# Investigating the role of data in the teaching of statistical modelling – a literature review

**Takashi Kawakami1 and Jonas Bergman Ärlebäck2**

1Utsunomiya University, Japan, 2Linköping University, Sweden

This study develops and discusses a framework for identifying the role of data in empirical research studies, focusing on the teaching of statistical modelling based on a literature review. A theoretically derived framework that recognises three distinct roles of data in teaching statistical modelling was used to analyse 63 papers. The findings show that data are assigned three different roles: data as a source of models, data as evidence to test model validity, and data as a storyteller of the data-generating process. The results of the analysis are further discussed in relation to three rationales for teaching statistical modelling: competency oriented, content oriented, and socially oriented. Directions for further theoretical and empirical lines of enquiry regarding the teaching of statistical modelling are suggested.

## Introduction

The importance and role of data are expanding in our data-driven society (e.g. Burrill & Pfannkuch, 2023; Lesh et al., 2008), demanding citizens to be productive in their personal and professional lives and to participate in society in a well-informed and critical way (Gal & Geiger, 2022). One approach to support students in coping with this data-intensive world is to develop their data literacy—the ability to describe, reason, analyse, argue and make decisions using data (OECD, 2019), which is necessary for all citizens.

Additionally, the recent rapid technological developments have diversified how data can be collected, managed and organised in large quantities and how the quality of the collected data can be improved (Burrill & Pfannkuch, 2023). Consequently, there is a need to start rethinking the concept of data and its role in mathematics and statistics education. This means, for example, considering non-traditional data types such as spatial data, video and images (Lee & Wilkerson, 2018). Another consequence is an increased emphasis on data science education (Burrill & Pfannkuch, 2023). One of the questions that arise in these evolving educational contexts is how school mathematics (including statistics education) can contribute to support students in developing the required competencies in data science and data literacy.

Previous research in mathematics and statistics education has explored the role and use of modelling, which puts data at the core of activities in educational contexts (Ärlebäck & Kawakami, 2023; Frejd & Ärlebäck, 2021; Kawakami & Ärlebäck, 2022). Notably, this line of inquiry has, in addition, (i) broadly pointed out the importance of data in mathematical and statistical modelling activities (e.g. Blomhøj & Jensen, 2007; Pfannkuch et al., 2018); and, as elaborated below, (ii) identified three rationales for teaching statistical modelling (*competency oriented*, *content oriented*, and *socially oriented*) (Ärlebäck & Kawakami, 2023; Kawakami & Ärlebäck, 2022). However, the role of data in developing competencies, concepts, and insights through statistical modelling with real-world and societal contexts, has not been systematically investigated and articulated. This study aimed to (ⅰ) develop and discuss a framework for identifying the roles data have been reported to play in empirical studies that investigated the teaching of statistical modelling based on a literature review; and (ⅱ) to provide an overview of the distributions of these roles.

## Theoretical framework and research question

### Statistical modelling and three rationales for teaching statistical modelling

No widely accepted definition of statistical modelling exists in statistical education. For the purposes of this review, we adopted the general definition of statistical modelling provided by Langrall et al. (2017) in their review of statistical education research:

Modeling in statistics refers to any one of a number of practices: the development of a distribution (empirical or descriptive model) from data; the process of creating a theoretical (probability) model from an empirical model; and the practice of sampling from a theoretical model (simulation). (p. 502)

Pfannkuch et al. (2018) opined that statistical modelling lies on a spectrum between being solely data-driven and being solely theory-driven. One end of this spectrum, called ‘data modelling with graphs’ by Pfannkuch et al. (2018, p. 115), involves making and/or using graphs and other representations (such as empirical models of the distribution) to make informal statistical inferences based on a good fit with data (Makar & Rubin, 2009). The other end of the spectrum is characterised as ‘chance modelling with mathematical theoretical distributions’ (Pfannkuch et al., 2018, p. 115), which involves making statistical inferences based on probability distribution (as theoretical models of the distribution) or simulations. In statistics education, various statistical modelling activities (such as data modelling, software modelling, exploring the output of pre-built models, and model recognition), are practised within the spectrum between these two ends (Pfannkuch et al., 2018).

Based on a systematic literature review of 48 peer-reviewed empirical research papers on teaching statistical modelling in the context of mathematics and statistics education, three rationales for teaching statistical modelling have been identified (Ärlebäck & Kawakami, 2023; Kawakami & Ärlebäck, 2022). These rationales for teaching statistical modelling are *competency oriented*, *content oriented*, and *socially oriented* (Table 1), and they capture different foci and emphasis with respect to the systematic design of statistical modelling tasks as well as the appropriate positioning statistical modelling practices within mathematics curricula. Statistical modelling can be comprehended holistically as a component of statistical competencies that integrates aspects of statistical content and shaping and influencing real-life, social and societal decision-making.

|  |  |  |
| --- | --- | --- |
| **#** | **Rationale** | **Description** |
| R1 | Competency oriented | This rationale aims to develop statistical competencies, including statistical literacy, reasoning, and thinking, as well as statistical processes such as statistical inquiry and informal statistical inference. It highlights the applicability of statistical modelling in solving real-world problems while emphasising the use of actual and authentic data to stress the applied nature of statistics. |
| R2 | Content oriented | This rationale promotes the learning of statistical knowledge and concepts, such as variability, distribution, sample, and sampling, along with knowledge and concepts related to statistical models and modelling. Statistical modelling is viewed as an epistemic practice of statistics and a pedagogical tool. |
| R3 | Socially oriented | This rationale promotes decision-making in the real world, social, and societal contexts where data are embedded and perform a crucial function. Furthermore, it emphasises the need to develop a critical understanding of the use and role of statistics, statistical models, and modelling in such contexts. Statistical modelling is viewed as both a means and an object of social criticism and decision-making based on data. It puts particular emphasis on addressing social issues and contexts. |

Table 1. The three rationales for teaching statistical modelling (Ärlebäck & Kawakami, 2023; Kawakami & Ärlebäck, 2022)

### The potential use of data in statistical modelling

Data are essential in statistical modelling and therefore, in order to achieve the educational goals of statistical modelling based on Table 1, the use of data appropriate to the goals needs to be incorporated into the design of tasks and lessons. According to Shibata (2015, p. 8), the broad dictionary meaning of data is ‘information on which inferences are based’, while the narrower meaning refers to ‘a sequence of values of a variable or a sequence of pairs of values of multiple variables’. In statistics education in general, data are commonly utilised in the latter sense, encompassing qualitative data (e.g. textual information such as blood type), and quantitative data (e.g. numerical information such as height and temporal data in the form of time series data acquired along changes, for instance, daily temperatures) (Bargagliotti et al., 2020). Furthermore, real data, realistic data or simulated data are employed in statistics education.

It is important to note that data represent ‘numbers with context’ (Cobb & Moore, 1997, p. 801). This means that either a real-world, concrete, or theoretical context underlies the data, and it is imperative to understand and interpret data in one of these contexts when handling such information. Furthermore, data have a signal and noise structure (Konold & Pollatsek, 2002). Signal refers to regular patterns or central tendencies in the data, whereas noise refers to irregular random patterns or errors in the data. Data are subjected to mathematisation to unveil or abstract its underlying structure for solving real-world problems, accompanied by sources of statistical and mathematical knowledge and concepts (Lesh et al., 2008; Wild & Pfannkuch, 1999). Hence, statistical modelling demands ongoing transitions between context, data, and the (statistical) model, as depicted in Figure 1 (Wild & Pfannkuch, 1999). Figure 1 illustrates that, on the one hand, data are generated from the context, and (statistical) models are generated from the data. On the other hand, data are interpreted and validated based on the context, and (statistical) models are interpreted and validated based on the data.

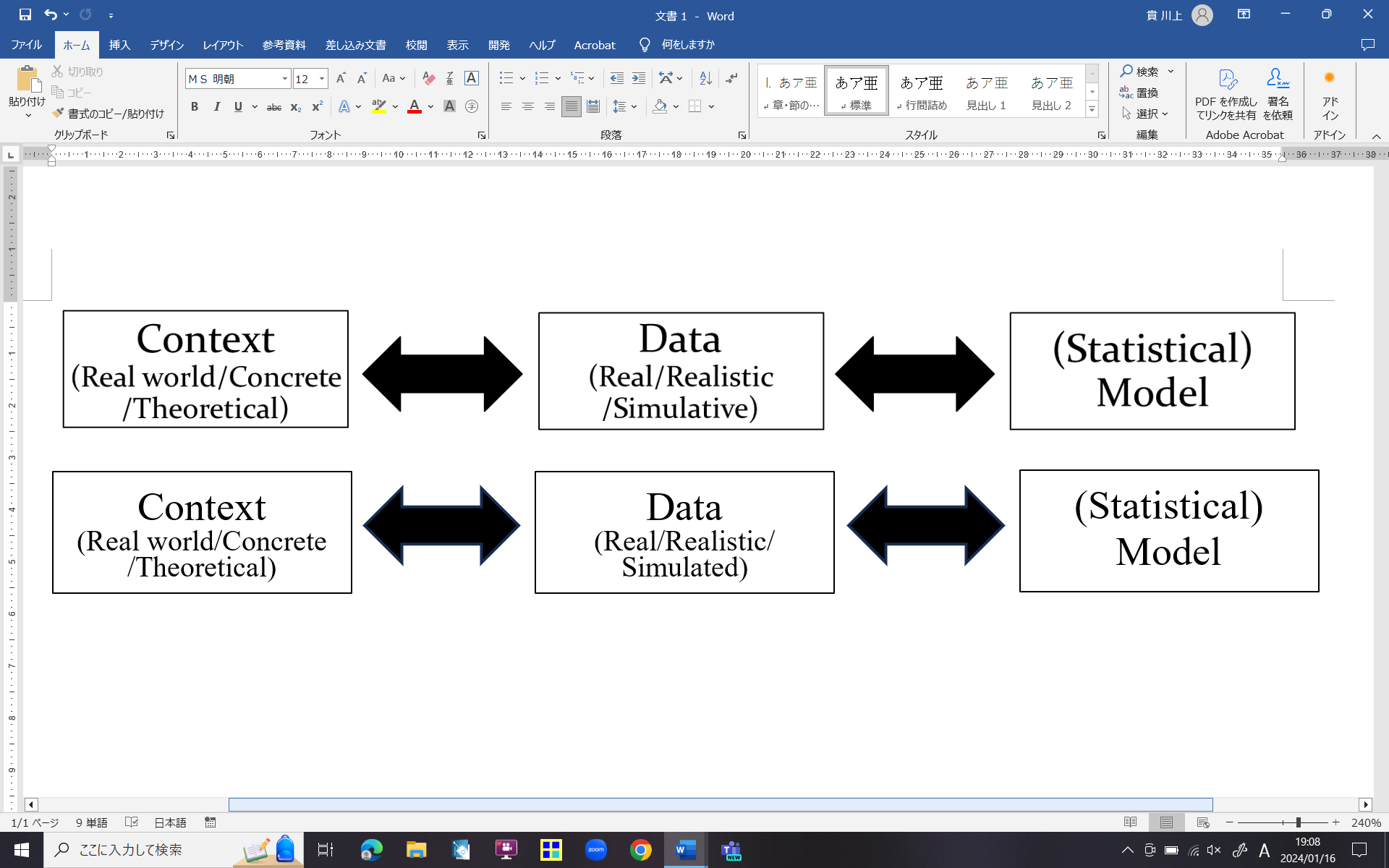


Figure 1. Relationship between context, data, and (statistical) model (Adapted from Wild and Pfannkuch, 1999)

Data play key roles in generating, validating, and modifying models in statistical modelling or modelling in general. Dunn and Marshman (2020), identified two roles of real data in model development in educational settings: (a) to validate the model’s structure or parameter values; and (b) to estimate the model’s structure or parameter values. These two roles correspond to the arrows between the data and the (statistical) model in Figure 1: data serveas *the source of the model*, and are *the evidence against which the model’s validity can be tested*. Moreover, data can also be considered as a model of phenomenon/theory (Wilkerson & Laina, 2018). The investigator does not always collect data, but often acquires them from other sources, such as public datasets. Simulated data are generated using simulation software that embeds the theory of probability distributions. As the data used in the classroom may already have been processed, ‘[m]aking meaning from someone else’s data requires being aware of the measurement process that generated data’ (Rubin, 2020, p. 156). Adopting such a mindset cultivates critical thinking regarding data and (statistical) models, which may result in modifications to the data collection and (statistical) models. The third role of data can be regarded as *a storyteller of the data-generating process*. This role corresponds to the arrows connecting the context and the data in Figure 1. Thereby, we adopted the following three categories of potential roles of data in statistical modelling (Table 2), with the aim of facilitating of the purposeful use of data according to the three rationales for teaching statistical modelling.

|  |  |
| --- | --- |
| **Role** | **Description** |
| Role A | Data as source of models (Data→Model) |
| Role B | Data as evidence by which to test model validity (Data←Model) |
| Role C | Data as a storyteller of the data-generating process (Context⇄Data) |

Table 2. The potential roles of data in statistical modelling

### Research question

Given the theoretical framing of the different roles data can play in statistics education research as summarised in Table 2, we formulated and addressed the following research question: *How are the roles of data (as a source, as evidence, as a storyteller) distributed in empirical research on statistical modelling that invokes different rationales for teaching statistical modelling?*

## Methodology

In Kawakami and Ärlebäck (2022), we identified empirical research literature on statistical modelling in mathematics and statistics education that invokes different rationales for teaching it. The literature dataset identified from this prior study was used to conduct a systematic review of the role of data in teaching statistical modelling. The literature was selected because it demonstrated a rationale for the use of statistical modelling in mathematics and statistics education and also focused on data, as indicated in the paper selection process described below (Figure 2), making it suitable for answering the research question. The literature dataset includes peer-reviewed research literature that used or investigated statistical modelling from the following influential mathematics education journals: *Educational Studies in Mathematics* (ESM) (1968–July 2021), *ZDM: Mathematics Education* (ZDM) (1997–July 2021), *Mathematical Thinking and Learning* (MTL) (1999–July 2021), *Journal for Research in Mathematics Education* (JRME) (1970–July 2021), and *Journal of Mathematical Behavior* (JMB) (1995–July 2021). Additionally, research papers from the following internationally recognised journals in statistics education were included: *Statistics Education Research Journal* (SERJ) (2002–July 2021) and *Journal of Statistics Education/Journal of Statistics and Data Science Education* (JSE/JSDE) (1993–July 2021). The varying start dates of the journal search is due to the fact that the journals’ first issue appeared in different years. Notably, special issues on statistical modelling research were found in ZDM Vol. 50, No. 7 and SERJ Vol. 16, No. 2.

We selected eligible papers through the journals’ databases, including Springer Link, JSTOR, Taylor & Francis Online, Science Direct and the IASE website. We extracted original papers that contain the terms ‘data’, ‘statistics’, or ‘statistical’, and ‘model’ or ‘mode[l]ling’ in the article’s title, keywords or abstract. Upon closer inspection, we excluded papers that contained these terms but did not focus on empirical research. We also excluded papers focusing solely on probability modelling or on modelling learning and cognition to ensure a focused discussion. A total of 63 papers were extracted, including 7 ESMs, 16 ZDMs, 10 MTLs, one JRME, one JMB, 14 SERJs and 14 JSEs/JSDEs[1]. Figure 2 illustrates the paper selection.

For the systematic review, we read all 63 identified papers and categorized the intended function or role of data in the teaching statistical modelling in these papers by coding them into one or more of the three categories designated as Roles A, B, and C in Table 2. The coding criteria were based on the explicit description(s) of the function or data in the papers. We focused on (1) the description of the study’s purpose and position, (2) the intentions and purposes of the utilised teaching materials, curriculum, and teaching practices, as well as (3) the research questions presented in the papers. Table 3 presents examples of such descriptions and their coding. The first author conducted the first analysis, which included classification, and the second author independently checked the assigned papers and the analysis. Where discrepancies occurred, the authors discussed and resolved them. In Kawakami and Ärlebäck (2022), we identified the rationales in Table 1 for using statistical modelling in the literature dataset, but did not analyse the roles of data in empirical research on statistical modelling that invokes different rationales for teaching statistical modelling. To answer the research question, the current study cross-tabulated the roles of data with the identified rationales in the literature dataset.

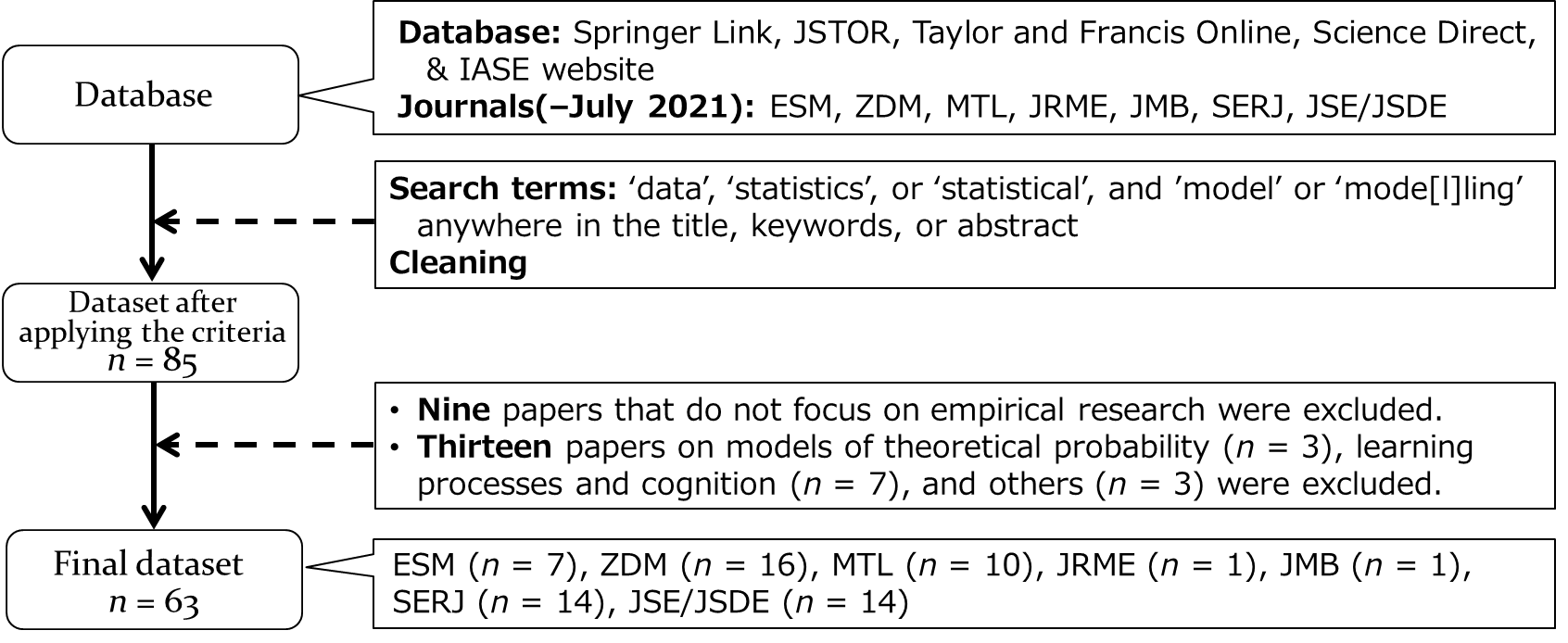


Figure 2. Paper selection process (Adapted from Kawakami and Ärlebäck, 2022, p. 1112)

|  |  |
| --- | --- |
| **Category** | **Examples of the descriptions in the papers** |
| Role A | * A pedagogical modelling approach that views real data as a source for a model of a situation in the real world can serve as a bridge between data and probability … (Aridor & Ben-Zvi, 2017, p. 41) * For statistical modelling, data from real world systems are a key part of building models, where the purpose is to create models that mimic random behaviour in real world systems … (Patel & Pfannkuch, 2018, p. 1198) |
| Role B | * The effectiveness of this model in producing the required output is evaluated by using graphs to compare the generated data with the initial data that the pupils collected or explored, in the style of AG [Active Graphing]. (Ainley & Pratt, 2017, p. 20) * Model evaluation drives model revision, which generally increases the explanatory power of a model and the scope of application as the model is fine-tuned to account for new data (i.e. evidence in scientific inquiry). (Doerr et al., 2017, p. 89) |
| Role C | * Typically, these stories are conceptualized as embedded within data, to be discovered or unlocked by learners. (Wilkerson & Laina, 2018, p. 1224) * In particular, the inferential explanations were grounded in the knowledge of context that was used as a story behind the data as well as enabling students to go beyond the data to make inferences. (Kazak et al., 2021/2023, p. 30) |

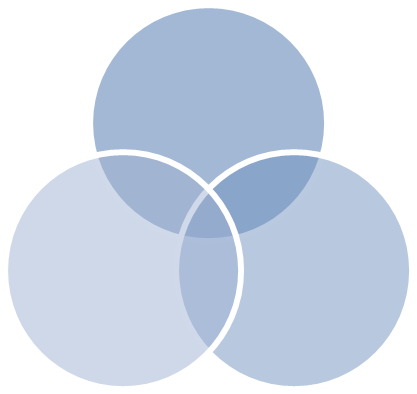
Table 3. Examples of the descriptions for determining categories Roles A, B, and C

## Results

We present the results of the literature review in the research question, organised in two sections: *the role of data discerned in empirical research on statistical modelling*, and *the role of data in terms of different rationales for the teaching of statistical modelling*.

### Role of data discerned in empirical research on statistical modelling

Figure 3 summarises the ways data were used in the 63 papers examined, with the majority (45%) of the papers employing data in tandem with both Roles A and B simultaneously. The second largest proportion of papers employed data in tandem with all three roles (24%). When examining the frequency with which data were used according to the three individual roles in the papers, it transpired that Role A was the most common (97%), followed by Role B (71%) and Role C (41%).



**Role A**

**Role B**

**Role C**

**9**

**28**

**9**

**2**

**15**

(Ex.) Hjalmarson et al. (2011), Mulligan (2015), Fielding-Wells (2018), Viirman & Nardi (2021)

(Ex.) Lesh et al. (2008), Dvir & Ben-Zvi (2018), Fergusson & Pfannkuch (2020), van Dijke-Droogers et al. (2021)

(Ex.) Wilkerson & Leina (2018), Kazak et al. (2021/2023), Zapata-Cardona & Martínez-Castro (2021/2023)

(Ex.) Ainley & Pratt (2017), Doerr et al. (2017), Aridor & Ben-Zvi (2018), Biehler et al. (2018), Makar & Allmond (2018), Patel & Pfannkuch (2018)

Budgett & Panucho (2018), Podworny & Biehler (2021/2022)

Figure 3. Role of data discerned in the literature (*n* = 63)[1]

Upon scrutinising the 26 papers using data in line with Role C (data as a storyteller of the data-generating process, Context⇄Data) more closely, we found a split between studies that focused on transitions between real-world contexts and data (*n* = 19), and studies that emphasised transitions between theoretical contexts (the world of probability distributions) and data (*n* = 7).

The former group of papers (*n* = 19) typically focused on data modelling, creating empirical or descriptive models of distributions from real or realistic data, or were included as part of the overall activity. However, details of the intended role of data in these papers were varied: raising awareness of uncertainty (*n* = 2); critical examination of data (*n* = 4); making sense of the data (*n* = 3); facilitating model development (*n* = 1); understanding of problem situations (*n* = 8); and understanding the causes of variation (*n* = 1).

The latter group of papers (*n* = 7) aimed to use simulation software to generate simulated data from a model that resembled the behaviour of the population used in the examples in the papers, with a focus on modelling chance, where the role of the data was to facilitate learners’ understanding of the process by which a sample is generated from the population.

### Role of data in terms of different rationales for the teaching of statistical modelling

Table 4 presents a cross-tabulation of the 63 papers analysed, showing how the rationales for teaching statistical modelling identified by Kawakami and Ärlebäck (2022) relate to the roles of data identified in the papers. Rationale R1 (competency oriented), prevalent in 76% of the papers analysed, used data in all three of Roles A, B and C. For papers that used rationale R2 (content oriented), all but two (95%) used data in line with Role A. Half of the 10 papers that used rationale R3 (socially oriented) adopted Role C.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Role A** | **Roles A and B** | **Roles A and C** | **Roles B and C** | **All Roles** | **Total** |
| **R1** | 3 | 11 | 2 |  | 3 | 19 |
| **R2** | 2 | 9 |  |  | 2 | 13 |
| **R1 and R2** | 3 | 5 | 2 | 2 | 9 | 21 |
| **R1 and R3** |  | 1 | 3 |  | 1 | 5 |
| **R2 and R3** | 1 |  | 1 |  |  | 2 |
| **R1, R2, and R3** |  | 2 | 1 |  |  | 3 |
| **Total** | 9 | 28 | 9 | 2 | 15 | 63 |

Table 4. Cross-tabulation of the role of data and three rationales for teaching statistical modelling discerned in the literature (*n* = 63)

## Discussion and conclusion

The result presented in Figure 3, namely that the most common use of data in the papers analysed was according to Roles A and B (data as a source for models and as evidence to test model validity), corresponds with the observation made by Cobb and Moore (1997, p. 810) that ‘[s]tatistics in practice resembles a dialogue between models and data’. Moreover, regarding the second-most common use of data in the papers analysed, all three roles were in line with the conclusion by Pfannkuch et al. (2018, p. 1116), which recognised data as central in the teaching of statistical modelling: ‘[C]ontexts interact and play an important role in promoting students’ learning and reasoning from data’.

Table 4 reflects that the papers in which R1 (competency oriented) was the only rationale identified (*n* = 19), included Roles A and/or B. Data used in accordance with Role C featured in only five papers (26%), and then never as the only role. This result is also in line with Cobb and Moore’s (1997, p. 810) finding: ‘Statistics in practice resembles a dialogue between models and data’. If one of the aims of mathematical modelling is the ability to move between the real world and the mathematical world (e.g. Blomhøj & Jensen, 2007), then one of the aims of statistical modelling may be the ability to move between the data world and the model world (Pfannkuch et al., 2018).

The fact that almost all the papers in which R2 (content oriented) was the only rationale for teaching statistical modelling had adopted Role A, illustrates the importance of the learner’s confidence in making models from the data when constructing statistical contents.

In the socially oriented papers (R3), using data in line with Role C, the focus was on making sense of and critically examining data and the context behind the data. With the growth of big and social data, critical examination of the processes by which real data are generated (e.g. Wilkerson & Laina, 2018) will become increasingly important.

This paper presented a framework for identifying the role of data in empirical research studies, focusing on the teaching of statistical modelling based on a literature review. The framework provides a perspective to facilitate researchers and teachers to identify the purposeful use of data according to the rationale for teaching statistical modelling and to compare the positions and characteristics of existing statistical modelling practices. Given the diverse ways in which high-quality data can be collected, managed and organised in large quantities (Burrill & Pfannkuch, 2023), and the need to rethink the concept of data and its role in mathematics and statistics education (Lee & Wilkerson, 2018), this analysis could be extended by conducting a more qualitative study of the role and types of data actually used in teaching statistical modelling. This would provide a more nuanced picture and suggest directions for further theoretical and empirical lines of enquiry regarding the teaching of statistical modelling.

### Note

1. For a list of the full references of all 63 papers in the literature data set, see <https://drive.google.com/file/d/1H1eLQ6OGJSq3XQE5pmFto5PSExCF2uq_/view?usp=sharing>

## References

Ainley, J., & Pratt, D. (2017). Computational modeling and children’s expressions of signal and noise. *Statistics Education Research Journal*, *16*(2), 15–37. <https://doi.org/kvtj>

Aridor, K., & Ben-Zvi, D. (2017). The co-emergence of aggregate and modelling reasoning. *Statistics Education Research Journal*, *16*(2), 38–63. <https://doi.org/kvtk>

Ärlebäck, J. B., & Kawakami, T. (2023). The relationship between statistics, statistical modelling and mathematical modelling. In G. Greefrath et al. (Eds.), *Advancing and consolidating mathematical modelling: Research from ICME-14* (pp. 293–309). Springer.

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. A. (2020). *Pre-K-12 guidelines for assessment and instruction in statistics education Ⅱ (GAISEⅡ)*. American Statistical Association.

Blomhøj, M., & Jensen, T. H. (2007). What’s all the fuss about competencies? In W. Blum, P. Galbraith, H.-W. Henn & M. Niss (Eds.), *Modelling and applications in mathematics education. The 14th ICMI study* (pp. 45–56). Springer.

Burrill, G., & Pfannkuch, M. (2023). Emerging trends in statistics education. *ZDM Mathematics Education*. [https://doi.org/10.1007/s11858-023-01501-7](%20https:/doi.org/10.1007/s11858-023-01501-7)

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, *104*, 801–823. <https://doi.org/10.2307/2975286>

Doerr, H. M., delMas, R., & Makar, K. (2017). A modeling approach to the development of students’ informal inferential reasoning. *Statistics Education Research Journal*, *16*(2), 86–115. <https://doi.org/10.52041/serj.v16i2.186>

Dunn, P. K., & Marshman, M. F. (2020). Teaching mathematical modelling: a framework to support teachers’ choice of resources. *Teaching Mathematics and its Application*, *39*, 127–144. <https://doi.org/10.1093/teamat/hrz008>

Frejd, P., & Ärlebäck, J. B. (2021). Connections and synergies between the teaching and learning of statistics and modelling: A pilot study. In F. K. S. Leung et al. (Eds.), *Mathematical modelling education in east and west* (pp. 607–617). Springer.

Gal, I., & Geiger, V. (2022). Welcome to the era of vague news: A study of the demands of statistical and mathematical products in the COVID-19 pandemic media. *Educational Studies in Mathematics*, *111*(1), 5–28. <https://doi.org/10.1007/s10649-022-10151-7>

Kawakami, T., & Ärlebäck, J. B. (2022). Characterising rationales for using statistical modelling in education research from a mathematical modelling perspective: A review. In J. Hodgen, et al. (Eds.), *Proceedings of CERME12* (pp. 1109–1116). Free University of Bozen-Bolzano and ERME. Retrieved from <https://hal.science/hal-03759036>

Kazak, S., Fujita, T., & Turmo, M. P. (2021/2023). Students’ informal statistical inferences through data modeling with a large multivariate dataset. *Mathematical Thinking and Learning*, *25*(1), 23–43. <https://doi.org/10.1080/10986065.2021.1922857>

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signal in noisy processes. J*ournal for Research in Mathematics Education*, *33*(4), 259–289. <https://doi.org/dh82nb>

Langrall, C., Makar, K., Nilsson, P., & Shaughnessy, J. M. (2017). Teaching and learning probability and statistics: An integrated perspective. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 490–525). NCTM.

Lee, V. R., & Wilkerson, M. (2018). Data use by middle and secondary students in the digital age: A status report and future prospects. *Commissioned Paper for the National Academies of Sciences, Engineering, and Medicine, Board on Science Education*, Committee on Science Investigations and Engineering Design for Grades 6-12. Washington, D.C.

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>

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, *8*, 82–105. <https://doi.org/10.52041/serj.v8i1.457>

OECD (2019). *OECD future of education and skills 2030 concept note*. Retrieved from <https://www.oecd.org/education/2030-project/contact/>

Patel, A., & Pfannkuch, M. (2018). Developing a statistical modeling framework to characterize Year 7 students’ reasoning. *ZDM Mathematics Education*, *50*(7), 1197–1212. <https://doi.org/10.1007/s11858-018-0960-2>

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

Rubin, A. (2020). Learning to reason with data: How did we get here and what do we know? *Journal of the Learning Sciences*, *29*, 154–164. <https://doi.org/ggrwz9>

Shibata, R. (2015). *Data assay and data science*. Kindai kagaku sha Co.,Ltd. (in Japanese)

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(1), 223–265. <https://doi.org/10.1111/j.1751-5823.1999.tb00442.x>

Wilkerson, M. H., & Laina, V. (2018). Middle school students’ reasoning about data and context through storytelling with repurposed local data. *ZDM Mathematics Education*, *50*(7), 1223–1235. <https://doi.org/10.1007/s11858-018-0974-9>