lead to models validated against new electronic data, D_2. Similar moves were activated when questioning which triangle, and what properties, should be modelled. At this point, the teacher organised a plenum session sharing ideas across groups. Students from the group we followed got confident that triangles were the most promising idea to pursue for lesson two. The mapping of the processes can be seen in the QA-map in figure 7.

The dashed lines indicate that the students e.g. jumped from the question of which points to choose, Q_2, to what mathematical object, Q_3, acknowledging that the choice of points is related to the mathematical object. We see that students go back and forth between digital data, D_2, and functions, trying to make sense of them as descriptions of facial traits and shapes. Doing this, they draw on existing knowledge, $A^0_{3,1}$ and $A^0_{3,2}$, and are seeking to model the system, mainly through research and construction, without the study of works and new knowledge, though rejecting their choices based on prior incomplete praxeologies.

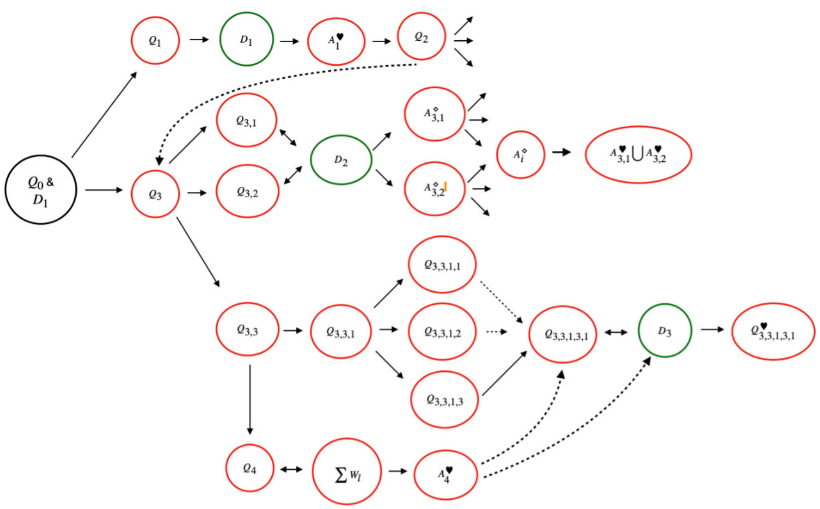

Figure 7: *The QA-map showing the work on facial recognition with students posing questions, visiting works, drawing on existing knowledge when developing their own answers.*

The second lesson starts with the students discussing, $Q_{3,3,1}$, what properties of triangles could be preserved? This happened in parallel with exploring the Q_4: how do facial recognition systems work in the real world? To develop answers, the student studied webpages from companies selling security systems, collecting information regarding what they promised the customers. We consider this a collection of works studied by the students, ΣW_1, which again led to the answer, $A_4^{\heartsuit}$: Most systems can recognise faces from different angles and distances, though the faces need to be relatively close to some camera. This became part of the milieu against which the following answers were validated. Thus, this process can be described by the following schema, where students studied works, decomposed this knowledge and reconstructed it into an answer regarding the main traits of such systems relevant for their modelling problem:

$$[S(X_1;y;Q_4) \Rightarrow \{W_l\}] \Rightarrow A_4^{\heartsuit}$$

This path can look like a side track to the main question regarding properties of triangles, but as indicated with dashed arrows in figure 7, the answer, $A_4^{\heartsuit}$, became crucial for developing new data and models.

Thus, from questioning properties, $Q_{3,3,1}$, the students name those based on prior learning (other A_i^0's), such as side length, $Q_{3,3,1,1}$, area, $Q_{3,3,1,2}$, inner angles of the triangle, $Q_{3,3,1,3}$. They consider how they can know if computers can differ one person from another, revisiting the data of a man's face on GeoGebra, D_2. They measure area, side lengths and angles, but seem in doubt. Then, they decide, to compare with one of their own faces, and see a difference, though not concluding anything, but to pursue the preservation of angles in a triangle between eyes and nose. They draw on $A_4^{\heartsuit}$, when choosing to take more pictures from different angles and distances, D_3 (right picture in figure 6). The dashed lines in figure 2 indicate the dialectic between questions, properties, existing answers, and knowledge on facial recognition, which leads them to discover that angles are preserved. A greater distance from the camera provides a smaller area and side length, but angles are the same.

$$[S(X_1;y;Q_{3,3,1,1}, Q_{3,3,1,2}, Q_{3,3,1,3}, Q_{3,3,1,3,1}) \Rightarrow \{A_i^0, \dots, A_4^{\heartsuit}, D_2, D_3\}] \Rightarrow A_{3,3,1,3,1}^{\heartsuit}$$

The reason for collecting the four questions in the didactic system, is that they are all explored against the milieu creating the need for new set of data D_3. There is a lot of going back and forth between studying the model they were developing, existing answers, new knowledge and data available. The data moves were again filtering, grouping, calculating and structuring the problem. When the lesson ended, they were considering collecting more data, D_4, to decide what was an acceptable deviation of decimals when calculating the angles from one picture to the next. The group never mentioned equilateral triangles, which can be interpreted as ways to model a certain step of the system considered.

Discussion and concluding remarks

Comparing the two cases, we conclude that for the first case study, initial data handed out to the students creates a need to study and question those, and that autonomy of recognising themselves in data makes the data and the milieu they are part of, relatable. In the second case, students are only generating data after the first exploration of those handed out. In the first case, data are quantitative. In the second, data are initially qualitative, but the requirement to choose three measures nudges students to translate pictures into points and data for regression and other calculations. Despite the differences between the cases, we see similar data moves activated in order to structure the problem and system to be modelled, develop partial answers, study and validate those, visualise students' thinking about the real-world problem and therefore the students' answers and models (Büscher, 2019; McLean et al., 2023). We see similar data moves linked to different techniques in the two cases, allowing students to develop different praxeologies. Therefore, the praxeological analysis combined with data moves, more clearly maps out the role data in inquiry-based modelling activities.

We see differences in the cases when considering the role of data in validating the developed models. In the second case, more data are produced, allowing the students either to reject their hypothesis of preserving angles in the triangle or create more data, to elaborate the model further. Here GeoGebra supports this activity, as it is easy to import facial pictures, do the calculations etc. In the first case, TinkerPlots allow students to explore statistical techniques, but there is no feedback from the system indicating meaningless visualisations (as those in figure 3). Without a developed logos for the techniques, students need the nudging from the teachers to revisit and study logos of the applied techniques in relation to the models considered. Here, the secondary students possess an autonomy in using their prior knowledge to reject regressions always leading to $R^2 = 1$. However, without the nudging from a teacher discovering this, they are not building the logos justifying this result. We see that Groshong's (2016) categorisations of modelling may hold true, but also that students in the second case go back and forth between deterministic and stochastic modelling and might have done the same between empirical and mechanical models, if not asked, to use three measures. Without the survey data set in the first case, students might have found other than statistical descriptors to answer the question on activity levels.

We see that data linked to the generating questions ignite the media-milieu dialectics, the question-answer dialectics, and the dialectics of study and research. Thus, we might hypothesise that data – qualitative or quantitative – is productive for modelling processes inviting students to engage in model construction. Still, the orchestration of the mentioned dialectics is crucial for the realisation of the learning potentials of such activities. Thus, to fully explore the potential, we might need to link data driven modelling and the use of digital technologies. We see potential for analysing data-driven modelling from the perspective of SRP and the Herbartian schema, as ATD enables analysis of actions and moves closely tied to the disciplines and data in terms of praxeologies, and how these in turn connect to reality as represented through media and milieu.

References

Ärlebäck, J. B., Frejd, P., & Doerr, H. M. (2021). Exploring students' models of sampling and inference with nominal variables. *Quadrante*, *30*(1), 158–177. https://doi.org/10.48489/quadrante.23655

Artigue, M., Bosch, M., Doorman, M., Juhász, P., Kvasz, L., & Maass, K. (2020). Inquiry based mathematics education and the development of learning trajectories. *Teaching Mathematics and Computer Science*, *18*(3), 63–89. https://doi.org/10.5485/TMCS.2020.0505

Bakker, A. (2018). *Design research in education: A practical guide for early career researchers*. Routledge. https://doi.org/10.4324/9780203701010

Barquero, B., Bosch, M., & Gascón, J. (2013). The ecological dimension in the teaching of mathematical modelling at university. *Recherches en didactique des mathématiques*, *33*(3), 307–338.

Barquero, B., & Jessen, B. E. (2020). Impact of theoretical perspectives on the design of mathematical modelling tasks. *Avances de Investigacion en Educacion Matematica*, *17*, 98-113. https://doi.org/10.35763/AIEM.V0I17.317

Berget, I. K. L. (2022). Mathematical modelling in textbook tasks and national examination in Norwegian upper secondary school. *Nordic Studies in Mathematics Education*, *27*(1), 51-70.

Blomhøj, M. & Kjeldsen, T. H. (2006). Teaching mathematical modelling through project work. *ZDM Mathematics Education*, *38*, 163–177. https://doi.org/10.1007/BF02655887

Bosch, M. & Winsløw, C. (2016). Linking problem solving and mathematical contents: the challenge of sustainable study and research processes. *Recherches en Didactique des Mathématiques*, *35*(3), 357–401.

Bosch, M., Florensa, I., Markulin, K., & Ruiz-Munzon, N. (2022). Real or Fake Inquiries? Study and Research Paths in Statistics and Engineering Education. In R. Biehler, M. Liebendörfer, G. Gueudet, C. Rasmussen, & C. Winsløw (Eds.), *Practice-Oriented Research in Tertiary Mathematics Education. Advances in Mathematics Education*. Springer. https://doi.org/10.1007/978-3-031-14175-1_19

Büscher, C. (2019). Students' development of measures. In G. Burrill & D. Ben-Zvi (Eds.), *Topics and Trends in Current Statistics Education Research: International Perspectives, ICME-13 Monographs* (pp. 27–50). Springer International Publishing. https://doi.org/10.1007/978-3-030-03472-6_2

Chevallard, Y. (2006). Steps towards a new epistemology in mathematics education. In M. Bosch (Ed.), *Proceedings of the Fourth Congress of the European Society for Research in Mathematics Education* (pp. 22-30). FUNDEMI IQS – Universitat Ramon Llull and ERME.

Chevallard, Y. (2008). Afterthoughts on a seeming didactic paradox. *Cadre du Colloque International'Efficacité y équité en éducation* (pp. 1-6).

Chevallard, Y. (2019). Introducing the anthropological theory of the didactic: An attempt at a principled approach. *Hiroshima Journal of Mathematics Education*, *12*, 71–114. https://doi.org/10.24529/hjme.1205

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, *104*(9), 801–823. https://doi.org/10.1080/00029890.1997.11990723

Danish Ministry of Education. (2017). *Matematik B – stx, august 2017, Bilag 112*. https://www.uvm.dk/-/media/filer/uvm/gym-laereplaner-2017/stx/matematik-b-stx-august-2017-ua.pdf

Danish Ministry of Education. (2019). *Matematik fælles mål*. https://emu.dk/sites/default/files/2020-09/GSK_F%C3%A6llesM%C3%A5l_Matematik.pdf

Danish Ministry of Education. (2022). *Vejledning til de tværgående kompetenceområder på de gymnasiale uddannelser.* https://www.uvm.dk/-/media/filer/uvm/udd/gym/pdf22/juni/220629-vejledning-til-de-tvaergaaende-kompetenceomraader-paa-de-gymnasiale-uddannelser.pdf

Edwards, D. & Hamson, M. (2007). *Guide to Mathematical Modelling.* 2nd Ed. Industrial Press, Inc.

English, L.D. (2012) Data modelling with first-grade students. *Educational Studies in Mathematics*, *81*, 15–30. https://doi.org/10.1007/s10649-011-9377-3

Fitzallen, N. E. (2013). Characterising students' interaction with TinkerPlots. *Technology Innovations in Statistics Education*, *7*(1). https://escholarship.org/uc/item/5pv5j7m2

Freixanet, M. J., Alsina, M., & Bosch, M. (2023). Statistics inquiry and the students' engagement in the change of paradigm. In P. Drijvers, C. Csapod, H. Palmér, K. Gosztonyi, & E. Kónya (Eds.), *Thirteenth Congress of the European Society for Research in Mathematics Education*. Alfréd Rényi Institute of Mathematics and ERME.

Frejd, P. (2013). An analysis of mathematical modelling in Swedish textbooks in upper secondary school. *Nordic Studies in Mathematics Education*, *18*(3), 59–95.

Frejd, P. & Bergsten, C. (2016). Mathematical modelling as a professional task. *Educational Studies in Mathematics*, *91*, 11–35. https://doi.org/10.1007/s10649-015-9654-7

Garcia, F. J., Pérez, J. G., Higueras, L. R., & Bosch, M. (2006). Mathematical modelling as a tool for the connection of school mathematics. *ZDM Mathematics Education*, *38*, 226–246. https://doi.org/10.1007/BF02652807

Gattuso, L., & Ottaviani, M. G. (2011). Complementing mathematical thinking and statistical thinking in school mathematics. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education* (pp. 121–132). Springer. https://doi.org/10.1007/978-94-007-1131-0_15

GeoGebra Team. (2023). *GeoGebra Classic* (Version 6.0) [Computer software]. GeoGebra. https://www.geogebra.org/

Groshong, K. (2016). Different types of mathematical models. In C. R. Hirsch & A. R. McDuffie (Eds.), *Annual Perspectives in Mathematics Education 2016: Mathematical modeling and modeling mathematics* (pp. 17–24). National Council for Teachers of Mathematics.

Hakamata, R., Otaki, K., Fukuda, H., & Otani, H. (2022). Statistical modelling in the Brousseaunian guessing game: A case of teacher education in Japan. In M. Trigueros, B. Barquero, R. Hochmuth, & J. Peters (Eds.), *Proceedings of the Fourth Conference of the International Network for Didactic Research in University Mathematics*. Leibniz Universität.

Hancock, C., Kaput, J. J., & Goldsmith, L. T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist*, *27*(3), 337–364. https://doi.org/10.1207/s15326985ep2703_5

Hansen, B., & Winsløw, C. (2011). Research and study course diagrams as an analytic tool: The case of bi-disciplinary projects combining mathematics and history. In Bosch, M., Gascón, J., Ruiz Olarría, A., Artaud, M., Bronner, A., Chevallard, Y., Cirade, G., Ladage, C. & Larguier, M. (Eds.) *3rd Conference of the Anthropological Theory of Didactics* (pp. 685–694). Centre de Recerca Matemàtica

Houston, K., Ireland, N., Galbraith, P., & Kaiser, G. (2009). *The first twenty-five years: The first century of the International Commission on Mathematical Instruction (1908–2008)*. http://www.icmihistory.unito.it (retrieved april 15, 2024)

Jablonski, S. (2023). Is it all about the setting?—A comparison of mathematical modelling with real objects and their representation. *Educational Studies in Mathematics*, *113*(2), 307–330. https://doi.org/10.1007/s10649-023-10215-2

Jessen, B. E. (2017). How to generate autonomous questioning in secondary mathematics teaching? *Recherches en Didactique des Mathematiques*, *37*(2/3), 217–245.

Jessen, B. E. (2019). Teachers learning to design and implement mathematical modelling activities through collaboration. In U. T. Jankvist, M. Van den Heuvel-Panhuizen , & M. Veldhuis (Eds.), *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education* (pp. 1182-1189). Freudenthal Group & Freudenthal Institute.

Jessen, B. E., & Kjeldsen, T. H. (2021). Mathematical modelling in scientific contexts and in Danish upper secondary education: are there any relations?. Q*uadrante*, *30*(2), 37-57. http://doi.org/10.48489/quadrate.23658

Jessen, B.E. (2024). Media and Milieu Dialectics and New Resources Available for Students. In I. Florensa, N. Ruiz-Munzón, K. Markulin, B. Barquero, M. Bosch, Y. Chevallard (Eds), *Proceedings of the 7th International Conference on the Anthropological Theory of the Didactic (CITAD7)* (pp. 293–302). Birkhäuser, Cham. https://doi.org/10.1007/978-3-031-55939-6_23

Kawakami, T., & Mineno, K. (2021). Data-based modelling to combine mathematical, statistical, and contextual approaches: Focusing on ninth-grade students. In F. K. S. Leung, G. A. Stillman, G. Kaiser, & K. L. Wong (Eds.), *Mathematical modelling education in East and West: International perspectives on the teaching and learning of mathematical modelling* (pp. 389–400). Springer International Publishing. https://doi.org/10.1007/978-3-030-66996-6_32

Kidron, I., Artigue, M., Bosch, M., Dreyfus, T., & Haspekian, M. (2014). Context, milieu, and media-milieus dialectic: A case study on networking of AiC, TDS, and ATD. In A. Birkner-Ahsbahs & S. Prediger (Eds.), *Networking of theories as a research practice in mathematics education* (pp. 153–177). https://doi.org/10.1007/978-3-319-05389-9_10

Konold, C., & Higgins, T. L. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 193–215). National Council of Teachers of Mathematics.

Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics*, *88*(3), 305-325. https://doi.org/10.1007/s10649-013-9529-8

Konold, C., & Miller, C. (2021). *TinkerPlots: Dynamic data exploration* (Version 3.3) [Computer software]. Data Description, Inc. https://www.tinkerplots.com/

Lesh, R., Middleton, J. A., Caylor, E., & Gupta, S. (2008). A science need: Designing tasks to engage students in modeling complex data. *Educational studies in Mathematics*, *68*, 113-130. https://doi.org/10.1007/s10649-008-9118-4

McLean, J.A., Brady, C., Jung, H., Dominguez, A., & Glancy, A.W. (2023). Seeing the Forest for the Trees: Investigating Students' Data Moves in a Citizen Science Based Model-Eliciting Activity. In Greefrath, G., Carreira, S., & Stillman, G.A. (Eds) *Advancing and Consolidating Mathematical Modelling. International Perspectives on the Teaching and Learning of Mathematical Modelling* (pp. 193–204). Springer, Cham. https://doi.org/10.1007/978-3-031-27115-1_12

Niss, M. (2018). National and international curricular use of competency based Danish "KOM project". In Y. Shimizu & R. Vithal (Eds.), *Proceedings of ICMI Study 24: School mathematics curriculum reforms: Challenges, changes and opportunities* (pp. 69–76). University of Tsukuba.

Niss, M. & Højgaard, T. Mathematical competencies revisited (2019). *Educational Studies in Mathematics*, *102*, 9–28. https://doi.org/10.1007/s10649-019-09903-9

Niss, M. & Jensen, T. H. (2002) *Kompetencer og matematiklæring—Ideer og inspiration til udvikling af matematikundervisning i Danmark* [Competencies and mathematical learning—ideas and inspiration for the development of mathematics teaching and learning in Denmark]. Copenhagen: The Ministry of Education. Available electronically—search for "Kompetencer og matematiklæring".

Noll, J., Schnell, S., Gould, R., & Makar, K. (2022). New ways of interacting with data, context, and chance in statistical modeling processes. *Mathematical Thinking and Learning*, *24*(4), 331–335. https://doi.org/10.1080/10986065.2021.1922855

Østergaard, C. H. (2022). An Inquiry Perspective on Statistics in Lower Secondary School in Denmark and Japan–An Elaboration and Modelling of the Anthropological Theory of the Didactic Through Two Statistics Classrooms. *European Journal of Science and Mathematics Education*, *10*(4), 529-546. https://doi.org/10.30935/scimath/12401

Østergaard, C. H., & Larsen, D. M. (2023). Questioning the world–A dialectic of study and research. In P. Drijvers, C. Csapodi, H. Palmér, K. Gosztonyi & E. Kónya (Eds.), *Thirteenth Congress of the European Society for Research in Mathematics Education* (CERME13) (No. 25). Alfréd Rényi Institute of Mathematics, ERME.

Østergaard, C. H. & Larsen, D. M. (submitted). More than inquiry – Analysing two inquiry approach in statistical education. *Statistics Education Research Journal.*

Pfannkuch, M., Ben-Zvi, D., & Budgett, S. (2018). Innovations in statistical modeling to connect data, chance and context. *ZDM Mathematics Education*, *50*(7), 1113-1123. https://doi.org/10.1007/s11858-018-0989-2

Serrano, L., Bosch, M., & Gascón, J. (2010). Fitting models to data: The mathematising step in the modelling process. In V. Durand-Guerrier, S. Soury-Lavergne, & F. Arzarello (Eds.), *Proceedings of the 6th Conference of the European Research on Mathematics Education* (pp. 2186–2195). ERME.

Verbisck, J., Bittar, M., Bosch, M., Barquero, B., & Benito, R. (2022). Study and research paths for statistics teacher education at secondary school level: An exploratory study. In S. Peters (Ed.) *Eleventh International Conference on Teaching Statistics* (ICOTS-11), IASE.

Wild, C. J. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–248. https://doi.org/10.1111/j.1751-5823.1999.tb00442.x

Winsløw, C. (2011). Anthropological theory of didactic phenomena: some principles of its use in the study of mathematics education. In M. Bosch Casabo, J. Gascón, A. Ruiz Olarría, M. Artaud, A. Bronner, Y. Chevallard, G. Cirade, C. Ladage, & M. Larguier (Eds.), *Un panorama de la TAD* (pp. 117-140). Centre de Recerca Matemàtica.

Winsløw, C., Matheron, Y., & Mercier, A. (2013). Study and research courses as an epistemological model for didactics. *Educational Studies in Mathematics*, *83*, 267–284. https://doi.org/10.1007/s10649-012-9453-3

Yin, R. K. (2013). Validity and generalization in future case study evaluations. *Evaluation*, *19*(3), 321–332. https://doi.org/10.1177/1356389013497081

Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Sage Publications.

Camilla Hellsten Østergaard

Camilla Hellsten Østergaard has served as an associate professor at University College Copenhagen since 2012. Her research centres on statistics education and mathematical modelling within the paradigm of questioning the world, with a particular focus on how students engage in modelling as a way of exploring phenomena. She further examines the role of artificial intelligence as an object of inquiry in mathematical. In addition, her work addresses teacher professional development in initial teacher education, particularly through the design and implementation of lesson study as a structured and collaborative approach to fostering inquiry-based mathematics teaching.

choe@kp.dk

Britta Eyrich Jessen

Britta Eyrich Jessen is an associate professor in didactics of mathematics at University of Copenhagen. Her research focuses on upper secondary mathematics with an emphasis on mathematical modelling. More recently this has moved in the direction of data science education and stochastics. Most of her research draws on theoretical frameworks rooted in French didactics. She is an expert on didactical engineering from the perspective of the Anthropological Theory of the Didactic (ATD), where she studies the ecology of Study and Research Paths and how to orchestrate media and milieu dialectics. She has been exploring the uses of generative AI as a resource for statistical modelling at university level. She runs professional development courses drawing on ATD and lesson study, and teaches at all levels from bachelor to PhD.

Britta.jessen@ind.ku.dk