EXPLORING PRIMARY SCHOOL STUDENTS' VALIDATION APPROACHES IN THE TASK OF MODELLING WITH DATA

Takashi Kawakami[1], Katsuki Akizawa[2]

[1]Utsunomiya University, [2]Utsunomiya University-Affiliated Elementary School

A data-driven society requires citizens, including young students, to develop the ability to validate data-based models, modelling processes, and data-generation processes. This study examined various validation approaches in a data-informed modelling task involving 30 Grade 6 students aged 11–12 years. Findings from the task-based questionnaire indicated that students validated the model itself, modelling processes, data, data-generation processes, and final products. Moreover, both deterministic and stochastic perspectives derived from data informed students' validation practices. These findings highlight the value of integrating rich, data-informed validation activities into primary school mathematical modelling.

INTRODUCTION

A data-driven society requires citizens, including young students, to validate models using mathematics and statistics, as well as modelling processes and data-based outcomes (Gal & Geiger, 2022). In addition, validating data-generation processes from real-world contexts and sources is essential to enable citizens to critically analyse mathematical and statistical information presented in mass media and public datasets (Wilkerson et al., 2021). Within this context, mathematical modelling education faces a new challenge: ensuring that students meaningfully engage in validation activities during data-informed modelling tasks in the early school years (Czocher, 2018; Geisler, 2021; Park et al., 2025). To address this challenge, the present study explored the potential for primary school students' validation approaches in a task involving *modelling with data* (MWD). In this study, MWD refers to generating, validating, and refining data-based models using mathematics and statistics to support predictions, decisions, and actions (English & Watson, 2018; Kawakami & Saeki, 2024).

Czocher (2018) defines validation in modelling as "comparing an obtained result with an (implicitly) anticipated result" (p. 144). The obtained result represents the object of validation, whereas the anticipated result encompasses the modeller's idealised image, logical expectations, and mathematical results or processes (e.g., predictions or mathematical expressions). Validation activities include verifying calculations, checking whether a solution satisfies the problem conditions, assessing the reasonableness of results, and evaluating the quality of the solution process itself (Czocher, 2018). Objects of validation vary and include assumptions, models, and both intermediate and final results (Czocher, 2018). Thus, validation in modelling has been reconceptualised as a complex, higher-order activity.

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. 779-786. PME.

Researchers have examined secondary and university students' validation approaches in MWD. Geisler (2021) reported that secondary school students struggled with the validation of their models and tended to validate measurements during experiments. Park et al. (2025) examined university students' validation approaches to models developed for data-informed modelling tasks. They found that open and authentic data-informed tasks encouraged students to validate not only the fit between data and models but also the plausibility of assumptions grounded in their extra-mathematical knowledge and prior modelling experiences. However, to the best of our knowledge, no study has examined what primary school students validate and how they engage in validation in MWD. This study addresses this gap in the literature.

THEORETICAL FRAMEWORK

To develop a framework for identifying primary school students' validation approaches, this study adapted three theoretical ideas: deterministic and stochastic perspectives on data; components of the MWD process, with a focus on mathematics and statistics; and data-generating processes grounded in context and data sources.

Data play a crucial role across various aspects of MWD, such as serving as model sources and reference points for testing and validating models. They possess two key epistemic characteristics. First, data consist of numerical values situated within real-world contexts (Cobb & Moore, 1997). Second, they have a structure comprising a deterministic aspect, which emphasises exact values and causal explanations with certainty, and a stochastic aspect, which emphasises variability (Innabi et al., 2023). Sources of variability include natural variation, measurement error, sampling processes, and random events. The deterministic and stochastic perspectives on data align with fundamental differences in data handling between mathematics and statistics.

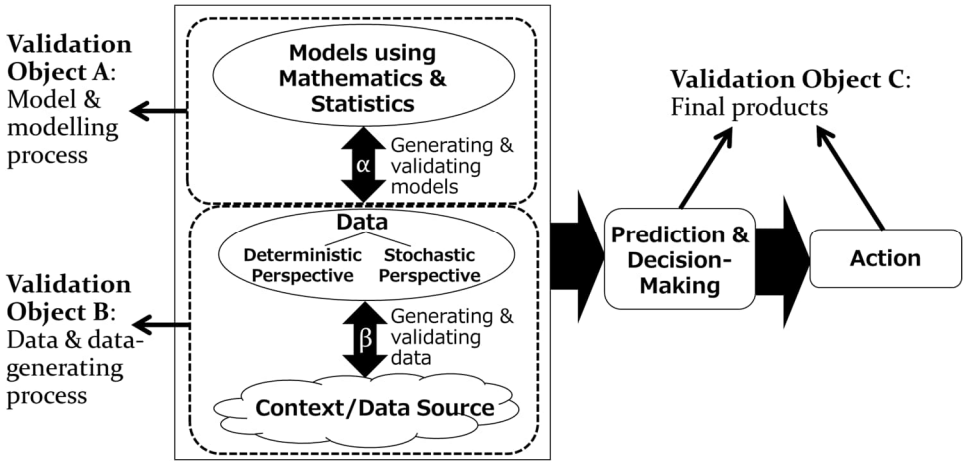
English and Watson (2018) proposed four components of the MWD process for primary school students, emphasising the integration of mathematics and statistics. The first component is "working in shared problem spaces between mathematics and statistics". This involves students integrating mathematical approaches or deterministic perspectives on data (i.e. operations and reasoning focused on invariant data structures) with statistical approaches or stochastic perspectives on data (i.e. operations and reasoning that focus on variability) in MWD (Kawakami & Saeki, 2024). The second component is "interpreting and reinterpreting problem contexts and questions", which involves students attending to the real-world context underlying the data. The third component is "interpreting, organising, and operating on data in model construction", which entails students generating, validating, and modifying data-based models (representations). The fourth component is "drawing informal inferences", which involves predicting unknown data from available data without relying on formal mathematical or statistical methods to support decision-making and action.

While MWD often begins with given data, the proliferation of unstructured and complex data in daily life has increased emphasis on generating data from contextual

sources as part of modelling, including identifying data attributes and measuring, structuring, and cleaning data (Fielding et al., 2025; Wilkerson et al., 2021). Validating the data-generating process is essential even for primary school students to critically evaluate reports and information derived from data-based models (Gal & Geiger, 2022).

Building on these theoretical ideas, this study developed the MWD framework to identify primary school students' validation objects and perspectives (Figure 1). The framework comprises key MWD elements, including context and data sources, data, models using mathematics & statistics, prediction & decision-making, and action, and key MWD processes, including transitions α and β. These elements and processes serve as perspectives for identifying primary validation objects, namely *models and modelling processes, data and data-generating processes, and final products*. *Deterministic and stochastic perspectives* derived from data at the core of MWD elements and processes were used to identify students' validation perspectives. In light of the MWD framework, this study addresses the following research question: *How do primary school students validate key MWD elements and processes by comparing their predictions and decision-making with outcomes after an action?*

Figure 1: MWD framework for identifying students' validation approaches



METHODS

Participants and lessons

To answer the research question, we conducted seven MWD lessons (45 minutes each) and a subsequent questionnaire-based validation using the same task as in the lessons with one Grade 6 class (n = 30; aged 11–12) in Japan and analysed students' questionnaire responses. The students had learned about proportionality and descriptive statistics. However, they had no prior experience with MWD. During the lessons, the students worked individually and in groups on the *Ring Decoration Task* (Figure 2), developing models to estimate the number of origami sheets needed to decorate the gymnasium's inner circumference with ring decorations for a thank-you party. The students were asked to repeatedly make assumptions, generate measures and structures, and represent data relevant to the ring decorations. As a result, the task constituted an open and authentic context for validation (Park et al., 2025).

Figure 2: Ring Decoration Task

For a thank-you party, you need to decorate the gymnasium's inner circumference with ring decorations.
We have 720 origami sheets for the decoration.
Will there be sufficient?

The lessons consisted of four main activities based on key MWD elements and processes (Figure 1). Activity 1 involved identifying the required information (e.g., the number of origami sheets and pieces, as well as the length of the gymnasium's inner circumference to be decorated) (*context/data source*). In Activity 2, the students created ring decorations and collected data, including the length and width of a single set of ring decorations and the number of ring decorations within that set (*transition* β). Activity 3 was a group activity in which the students used data to construct models predicting the number of sheets required (*transition* α). Activity 4 involved sharing representative group models and reaching a consensus on a classroom model (Figure 3), which used proportional reasoning to predict the width of one set of ring decorations and the number of origami sheets required for the gymnasium decorations (*prediction*). Based on the prediction, the students decided that there would be just enough origami paper (*decision-making*). The party was cancelled due to COVID-19-related school closures. Consequently, the researchers decorated the gymnasium with ring decorations and recorded video footage (*action*). The prepared origami paper was sufficient; however, approximately 4 m of ring decorations remained unused.

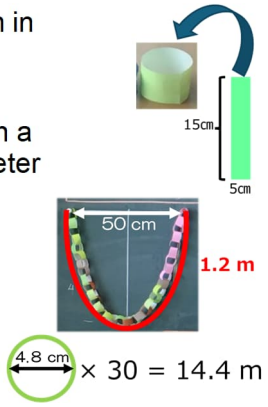
Figure 3: Classroom model

(a) Assume each piece used for one ring is a rectangle measuring 15 cm in height and 5 cm in width. We can get five pieces from one sheet of origami paper and 3,600 pieces from 720 sheets of origami paper.

(b) Given that the gymnasium's perimeter is 60 m, assume a ring set with a width of 50 cm (120 ring sets) is created to decorate the entire perimeter without any leftover material.

(c) A prototype ring set is made, and its length is measured to be approximately 1.2 m. Approximately 30 rings are used.

(d) Calculating the diameter of one ring gives 4.8 cm (15 ÷ 3.14). Attaching 30 rings without overlap gives 4.8 × 30 = 14.4 m. This is longer than 1.2 m and therefore sufficient.



Task-based questionnaire for validation after the lessons

The students were asked to watch a video of the gymnasium decorations and complete a validation questionnaire based on the Ring Decoration Task (Figure 2) at home. The questionnaire worksheet presented the task, context/data sources, and assumptions shared in the classroom (e.g., the number of origami sheets and pieces, the dimensions of an origami sheet and its pieces used for ring decoration, and the glue allowance for making the rings). The worksheet also presented the classroom model (Figure 3). Students could refer to the context/data sources, the data used, and the classroom model when comparing their predictions and decision-making with the actual decoration

results on the worksheet. They were asked to write their reflections after comparing their predictions and decisions with the decoration results by responding to the following prompt: Compare your predictions and decisions with the actual decoration results. Explain why the outcome matched or differed from the prediction.

Data collection and analysis

All respondents' texts were analysed quantitatively using a two-phase coding procedure based on the MWD framework (Figure 1). Lesson transcripts were used solely as supplementary materials to explain the findings. First, we identified texts that compared actual results or processes with anticipated results or processes (including idealised or assumed images) and coded the Validation Objects A (models, transition α), B (data, transition β), and C (predictions, decision-making, and decorating action). In the analysis, models included representations of assumptions, values calculated using the models, and intermediate calculations. Second, we coded each student's validation perspective as deterministic or stochastic. Validation oriented towards exact numbers and causal explanations with certainty was coded as deterministic, whereas validation oriented towards uncertainty and variability (natural, measurement, sampling, and random variability) was coded as stochastic. The first author conducted all analyses, and the second author reviewed them. Differences in interpretation between the authors were discussed, and the analysis was revised accordingly.

RESULTS

Objects validated by students in the data-informed modelling task

We identified three primary validation objects that the students evaluated. The first involved models and modelling processes (Validation Object A). For example, a student, Mai, validated the classroom model by focusing on proportional reasoning (e.g., "The number of ring decorations and the length of a set of ring decorations may not be proportional"). The validation objects included assumptions embedded in the classroom model (e.g., "the proportionality between the number of ring decorations and their total length"), as well as the accuracy of mathematical processing (e.g., "the calculation of the length of a set of ring decorations using these models"). In addition, this validation encompassed the process of constructing the classroom model based on the gymnasium's inner circumference.

The second validation object concerned data and data-generating processes (Validation Object B). For example, a student, Yuri, validated transition β by drawing on knowledge of measurement processes and error (e.g., "I think that the extra 4 m were too long, although there might be some errors due to slack or angles. Another possible reason could be that we could not set the glue line to 0 cm while making it."). The validation objects included specific data values (e.g., "the diameter of a ring decoration and the dimensions of a set of ring decorations used to generate the classroom model"). In addition, this validation encompassed the processes of creating a glue line for the

ring decorations, measuring the length of specific ring decorations, and determining the width of a set of ring decorations.

The third validation object was the final product (Validation Object C). For instance, a student, Fumi, validated the action (e.g., “When putting on the ring decorations, some were loose and others were not as loose, so it was not possible to put them on the same way every time”). The validation objects included the accuracy of predictions (e.g., “whether there would be sufficient origami paper”) and the appropriateness of the actions (e.g., “how to decorate and how far the ring decorations should hang down”).

Types of combinations of student validation objects

Table 1 lists six types of combinations of student validation objects. All students validated one or more of the following: models and modelling processes, data and data-generating processes, or final products. Types 1 and 2 were the most common (30%) and second most common (26.7%), respectively, and involved single validation objects. Approximately 40% of the students were categorised as Types 3, 4, 5, or 6, which involved multiple validation objects. An example of Type 6 is a student’s validation (Kana; Figure 4), in which the classroom model and data-generating processes (e.g., “making ring decorations and measuring the length and width of one set of ring decorations”), prediction and decision-making, and the action of actual decorating were validated using the knowledge of approximation and measurement error.

Table 1: Types of combinations for student validation objects (n = 30)

Type	Object	Description	#
1	A	Validating models only	9
2	C	Validating final products (prediction, decision-making, and action) only	8
3	A & B	Validating models and data/data-generating processes	4
4	A & C	Validating models/modelling processes and final products (action)	2
5	B & C	Validating data/data-generating processes and final products (prediction, decision-making, and action)	2
6	A, B, & C	Validating models, data/data-generating processes, and final products (prediction, decision-making, and action)	5

Figure 4: Kana’s validation text as an example of Type 6

The prediction of “enough” was correct. Additionally, precisely 120 sets were formed. The calculations [from the classroom model] provided a close (approximate) number but not a true one. Errors occurred within a set of ring decorations varying from 120 to 121 cm. The error was even greater when 120 sets were collected. It is no wonder that there were a few leftovers. However, I think that [the classroom model] was relatively accurate, albeit with some errors?! It shows the results of various ideas proposed by everyone in the class.

Deterministic and stochastic perspectives for validation

Twenty-seven students validated models and modelling processes, data and data-generating processes, and final products from deterministic, stochastic, or dual perspectives. Ten students validated models, data, data-generating processes, and final products from a deterministic perspective, using words such as “precisely”, “properly”, “firmly”, and “snugly”. Fourteen students validated these objects from a stochastic perspective, using words such as “scattered”, “misalignment”, and “error”. The sources of variability identified by the students were natural (n = 6), measurement (n = 11), and random-related (n = 2). Three students, one of whom was Kana (Figure 4), validated all models and modelling processes, data and data-generating processes, and final products from dual perspectives. Kana validated data-generating processes and actions using the word “error” and by considering uncertainty and measurement variability. A stochastic approach was followed to validate the classroom model. Additionally, Kana validated the accuracy of the models using the word “accurate” from a deterministic perspective, resulting in a relative validation of the models from dual perspectives.

DISCUSSION

We discuss two findings regarding the multifaceted and interconnected nature of validation demonstrated by Grade 6 students in a data-informed modelling task. First, a range of validation objects and validation perspectives was observed. The students validated one or more models and modelling processes (models, transition α), data and data-generating processes (data, transition β), and final products (prediction, decision-making, and action) when comparing predictions and decision-making with actual outcomes (Table 1). Moreover, both deterministic and stochastic perspectives derived from data informed their validation practices. On the one hand, the multifaceted nature of validation in MWD was enhanced by placing data at the core of the modelling process, as deterministic and stochastic perspectives derived from data were evident across validation objects. On the other hand, this multifaceted nature entails substantial cognitive demands (Czocher, 2018; Geisler, 2021; Park et al., 2025). Indeed, fewer students validated multiple objects than a single object (Table 1), and only three students validated from both deterministic and stochastic perspectives. As illustrated in Figure 4, Kana validated multiple objects from dual perspectives by reflecting on the lesson activities and decoration context recorded in the video. This finding aligns with Park et al.’s (2025) observation that students’ validation is grounded in prior experiences. These results highlight the role of real-world contexts and hands-on modelling experiences in multifaceted validation, as well as the need for appropriate teacher interventions to support such validation among primary school students.

Second, connections and interactions among the validation of models and modelling processes, data and data-generating processes, and final products were evident in the students’ validation activities (Table 1). The validation of context and data sources was connected to other forms of validation evident in the students’ texts. This finding supports Wilkerson et al.’s (2021) claim that data-generating processes function

purposefully within modelling processes. Specifically, as illustrated by Kana's case (Figure 4), eight students in Types 3, 4, and 6 noticed variability (e.g., measurement errors) when validating data-generating processes and/or actions, reinterpreted the classroom model that assumed proportionality, and evaluated the model stochastically by considering error and uncertainty. The results indicate that interconnected validation activities in MWD can foster blended mathematical and statistical reasoning, which is indispensable for citizens living in disruptive times (e.g., Gal & Geiger, 2022).

This study illustrates primary school students' potential for diverse validation approaches, as well as cognitive demands involved in MWD, which have not previously been documented. The findings can contribute to the design of rich validation activities and teacher support for MWD. However, as the analysis was based on a task, a questionnaire, and a small sample size, further research, including other tasks, interviews, and implementation across additional classes, is necessary.

References

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, *104*(9), 801-823.

Czocher, J. A. (2018). How does validating activity contribute to the modeling process? *Educational Studies in Mathematics*, *99*(2), 137-159.

English, L., & Watson, J. (2018). Modelling with authentic data in sixth grade. *ZDM–Mathematics Education*, *50*(1-2), 103-115.

Fielding, J., Makar, K., & Ben-Zvi, D. (2025). Developing students' reasoning with data and data-ing. *ZDM–Mathematics Education*, 57(1), 1-18.

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.

Geisler, S. (2021). Data-based modelling with experiments–Students' experiences with model-validation. In M. Inprasitha et al. (Eds.), *Proc. 44th Conf. of the Int. Group for the Psychology of Mathematics Education* (Vol. 2, pp. 330-339). PME.

Innabi, H., Marton, F., & Emanuelsson, J. (2023). Sustainable learning of statistics. In G. F. Burrill, L. de Oliveria Souza, & E. Reston (Eds.), *Research on reasoning with data and statistical thinking: International perspectives* (pp. 279-302). Springer.

Kawakami, T., & Saeki, A. (2024). Roles of mathematical and statistical models in data-driven modelling: A prescriptive modelling perspective. In H.-S. Siller, et al. (Eds.), *Researching mathematical modelling education in disruptive times* (pp. 595-605). Springer.

Park, J., Czocher, J. A., & White, R. T. (2025). Exploring STEM students' model validation across two data-informed task environments. *International Journal of Science and Mathematics Education*. (online puplished)

Wilkerson, M. H., Lanouette, K., & Shareff, R. L. (2021). Exploring variability during data preparation: a way to connect data, chance, and context when working with complex public datasets. *Mathematical Thinking and Learning*, *24*(4), 312-330.