Educ Stud Math (2018) 97:163–183
https://doi.org/10.1007/s10649-017-9791-2

The role of perceptual similarity, context, and situation when selecting attributes: considerations made by 5–6-year-olds in data modeling environments

Aisling Leavy[1] · **Mairead Hourigan**[1]

Published online: 18 November 2017

Abstract Classroom data modeling involves posing questions, identifying attributes of phenomena, measuring and structuring these attributes, and then composing, revising, and communicating the outcomes. Selecting attributes is a fundamental component of data modeling, and the considerations made when selecting attributes is the focus of this paper. A teaching experiment involving 2 teacher educators and 25 pre-service teachers (PSTs) was carried out with 24 young children (5–6-year-olds) as part of a 4-day data modeling investigation. Although perceptual features of the data influenced initial approaches to attribute selection, considerations of the problem situation influenced a shift from the perceptual and towards consideration of attributes such as taxonomy, habitat, behavior, and diet. Expertise in the data context (animal kingdom) and ability to collaborate and negotiate within groups supported children in their ability to switch attributes, attend to multiple situations presented by the problem, and modify and extend their categorizations of data.

Keywords Data modeling · Attribute selection · Statistical inquiry · Young children · Teaching mathematics · Statistics · Elementary education

1 Introduction

Advancement of modern technologies has resulted in children gaining access to data at younger ages. Consequently, there arises the need to support the development of young children's statistical reasoning and thinking. Indeed, recent research exploring data modeling

✉ Aisling Leavy
aisling.leavy@mic.ul.ie

Mairead Hourigan
mairead.hourigan@mic.ul.ie

[1] Department of Language, Literacy and Mathematics Education, Mary Immaculate College, University of Limerick, South Circular Road, Limerick, Ireland

acknowledges the nascent potential of young children to engage in the types of reasoning and thinking that provide the foundations for later statistical thinking, reasoning, and literacy. However, the paucity of such research in the early years means we are only beginning to understand young children's statistical capacity. The authors sought to build on previous research, examining young children's strategies when engaged in a data modeling environment. In particular, we explore the considerations young learners make when selecting attributes to categorize objects in response to a problem (what is commonly referred to as *problem context*) and explore the influence of the situations arising from the problem context (Brousseau, 1997) on their selection of attributes and ultimately on the outcomes of these decisions.

2 Background to this study

2.1 Data modeling

Historically, models and modeling have a variety of meanings in relation to mathematics. Mathematical modeling is usually carried out in an effort to describe and represent our beliefs about how aspects of the world or a system function. Thus, they are representations or descriptions of reality that move beyond the real-life situation or external world and examine its structural features through mathematics. Mathematical modeling serves several functions. Models serve to develop scientific understanding by producing descriptions of behaviors and results, explanations for why particular results occurred, and predictions for future behaviors or results. Several different categorizations of mathematical models exist. For example, determinist and stochastic models differ in terms of the outcome predicted whereas empirical and deterministic models differ in terms of the degree to which they take into account the mechanisms through which change occurs within the system being modeled.

Within statistics, data modeling is what statisticians and others do when they reason statistically and try to make sense of real situations (Wild & Pfannkuch, 1999). Unlike traditional school statistics where the focus is frequently on teaching routine procedures and applying these procedures to a set of decontextualized data, data modeling is a more fluid and open process where definition of the data and its attributes are in flux and under negotiation. Lehrer and Schauble (2007) describe data modeling as

> deciding which aspects of the world are relevant to the conceptual model, how best to measure them, how to structure and represent the resulting measures, and then how to make inferences that are situated within and informed by knowledge about the qualities of those measures and representations. (p. 150)

By their very nature of being developmentally appropriate, classroom data modeling experiences do not exactly mirror data modeling demands in the professional world, although attempts by curriculum developers can enhance the likelihood that the demands mimic what ordinarily would be encountered. Thus, classroom data modeling involves children directly in posing questions, identifying attributes of phenomena, measuring and structuring these attributes, and then composing, revising, and communicating the outcomes. The problem context is a key factor to consider in the design of successful learning environments for young children (Kinnear, 2013; Leavy & Hourigan, 2016b); in particular, data modeling environments require the presentation of a rich problem context that has embedded within it the stimuli for the desired statistics (Hourigan & Leavy, 2016; Kinnear, 2013; Leavy & Hourigan, 2016a). Such experiences of modeling are described by English (2010) as “vehicles for students to construct significant mathematical ideas and processes rather than simply apply previously taught procedures” (p. 26).

Most research on data modeling has focused on secondary and tertiary level learners; however, some recent research has focused on young children and identified a number of critical elements of data modeling environments. Components identified as important include the use of driving questions (problem contexts) which lead to investigations of relevant and meaningful phenomena (Kinnear, 2013; Leavy, 2015; Leavy & Hourigan, 2015; Lesh & Doerr, 2003) and in turn the development of interest and ownership of ideas and artifacts through small group work and discussions (DiSessa, Hammer, Sherin, & Kolpakowski, 1991). The provision of opportunities to decide what is “worthy of attention” (Hanner, James, & Rohlfing, p.100) through engaging with rich complex data (English, 2010, 2012), classifying and structuring data (Lehrer & Schauble, 2007), representing and displaying data (Hanner, James, Rohlfing, 2002; Lehrer & Schauble, 2007) in addition to providing opportunities to invent and revise their own models (Lehrer & Schauble, 2000) have all been identified as important components of the data modeling process. The role of technology in supporting the representation and analysis of data in modeling environments has also been documented with younger children (Cobb, McClain, & Gravemeijer, 2003; Hancock, Kaput, & Goldsmith, 1992; Lehrer, Kim, & Schauble, 2007; Paparistodemou & Meletiou-Mavrotheris, 2008).

Role of culture Any discussion of cognitive processes that underpin data modeling, such as classification and concept formation, needs to acknowledge the critical role that culture plays in shaping cognitive structure. Recognition of the role of culture came about as a result of the seminal work of Luria and Vygotsky, most notably Luria’s 1931 expedition to Uzbekistan. The purpose of the expedition was to investigate the influence of culture (i.e., changing economic, educational, and social conditions) on the development of higher cognitive functions. Although the conclusions of these studies, which provide earliest insights into the role of culture as a determinant of cognitive processes, are controversial, they have been supported by contemporary research on a diversity of cultural groups around the world. An overview of these studies can be found in Fletcher-Janzen, Strickland, and Reynolds (2000) and Uzzell, Ponton, and Ardila (2013).

Generating and selecting attributes Initial experiences with data modeling include the “creation, analysis, and revision of data classification models. A fundamental element in creating these models is selecting attributes and classifying items according to these attributes” (English, 2012, p. 16). This ability to focus on attributes of data demands that children attend to the qualities of items rather than the items themselves. Thus, rather than identifying a data set as consisting of four apples and two strawberries and three pears, a child might refer to the data as fruit that grow on trees and fruit that do not grow on trees. Knowledge of the phenomenon under investigation guides the selection of attributes and supports the types of conversation and negotiation that take place within a group when selecting attributes.

Selecting attributes is not a trivial task. It is described by Lehrer and Schauble (2007) as, “seeing things in a particular way, as a collection of qualities, rather than intact objects” (p. 154). Studies have shown that although there is the tendency for young children to draw on the perceptual similarity between objects when selecting attributes to categorize objects (Gentner & Namy, 1999; Samuelson & Smith, 2005), there is also evidence of their ability to focus on more taxonomically relevant features (such as function) when selecting attributes to focus upon (Blair & Somerville, 2009; Gelman & Markman, 1986). A three-year longitudinal study of 6-year-olds engaging in data modeling activities provides unique insights into the ways children generate attributes (English, 2010, 2012; Kinnear, 2013). The overarching context of investigation in these studies was recycling, and this context informed the selection of attributes. Children first selected attributes focusing on perceptual aspects such as material (plastic, paper, and cardboard) and shape (circle, rectangle, and square). However, when asked to consider whether their attributes should be changed, over half the groups changed their attributes and focused on a wider range of less-visible attributes such as heavy/light and hard/not hard; this demonstrated their ability to "switch their attention from one item feature to another" (English, 2012, p. 27). This ability to focus on less perceptible features of objects when identifying attributes is also reported by Gelman (2006) and Kinnear (2013). Such studies which report on the diminished role of perceptual features in informing attribute generation are promising and depart from the predominant belief that children rely on resemblance and perceptual similarity of objects in their construction and use of attributes.

Furthermore, young children demonstrate a keen awareness of the need to define attributes and distinguish between categories (Kinnear, 2013) in addition to the ability to modify and expand the list of attributes being used. Children also demonstrate the ability to discern objects that display more than one attribute and to suggest strategies to deal with these objects. For example, one group in a study by English (2010) engaged in discussion around why a plastic bag displayed the attributes of both "reuse" and "throwaway" culminating in one student suggesting that half the bags be placed in each category.

Despite the abilities demonstrated by young children in some studies, other research suggests that this ability to switch attention to a selected attribute and avoid or eliminate attention to other attributes that may be visually compelling is not trivial for children. Furthermore, the task of deciding what is "worthy of attention" (Hanner et al., 2002, p. 100) has been shown to be challenging for some children. In their study of elementary students' data classification models, Hanner et al. (2002) found that first and second graders tended not to communicate these attributes and assumed that whatever they did to construct categories would be apparent to everyone; they sorted objects as if the task and selected attributes were self-evident. Problems with describing attributes were also noticed in situations where children used imprecise definitions or selected descriptors that could not be evaluated by examination of the objects. Children also demonstrated the tendency in the initial stages of the study to jump from one attribute to another as they moved through the objects indicating that they were regarding the objects, "as instances rather than collections of attributes" (p. 102). Lehrer and Schauble (2000) also revealed initial difficulties in children working together as a group and little or no efforts to reconcile differences of opinion.

There is some research which points to conditions that may support the engagement of young learners when selecting attributes. DiSessa et al. (1991) identified ownership and competence as important factors that support engagement. Ownership of "ideas and artifacts" was considered critical to maintaining interest. Furthermore, DiSessa et al. argue that if learners are competent in the subject matter (what is commonly referred to as *data context*), then it is far easier to engage in discussion around that subject matter. Studies of young children have shown the role of *expertise* in informing attribute selection was influential in group deliberations, where those with expertise or extensive knowledge in a domain differ from less expert children in terms of their ability to structure or organize knowledge in that particular domain. Similarly, other studies have shown that when selecting attributes and making similarity judgments, pre-school children who had expertise in dinosaurs (Chi & Koeske, 1983) and shorebirds (Johnson & Mervis, 1994), when compared to their non-expert peers, were more likely to draw on functional features (such as defense mechanisms and locomotion) as opposed to perceptual or morphological features. These studies provide compelling evidence of the important role played by knowledge of the subject matter (data context) in informing children's decisions.

In summary, whereas research with young children working in data modeling environments provides evidence of their natural propensity for attending to and categorizing attributes, there is also evidence of difficulties experienced by children. There remains a lack of consensus around the role played by perceptual features in attribute selection and in the ability of children to switch attention and focus on different attributes. Furthermore, we are only beginning to understand the conditions under which attribute selection is supported and the role played by problem context (i.e., the nature of the task itself) and the data context (i.e., expertise of the learner) in supporting children when attending to and selecting attributes. The development of such insights is critical in informing the design of tasks and experiences that support young children in attending to critical aspects of data and in developing their statistical thinking, reasoning, and literacy.

3 This study

This study explored the considerations young children make when selecting attributes to categorize data while participating in a 4-day data modeling investigation. The investigation was set within the context of designing a zoo. The data (3D models of animals) presented differ from traditional sorting tasks found in early years mathematics. Common early mathematical activities found in national curricula and textbooks require children to sort objects that possess one, or at most, two, explicit defining features, e.g., materials such as compare bears can be classified by either size (small, medium, large) or color (red, blue, green). In contrast, data in this study were multi-attribute and messier. By messier we mean that the objects (animals) do not neatly fall into categories. They may have attributes in common with more than one classification group, for example, sheep, cows, and lions may be considered land animals. However, lions are also wild animals just like snakes and gorillas; whereas sheep and cows are generally domestic farm animals. These same animals fall into different groupings again if we consider color, diet, speed, and a variety of other attributes.

In contrast to recent studies with young children in data modeling situations (English, 2010, 2012; Kinnear, 2013), children were not presented with predetermined categories—they were required to determine the categories by classifying objects based on attributes they selected themselves. By not pre-determining the categories, the influence of perceptual features of the objects in guiding the attribute selection was considered likely. When referring to perceptual features, we refer to measurable aspects of an object as determinable by the visual system (e.g., color, size, orientation); these aspects varied among the objects presented to children.

We were interested in the degree to which children relied on the perceptual features and whether they could move beyond the perceptual when selecting attributes. Our research questions were the following:

1. What is the role of perceptual features when selecting attributes?

2. What considerations did children take into account when responding to the problem context?

4 Research approach

The three-tiered teaching experiment (Lesh & Kelly, 2000) design facilitated investigation of the developing knowledge and understandings of the researchers, pre-service teachers (PSTs), and the children they taught. Three-tiered teaching experiments are a form of research design involving teams of researchers and teachers working together in the natural setting of a classroom. Lesh and Kelly (2000, p. 197) explain that

> Tier 1 of such projects may be aimed at investigating the nature of students' developing knowledge and abilities: Tier 2 may focus on teachers' developing assumptions about the nature of students' mathematical knowledge and abilities: and, Tier 3 may concentrate on researchers' developing conceptions about the nature of students' and teachers' developing knowledge and abilities'.

This paper focuses on just one of these tiers—the children's developing knowledge and abilities.

4.1 Participants

Participants were an intact multi-grade class of 24 primary school children. They were aged five to six and were in their second and third year of education in a local primary school. The school system in Ireland consists of 8 years of primary education and 5 years of secondary education. While the compulsory school starting age is six, this represents the maximum age at which children start school. Many children attend school younger than this age with the majority of 5-year-olds and half of 4-year-olds enrolled in primary schools in Ireland (Darmody & Smyth, 2012).

The profile of the children participating in this study was diverse in terms of gender and socio-economic and ethnic backgrounds. The participants were following the revised primary mathematics curriculum (NCCA, 1999), where one of the strands of study is "Data." However, on request from the researchers, the classroom teacher had not yet addressed this strand prior to the study.

4.2 Method

The researchers, who are teacher educators, worked with 25 pre-service teachers on the design and implementation of a unit of instruction on statistics, which engaged children in a cycle of statistical inquiry modeled broadly on the PPDAC cycle (Wild & Pfannkuch, 1999). Instruction consisted of a sequence of four lessons (see Table 1) taught on consecutive days to the same class of 5–6-year-old children; each lesson lasted approximately 60 min. PSTs were assigned to one of four groups, each of which focused on one of the four lessons. While all PSTs attended all planning and debriefing sessions, they only attended the lesson to which they were assigned. The study described here focuses on the insights we gained during the second lesson, which focused on sorting data based on the identification of attributes. A total of 7 PSTs worked on the design and implementation of this lesson.

Table 1 Lesson design groups

Group	Lesson order	Focus
1	1	Data generation and collection
2	2	Identifying attributes
3	3	Structuring and representing data
4	4	Informal inference

During the *research and preparation* stage, groups researched the literature relating to the foci of the four lessons. Thorough lesson and task design were promoted through the use of Ertle, Chokshi, and Fernandez's (2001) lesson note format. This format supported a focus on the task content, teacher questions, expected student reactions, and teacher responses to those reactions. The *teaching* stage involved each lesson being taught in the local primary school. During each lesson, one PST took the role of "class teacher" (referred to as class-PST) and coordinated the lesson. All PSTs within the respective groups (referred to as group-PST) worked with a mixed ability group of 3–4 children and facilitated interactions and conversations within the group. The role of the class-PST and the group-PST was primarily to pose questions in an effort to reveal children's reasoning and justification for the selection of attributes. Hence, the child was foregrounded—it was the child who made suggestions and proposed ideas and approaches and the class-PST and group-PST who facilitated the children in meeting their goals.

The problem context In conceptualizing the problem, we draw on Brousseau's (1997) notion of situation:

> The modern conception of teaching ... requires the teacher to provoke the expected adaptation in her students by a judicious choice of problems that she puts before them. These problems, chosen in such a way that the students can accept them, must make the students act, speak, think, and evolve by their own motivation. (p. 30)

It was this "judicious" choice of problem that required considerations relating to both students and their learning, what Brousseau calls the "fundamental situation" (p. 24) and resulted in design of the task. In our purpose-made video, Zach the Zookeeper (an actor) posed the following problem to children:

> *I am designing a new zoo. Unfortunately, there aren't enough enclosures in my zoo to keep all the animals separate and on their own, so we need to group them. So I am going to ask you all to look at the animals and help me to sort them. So remember, work together with your friends and talk about how we can sort the animals into different groups.*

The problem can be considered to consist of two different situations: The zoo situation and the sort situation. The *zoo situation*, presented by Zach at the beginning of the problem, requests children "to regroup some animals in the same enclosure" and thus refers to the dilemma of the limited number of enclosures. The *sort situation*, presented at the end of the problem, requires children to "look at the animals and help me to sort them" according to attributes the children select themselves.

Certain considerations were made by the researchers in relation to the problem. We wanted to explore the attributes that were salient to children when engaging in an open sort; thus, we did not want to place any parameters on the number of enclosures that Zach could construct. Similarly, we did not want physically to represent the enclosures, with string or a structure of some sort, due to the possibility that this may inform the sorting situation and "override" the attributes that were inherent to the animals. For example, the length of string and thus the area of the enclosures may have been used by children to determine the number or size of animals to place within an enclosure and thus detract from the focus on salient attributes.

Children were divided into six groups. Each group was presented with a selection of miniature plastic 3D models of animals. These were used as physical models to support children in referring to the real-world animal in order to select attributes to classify them into groups. Care was taken in determining the selections of animals to ensure animals varied in distinctive perceptual features (for example, color and pattern, presence of wings, beaks, fins, tails, feathers, scales, hooves, etc.), behaviors, habitats, and other classifications (such as diet). Only one of each type of animal was presented. In an effort to develop as robust insights as possible into children's attribute selection, two different groupings of animals were presented. Selection A contained an elephant, dog, snake, sheep, seal, cow, lion, horse, and pelican. Selection B consisted of a duck, horse, pig, zebra, cow, penguin, giraffe, polar bear, cheetah, hippopotamus, and chimpanzee. Three groups of children received the first selection, and the remaining three groups received the second selection. The class-PST encouraged children to think about the problem, share their ideas, and group the animals according to the selected attributes.

Role of the pre-service teachers The work of each group of children was facilitated by a group-PST. In preparation for the study, all PSTs worked with the researchers to identify animals to be included in the two selections. They made predictions of possible classifications that children might produce and generated and recorded categorizations of groupings of animals until they believed they had exhausted all possibilities. Thus, PSTs had extensive knowledge of the data set and of expected categorizations children might construct. They also had clear guidelines regarding their role and responsibilities as facilitators. As the children were very young, a primary role of the group-PST was initially to develop rapport with the group, review the group's understanding of the problem, and ensure the group stay focused on solving the problem. In maintaining focus, group-PSTs had responsibility for guiding rather than directing. Thus, each had a list of guiding questions and prompts that were suitable for use with their group; they were encouraged to adhere to these. While guiding the group through the task, they asked questions to seek clarification around the both the zoo and sort situations (What is Zach's problem? What have we to do?) and around the sorting decisions made by children (What animals would be in that group? Why do you think that? Why is that animal in that group?). The prompts also provided opportunities for all children to participate (Do you have an idea? Does everyone agree?) and maintain focus on the sorting (Show me what animal would be in that group). Group-PSTs were aware of the importance of not leading the children in the selection of attributes. They were encouraged to use strategies such as re-voicing in an effort to clarify a child's idea so that other children could engage with it (You don't think the duck belongs there?). Thus, the regular interventions of group-PSTs within the groups were critical in maintaining the attention of the children to the problem, in revealing childrens' thinking and in guiding focus towards both the zoo situation and the sort situation.

4.3 Data collection and analysis

PSTs were participant observers focusing on children's responses. Dialog within each of the six groups was captured digitally, downloaded, and transcribed. A video crew recorded the lesson and focused on one group of children during the cooperative group work involving attribute selection and the ensuing sorting. Subsequently, group-PSTs constructed analyses and reflections of the attribute selections within their own group informed by the notes, observations, and transcripts of the group conversation.

One researcher made the first pass through the transcriptions with a view to identifying the different attributes selected across the cohort of children (research question 1). A list of 32 distinctly different attribute-groupings was generated (see Table 2). Both researchers then re-examined the transcripts and attribute-groupings for accuracy of coding and to ascertain the general features of these attribute-groupings (for example, ascertaining how many were derived from perceptual features of the animals).

The second pass through the data focused primarily on the activity within each of the groups. The analysis focused on locating incidences where attributes were switched during a sort and occasions where disagreements arose within groups relating to attribute selection and subsequent sorting. Data were also examined with a view to examining the influence of the problem context (in particular the zoo situation and sort situation) and data context (knowledge of animal kingdom) on group's approaches to solving the problem (research question 2).

Efforts were made to ensure validity of the process through the triangulation of data arising from the transcriptions of work within each group, from pre-service teacher reflections, research reflections, and analysis of the video. Agreement between the researchers was reached regarding the list of attributes and the main categories (see Table 2).

Table 2 Attributes generated to categorize animals

Perceptual features	*Temperament*
Black+white/not black and white	Scary/not scary/kind of scary
Black+white/yellow/brown	Might eat me/will not eat me
Gray/green/brown/white	Not dangerous/dangerous/can be dangerous
Black and white/brown/spotty	Good/bad/sometimes good and sometimes bad
Big/small	Tough/kind/in the middle
Fat/kind of thin	Kind/helpful/greedy/dangerous
Big/medium/small	Animals that get along together/do not
Wide/not wide	Kind/good and mean/mean
Large/medium/small	
Tails/no tails	
Rough skin/soft skin	
Feathers/no feathers	
Feathers/soft/smooth/furry/wool/scales	
Geography and habitat	*Diet and Behavior*
African/Ireland/South Pole	Fast/slow
Wild animals/farm animals	Vegetarians/carnivores
Swim in the sea/do not swim in the sea	Eat meat/eat plants
Like water/do not like water	Eats grass/does not eat grass
	Grass eaters/meat eaters/fish eaters
Other	
soft soft+fast/soft+slow/rough+fast/rough+slow	
interesting/not interesting	

5 Results

5.1 Selecting attributes

Across the groups, animals were sorted in 32 different ways based on the attributes selected. Table 2 presents a list of the attributes children used to sort the animals. The attributes selected generally fell into four main categories: *perceptual features*, *geography/habitat*, *temperament*, *and diet and behavior*. Our analysis also revealed that these grouping arrangements for the most part took account of the sort situation (rather than the zoo situation).

A proportion (25%) of sorts focused on an attribute and its complement (*x/not x*) whereas the remainder assigned all animals to discretely named categories (*x/y/z*). These different approaches are apparent when focusing on those groups who initially selected color as an attribute. Some groups who focused on *x/not x* categorized animals according to whether they were black and white (zebra, penguin) or not black and white (remainder of animals). In contrast, another group categorized their animals based on whether they were brown (chimpanzee, horse), black and white (zebra, penguin, cow), or spotty (cheetah, giraffe, pig, duck).

The role of perceptual features is evident in the number of attributes that focused on *appearance* (see Table 2). The color, size, and texture of animals were the predominant attributes selected. Although perceptual features of animals accounted for 13 of the 32 categorizations, only one group of children (group 6) limited their sorts primarily to perceptual features; this group focused on perceptual features for six of their seven sorts. Even though the perceptual features of animals were observable, their determination was not uncomplicated and was at times contentious. For example, where criteria were relative (such as in the case of size) children often had to provide a definition or parameters upon which to guide their decision-making. For example, in group 6, a contentious perceptual feature was pattern, in particular what constituted a "dotted" animal. Whereas cheetahs neatly fit into the dotted category, the category name was changed to "patches" to allow the inclusion of the pig into the category. This close attention to perceptual descriptors was also evident when assigning animals to different types of skin texture. In another group, the original categories of soft and rough were soon extended, after debate and consideration of the group features, to include feathers, soft/smooth, fur, wool, and scales.

The remaining categories depended on the selection of attributes that were not visible to children and drew on their knowledge of the animal kingdom—these were *geography/habitat*, *temperament*, and *diet and behavior*. These categories accounted for 17 of the 32 categorizations. These attributes drew on the children's knowledge and observation of animals that they had gathered from sources such as books, TV documentaries, zoos, and farms.

Two of the selected attributes did not fit into the four categories and consisted of a combination of attributes (soft+fast/soft+slow/rough+fast/rough+slow) or involved attributes that could not be evaluated by examination of the objects or by knowledge of the animal behavior of habitat (interesting/not interesting). The selection of the attribute "interesting" resulted in some disagreement between children when determining whether particular animals were interesting or not.

[Elephant, dog, snake, sheep, seal, cow, lion, horse, pelican] *group* 2

Aidan	We're doing interesting and not interesting. The elephant is interesting, the seal is *not* interesting. A dog is interesting…it cares for you.
Group-PST	Okay, do you think interesting and not interesting is a good way to group them? Will everyone agree with you?
Aidan	A snake is not interesting.

Group-PST	Why not, Aidan?
Aidan	A snake is really small and boring.
Olivia	Can I put this one here? A cow is interesting. And a sheep is.
Aidan	No no, Olivia.
Olivia	Yes it is Aidan, I want to put him there.
	(Children argue among themselves .. Olivia finds it interesting and Aidan does not)
Aidan	Okay then, it [pointing to the sheep] is interesting.
Group-PST	Why do you think that?
Aidan	Because it gives you wool and stuff. And you can make clothes out of wool. And that makes it helpful.
Group-PST	Do you think the sheep is interesting as well, Olivia, or do you think it's not interesting?
Olivia	Yes because you can make clothes. Yeah.
Group-PST	Okay, what about these here, what about the dog? What is not interesting about the dog? [pointing to the dog who is in the 'not interesting' category]
Olivia	Um I don't know. I didn't put him in there. Aidan put the dog there.
Group-PST	Aidan, what isn't interesting about a dog? Ye need to agree on these, working as a team, remember?
Aidan	Well then a dog is interesting.

Analysis of the initial transcripts reveals a preoccupation with the sort situation and little consideration of the zoo situation. This is the only example where the sorting attribute was open to interpretation rather than being defined by perceptual appearance, habitat, or some other measurable characteristic. It is also interesting to note the absence of egocentric reasoning that may be associated with 5–6-year-old children; what we see instead is children engage in data-based reasoning and justification through negotiating the categorization of animals based on their experiences and knowledge of animals.

5.2 Consideration of the "sort situation," i.e., classifying data (animals) according to attributes

Following identification of attributes, children classified animals into groups. Approximately 50% of the sorts resulted in two subgroups (rough/soft skinned animals), one third resulted in three subgroups (big/medium/small animals) and the remaining had more than three groups (gray/green/brown/white animals).

The classification was not always straightforward and required discussion and negotiation. Across the groups, children demonstrated the ability to listen to each other, consider and assimilate ideas and feedback, and generate shared categorizations that had been socially mediated and negotiated. While on some occasions, group members negotiated solutions, at times, these solutions were reached by one child being convinced by the argument of another and conceding their original position. On other occasions, however, this was not possible. Three distinct scenarios occurred that resulted in complex negotiation between group members. Together, these instances demonstrate that young children have the capacity to build upon each other's ideas resulting in shared ownership of the emerging categorizations.

Scenario 1: when a data value possesses more than one of the selected attributes In every group, children discussed whether animals possessed the selected attributes. In the following transcript, we see Darren (group 1) suggest the attributes "good/bad" as a way to categorize the animals. The other children readily accept these attributes and continue to negotiate the positioning of animals that do not fall readily within the categorization as "sometimes good and sometimes bad." Four of the children comment on the positioning of the horse and provide examples of horse behavior that justify it straddling the "good/bad" dichotomy.

[Elephant, dog, snake, sheep, seal, cow, lion, horse, pelican] *group* 1

Darren	The seal and the pelican over here and then these are all the bad ones. And these are the good ones (the seal and pelican) and the lion goes here (in with the bad ones).
Group-PST	And what about the horse where will that go?
Darren	The middle
Mia	Sometimes it's bad and sometimes it's good
Group-PST	And what about the pelican?
Otille	The bad ones
Mia	What about the middle because sometimes it might be ok?
Group-PST	What else would you put in the middle then Mia?
Mia	The horse
Melios	Yeah with the pelican. Sometimes it [the horse] will knock you over and make you fall on the ground.
Ottlie	The middle. Sometimes they get a bit scary and they kick you.

Here, we see the construction of a category that recognizes that animals might possess both attributes "good/bad"; this facilitated all animals in the collection to be classified. This strategy also occurred in two other groups; one group generated the category "can be dangerous" when sorting animals using the attributes "dangerous/not dangerous" and similarly another group made the category "in the middle" to place animals that did not readily classify as "tough" or "kind." For all of these cases, the behavior of the animal varied and matched both original selected attributes.

Scenario 2: when the selected attribute is ambiguous At other times, however, the attribute itself was ambiguous and this resulted in uncertainty around classifying particular animals. In group 5, the attributes "large/medium/small" were selected. The group had been struggling to decide how big an animal had to be in order to classify it as "large" when one member Polina made the suggestion that in order for an animal to be assigned to the "large" category, it had to be "bigger than children." The development of this working definition of the attribute, that served to assign parameters to the attribute, facilitated easier classification of the animals.

Examination of the conversation within group 6 around classification of the duck provides insights into the ability of young children to negotiate relatively complex categorizations. In the transcript, we can see that sorting using the attributes fast/slow was straightforward enough until consideration of the duck. As children could not come to agreement on the duck and his behavior, they invoked another rule and classified him with a similar animal—the penguin, as they were both birds. This strategy indicates an ability to shift focus and apply new criteria on which to base a judgment.

[Duck, horse, pig, zebra, cow, penguin, giraffe, polar bear, cheetah, hippopotamus, chimpanzee] *group* 6

Group-PST	Ok, does anybody else have any other way of sorting the animals?
Leah	I think you could put the fast ones with the fast ones and the slow ones with the slow ones.
Sian:	Cheetahs are very fast.
Tomi:	Cheetah is the fastest animal in the world.
Group-PST:	Ok, so let's decide what is fast and what is slow.
Children:	Cheetah cheetah cheetah.
Sheena:	Are these fast? (hippo)
Sian:	Slow.
Caithyln:	Slow.
Hannah:	That's (penguin) so slow
Sian:	That's walking slowly to the slow pile ("walks" the penguin to the slow pile)
Tomi:	Slow. (placing cow with slow animals)

Group-PST:	So Leah, you have the fast animals do you?
Leah:	This one is fast. (moving bear from slow to fast)
Group-PST:	Does everybody agree?
Sheena:	A duck is not a fast animal.
Leah:	It flies fast.
Group-PST:	Ok, Leah and Sheena, you must decide where you are going to put it*.
Leah:	But it flies fast.
Sheena:	But it doesn't fly often.
Leah:	It's a bird
Group-PST	So if we said it was a bird, would you think it was fast?
Sheena:	No.
Sian:	Giraffes are fast (moves it to the fast group)
Group-PST:	They (giraffes) are fast yes? [children nod]. Ok, so we are still not happy about where the bird is?
Caithlyn:	He could go with the penguin. A penguin is a bird.
Sian:	(demonstrating the bird flying) He can fly fast.
Group-PST:	So do we agree that the duck will go with the penguin?
Children:	Yaaa.
Tomi:	Yes yes yes

*In the transcript, the group-PST strongly encourages the children to come to a decision. The emphasis on placing an animal in a category is considered beyond the communicated task of group facilitator as outlined during their preparation for the lesson

Scenario 3: when attributes are tenacious Some children demonstrated difficulty switching attention from one attribute to a new attribute. For these children, certain attributes were almost overpowering and caused distraction when sorting. A number of outcomes occurred in these situations. In some groups, the group-PST reminded the children of the selected attributes; this reminder provided adequate support to redirect their attention and support them when sorting. However, in group 2 while the children were considering the attributes "fast/slow," Paul was unable to switch attention and remained unhappy with the outcome of the sorting. He continuously referred to previously discussed attributes, rough (gray wrinkly skin of elephant and rhino) and soft (furry animals), while the remainder of the group were focusing on fast and slow animals. In response, the children themselves negotiated a solution by blending the attributes being used to classify animals resulting in four possible categories of animals: soft+fast, soft+slow, rough+fast, and rough+slow. The children's ability to blend the attributes is noteworthy given their young age and lack of data modeling experiences and may point to evolving understandings of multivariate data.

5.3 Consideration of the "zoo situation," i.e., grouping animals into the same enclosure

In many of the categorizations described earlier, we observed a focus on the sort situation with the relative neglect of the zoo situation. However, there were occasions when following placement of animals into groups according to a selected attribute (i.e., focus on the sort situation), the zoo situation (i.e., the criteria that animals categorized in the same group would be contained in the same enclosure) became a critical factor for children when considering the appropriateness of their categorizations. It is interesting to note that while some groups progressively paid more attention to the "zoo situation" over the course of the task (in later sorts), some groups attended to it from the start. In these cases, children became concerned that some animals would be placed in enclosures with animals that would eat or harm them. The coordination of attention to both "situations" resulted in the generation of a number of different approaches to address and account for these concerns.

Approach 1: breaching the categorization This strategy occurred in group 3, which were using the attribute "eats meat/eats plants" to categorize animals resulting in the penguin being assigned to the category "eats meat" because penguins eat fish. However, one child realized that, once categorized, the penguin would now be in the same group (and hence enclosure) as the polar bear (who also eats meat). The child decided to move the penguin to the "eats plants" group of animals (even though the penguin did not possess this attribute) to protect it from the polar bear.

[Duck, horse, pig, zebra, cow, penguin, giraffe, polar bear, cheetah, hippopotamus, chimpanzee] *group* 3

Séan	The ones that can eat people can be over here.
	[Children group the polar bear, penguin and cheetah]
Group-PST	Okay if these ones can eat people, what will these ones be?
Séan	These ones might be vegetarians.
Group-PST	So they don't eat people?
Séan	Yeah.
Group-PST	Why is the penguin in the ones that can eat people?
Séan	Because it can eat fish instead.
Group-PST	Are fish people?
Séan	Fish are meat.
Group-PST	Yeah. But can we leave the penguin in this group? Will the polar bear get him or will he not?
Peig	I think I'll move it (the penguin) so it can't get eaten.
	[moves penguin out of the "eats meat" category]
Group-PST	We have a chimpanzee, does that eat meat?
Séan	Well it kills people. I saw that in a movie.
Group-PST	Okay. A hippo?
Deirdre	Sometimes he can.
Séan	[moves the hippo to the "eats meat" group]
Group-PST	A bear?
Peig	A polar bear it can eat a seal.
	[moves the polar bear to the "eats meat" group]
Group-PST	Okay name the ones now that eat meat - tell me what's in that group?
Fiona	A chimpanzee, a hippo, a bear and a cheetah.
Group-PST	Okay so and what about these ones here what are they [pointing to the other collection of animals]?
Séan	Herbivores.

Therefore, consideration of the "zoo situation," and its possible detrimental outcome for the penguin, lead children to breach the parameters of the sort situation and categorize the penguin as an herbivore. Thus, we see the zoo situation supplant the sort situation. This "supplanting" of the sort situation is revealed also in the activity of group 1. Examination of their transcript reveals that although initial discussions attend only to the sort situation, disagreement around categorization of the dog as a meat or plant eater is resolved through consideration of the zoo context.

[Elephant, dog, snake, sheep, seal, cow, lion, horse, pelican] *group* 1

Group-PST	Can we think of another way to sort our animals?
Melios	I know meat eaters and non-meat eaters.
Mia	The seal eats fish not meat… [putting it in non-meat eater side]
Melios	Elephants eat leaves not meat don't they?
Group-PST	Yeah I think they do. What about the rhino do they eat meat or not?
Melios	Meat eaters I think. Snakes are meat eaters they eat rats and mice and things

Group-PST	Do they? What about our sheep where is he going to go?
Ottlie	Grass
Group-PST	And Darren where's the horse going to go?
Darren	Over here [meat]
Mia	No, the grass side. Horses like grass.
Group-PST	What else might go in our grass group then?
Darren	The cow
Melios	The rhino
Ottlie	The elephant
Melios	The elephant eat leaves not grass
Mia	The dog
Children	No [moving dog back from grass group to meat eaters group]
Darren	They eat ham
	[Mia moves dog back to grass group]
Darren	Mia stop!
	[Darren moves the dog back to meat eaters group]
Group-PST	And what about the seal?
Darren	These eat fish (pointing to the pelican too)
Ottlie	Yeah these both eat fish… maybe the middle.. the ones that eat fish in the middle
Group-PST	Maybe the middle? Mia what about the seal what do you think?
Mia	I think it's going to go in the middle
Ottlie	Because the seal eats fish like the pelican
Group-PST	Will we take a picture of that for Zack and see what he thinks?
Melios	Before you take the picture.. do lions eat dogs?
Group-PST	Oh wait now Melios has a very good point. He might have spotted a bit of a problem. He wonders if lions eat dogs…
	What do we think? Do lions eat dogs?
	[Mia moves dog back to grass group]
Ottlie	I think they do because they are very greedy and a bit dangerous too

Thus, we see the forced categorization of the dog as a plant eater (sort situation) due to the safety issues (zoo situation) arising from classifying him as a meat eater and thereby placing him in the same enclosure as the lion. The contribution from Melios suggests a further refinement of the "non-meat eaters" to grass and leaf eaters. This indicates attention to the zoo situation where an enclosure should ideally include grass for grass eaters and foliage for leaf eaters. Ottlie, building on a comment from Darren who refers to the seals as fish eaters, constructs a sub category to distinguish "fish eaters" from the general category of "meat eaters." This is readily accepted by the children in the group.

Approach 2: selecting new attributes In other groups, children disregarded their initial attribute selection when it resulted in groupings that posed a challenge for the zoo situation. In these situations, children were flexible in switching their attention between attributes. The strong influence of the zoo situation on children's judgment of the appropriateness of the categorizations was evident in the activity of group 4. In the transcript below, we can see that one child (Peter) demonstrated very specific and sophisticated knowledge of the "data context" (i.e., animal kingdom). He was acutely aware of the dangers associated with placing the cheetah with animals it might eat. We can see that Peter was not satisfied with the outcome of the first sort "big/small." Hence, he believed the only solution was to use different attributes on which to base the sort resulting in a categorization based on "animals that get along together" and "animals that don't get along together"; perhaps pointing to a nascent understanding of ecosystem. Hence, his consideration of the zoo situation influenced the selection of new sorting attributes and ultimately placement of the cheetah in an enclosure of its own.

[Duck, horse, pig, zebra, cow, penguin, giraffe, polar bear, cheetah, hippopotamus, chimpanzee] *group* 4

Mia	The giraffe, cow and zebra are in one group, I think we can divide them into big and small animals and the rest of these animals go in the small group
Peter	Wouldn't the cheetah eat the others? The Cheetah will try to eat them.
Group-PST	Peter do you think the way Mia organised them would work?
Peter	No because the cheetah would eat them all
Group-PST	Peter will you explain what you were thinking?
Peter	This one goes here because they won't hurt each other so they can go here. The hen and the polar bear will be ok together. Cheetah is on its own.
Group-PST	Ok Peter so you think that these groups of animals will all get along with each other?
Peter	Ya that's the way I grouped them. The ones that get along together and the ones that don't. [children group animals according to this arrangement]

Peter remained concerned throughout the task. When the group moved on to sort by country of origin, he expressed his concern that the penguin would be in danger of being eaten by the polar bear.[1]

Group-PST	Can anyone else see a different way of sorting the animals?
Kate	We could group them into countries. We could put horse, cow and duck for Ireland. Zebra is an African so we can put the gorilla here and the hippo here. The Cheetah is English so we put it over here [placing it with the "Ireland" group].
Anna	And then the Penguin and Polar Bear can go in the kind of cold area
Group-PST	So can you remind us what the groups are so we will not forget for Zach?
Kate	African, Ireland and the cold ones
Group-PST	And the Cheetah you placed that in the Irish one tell me about that?
Kate	He should be with the Africans I think, I'll move him
Anna	Ya because there are no cheetah's in Ireland
Group-PST	What do you all think of this way of grouping the animals?
Peter	I don't quite like it because you don't know if they are going to get along perfectly.
Mia	I think it is a good way
Anna	I like it because I know which countries they live in.
Group-PST	Peter if you went to this zoo and the animals were arranged like this, do you think you would like it?
Peter	No because if these animals lived together they might kill each other
Kate	No, polar bears don't eat penguins they like fish more than meat, so I think they would be alright together!

While in group 4 (above), Peter demonstrated an exceptional level of awareness from the beginning of the task—within his group, he was a lone voice of warning. However in other groups, there was evidence of more group members' awareness of and reaction to the "zoo situation." For example, in the transcript from group 6, we can see that the first two categorizations (bird/not bird, patterns/not patterns) were discarded when the emerging categorization resulted in placing an animal in danger of being eaten by another animal.

[Duck, horse, pig, zebra, cow, penguin, giraffe, polar bear, cheetah, hippopotamus, and chimpanzee] *group* 6

Group-PST	Does anybody have a way of sorting the animals because we really need to help Zach out today.
Caithlyn	Em, put the penguin with the duck.
Group-PST	Why would you put the penguin with the duck?
Caithlyn	Because they are both types of birds.
Tomi	But what if the penguin eats the duck?

[1] The children appeared unaware that polar bears are arctic animals and penguins reside in a number of regions (typically the southern hemisphere, sometimes Antarctic, and sometimes residing in tropical islands). They are not all cold weather birds.

Caithlyn	Yeah. We can't do that.
Tomi	We could put the ones with dots together and the one with no dots together.
Group-PST	So you pick out which ones you want to put together?
Caithlyn	[when the pig is chosen and placed in the dotted group] That is not a dot.
Group-PST	You don't think that that is a dot?
Caithlyn	No.
Sian and Leah	No.
Hannah	We could call them patches? Or patterns?
Group-PST	Would you like to sort the animals by patterns?
Children	Oh yeah.
Hannah	Like the zebra the cow...
Sian	No because the cheetah might eat the giraffe.
Group-PST	Why do you think that the leopard might eat the giraffe?
Sian	Because cheetah are wild.
Hannah	Yeah they love meat.
Sheena	I think that we could put the farm animals with the farm animals and the wild animals with the wild animals.

Again in this transcript, we see that the "zoo situation" takes precedence over the "sort situation." The initial attributes chosen (arising from attention to the sort situation) were reviewed and revised once consideration of the zoo situation revealed a problematic grouping of animals in relation to safety.

Thus, some groups demonstrated the ability to shift focus from the sort situation to the zoo situation and reconsider and make changes to the attributes used to categorize animals. Although the children did not for the most part prioritize the "zoo situation" they gave it increasing consideration over the course of the task. This pattern of behavior, demonstrated in many of the groups, provides evidence that perceptual features can be overridden as attributes if the resulting categorizations in some way contravene the initial task requirements. In other words, children demonstrated the ability to coordinate both the zoo and sort situations.

6 Discussion

During this data modeling investigation involving the design of a zoo, children were presented with a problem consisting of two different situations. They were required to group animals into enclosures (zoo situation) based on the identification of attributes of the animals/objects (sort situation) that would be used as criteria for sorting.

Analysis of the data revealed that children readily responded to the sort situation by identifying and communicating attributes to classify the data (animals). Although children were only 5–6 years old, they were aided by the selection of a data context (i.e., animal kingdom) for which they had some level of expertise. Thus, they were able to engage in focused and deliberate conversation on the subject of animals. Whereas some children drew on extensive knowledge of the animal kingdom to inform their attribute selection, others referred to experiences they had garnered from visits to zoos and farms (seeing dogs chase cows on a farm) and from watching TV (documentaries, movies). The attribute selection and subsequent classification suggest that children demonstrate the ability to identify common attributes shared among the data rather than merely identify the various attributes of individual data values. It is important not to underestimate this ability to view a collection of data "as an entity" (Hancock, Kaput, & Goldsmith, 1992), particularly because the data presented in this study were complex and multi-attribute—they were not clearly defined and did not possess readily discernible features that were shared among the group. The problem context required children to view the objects, each of them different, as a collection. Thus, children needed to move from a focus on discrete and separate animals to a collection that share qualities; these qualities were, in turn, used to classify the data. The identification of the shared qualities or attributes required that they disregard other attributes that may be more salient, a task that has been identified as difficult and "counterintuitive" for children (Lehrer & Schauble, 2007, p. 154). A small number of children in this study demonstrated difficulty disregarding attributes; however, the majority demonstrated relative ease in identifying and implementing relevant attributes as well as proficiency in communicating and justifying these decisions. These findings are in contrast to some of the difficulties experienced by students in the Hanner et al. (2002) study. Firstly, the issue of selecting an attribute that could not be evaluated by observing the data (e.g., interesting/not interesting) or knowledge of the data context (i.e., animal kingdom) was much less of an issue in this study when compared to the findings of Hanner et al. Furthermore, when selecting attributes, some children in our study (see Polina described in Scenario 2) demonstrated an ability to qualify, or provide precise descriptors, a capacity that was deemed problematic for similar aged children (Hanner et al.).

The findings also demonstrate that the young children in this study possess "pragmatic expertise" or the ability to listen to the arguments of others (DiSessa et al., 1991, p. 153). In some cases, the attribute suggested by one child was "taken-as-shared" (Cobb, Wood, Yackel, & McNeal, 1992) as the other members of the group accepted and implemented it without question. However, as the results suggest this was not always the case, and much negotiation was required during the modeling process. The remaining part of the discussion section is structured to respond to the initial research questions.

6.1 Research question 1: What is the role of perceptual features when selecting attributes?

We were interested in the role of perceptual features when considering attributes that were "worthy of attention" (Hanner et al., 2002). Perceptual resemblance in data modeling has been regarded as a bridge between form and function that helps move children "from literal similarity to analogical mapping of systems of relationships" (Lehrer & Schauble, 2007, p. 155). Children were closely attuned to perceptual features and perceptual resemblance accounted for 40% of the categorizations. While there was the initial tendency in some groups to focus on perceptual features, attribute selection shifted to more imperceptible features such as diet, behavior, and habitat. As evident from the findings, many initial categorizations based on perceptual features were abandoned when considerations of animal safety were involved. This suggests that the shift away from perceptual features may have been motivated by the zoo situation and guided by the expertise of children in relation to the data context (i.e., animal kingdom). This ability to change attributes was also a finding in the work in English's (2012) study which reported that over half of the 6-year-old children changed the attributes they used to categorize their data. English refers to this ability to switch their attention to different features of data as "lifting away from the plane of activity" (Lesh & Lehrer 2003, p. 377). Hence, this study concurs with other studies in the finding that young children can move beyond perceptual resemblance and consistently create and apply attributes based on unseen properties.

6.2 Research question 2: What considerations did children take into account when responding to the problem context?

A number of considerations came into play and performed critical, yet different roles, in switching attention to different attributes which subsequently informed the categorizations. Knowledge of the data context, in this case the animal kingdom, supported children in identifying attributes and categorizing objects (for example, knowledge of animal habitats was used to group African and Antarctic animals together). However, many initial models, which attended primarily to the "sort situation" aspect of the problem, were frequently considered inadequate and not fit-for-purpose in terms of solving the problem posed (problem context) and resulted in children revising, extending, and sometimes even discarding the models in favor of models that were perceived to address the problem better. For example, one group's initial categorization of animals into African and Antarctic was disbanded once they considered the outcomes for the animals involved—in this case, that the penguin would be in danger from being placed in the same enclosure as the polar bear. It appears that the "zoo situation" component of the problem (regarding how to choose animals to reside within a limited number of enclosures) was sufficiently compelling and meaningful to children that the models they chose were subject to constant revision and modification. The considerations of animal welfare, which frequently served as the basis for revising models, suggest that although knowledge of the data context (animal kingdom) was used to select attributes and in turn construct models, it was the "zoo situation" which served to assess the suitability and fit of the model. Interestingly, whereas the "zoo situation" served to challenge the models created, it was knowledge of the data context that further informed modification of these groupings to take account of the zoo situation. Therefore, the data context (knowledge of the animal kingdom) served a dual role in this regard. Consideration of the connection between the data context and the problem context was also evident in the study by Kinnear (2013) and Kinnear and Clarke (2016) where children took into consideration the needs of the character presented in the picture storybook when solving the problem presented.

7 Conclusions

In conclusion, there is a dearth of research examining data modeling in the early years. This study contributes to and extends previous research as it examines young children's modeling potential, in particular their approaches to attribute selection and implementation in a realistic problem context with which they are knowledgeable.

The research has a number of limitations. Firstly, this is a case study of one class of 5–6-year-olds; hence, the results of this study cannot be generalized to all children of this age. However, there is potential for further study to examine additional age groups in a variety of educational settings. The second limitation relates to the use of the model 3D animals. The data context was selected in light of children's relative expertise with the animal kingdom, and it was assumed that classifications were based on the real-world referents that were represented by the models. This assumption may not have held for all children, particularly those with limited experience of the animal kingdom. Thus, we acknowledge the findings of research that indicates a complex interplay between perceptual features (physical features of the objects themselves) and the knowledge that these features activate (Gelman, Chesnick, & Waxman, 2005; Petersen & McNeil, 2013). On a third note, although our data does not identify situations where group-PSTs assumed a more active role than that of a facilitator, it is possible that they may have influenced the attribute selection within groups due to non-verbal cues or otherwise. Finally, another issue is the relatively short time the children engaged with data modeling. The extension of the study over a longer duration would facilitate more thorough analysis and reap interesting findings where children's modeling could be tracked over a number of years. Another potential avenue of further study is to examine the influence of the data context (knowledge of subject) on young children's modeling ability—to compare differences in children's capacity in response to a variety of problem contexts.

Acknowledgements We gratefully acknowledge the work of all reviewers and editors of this manuscript who invested their time and energy in providing insightful and constructive feedback on earlier versions of this paper.

Funding The research was supported by a faculty seed-funding grant [SF16-101] from Mary Immaculate College.

References

Blair, M., & Somerville, S. C. (2009). The importance of differentiation in young children's acquisition of expertise. *Cognition*, *112*, 259–280.

Brousseau, G. (1997). *Theory of didactical situations in mathematics*. Dordrecht: Kluwer.

Chi, M. T. H., & Koeske, R. D. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology*, *19*, 29–39.

Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction*, *21*(1), 1–78.

Cobb, P., Wood, T., Yackel, E., & McNeal, B. (1992). Characteristics of classroom mathematics traditions: An interactional analysis. *American Educational Research Journal*, *29*, 573–604.

Darmody, M., & Smyth, E. (2012). Exploring school and classroom environments in Irish primary schools. *Children, Youth and Environments*, *22*(1), 178–197.

DiSessa, A., Hammer, D., Sherin, B., & Kolpakowski, T. (1991). Inventing graphing: Metarepresentational expertise in children. *Journal of Mathematical Behavior*, *10*(1), 117–160.

English, L. D. (2010). Young children's early modelling with data. *Mathematics Education Research Journal*, *22*(2), 24–47.

English, L. D. (2012). Data modelling with first-grade students. *Educational Studies in Mathematics*, *81*, 15–30.

Ertle, B., Chokshi, S., & Fernandez, C. (2001). Lesson planning tool. Retrieved November 28 2016, from www.tc.columbia.edu/centers/lessonstudy/doc/Lesson_Planning_Tool.pdf

Fletcher-Janzen, E., Strickland, T. L., & Reynolds, C. (2000). *Handbook of cross-cultural neuro-psychology*. New York: Springer Science & Business Media. https://doi.org/10.1007/978-1-4615- 4219-3

Gelman, S. A. (2006). Early conceptual development. In K. McCartney & D. Phillips (Eds.), *Blackwell handbook of early childhood development* (pp. 149–166). Malden: Blackwell.

Gelman, S. A., Chesnick, R., & Waxman, S. R. (2005). Mother-child conversations about pictures and objects: Referring to categories and individuals. *Child Development, 76*, 1129–1143.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*, 183–209.

Gentner, D., & Namy, L. L. (1999). Comparison in the development of categories. *Cognitive Development*, *14*, 487–513.

Hancock, C., Kaput, J. T., & Goldsmith, L. T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist*, *27*(3), 337–364.

Hanner, S., James, E., & Rohlfing, M. (2002). Classification models across grades. In R. Lehrer & L. Schauble (Eds.), *Investigating real data in the classroom* (pp. 99–117). New York: Teachers College.

Hourigan, M., & Leavy, A. M. (2016). Practical problems: Introducing statistics to kindergarteners. *Teaching Children Mathematics*, *22*(5), 283–291.

Johnson, K. E., & Mervis, C. B. (1994). Microgenetic analysis of first steps in children's acquisition of expertise on shorebirds. *Developmental Psychology*, *30*, 418–435.

Kinnear, V. (2013). *Young children's statistical reasoning: A tale of two contexts* (Unpublished doctoral dissertation). Queensland University of Technology, Brisbane, Australia.

Kinnear, V., & Clarke, J. (2016). Young children's abductive reasoning about data. In *Proceedings of the 13th International Congress on Mathematical Education. Hamburg, July 24–31*. Retrieved from http://icme13.org/files/ICME13-Programme356_low36.pdf

Leavy, A. M. (2015). Looking at practice: Revealing the knowledge demands of teaching data handling in the primary classroom. *Mathematics Education Research Journal*, *27*(3), 283–309. https://doi.org/10.1007/s13394-014-0138-3

Leavy, A. M., & Hourigan, M. (2015). Motivating inquiry in statistics and probability in the primary classroom. *Teaching Statistics*, *37*(2), 41–47. https://doi.org/10.1111/test.12062

Leavy, A. M., & Hourigan, M. (2016a). Crime scenes and mystery players! Using interesting contexts and driving questions to support the development of statistical literacy. *Teaching Statistics, 38*(1), 29–35. https://doi.org/10.1111/test.12088

Leavy, A. M., & Hourigan, M. (2016b). Using lesson study to support knowledge development in initial teacher education: Insights from early number classrooms. *Teaching and Teacher Education, 57*, 161–175. https://doi.org/10.1016/j.tate.2016.04.002

Lehrer, R., Kim, M., & Schauble, L. (2007). Supporting the development of conceptions of statistics by engaging students in measuring and modeling variability. *International Journal of Computers for Mathematical Learning*, *12*, 195–216.

Lehrer, R., & Schauble, L. (2000). Inventing data structures for representational purposes: Elementary grade students' classification models. *Mathematical Thinking and Learning*, *21*(1&2), 51–74.

Lehrer, R., & Schauble, L. (2007). Contrasting emerging conceptions of distribution in contexts of error and natural variation. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 149–176). New York: Taylor & Francis.

Lesh, R., & Doerr, H. M. (Eds.). (2003). *Beyond constructivism: Models and modeling perspectives on mathematics teaching, learning, and problem solving*. Mahwah: Lawrence Erlbaum Associates, Inc..

Lesh, R., & Kelly, A. (2000). Multitiered teaching experiments. In A. Kelly & R. Lesh (Eds.), *Research design in mathematics and science education* (pp. 197–230). Mahwah: Lawrence Erlbaum Associates.

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

Paparistodemou, E., & Meletiou-Mavrotheris, M. (2008). Enhancing reasoning about statistical inference in 8 year-old students. *Statistics Education Research Journal*, *7*(2), 83–106.

Petersen, L. A., & McNeil, N. M. (2013). Effects of perceptually rich manipulatives on preschoolers' counting performance: Established knowledge counts. *Child Development*, *84*(3), 1020–1033.

Samuelson, L. K., & Smith, L. B. (2005). They call it like they see it: Spontaneous naming and attention to shape. *Developmental Science*, *8*, 182–198.

Uzzell, B. P., Ponton, M., & Ardila, A. (Eds.). (2013). *International handbook of cross-cultural neuropsychology.* Mahwah: Lawrence Erlbaum.

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry (with discussion). *International Statistical Review*, *67*(3), 223–265.