HOW MODEL AND MODELLING APPROACHES CAN PROMOTE YOUNG CHILDREN'S STATISTICAL REASONING

Takashi Kawakami
Utsunomiya University, Japan
t-kawakami@cc.utsunomiya-u.ac.jp

The need to overcome the boundaries between statistics education and model and modelling approaches has been increasingly emphasised. These approaches spotlight children's models and the action to create, modify, and apply models connected to real-world contexts. This study illustrates how model and modelling approaches encourage young children (aged 7–8 years) to focus on the views of distribution, including the relationship among the elements of distribution, leading to statistical reasoning. To model variation and formulate decisions with data, the children individually created models, collectively made up a dot plot, and then individually evaluated and revised their initial models. Articulation and changes in the children's statistical reasoning, associated models of distribution, and creation processes in collaborative and individual modelling are discussed in relation to the roles of inscriptions.

INTRODUCTION

Statistical reasoning is regarded as an essential elementary attainment in today's information and knowledge-based society, in which people face an enormous amount of data and form decisions amid uncertainty. This skill involves the ability to make sense of and explain statistical information, procedure, and processes by connecting several statistical concepts and ideas (Ben-Zvi & Garfield, 2004). It focuses on the construction and utilisation of fundamental statistical concepts (i.e. variation and distribution). Previous literature (e.g. Konold et al., 2015) has indicated that developing statistical reasoning takes time and must begin in the earliest years of schooling. Limited research exists, however, on developing young children's statistical reasoning, including informal inferential reasoning (e.g. Ben-Zvi & Sharett-Amir, 2005; Makar, 2016).

Recent research both in statistics and mathematics education communities has increasingly highlighted the contribution of model and modelling approaches to foster children's statistical reasoning (e.g. English, 2014; Fielding-Wells & Makar, 2015; Kawakami, 2017; Lehrer & Schauble, 2004). Nevertheless, the interdependent relationship between the characteristics of model and modelling approaches and the development of statistical reasoning remains poorly understood. This study illustrates how model and modelling approaches encourage young children (aged 7–8 years) to convey articulately, promote, and apply statistical reasoning, through a teaching experiment on the creation of dot plots to model variation and make decisions with data.

BACKGROUND

A fundamental component of statistical reasoning, distribution is at the heart of statistics (Lehrer & Schauble, 2004). Bakker and Gravemeijer (2004) demonstrated the emergent features of distribution, such as *centre*, *spread*, *density*, and *skewness*. The aggregate lens of distribution enables the organisation and perception of data variation as a unity with these emergent properties (Bakker & Gravemeijer, 2004; Konold et al., 2014). Previous research (e.g. Ben-Zvi & Sharett-Amir, 2005; Konold et al., 2014) indicates the difficulties of children in developing an aggregate view of distribution for and the necessity of fostering young students' informal views of distribution and gradually developing their aggregate view.

Recent research has indicated the role of model and modelling approaches in developing this aggregate view and fostering statistical reasoning (e.g. English, 2014; Lehrer & Schauble, 2004). A *model* refers to a representation of structure in a given system from the cognitive perspective (Hestenes, 2010). Hestenes (2010) defines a *system* as a set of related objects and the *structure* of a system as a set of relationships among the objects in the system. It is crucial that children create their own models of distribution that form their aggregate view by selecting and connecting the relevant elements of the distribution. The modelling process, in which various models are created and improved by switching between the physical world and model world, can be the vehicle to construct statistical concepts, such as distribution and variation, from real-world

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

contexts (e.g. Kawakami, 2017; Lehrer & Schauble, 2004).

Furthermore, model and modelling approaches can highlight inscriptions as models and the action to create, validate, modify, and apply inscriptions connected to real-world contexts (Lehrer & Lesh, 2003). Inscriptions include drawing, maps, diagrams, text, recordings from instruments, mathematical formalisms, and even physical models (Lehrer et al., 2001). These inscriptions reflect the subject's understanding and thinking. Referring to Lehrer and Lesh (2003), two roles of inscriptions in mathematical reasoning are summarised: (1) Tools for reasoning and triggers for conceptual understanding, and (2) Mediations of reasoning. Concerning the former, children invent, change, and use inscriptions as thinking tools for investigating and solving the problem, leading to new ideas. Model and modelling approaches can create opportunities to invent and evaluate inscriptions. As for the latter, inscriptions are recruited to create and articulate mathematical arguments in the classroom and otherwise. This study adapts these roles of inscriptions in mathematical reasoning to the context of statistical reasoning to illustrate the interdependent relationship between the generation of view of distribution in collaborative modelling and the creation and modification of models (inscriptions) of distribution in individual modelling.

METHODOLOGY

This study addresses the following research questions: *(1) How do children generate their view of distribution in collaborative modelling? (2) How do children verify and modify their initial models of distribution in individual modelling after participating in collaborative modelling?* Children's individual and group works were considered in a teaching experiment on the creation of dot plots. The 7- to 8-year-old participants comprised 27 second-grade children (15 males and 12 females). The lessons were delivered in a public primary school in Japan. Although the children had learned about picture graphs of categorical data, they were inexperienced with numerical data. The real-world context treated in the lessons was the children's real data of softball throws in a physical fitness test. The authentic context encourages the children's interest in statistical information around them (English, 2014; Lehrer & Schauble, 2004), as well as its utility in evaluating their ball-throwing ability and then motivating them to enhance it.

The teaching experiment consisted of two 45-minute lessons. These lessons were designed such that the children could individually create models related to dot plots, and then collectively compare multiple models, evaluate and revise their initial models to organise variation, and form decisions based on the data. The children were each provided with a tablet loaded with each child's handball-throwing (values of distance) records matched to the smiley face icons. The children could create and modify models by moving the icons; their responses were recorded automatically. By saving the worksheets on their tablet, the children were able to refer to and compare their models with their peers' models. In the first lesson, the teacher placed the icons bearing each child's record of ball throwing on the blackboard and then asked the children, 'Compared to everyone in the class, which term best describes your ball-throwing record: good, typical, or not good?' The children noted that they could not compare properly because the icons were randomly placed. Therefore, the teacher set up the task: *How would you organise your classmates' records comprehensively?* The children arranged the data on their respective tablet. Then, the teacher demonstrated the typical models on the children's tablets. Subsequently, the teacher encouraged them to compare different models and modify them collectively to create a better representation, such as dot plot, that would facilitate easier understanding of the tendency of the ball-throwing records' distribution. In the second lesson, the children collectively created a dot plot. The children were asked to evaluate their own ball-throwing record and set goals for next year's ball-throwing test based on the dot plot.

Before the first lesson and after the second lesson, the children completed the pre- and post-tasks to investigate the change in children's models of distribution and their processes of model creation through the lessons. The contexts of the tasks adopted losing milk teeth in Ben-Zvi and Sharett-Amir (2005). Both tasks involved the following. (1) *Organize the number of lost milk teeth so that your classmates can understand around how many teeth 30 second-grade children lose*. (2) *Answer around how many teeth 30 second-grade children lose*. The children were asked to arrange the data on their tablet where they could refer to earlier inscriptions in the pre-task.

The data collection comprised transcripts of videotaped class discussions and individual tablets' screens recorded by BB Flash Express, children's inscriptions (i.e. graphical representations, sketches, and comments), teacher's inscriptions on the blackboard, and field notes. The data were coded and examined for patterns and trends according to which elements of distribution (e.g. centre, spread, density, and shape) the children focused on and how they connected the elements. The analysis focused on the sequence of creating dot plot in the classroom (RQ1), the change in children's models in the pre- and post-tasks (RQ2), and the change in model creation in the pre- and post-tasks (RQ2).

RESULTS

Generation of view of distribution by referring to created models and its creation process (RQ1)

Figure 1 shows the sequence of creating a dot plot in the classroom: (a) sharing the typical classmate's models (grouping and frequency types) and reading the features, (b) comparing different models, (c) sharing the process of inventing a frequency-type model, (d) reading the features of distribution and linking them with 'typical records', (e) modifying the frequency-type inscription, and (f) examining the values that were the 'typical records'. The teacher helped the children elicit elements of distribution from created models and the creation process.

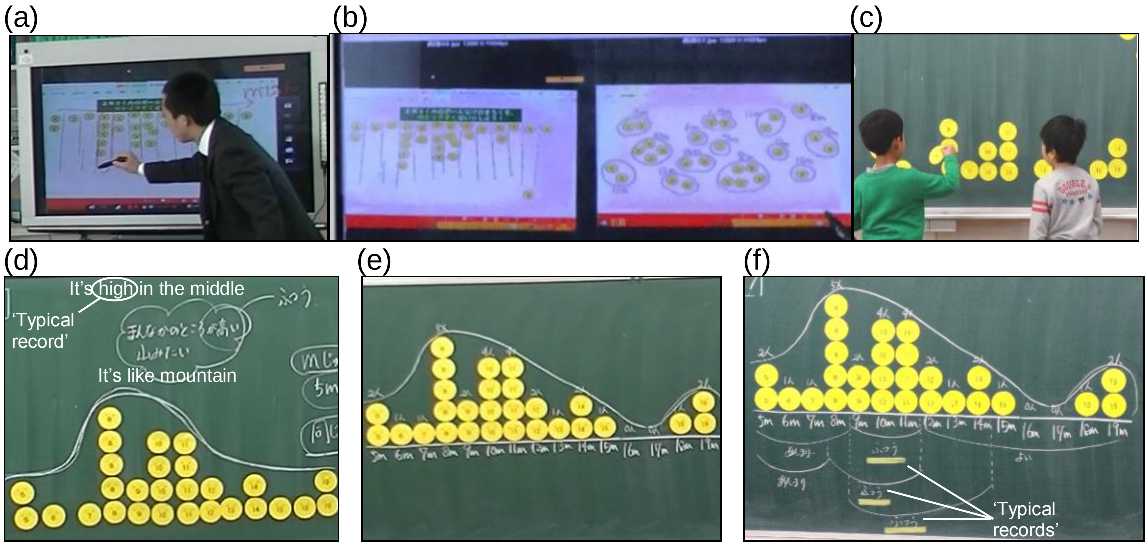

Figure 1(a)–(f). The sequence of creating a dot plot in the classroom

In phases (a) and (b), the teacher helped the children elicit elements of distribution by focusing on and contrasting the invented models. The teacher tackled the grouping- and frequency-type models and then asked the children regarding the models' arrangement and comprehensibility. Then the children observed what values a model took (*spread*) and how often it took these values (*density*) (e.g. '8 m is the most frequent'). These viewpoints of distribution became more explicit in phase (c), creating the frequency-type model (e.g. 'Arranging horizontally according to value order', 'Arranging with one-to-one correspondence with the icons'). In phase (d), the teacher helped the children elicit aggregate views of distribution by looking at the overall features of the frequency-type model. The children noticed the distribution shape and paid attention to the entire central range, 'modal clump' (Konold et al., 2002), by linking the shape with 'typical records':

Teacher: Do you notice something about the whole? (Pointing at the frequency-type model on the blackboard as shown in Figure 1d)
[Dai started tracing with his finger as he drew the distribution shape.]
Teacher: Dai, what is it?
[Other children also started tracing with their fingers.]
Dai: It is like a mountain.
Yae: It is high in the middle.
[The teacher drew the shape of distribution on the blackboard as shown in Figure 1d.]
Teacher: The number of people is getting bigger; it is getting less like this. (Tracing the shape)
Dai: Ah! 'Typical record' means the high place!

In phase (e), the children noticed the values with no frequency and added a horizontal axis and scale of distance with the help of teacher by focusing on range and frequencies. Then they focused on the distribution shape again (e.g. ‘It’s like an island’). In phase (f), the teacher asked the children to decide on ‘typical record’ based on the created dot plot. The children focused on the modal clumps and connected it to the shape of distribution (e.g. ‘8 m to 11 m are typical, as the numbers appear often and are in the place where the mountain is going up’). Thus, they generated an aggregate view of distribution, connecting centre, spread, shape, and density, by referring to the created models and their creation process inscribed on the blackboard.

Modifying or remaking initial models by referring to generated view of distribution (RQ2)

The children’s models of distribution in the pre- and post-tasks were categorised into four types, including elements of distribution. Table 1 shows the results of the children’s models in each task. Overall, 12 children could connect the elements of distribution in the post-task, whereas only one child was able to do so in the pre-task. For *Model B* in each task, the children grouped the icons for each value or subjective record goodness (i.e. ‘not good’, ‘typical’, and ‘good’) without using a baseline. Their models included only a single aspect of the distribution, the interval (*spread*) or the frequency of icons or values (*density*). Concerning *Model C* in each task, the children arranged the icons according to the value or frequency vertically or horizontally with baseline, but they did not indicate the centre of distribution based on the created models (e.g. ‘Thirty second-grade children lost around 7 to 8 teeth’). Their models included several aspects of distribution, what values it takes (*spread*), and how often it takes these values (*density*), but did not connect the aspects. For *Model D* in each task, the children made dot plots and surrounded the distribution like a mountain (*shape*) or surrounded the modal clump (*centre and spread*), and then answered the tendency of distribution connecting these aspects (e.g. ‘Thirty second-grade children lost around 7 to 9 teeth, because the place where the mountain is going up’). A decrease in *Models A*, *B*, and *C* and an increase in *Model D* indicated that the dot plot created in the classroom and viewpoints of distribution embedded in the model facilitated individual children’s multifaceted views of distribution.

Table 1. Types of children’s models of distribution in the pre- and post-tasks (N = 27)

Model	Characteristics	Example of Children’s Models	Frequency (n)	
			Pre	Post
A	No elements of distribution	No inscription/failure to arrange icon to display the data	4	1
B	Single elements of distribution	Grouping icon for each value	12	7
C	Separated elements of distribution	Arranging in vertical or horizontal to the frequency order/value order, but not connecting the elements of distribution explicitly	10	7
D	Connected elements of distribution	Arranging in vertical or horizontal to the value order connecting the elements of distribution explicitly	1	12

A successful case (Kouta) illustrates the change in model creation process through lessons. His model changed from *Model B* to *Model D*. In the pre-test, he initially collected the values of 8 teeth and arranged them in vertical order (Figure 2a). Then he arranged values of 9 and 7 teeth, and surrounded each value (Figure 2b). He focused on the frequency (*density*). Subsequently, he arranged it to the frequency order (except for 13 teeth) (Figure 2c) (taking around 11 min so far). He then answered, ‘Thirty second-grade children lost around 8 to 9 teeth because the values of 8 teeth are higher than the values of 7 teeth’. In the post-task, he initially moved a value of 3 teeth to the lower left of the screen (Figure 3a), saying ‘The value of 3 teeth is little’. He focused on the minimum value (*spread*). In fact, after this, he confirmed the maxim value: ‘There is a person who lost 13 teeth’. Then he looked back at the model created in the pre-task (Figure 2c) and confirmed the frequency of the value of 3 teeth (*density*), as shown in Figure 3b. This action reflected the moving and piling up of the other value of 3 teeth (Figure 3c). He looked back at the model created in the pre-task (Figure 2c) again; he rearranged the values of 3 teeth neatly and arranged the values of 6 teeth (Figure 3d). He confirmed what values it took (*spread*) and remembered the process of creating a dot plot (e.g. Figure 1c) from the created model in the pre-task. In arranging according to the values and piling up a value of 8 teeth (Figure 3e), he focused on the peak of distribution (*centre*): 'The number of values of 8 teeth is a lot'. Lastly, he prepared a dot plot (taking around 6 min so far), surrounded the modal clump from 7 to 9 teeth (centre and spread), and provided the following description: 'There are many here' (Figure 3f), and 'Thirty second-grade children lost around 7 to 9 teeth'. Although he did not describe the horizontal scale, he explicitly connected the density, spread, and centre by referring to his initial model, the classroom's model, and their creation processes.

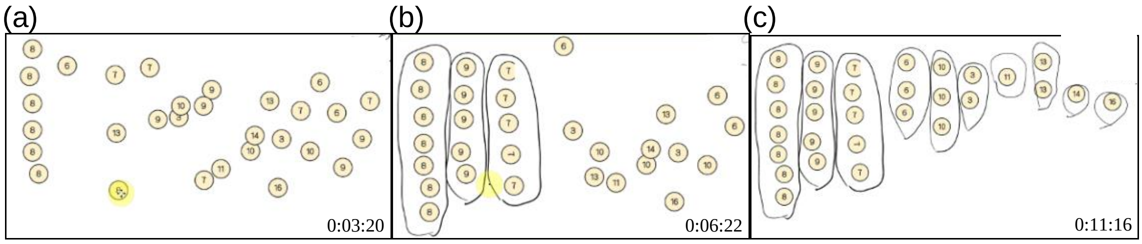

Figure 2(a)–(c). Kouta's process of model creation in the pre-task (*Model B*) with elapsed time

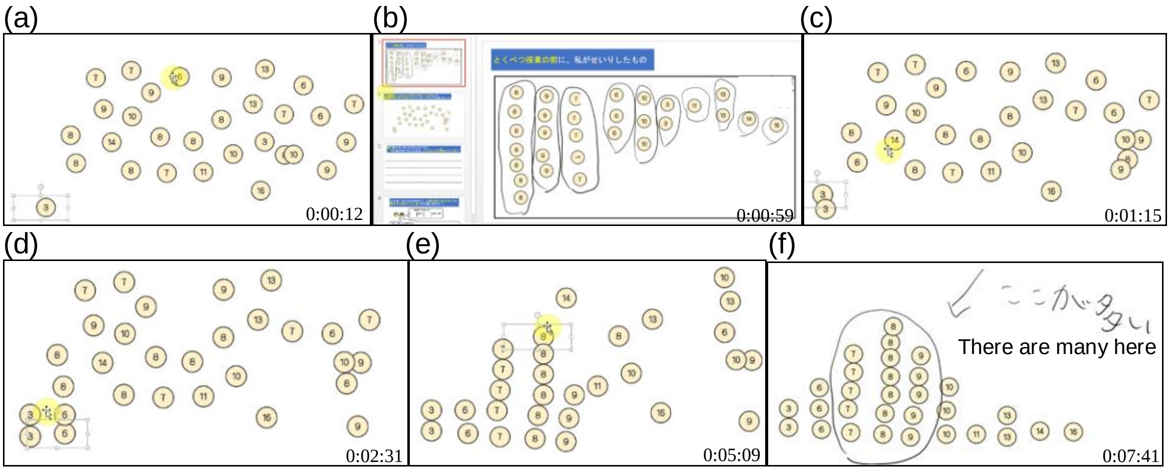

Figure 3(a)–(f). Kouta's process of model creation in the post-task (*Model D*) with elapsed time

DISCUSSION

This study illustrated how model and modelling approaches encouraged young children to focus on the underlying statistical concepts behind real data and the relationships among concepts leading to statistical reasoning. The analysis of dot plot creation sequence in the classroom illustrated that the children elicited and shared a view of distribution and then developed an aggregate view by referring to models on the blackboard (e.g. Figure 1d) and the creation process illustrated on the blackboard (e.g. Figure 1c). Their aggregate view was formed after noticing the shape of distribution (e.g. Bakker & Gravemeijer, 2004). Inscriptions as models helped the children elicit statistical concepts and articulate on their statistical reasoning. This result strengthens the two roles of inscriptions in mathematical reasoning (Lehrer & Lesh, 2003), demonstrating that it extends to the development of statistical reasoning.

The change in children's models of distribution through lessons, namely, a decrease in *Models A*, *B*, and *C* and an increase in *Model D* (see Table 1), indicated that the view of distribution generated and shared in the classroom transferred to the individual lens of distribution. Inscriptions were combined with the view of distribution and articulated collective statistical reasoning to individual statistical reasoning. Kouta forms an association between the classroom's models that inscribed modal clumps, as shown in Figure 1f, and his final model that inscribed modal clump, as shown in Figure 3f. He was able to reason about distribution well, although he was not so in the pre-task. In addition, the change in model creation process confirmed the above results. In Kouta's case, initial action was different between the pre-task (Figure 2a) and post-task

(Figure 3a). In the post-task, he paid attention to what values it took (*spread*) and how often it took these values (*density*), which were generated and shared in the lessons, from the beginning (see Figure 3a, 3b). It is interesting that he articulated his earlier inscription to his later inscription, confirmed the view of distribution, and remembered the creation processes (e.g. Figure 3b, 3d), leading to promoted statistical reasoning (see Figure 3e, 3f). This result suggests that inscriptions as models could convey not only the classroom's statistical-reasoning process but also individual statistical-reasoning process, and model and modelling approaches can highlight the successive change processes in children's inscriptions. Although the reported findings are limited to a small sample, they suggest that models and modelling approaches can activate the interdependence of creating graphs and fostering an aggregate view of distribution, leading to statistical reasoning.

There is value in creating graphs for young children. Previous research (e.g. English, 2014; Makar, 2016) has indicated that inventing representations might better facilitate children's sense-making of data and understanding of the benefits of representations and the value of summarising and organising data. The present study confirms the indication, based on Kouta's comment after lesson 2: 'I think that graphs are useful because the typical area is like a mountain'.

REFERENCES

Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi, & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 147–168). Dordrecht: Kluwer Academic.

Ben-Zvi, D., & Garfield, J. (2004). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In D. Ben-Zvi, & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 3–15). Dordrecht: Kluwer Academic.

Ben-Zvi, D., & Sharett-Amir, Y. (2005). How do primary school students begin to reason about distributions? In K. Makar (Ed.), *Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking, and Literacy*. Brisbane: University of Queensland.

English, L. D. (2014). Promoting statistical literacy through data modelling in the early school years. In E.J. Chernoff, & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 441–457). Dordrecht: Springer.

Fielding-Wells, J., & Makar, K. (2015). Inferring to a model: Using inquiry-based argumentation to challenge young children's expectations of equally likely outcomes. In A. Zieffler & E. Fry (Eds.), *Reasoning about uncertainty: Learning and teaching informal inferential reasoning* (pp. 1–27). Minnesota: Catalyst Press.

Hestenes, D. (2010). Modeling theory for math and science education. In R. Lesh, P. Galbraith & C. Haines (Eds.), *Modeling students' mathematical modeling competencies* (pp. 13–41). New York: Springer.

Kawakami, T. (2017). Combining models related to data distribution through productive experimentation. In G.A. Stillman et al. (Eds.), *Mathematical modelling and applications: Crossing and researching boundaries in mathematics education* (pp. 95–106). Cham: Springer.

Konold, C., Robinson, A., Khalil, K., Pollatsek, A., et al. (2002). Students' use of modal clumps to summarize data. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*. Voorburg: IASE.

Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics*, *88*(3), 305–325.

Lehrer, R., Schauble, L., Petrosino, A. J. (2001). Reconsidering the role of experiment in science education. In K. Crowley et al. (Eds.), *Designing for science: Implications from everyday, classroom, and professional settings* (pp. 251–278). Mahwah, NJ: Lawrence Erlbaum Associates.

Lehrer, R., & Lesh, R. (2003). Mathematical learning. In W. Reynolds & G. Miller (Eds.), *Comprehensive handbook of psychology* (Vol. 7, pp. 357–390). New York: John Wiley.

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal*, *41*(3), 635–679.

Makar, K. (2016). Developing young children's emergent inferential practices in statistics. *Mathematical Thinking and Learning*, *16*(1), 1–24.