



HAL
open science

Constraints and conditions for teaching data science as part of mathematical modelling in upper secondary mathematics

Britta Eyrich Jessen

► **To cite this version:**

Britta Eyrich Jessen. Constraints and conditions for teaching data science as part of mathematical modelling in upper secondary mathematics. Thirteenth Congress of the European Society for Research in Mathematics Education (CERME13), Alfréd Rényi Institute of Mathematics; Eötvös Loránd University of Budapest, Jul 2023, Budapest, Hungary. hal-04415835

HAL Id: hal-04415835

<https://hal.science/hal-04415835>

Submitted on 24 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Constraints and conditions for teaching data science as part of mathematical modelling in upper secondary mathematics

Britta Eyrich Jessen

University of Copenhagen, Denmark; britta.jessen@ind.ku.dk

In this paper we address recent years' attempt to include data science in school mathematics treated from the perspective of mathematical modelling. In our analysis we draw on the notion of scale of levels of co-determinacy from the Anthropological Theory of the Didactic (ATD) to identify the options or limitations for implementing data science in mathematics education as part of mathematical modelling. We analyse trends in educational debate, curriculum documents, textbooks and teaching material developed for Danish upper secondary school and called 'Data analysis and Cancer'. We identify fruitful conditions and delimiting constraints for merging mathematics and data science in upper secondary mathematics.

Keywords: Mathematical modelling, data science, technology, anthropological theory of the didactic, upper secondary mathematics.

Introduction

During recent years the nature of mathematical modelling in school and relations to modelling in workplaces and universities have been addressed within mathematics education research e.g. Frejd and Bergsten (2016) who investigated the external didactic transposition of mathematical modelling. According to Chevallard and Bosch (2020) the didactic transposition is characterised as the “transformations an object or a body of knowledge undergoes from the moment it is produced, put into use, selected, and designed to be taught until it is actually taught in a given educational institution” (Chevallard & Bosch, 2020, p. 214). This kind of analysis enlarges the empirical unit of analysis when we pay attention to the nature, the origin and the initial function of the mathematics that is being taught and learned in school (Bosch, 2015). To fully grasp the meaning of mathematical modelling as a still developing concept, we need to keep analysing its development. Frejd and Bergsten conclude that for current scholarly modelling “model construction is usually teamwork, in particular for larger projects, and the use of computer support is central” and furthermore that “the interdependence of mathematical modelling and technology embedded in a social context” (p. 28, 2016). The communication with clients is crucial when negotiating what is actually the modelling problem to be addressed and to what extent the produced models serve the intended purpose. However, these aspects are part of what creates a gap between scholarly knowledge and knowledge to be taught in school, though they may not be particular for current mathematical modelling. Jessen and Kjeldsen (2021) have made an additional study of based on three historic cases of mathematical modelling and how they link to school modelling. They conclude that:

The curriculum emphasises the importance of being able to work across disciplinary boundaries and to apply mathematics when addressing atypical problems. Students should recognise the role played by modelling in various contexts. These aims could nurture activities inviting the students to develop practices similar to those presented in our historic cases such as drawing on analogies, drawing on knowledge from the target discipline, developing (learning) new mathematics, and

discuss the epistemic value of the model from the different disciplinary viewpoints. (Jessen & Kjeldsen, 2021, p. 53)

These learning potentials overlap with recent years discussion of implementing data science in mathematics education (Boaler & Lewitt, 2019). Anderson and Register (2023) argue that:

given the mathematical nature of data science and its prevalence in the social, economic, and political aspects of globalized society, it has become imperative that students become proficient in its techniques and algorithmic ways of thinking. (Anderson & Register, 2023, p. 3)

They argue for a need to adjust mathematics curriculum to modern forms of analysing real world situations and study aspects of how to prepare future teachers for that. In higher education elements of data-generated modelling and data science has been explored to bridge statistics education for business degrees with future practice of the students. Markulin, Jessen and Florensa (in press) analyse statistical modelling activities, where the problem and the final solutions are framed and negotiated with employees from private companies. They conclude that this organisation of modelling activities leads to very different interdisciplinary collaborations and that it promoted 'adidacticity' meaning that "the existence of an external clients can favour the conditions under which the development of a final answer becomes the central activity" (Markulin et al., in press, p. 8). However, satisfying answers may reflect current practices in real life rather than covering specific course content. This represents a restriction for implementing this approach to teaching more broadly in the program and in schools. Still, Boaler and Lewitt (2019) argue in that mathematics curricular are similar to those of the 1950s and argue that very few data scientists use school algebra nor calculus in their daily jobs. Therefore, they propose "data and its analysis at the center of high school mathematics" and moreover that:

teaching students to make data-based arguments will endow them with many of the same critical-thinking skills they are learning today through algebraic proofs, but also give them more practical skills for navigating our newly data-rich world. (Boaler and Lewitt, 2019, p. 1)

Conrad and Mazzeo (2022) report how classic mathematics courses have been replaced by courses covering introductions to data science, but these do not grant access to higher education within data science. They argue for rethinking curriculum changes in ways, where "the most equitable policies will not include a wholesale redefinition of what constitutes mathematics relevant to students' long-term success, in school and beyond." (Conrad & Mazzeo, 2022, p. 2). Thus, not consider data science as a topic substituting existing content in mathematics but explore to what extent it can taught and learned within existing frames of upper secondary mathematics. An initial step in this direction has been taken by Heineman and colleagues (2018), when they designed a data science curriculum for upper secondary school. They define data science "as an interdisciplinary field [which] can be defined as the art of generating information and knowledge from data in various forms." (Heinemann et al., 2018, p. 1), drawing on techniques from computer science and mathematics, in particular stochastics. It must nurture students' development of data literacy being "the ability to collect, manage, evaluate, and apply data, in a critical manner" (Ridsdale et al., 2015, p. 3) and argue that "the definition in itself resembles process models for data analysis" (Heinemann et al., 2018, p. 2). They depict the process

as shown in Figure 1, where the real world delivers some information or data which is displayed in ways where it can be treated (using techniques from computer science and mathematics) in order to be represented in new ways allowing the data scientist or modeller to interpret the outcome and deliver new information to the real world. The whole process being driven by a question from the real world (Anderson & Register, 2023). The translation from the real world to mathematics or data science and back resembles the processes of modelling cycles, where

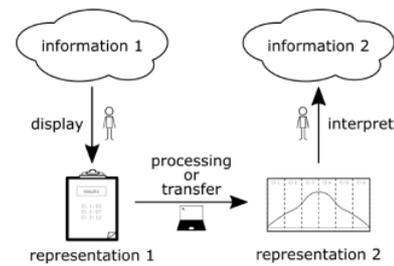


Figure 1: The model process in data science (Heinemann et al., 2018, p. 3)

driving the process through real world questions resembles the dynamics of modelling through study and research paths (see Barquero & Jessen, 2020), though with technology playing an explicit role. Thus, to consider data science from a mathematical modelling perspective might further our knowledge regarding how, when, and why to include technologies for mathematical modelling. We need more research on this, according to Siller and colleagues (2022), who have analysed existing research and conclude that “most studies were based on what might now be considered conventional digital tools [...] (e.g. computers), rather than new and emerging technologies (e.g. augmented and virtual reality)” (Siller et al., 2022, p. 4). Jessen (2022) on the other hand argue how conventional technologies can be considered the ‘milieu’ representing the real world against which the answers must be validated (adidacticity). However, this requires more open modelling problems, designed for the students to have multitude solution strategies relating to both epistemic and pragmatic uses of relevant technology. Hence, to explore data science from the perspective of mathematical modelling might develop our knowledge regarding the potential role for data science in upper secondary curricular as well as the role of computers and technology in mathematical modelling. This leads to the research question of this paper: What are the constraints and conditions for teaching data science as part of mathematical modelling? And based on the concrete teaching materials, what does students learn from and about mathematical modelling, when engaged with data science activities? Below we will outline the theoretical frameworks which defines our meaning of constraints, condition and learn from and with mathematical modelling.

Theoretical framework

The analysis draws on the notion of ‘scale of levels of co-determinacy’ from the Anthropological theory of the Didactic (ATD). Chevallard (2002) proposes the notion for the study of the constraints and condition regarding knowledge to be taught in school systems. The scale is defined as presented in Figure 2.

Civilisation ⇌ Societies ⇌ Schools ⇌ Pedagogies ⇌
Disciplines ⇌ Domains ⇌ Sectors ⇌ Themes ⇌ Questions

Figure 2: The scale of levels of co-determinacy (cf. Chevallard, 2002, p. 10)

We draw on the methodology outlined by Jessen et al. (2019) analysing the scale of levels of co-determinacy through six aspects or questions: 1) how ministerial guidelines and curriculum is developed, which includes the influx from the higher levels in Figure 2, 2) if the curriculum support

teaching and learning of the specific content knowledge aiming at the lower levels of Figure 2, e.g. mentions data science, 3) how textbook materials are developed and what is the legitimacy of those with respect to teaching practice as indicators of school, pedagogy and discipline, 4) how is mathematics assessed and how does this relate to further education, also addressing the lower levels, 5) to what extent does there exist professional development structures for implementing new initiatives (indicating explicit support from higher levels to lower ones) and finally 6) to what extent does the role of technology in the classroom support the learning (of data science and mathematical modelling)? (Jessen et al., 2019). In our analysis, we will focus on the descriptions of mathematical modelling in the curriculum documents and relate those to data science. When reaching the level of discipline and below, we will mainly focus on the situation in Denmark, here we draw on two approaches to mathematical modelling; the modelling cycle and Study and Research Paths (SRP), both lenses for designing and analysing modelling activities (Barquero & Jessen, 2020). The modelling cycle being described as a translation between the real world and mathematics, where the modelling process is divided into subphases: mathematise, mathematical work, interpret and validate. SRP is a modelling activity initiated from a generating question leading to dialectics between questions and answers as well as media and milieu (see Barquero & Jessen, 2020, Jessen, 2022).

Analysis of higher levels of co-determinacy

These levels are addressed by the first aspect and questions they raise and we see that the curriculum documents reflect well the broader debate concerning mathematics education enabling students to model the real world, where data science has gained attention. Heinemann et al. (2018) claims that “data science becomes more and more important in industrial and economic automation processes, in marketing and monitoring, e.g. in politics, so that data science permeates all areas of life” (Heinemann et al., 2018, p. 1). Boaler and Lewitt claim that modern uses of mathematics in societal contexts, private companies and research rarely is based on current upper secondary mathematics. Andersson and Register (2023) emphasise that “the ability to predict human behavior has reshaped the world economy into a data driven one in which leading entities are those that have the financial, technological, and human resources to be proactive” (Anderson & Register, 2023, p. 1). These views align with the purpose of upper secondary mathematics in Denmark (students age 16-18) formulated as students must know how sciences, technology, society and cultural questions are influenced by mathematics and prepare students to take active part in society by constructing innovative solutions to new problems. Furthermore, the document states that students should become able to:

apply functions to construct models describing data and knowledge from other disciplines, analyse mathematical models, create simulations, predictions and reflect upon the idealistic nature of models and their domains [...] including the treatment of more complex problems. (Danish Ministry of Education, 2017, p. 1)

In Denmark curriculum reforms are written by people appointed by Ministry of Education representing the ministry, upper secondary teachers and universities. The group responds to a request from politicians and other stake holders when reforming curricula. In the latest reform a renewed emphasis was put on stochastics from several stake holders (Jessen et al., 2019), which aligns with the proposal of Boaler and Lewitt (2019), though not explicitly implementing data science of any

kind. Still, we consider stochastics combined with explicit aims for mathematical modelling of real world and other disciplines favourable conditions for data science. Though not falling into the traps listed by Conrad and Mazzeo (2022), who outlines the challenges of not fully analysing the situation and impact of changing the content of core mathematics curriculum of upper secondary. Thus, we see delimited but favourable conditions stemming from the levels of civilisation, society and school. The above links to the second aspect of content knowledge and competencies listed in curriculum and ministerial guidelines. As found by Jessen et al. (2019) and Jessen & Kjeldsen (2021) there are room for students working with different approaches to modelling of real-world situations. Jessen and Kjeldsen finds that ministerial guidelines implicitly propose to teach students the phases of a modelling cycle as content, but also invite students to across domains “learn to engage with new atypical modelling problems” (2021, p. 48). This is considered a favourable condition for SRP and stochastics education as described by Markulin et al., (in press) in terms of SRP. Addressing the third aspect concerning textbook materials, the favourable conditions mentioned above are not evident. The textbooks analysed, are closer to ‘applicationism’ (Jessen, 2019) rather than offering the students opportunities for exploring the creative process of model development (Jessen & Kjeldsen, 2021). Still, they are considered potentially rich resources as media for SRPs. The textbooks are written by experienced upper secondary school teachers, where some takes part in developing curriculum. Thus, the textbooks have legitimacy with respect to practice, but data science is not covered and a delimited view on modelling presented. Additionally, two projects initiated by different groups of data science researchers at University of Copenhagen (UCPH) have developed teaching materials for upper secondary school supported by a private foundation, which is a stake holder promoting data science education throughout the school system. One project called ‘Data analysis and Cancer’ (2019) was developed by UCPH researchers in collaboration with an upper secondary school teacher. Currently, a new project has been initiated from UCPH in collaboration with three other upper secondary mathematics teachers called ‘Data Expeditions’. Thus, the domain of data science is slowly emerging in upper secondary mathematics with limited propagation. Data science is neither promoted as part of the high-stake exit examination functioning as part of entrance to higher education, the fourth aspect. When the curriculum reform was implemented in 2017, professional development was offered unsystematically addressing stochastics, inquiry-based approaches to education and discrete mathematics (Jessen et al., 2019). An evaluation of the implementation report that these parts of the reform has been the most challenging the teachers. The project group of Data Analysis and Cancer (2019) has offered professional development for teachers based on the teaching materials, which is a favourable condition for implementing this specific data science activity in classrooms. The lack of formal structures for teacher collaboration in Denmark is delimiting development (Jessen et al., 2019). The last aspect concerns technology for modelling and data science. We have full allowance of computers in upper secondary mathematics, and it is allowed for parts of the exit examination. The dominating tools are TI Nspire, Maple and GeoGebra. These have proved rich for experimental work within mathematical modelling (Jessen, 2022), but often they are reduced to advanced calculators in the classroom with negative effects on students learning (Jessen et al., 2019). Furthermore, the mentioned programs are not produced for data science and leaves the teacher with the design task of how to employ them or choose additional programs. In the following we will analyse the concrete teaching materials from the perspective of mathematical modelling.

Analysis of lower levels of co-determinacy

We adopt the methodology for analysis of teaching materials presented by Jessen and Kjeldsen (2021). The teaching material is developed interdisciplinary teaching across Biotechnology and Mathematics taught at advanced level, though the material is produced so you can ignore mathematics and solely rely on data science technology available online. The material is a written document (126 pages) consisting of articles addressing specific parts of biology (DNA, genes, cancer etc.), medicine and treatments (precision medicine), some chapters introducing the technological tools called cBioPortal (to access data and 'black box' data analysis), and a more formal introduction to statistical analysis. The material provide links to videos with researchers explaining some of the content. The articles have list of references covering additional article suitable for students. The materials also cover classroom exercises and assignments to be handed in to teachers, and finally students are offered to assess their learning progress through quizzes. Therefore, the material offers favourable conditions for getting started with interdisciplinary teaching of data science, also for teachers new to the domain.

The mathematics articles start with 'Correlations and statistical analysis' where the focus is Pearson correlation applied on protein and mRNA expression. This introduces students to conditions needed to be fulfilled such as linearity, independence and 'no extreme outliers'. All notions beyond mathematics curriculum. The chapter ends with exercises imitating the examples of the chapter, and 'extra material' introducing Spearman correlation. The first mathematics assignment is addressing breast cancer and correlation between mRNA expression and protein levels through statistical analysis. The assignment is formulated as 12 subtasks to be completed and linked with earlier biotechnology assignment. An example of a task is: "Make linear regression and determine important statistical estimators (e.g., correlation). Does this make sense?" (Data analysis and cancer, 2019, p. 81). The article has an appendix containing a manual for exporting data from the cBioPortal and import it to Excel, which is the proposed technology for carrying out the analyses. Students will most likely know this tool from lower secondary. The following article introduces χ^2 -test and comparison of probabilities. This notion was earlier mandatory for upper secondary mathematics but replaced by binomial test. Thus, hypothesis testing is known to teachers and students, and teachers have taught this topic previously (Jessen et al., 2019). χ^2 -test was omitted because it was often taught as techniques with no reasoning behind. The article too contains additional notions, examples and exercises leading to the second assignment on brain cancer and the impact of mutations. The last article introduces survival statistics through Kaplan-Meier curves and log-rank test, concepts beyond curriculum. The last assignment concerns ovary cancer and survival statistics.

The mathematisation process is carried out for the students in the teaching material. The biotechnology articles explain the functioning of cell biology and how to study and measure it in real life. The measurements lead to data productions and shapes the mathematisation process of the situations into a statistical domain, here considered a subset of mathematics. The biotechnology articles propose to outsource the data analysis to cBioPortal hindering the validation or critically review of the outcome. To some extent this mimics the process depicted in Figure 1. The statistics chapters offer to grey or white box the cBioPortal analyses employing statistical notions in Excel. We could argue that this resembles the phases of working mathematically and interpreting within the

modelling cycle (e.g. see Barquero & Jessen, 2020), though other phases are not present or unfolded. In this sense the material matches parts of curriculum asking students to work with authentic data, model real world situations, use technology and carry out hypothesis testing. However, the transmission of knowledge approach to teaching and learning is akin to most Danish textbooks and does not explicitly offer exploratory approaches for the students to develop autonomous relation with the content knowledge or modelling as e.g. offered by SRP (Barquero & Jessen. 2020), though the material may work well as media for an SRP.

Discussion and concluding remarks

Returning to the research questions of this paper, we see both favourable conditions and limiting constraints for teaching data science through mathematical modelling in upper secondary mathematics in Denmark. Data science may broaden the perspectives on technologies which can be used for modelling in biological contexts, which is part of the research Siller et al. (2022) argue we need to pursue. Though these first attempts seem to give rise to new challenges with black boxing the mathematical content. Further obeying societal demands of teaching upper secondary students data science may lead to “a loose collection of topics preliminary to data science” (Conrad & Mazzeo, 2022, p. 1) and does not seem worth striving for. On the contrary, we need to develop task design tools for modelling and data science, which further students’ autonomous relation with either the content knowledge (mainly stochastics, cf. Andersson & Register, 2023) or the notion of modelling that being data-generated modelling or theory-generated modelling (cf. Frejd & Bergsten, 2016). From our case study, students were not offered the opportunity of exploring atypically problems, as within SRP, nor were they offered to learn about modelling correlations within cell biology and the steps of such a modelling process, e.g. described through the modelling cycle. The unexplored linkages between intentions of the teaching materials and task design perspective (e.g. Barquero & Jessen, 2020) calls for further research. Teaching materials for data science may serve as a starting point research capturing potentials of developing students’ modelling skills, competencies and knowledge together with constructing coherent knowledge regarding domains of stochastics, mathematics, computer science and other school disciplines. Thus, rich collaborations between data scientist and didacticians are needed, and we may need to go beyond the domain of stochastics and consider e.g. machine learning, exploring the full potential or challenges of data science.

References

- Andersson, C. H., & Register, J. T. (2023). An examination of pre-service mathematics teachers’ ethical reasoning in big data with considerations of access to data. *The Journal of Mathematical Behavior*, 70, 101029. <https://doi.org/10.1016/j.jmathb.2022.101029>
- Barquero Farràs, B., & Jessen, B. E. (2020). Impact of theoretical perspectives on the design of mathematical modelling task. *Avances de investigación en educación matemática*, 17, 98–113. <https://doi.org/10.35763/AIEM.V0I17.317>
- Boaler, J., & Levitt, S.D. (2019). *Opinion: Modern high school math should be about data science — not Algebra 2*. LA Times. <https://www.youcubed.org/wp-content/uploads/2019/10/LA-times-op-ed.pdf>.

- Bosch, M. (2015). Doing research within the anthropological theory of the didactic: The case of school algebra. In *Selected regular lectures from the 12th International Congress on Mathematical Education* (pp. 51–69). Springer. https://doi.org/10.1007/978-3-319-17187-6_4
- Chevallard, Y. (2002). Organiser l'étude 3. Écologie & régulation [Organising Study 3. Ecology & regulation]. *Actes de la Pensée Sauvage, Grenoble, 11*, 41–56.
- Chevallard, Y., & Bosch, M. (2020). Didactic transposition in mathematics education. *Encyclopedia of mathematics education*, 214–218. Springer.
- Conrad, B. & Mazzeo, R. (2022). Editorial: *Mathematicians and the high school math curriculum*. IMU News, 111. International Mathematics Union. <https://www.mathunion.org/imu-news/archive/2022/imu-news-111-january-2022>
- Data analysis and Cancer (2019). *Data analysis as the foundation for cancer research and development of precision medicine*. University of Copenhagen. https://dataanalyseogkraeft.ku.dk/Dataanalyse_og_kraeft.accessible.pdf
- 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>
- Heinemann, B., Opel, S., Budde, L., Schulte, C., Frischmeier, D., Biehler, R., Podworny, S., & Wassong, T. (2018). Drafting a data science curriculum for secondary schools. In *Proceedings of the 18th Koli Calling International Conference on Computing Education Research (Koli Calling '18)*. Association for Computing Machinery, NY. <https://doi.org/10.1145/3279720.3279737>.
- Jessen, B., Otaki, K., Miyakawa, T., Hamanaka, H., Mizoguchi, T., Shinno, Y., & Winsløw, C. (2019). The ecology of study and research paths in upper secondary school: The cases of Denmark and Japan. In *Working with the anthropological theory of the didactic in mathematics education* (pp. 118–138). Routledge.
- Jessen, B. E., & Kjeldsen, T. H. (2021). Mathematical modelling in scientific contexts and in Danish upper secondary education: are there any relations? *Quadrante, 30*(2), 37–57. <https://doi.org/10.48489/quadrante.23658>
- Jessen, B. E. (2022). Characterising the roles of digital resources in mathematical modelling. In J. Hodgen, E. Geraniou, G. Bolondi, & F. Ferretti (Eds.), *Proceedings of the Twelfth Congress of the European Society for Research in Mathematics Education (CERME12)*. Free University of Bozen-Bolzano and ERME. <https://hal.science/hal-03759024/>
- Markulin, K., Jessen, B., & Florensa, I. (In press). Cross disciplinary characteristics of study and research paths: statistics for business administration. In Proceedings of INDRUM22.
- Ridsdale, C., Rothwell, J. Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelley, D., Matwin, S. and Wuetherick, B. (2015). *Strategies and best practices for data literacy education: Knowledge synthesis report*. Dalhousie University.
- Siller, H-S, Cevikbas, M., Geiger, V., & Greefrath, G. (2022). The role of digital tools in mathematical modelling research. In C. Fernández, S. Llinares, A. Gutiérrez, & N. Planas (Eds.) *Proceedings of the 45th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 152–155). PME.