MATHEMATIZATION USING FUNCTIONS IN DATA-BASED MODELLING BY UPPER SECONDARY STUDENTS

Stefanie Rach[1], Marielena Menzel[2], Sebastian Geisler[2], Timo Kosiol[1]

[1]OVGU Magdeburg, [2]University of Potsdam

Mathematizing real-world problems is a crucial skill in modern society, yet challenging for many students. We examine variations in students' mathematization processes and the conditions under which they successfully formulate a mathematical function for a real-world problem represented by data. In this experimental study, 561 upper secondary students model the decay of beer foam using one of the following data types: self-collected experimental data, provided experimental data, or smoothed data. Additionally, we collected students' mathematics grades, and assessed their content knowledge about functions. The findings suggest significant differences in students' mathematization approaches, partially related to their content knowledge and the data type provided. We discuss strategies to enhance this crucial skill.

INTRODUCTION

Describing real-world situations through mathematical concepts is an essential skill for active citizenship in modern society. As data usage becomes increasingly pivotal, data-based modelling is central to mathematics education. This study examines three key aspects that may affect students' ability to create a mathematical model, particularly a function, that accurately describes the data. First, we analyse students' approaches and the errors they encounter when constructing such a function. Second, we explore how the type and source of data affect this process. Third, we identify student prerequisites that may facilitate successful mathematization. Based on our findings, we propose strategies to enhance support for students in their modelling process.

THEORETICAL BACKGROUND

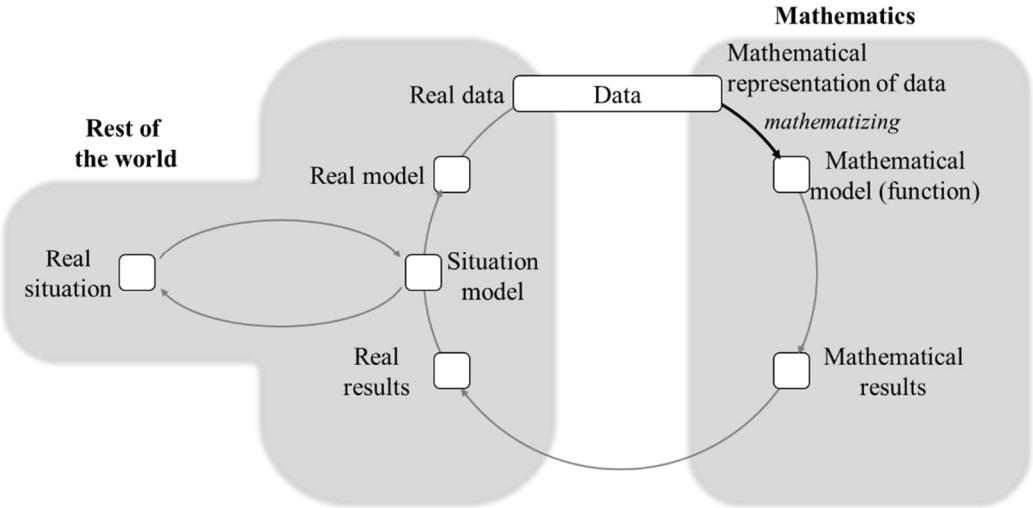
Data-based modelling as an activity

Mathematical modelling is a fundamental activity in mathematics education. This study emphasizes data-based modelling, wherein real-world situations are represented by data, often mathematized using a function (Mohr & Ufer, 2023; Stillman & Brown, 2021). Modelling real-world situations is frequently depicted through a modelling cycle. Initially, the individual comprehends and structures the situation presented by the data, subsequently creating a mathematical model to work with mathematically. The ensuing mathematical results are then interpreted, validated, and applied to real-world contexts (Leiss et al., 2024, Figure 1). The first critical transition from the real world to the mathematical domain, known as mathematization, is vital for successful modelling processes because the resulting mathematical model serves as the

2026. In H. Krzywacki, M. Hähkiöniemi, E. Haataja, M.S. Hannula & P. Hästö (Eds.). *Proceedings of the 49th Conference of the International Group for the Psychology of Mathematics Education: Research Reports*, pp. 1203-1210. PME.

foundation for subsequent steps. We define mathematizing as the process of selecting an appropriate mathematical representation of real-world data, which means constructing a mathematical model based on a real-world scenario.

Figure 1: Modelling cycle adapted on Leiss et al. (2024)



In data-based modelling (Michelsen, 2006), a function serves as a mathematical model to represent the relationship between two numerical variables, enabling the prediction of outcomes in specific situations. A two-step approach can be effective when constructing a function: initially, selecting the type of function, such as linear, quadratic, or exponential (function type selection), and subsequently determining the function parameters (parameter specification).

Data type and personal characteristics in relation to data-based modelling

Previous studies have highlighted that students encounter significant challenges in developing mathematical models (e.g., Frejd & Ärlebäck, 2011). Jankvist and Niss (2020, p. 487) describe mathematization as a “stumbling block”. Theoretical analyses and qualitative research suggest that successful mathematization requires students to possess mathematical knowledge for making assumptions about a given situation. This also involves utilizing extra-mathematical knowledge to identify key aspects of the situation and the relationships among these aspects (Djepaxhija et al., 2015). When creating a mathematical model from data, students may draw upon both the situational context and the data itself. For instance, when modelling the decay of beer foam, students might focus on the deceleration of the decay process to choose an appropriate mathematical function and determine its parameters, guided by the data.

There are two types of data: smoothed data, where data points align perfectly with a mathematical function, and real data, where measurement errors cause discrepancies between data points and plausible functions. Engel (2018) advocated for using real data, potentially generated by students, as a foundation for modelling processes, given that real data enhances the authenticity of the task. When handling real data, students must integrate mathematical, contextual, and statistical approaches to account for

variations in data that are randomly distributed (Kawakami & Minero, 2021). When students generate the data themselves, they become familiar with the data's origin, which may support a more accurate understanding of the real-world context and the mathematization process (Jankvist & Niss, 2020). In contrast to real data, because smoothed data lacks measurement errors, all data points fit precisely to a single function, which students then use to develop the mathematical function. Thus, we hypothesize that students' processes for creating mathematical models vary depending on the type of data—smoothed or real.

Regarding student characteristics, students must apply their content knowledge when mathematizing real-world scenarios—such as understanding the behaviour of a function's graph to determine which function type best suits a given situation, or comprehending the meaning of parameters in linear or exponential functions to appropriately adjust these parameters. The study by Mohr and Ufer (2023) in a laboratory setting provides evidence supporting the assumed relationship between data-based modelling skills and knowledge of functions.

RESEARCH QUESTIONS (RQ)

1. To what extent do student characteristics, such as content knowledge and prior mathematics achievement, and the type of data predict whether students select an exponential function to model the decay of beer foam?
2. What are the typical approaches and errors students encounter when mathematizing the decay of beer foam?
3. To what extent do student characteristics, such as content knowledge and prior mathematics achievement, and the type of data predict how successfully students specify the parameters of an exponential model to mathematize the decay of beer foam?

METHODS

Sample and modelling task

The sample consists of $N = 561$ students in grades 10 and 11 from 27 classes in Germany (52% female, $M_{age} = 16$ years). This study is part of the project *Experiments to foster modelling competences and motivation in mathematics* (Ex2MoMa, Jonscher et al., 2024). In this project, students engaged in three modelling tasks during three consecutive mathematics lessons. Each real-world situation can be described using either a linear or an exponential function. Each class was randomly assigned to one of three conditions, which differed only in the type of data provided for the modelling tasks: provided smoothed data ($n = 186$ students), self-collected experimental data ($n = 215$ students), and provided real data ($n = 160$ students). The function f that best fits the relationship between the height of the beer foam and the time x in the context of smoothed data is $f(x) = 23.5 \cdot 0.74^x$. For provided real data,

the function is $f(x) = 22.3 \cdot 0.78^x$. For self-collected experimental data, the best-fitting function is determined based on the data gathered by the students.

Design

To address the research questions, we examined students' modelling processes during the "beer foam decay" task in the second lesson. An exponential function is suitable for modelling the initial phase of beer foam decay. Content knowledge was assessed using a test, and prior achievement was indicated by students' mathematics grades, both administered on the first lesson.

Students worked in pairs on the modelling problems and documented their work independently. The learning took place in a self-regulated environment, organised by trained research assistants, and supported by a guiding manual designed to assist with the determination of parameters of (linear and) exponential functions using multiple data points. Due to students' limited understanding of algorithms and statistics, least-squares techniques were not included in the learning environment. Instead, students could estimate the function parameters using multiple data points and could employ a GeoGebra applet with restricted functionality to graphically optimize the function.

Instruments and analysis strategy

Content knowledge and prior achievement. The test used to assess content knowledge comprises seven items (see Figure 2, for an item example). The items require students to identify parameters of a linear or exponential function, given a table or a graph, and to match a graph to a given equation (see Jonscher et al., 2024). We scaled the test using item-response theory: WLE = .62 and z-standardized it. The mathematics grade ranges from 1: very good to 6: insufficient (*M* = 2.58, *SD* = 1.09).

Figure 2: Item example for the content knowledge test on functions

An exponential function f is given in the value table. State the corresponding functional equation for f.

x	0	1	2	3
$f(x)$	1	3	9	27

$f(x) =$ _____

Function type selection. To analyse which mathematical models students created, we first coded the type of function each student used to describe their data. We classified these into exponential functions, linear functions, other types of mathematical objects, or an absence of a mathematical model. Quantitatively, we explored the impact of student characteristics and the type of data on whether students constructed an exponential model (RQ 1). Additionally, we qualitatively categorized some of the approaches and errors that occurred (RQ 2).

Parameter specification. To address RQ 3, we focus on students who worked with real data and developed an exponential model. The use of real data is crucial for data-based modelling, and the quality of parameter specification can only be compared among those who established an exponential model. We employed the least squares method

to determine which exponential function best fits the data in each case. Additionally, we used the residual sum of squares (RSS) to evaluate the quality of parameter specification, as it measures the model's fit. In the condition using provided real data, the RSS of the best-fitting function is constant (RSS = 9.91), whereas in the experimental condition, the RSS of the best-fitting functions varies (M_{RSS} = 5.25, SD_{RSS} = 4.78; range: 0.28–22.80). To account for this variation, we computed the difference between the RSS of the function developed by the student and the RSS of the best-fitting function. The lower this value, the better the function fits the data.

We employed general linear models in R (version 4.3.0) using family=binomial (RQ1) and family = Gamma (link="log") (RQ3). The binomial model accounts for the dichotomous outcome in RQ1, while the log link accommodates the non-normal distribution of the outcome variable in RQ3. In both analyses, we examined how student characteristics—specifically content knowledge and prior achievement—and the type of data predict model selection or parameter specification.

RESULTS

Predicting function type selection (RQ1)

Table 1 presents the number of students across the three conditions who utilized an exponential function, a linear function, or any other type of mathematical object to describe the decay of beer foam.

Table 1: Relation between the function type and data type

Function type	Self-generated experimental data	Provided experimental data	Smoothed data
Exponential	183	135	173
Linear	10	9	0
Other	1	4	6
No answer	21	12	7

In terms of student variables, content knowledge significantly predicts model selection (Estimate = 0.58, $p < .01$). This finding suggests that the odds for selecting an exponential model increase by 1.79 for students with content knowledge one standard deviation above the mean, controlling for data type and grade. Descriptively, working with real data appears to reduce the likelihood of selecting an exponential model, regardless of whether the data was self-generated (Estimate = -0.70, $p = .12$) or provided (Estimate = -0.84, $p = .07$) when compared to modelling with smoothed data. However, these results do not reach statistical significance. Additionally, we cannot show that the last grade predicts the selection of the function type significantly (Estimate = -0.24, $p = .15$, last grade with 1 best, and 6 worst grade).

Approaches and errors when creating a mathematical model (RQ2)

Figure 3 illustrates two typical student responses that were classified as “other” because the students neither utilized an exponential nor a linear function with a non-zero slope. The first example, provided by a student in the given experimental condition, likely intended to establish an exponential function but instead documented a null function (Figure 3). In the second example, from the smoothed condition, a variable is missing. The third example, originating from the real-data condition, reflects a solution that failed to account for the gradual disappearance of beer foam over time (Figure 4). The fourth example (Figure 4) exemplifies a common error: students correctly identified the function type but miscalculated the parameter as 0.86. This error frequently occurred because data were collected at 0.5 intervals, and multiple students failed to adjust the decay factor to account for this interval.

Figure 4: Examples of commonly misfitting mathematical models (category: exponential)

Figure 3: Examples of commonly misfitting mathematical models (category: other)

Predicting parameter specification (RQ3)

To mitigate the impact of outliers in our dataset, we applied the interquartile range (IQR) method (Barbato et al., 2011) and excluded students who formulated a linear function or other objects from the subsequent analysis. The analysis was thus conducted on 163 students in the experimental condition and 115 students in the provided real data condition. Our findings indicated that no predictor accounted for the difference between the residual sum of squares (RSS) of the student model and the best-fitting model. Neither content knowledge (Estimate = 0.03, p = .73), nor the last grade (Estimate = 0.07, p = .41), nor the type of data (Estimate = 0.00, p = 1.00) significantly predicted the accuracy of parameter specification.

DISCUSSION

This study investigated the mathematical functions students choose to model a real-world situation, specifically the decay of beer foam. The findings indicate that many students correctly identified the appropriate function type to model the data, with those working with smoothed data exhibiting greater success in selecting a suitable function type. Additionally, factors were identified that impede students’ ability to adequately mathematize real-world scenarios: insufficient conceptual knowledge of variables and

functions, as well as calculation problems (Jankvist & Niss, 2020). Common errors arise in the use of variables and equations (see Figure 3) and in the selection of function parameters (see Figure 4). The latter error may stem from misrepresentation of the situation, a lack of understanding regarding the parameter's meaning within the equation, or calculation errors. Notably, a higher proportion of students in the experimental data condition, compared to those working with smoothed data, failed to provide any function. This might result from increased demand arising from the additional experiment or from the absence of similar tasks in mathematics lessons.

Content knowledge plays a dual role in the modelling process. While it appears crucial for selecting an appropriate function type, it is less critical for specifying parameters, partially corroborating the findings of Mohr and Ufer (2023). This outcome can be explained by the observations of Jankvist and Niss (2020), who argue that students often lack a complete understanding of the situation and fail to make appropriate assumptions, thereby hindering the mathematization process. Consequently, even when students possess adequate content knowledge, they may struggle to apply it effectively (see also Djepaxhija et al., 2015).

This research has certain limitations related to the data collection process and the selection of specific student characteristics. Since the edited manual was collected after students completed the entire task, students had the opportunity to modify their functions through a validation process. Unlike the approach taken by Hankeln and Greefrath (2021), we assessed students' modelling skills holistically, considering their work on a complete modelling problem (see Cevikbas et al., 2022). This method aligns more closely with the natural learning process in mathematics. Furthermore, our study only considered cognitive prerequisites, excluding affective factors (see Kaiser & Schukajlow et al., 2022).

Future research could examine the processes students engage in when mathematizing real-world problems, examining the strategies they use to formulate functions, such as the use of a GeoGebra applet (Hankeln & Greefrath, 2021). This line of inquiry could also explore why content knowledge does not seem to significantly influence parameter specification. This research highlights that students can work with real data and select an appropriate function type, but they need support to specify function parameters.

Funding: This study is funded by the German federal ministry of education and research under grant No. 16MF1028A.

References

Barbato, G., Barini, E. M., Genta, G., & Levi, R. (2011). Features and performance of some outlier detection methods. *Journal of Applied Statistics, 38*(10), 2133-2149.

Cevikbas, M., Kaiser, G., & Schukajlow, S. (2022). A systematic literature review of the current discussion on mathematical modelling competencies: State-of-the-art developments in conceptualizing, measuring, and fostering. *Educational Studies in Mathematics, 109*(2), 205-236. https://doi.org/10.1007/s10649-021-10104-6

Djepaxhija, B., Vos, P. & Fuglestad, A. B. (2015). Exploring grade 9 students' assumption making when mathematizing. In K. Krainer & N. Vondrová (Eds.), *Proceedings of the Ninth Congress of the European Society for Research in Mathematics Education* (pp. 848-854). CERME 9.

Engel, J. (2018). *Anwendungsorientierte Mathematik: Von Daten zur Funktion: Eine Einführung in die mathematische Modellbildung für Lehramtsstudierende.* [Application-oriented mathematics: From data to function: An introduction to mathematical modelling for student teachers.] Springer Berlin Heidelberg.

Frejd, P., & Ärlebäck, J. B. (2011). First results from a study investigating Swedish upper secondary students' mathematical modelling competencies. In G. Kaiser, W. Blum, R. Borromeo Ferri, & G. Stillman (Eds.), *Trends in Teaching and Learning of Mathematical Modelling: ICTMA14* (pp. 407-416). Springer.

Hankeln, C. & Greefrath, G. (2021). Mathematische Modellierungskompetenz fördern durch Lösungsplan oder Dynamische Geometrie-Software? Empirische Ergebnisse aus dem LIMo-Projekt [Fostering Mathematical Modelling Competency via Solution Plan or Dynamic Geometry Software? Empirical Results of the LIMo-Project]. *Journal für Mathematik-Didaktik, 42*, 367-394. https://doi.org/10.1007/s13138-020-00178-9

Jankvist, U. T. & Niss, M. (2020) Upper secondary school students' difficulties with mathematical modelling. *International Journal of Mathematical Education in Science and Technology, 51*(4), 467-496. https://doi.org/10.1080/0020739X.2019.1587530

Jonscher, M., Menzel, M., Geisler, S., & Rach, S. (2024). Modelling performance using functions: Relation to person characteristics and different solution approaches. In T. Evans, O. Marmur, J. Hunter, & G. Leach (Eds.), *Proc. 47th Conf. of the Int. Group for the Psychology of Mathematics Education* (Vol. 3, pp. 113-120). PME.

Kaiser, G., & Schukajlow, S. (2022). Innovative Perspectives in Research in Mathematical Modelling Education. In C. Fernández, S. Llinares, A. Gutiérrez, & N. Planas (Eds.), *Proc. 45th Conf. of the Int. Group for Psychology of Math. Education, Vol. 1* (pp. 147-176). PME.

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* (pp. 389-400). Springer.

Leiss, D., Ehmke, T., & Heine, L. (2024). Reality-based tasks for competency-based education: The need for an integrated analysis of subject-specific, linguistic, and contextual task features. *Learning and Individual Differences, 114*, 102518.

Michelsen, C. (2006). Functions: a modelling tool in mathematics and science. *Zentralblatt für Didaktik der Mathematik, 38*, 269-280. https://doi.org/10.1007/BF02652810

Mohr, M., & Ufer, S. (2023). Measuring Data-based Modeling Skills in a Collaborative Setting. In M. Ayalon, B. Koichu, R. Leikin, L. Rubel, & M. Tabach (Eds.), *Proc. 46th Conf. of the Int. Group for the Psychology of Math. Education* (Vol. 3, pp. 355-362). PME.

Stillman, G. A., & Brown, J. P. (2021). Modeling the phenomenon versus modeling the data set. *Mathematical Thinking and Learning, 25*(3), 270-295.