THE ROLE OF CONTEXTS IN SUPPORTING EARLY STATISTICAL REASONING IN DATA MODELING

LUCÍA ZAPATA-CARDONA
Universidad de Antioquia, Colombia
lucia.zapata1@udea.edu.co

ABSTRACT

Data modeling is an essential activity in a data-driven society, but such a topic and how the context shapes it has received limited attention. This paper reports on research that investigated the role of context in supporting early statistical reasoning in the data modeling process. The data were collected throughout sessions in which young children (7 year-old) worked out problem activities designed to stimulate data modeling. The problem activities started by reading children's story books purposefully created as a strategy to provide contexts of interest. The stories were inscribed within culturally relevant contexts in which the characters deal with data in different formats. The data modeling problem activities were closely related to the stories described within the books. Special attention was put into the actions of organizing, structuring, visualizing, and representing data and the role of the context in the data modeling process. The main results suggest that the context of the problem activities for the data modeling process seems to facilitate statistical reasoning in young children. Additionally, the context of the problem activities helped participants to develop strategies to identify attributes of data, assess the model created, make sense of the data, and make informal inferences.

Keywords: *Statistics education research; Early childhood statistics; Story books; Data modeling; Statistical reasoning; Informal inferential reasoning*

1. INTRODUCTION

Data modeling has received limited or no attention in the early school statistics curriculum (English, 2010; Leavy & Hourigan, 2018), despite the need to educate informed citizens in a data-driven society. With the structure of today's world and the early access to computer technology, young children are exposed to multiple sources of data from social media, advertisement, and social statistics that could become the foundations for information and decision making if they are understood. Several researchers have indicated that reasoning about data could be stimulated when children are given the opportunity to create and assess their own models (English, 2010; Leavy & Hourigan, 2018; Lehrer & Schauble, 2007). Most classification assignments in school statistics, however, are based on attributes created by someone other than the student and in contexts that have little meaning to them. Research has suggested that statistical reasoning, the capacity to justify statistical processes and interpret results, could be promoted if learners are expected to create and revise different models from data (English, 2010; Leavy & Hourigan, 2018) within meaningful contexts (Campos, 2016; Makar, 2018).

In the statisticians' professional practice, context has proven to be an important element for statistical reasoning (Pfannkuch et al., 2016). In agreement with this statement, but in the educational field, some authors have shown that statistical ideas should not be presented in isolation because they could become atomistic, disintegrated, and impoverished leading to a lack of coherence in students' reasoning (Bakker & Derry, 2011). Other authors have suggested that context is an essential element in the design of learning experiences (Leavy & Hourigan, 2018). Makar (2018) suggested that the context could be a potent scaffold for children to engage with powerful statistical ideas. Despite its widely recognized importance in statistical practice, context has received very limited attention in school statistics in early childhood education (Leavy & Hourigan, 2018; Makar, 2018; Pfannkuch, 2011). Several scholars have proposed that data context provides meaning for data (e.g., Moore, 1990),

Statistics Education Research Journal, 22(2), Article 5. https://doi.org/10.52041/serj.v22i2.448

becomes the framing structure for data analysis and reasoning (Garfield & Ben-Zvi, 2007), and influences statistical sense making and reasoning processes (Kinnear, 2018). To develop rich statistical understandings, students must see how statistical concepts are related to one another, and make strong links to prior knowledge and previous experiences.

Taking into consideration the pertinence and importance of context in statistical reasoning and in the process of sense making, this study utilized meaningful contexts for data modeling from children's story books. In the field of statistics education, children's story books evoke an emotional connection with students, allow the development of mathematical ideas, function as a starting point for motivation of students, provide an important connection with prior knowledge, stimulate interest, and are a cognitive attraction (Kinnear, 2018). The research question to be answered is:

What is the role of context in supporting early statistical reasoning in the data modeling process?

2. THEORETICAL FRAMEWORK

2.1. DATA MODELING

Statistical models are statistical ways of representing and thinking about reality (Pfannkuch et al., 2016). Modeling in the field of statistics considers several processes, including interpreting and understanding a problem and its context; identifying, posing, and refining questions; collecting and organizing data; recognizing variation; creating models; and drawing conclusions and informal inferences from models generated (English, 2018). Such considerations imply that a simple tool, such as a graphical representation of data or a specific approach to structure data from a contextual phenomenon, could be considered key components in the modeling process (Konold et al., 2017).

In the research literature there are different interpretations of modeling. In this research, however, there is a special interest in data modeling that is considered to be a "fluid and open process where definition of the data and its attributes are in flux and under negotiation" (Leavy & Hourigan, 2018, p. 164). Such a process involves relevant actions that have to do with the attributes of data that are considered in the model, how they are measured, structured, represented, and used to make inferences (Lehrer & Schauble, 2007). Data modeling also involves displaying, summarizing, examining, and interpreting information to discover patterns and deviations from such patterns (English, 2010). Some authors value the specific actions that are carried out with data processing, but also the role of attribute selection and the communicative process in data modeling: "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" (Leavy & Hourigan, 2018, p. 164). In data modeling, the selection of attributes and their application to support the classification process is essential. That means, there is a focus on the attributes of the data rather than on the classification process itself. To be able to stimulate data modeling in school statistics, the designs that support teaching need to include tasks that expose students to multiple attributes in the data.

Early experiences with data modeling include reviewing data sets where there is a need to organize, classify, and represent according to a model. A critical element in creating models is determining the attributes and applying them in classification tasks. In those tasks, students make decisions about the characteristics that catch their attention and those that need to be ignored (Lehrer & Schauble, 2005; 2006). Most of the time, however, early education students are not exposed to these types of experiences, but rather are presented with situations in which the attributes are defined by someone other than the students. This oversimplifies the data modeling process, reduces the challenge for students, and diminishes possibilities for early statistical reasoning. Literature has shown that young children are able to create different representations for the same data, but it needs to be stimulated (English, 2010; Leavy & Hourigan, 2018). They need to have opportunities to define the data and its attributes.

2.2. EARLY STATISTICAL REASONING

Statistical reasoning is a skill that any citizen of a data-driven society should have to understand and informedly participate in society. Early childhood students are part of that data-driven society. They are constantly exposed to mass media, graphical representations of information, and social media. Hence, it is crucial to strengthen students' statistical reasoning to prepare them for the challenges faced as citizens.

Statistical reasoning is related to how people reason with statistical ideas and make sense of statistical information (Garfield, 2002; Garfield & Ben-Zvi, 2008). This involves understanding and explaining statistical processes and making interpretations based on data sets, graphical representations, and statistical summaries (Garfield, 2002; Gómez-Blancarte et al., 2021). Much statistical reasoning combines ideas about data and chance with abilities such as explanation, inference, verification, and demonstration. An important aspect of these types of reasoning is to understand the context in which statistical information is interpreted (Campos et al., 2013; Campos, 2016). Therefore, to aid the development of statistical reasoning, it is important whenever possible to work with real data, relate the data to the context, and promote the interpretation of the results.

One of the purposes of statistical reasoning is the possibility to draw conclusions from data, that is to make statistical inferences. Statistical inference is "justified by a probability model linking the data to a broader context" (Rossman, 2008, p. 5); however, with young children, who are not yet introduced to formal procedures, these inferences are explored from an informal point of view. Informal inference is the process of meaning-making and evidence-building from data without running formal procedures. An informal statistical inference is a statement that includes the principles of (1) generalizing beyond the data collected, (2) using the data as evidence when generalizing, (3) and applying probabilistic language that suggests uncertainty (Makar & Rubin, 2009). In this sense, the early statistical reasoning referred to in this study could be understood as a cognitive activity that seeks to establish generalizations based on the evidence and the uncertainty provided by the data, but without resorting to formal inferential processes.

2.3. THE CONTEXT IN THE STATISTICAL REASONING

Context is an essential component to promote statistical reasoning; however, it needs to be relevant to students (Zapata-Cardona, 2018). To illustrate this fact, Makar and Rubin (2009) described a thought-provoking case that shows the lack of student connection in a statistics class with a topic that was situated too far from the children's experiences and interests to engage their reasoning. In contrast, Langrall et al. (2006) presented an experience that revealed the context to be a central factor for students to engage in statistical tasks. Ben-Zvi and Aridor-Berger (2016) described a case in which participants wander between context and data worlds to develop their reasoning. They proposed that the context served as a scaffold for the development of informal reasoning by providing a common language that helps to better express the statistical ideas behind the tasks. Some authors have suggested that context is important in statistics learning because ideas are not atomized (isolated) but integrated (Bakker & Derry, 2011). Within rich context students have opportunities to use statistical ideas in holistic ways. In addition, when students are exposed to relevant cultural contexts, they gain important experiences with informal versions of advanced statistical structures that they could use later to build formal ideas (Makar, 2018). Research has revealed that early education students possess multiple conceptual resources that could be used to develop sophisticated ways of reasoning if appropriate learning experiences that challenge their intellectual activity are designed and implemented (English, 2010).

2.4. STORY BOOKS

The use of story books to create scenarios for teaching has proved to have several advantages. Some researchers in the field of mathematics education have suggested that story books could offer informal spaces for children to experience mathematical concepts and structures from which they can build more formal knowledge (van den Heuvel-Panhuizen & van den Boogaard, 2008). In this regard, van den Heuvel-Panhuizen and van den Boogaard (2008) showed indications of cognitive engagement and mathematical thinking of five-year-old children when a book was read to them.

Story books have also worked as motivators to offer children meaningful scenarios as bases for the construction of mathematical ideas (Elia et al., 2010). English (2012) suggested that students' motivation is activated in the storytelling process promoted by story books, which could have favorable implications for learning mathematics, achievement, and activity engagement. Story books have offered important connection to the context under study (van den Heuvel-Panhuizen & van den Boogaard, 2008) and provide students a useful link to prior knowledge. In the study by Heuvel-Panhuizen and van den Boogaard (2008), participants were confident producing mathematical related utterances because the story book proposed a familiar context that activated their mathematical thinking. In the field of statistics education, story books have been used as initiators of students' interest in a problem and as providers of contexts for statistical problem solving (Kinnear, 2018), as well as starting points for data modeling (English, 2010).

3. METHOD

This research has a qualitative approach since the purpose is to study how the context shapes participants' early statistical reasoning in data modeling activities. Observations and interviews of participants while solving different problem activities in multiple sessions were the main techniques to produce the information, which was complemented by participants' written work (Creswell & Creswell, 2018). The participants for this study were two first-grade seven-year-old male students in an out of school environment. They voluntarily accepted the invitation to be part of the study and received no other retribution than the pleasure of reading and talking about children's story books and facing the challenges of the proposed problem activities. The only reason to consider male participants was their availability. The participants had not received formal instruction in data modeling.

The study took place at the last month of the school year when the participants were at the end of their first grade. The initial purpose of this study was to test the potential of the story books and the activities designed for data modeling. The implementation of the activities, however, offered important information in terms of the influence of the context in the early statistical reasoning in the data modeling process.

3.1. DESIGN

Four children's story books were the foundations of this research study. They were purposefully designed to promote data modeling. Three were designed by the author and one was taken from English's (2010) research. Each story book provided a relevant and familiar scenario to engage the participants in a series of four, multi-component problem activities and to promote early statistical reasoning. Table 1 presents a summary of each children's story book and a brief description of the problem activities. The children's story books were read individually to each participant using on-line resources and subsequently, a series of problem activities were implemented.

Table 1. Summary of the children's story books

Story book	Summary	Problem Activities
Baxter Brown's Messy Room (English, 2010)	The story tells how a little white dog collects all sorts of sundries and accumulates them in his bedroom, and he gets lost among the piles of items that have been collected. Baxter Brown has to solve his messy problem and the participants were invited to help the dog figure out how to clean his room, considering classification of the items.	1. Forty-four items from the messy room are given to the participants to be classified in different ways. Several items are of the same kind. 2. Twelve items from the messy room are given to the participants to be classified using different attributes. All items are different. 3. A table reporting items collected on Monday, Tuesday and Wednesday is given to participants to make a prediction for Thursday. 4. A visual representation with a classification of items is given. The participants are asked to

		predict where an unknown item would more likely go in that classification.
Feliciano's Toys	The story tells the adventures of Feliciano, a boy with a chaotic collection of toys. Due to his messiness, the story has quite interesting and funny implications. Since the toys are all over the house, the family is late getting to destinations because they have to wait for Feliciano to organize a bit before leaving the house. Once, they went to the supermarket as the store was being closed and the family could not do their grocery shopping. Participants are invited to help Feliciano organize his toys using different attributes in the classification.	1. Twelve toys are given to be classified in three baskets according to different attributes. 2. Seventeen toys, some of them repeated, are given to be classified using different attributes. 3. A visual representation is given, and the participants are asked to predict what toy Feliciano is more likely to use next time.
Juanita's Birthday	The story tells how Juanita loses her first tooth at her sixth birthday party. During the party, she notices that many of her guests have lost teeth and asks them at what age they lost their first tooth. With the information given in the story, different tasks of organization, structuring and inference are proposed.	1. Participants are asked: Did the teeth fall out start at the same age for everyone in the party? 2. A visual representation with the information of the loss of milk teeth of the guests is given. Participants are asked to predict the teeth fall out for: (a) A five-year-old who has not started his/her teeth fall out, (b) a new guest that comes to the party whose age is unknown, and (c) an eight-year-old guest. 3. Participants are asked to predict the number of teeth loss for a six-year-old guest.
A Fantastic Magician	This story describes how a fantastic magician would take a colored pencil out of a hat, show it to the public and put it back in the hat. After ten trials he asks the audience several questions, "How many pencils are in the hat? What is the proportion of pencils in the hat? What color pencils are most likely to be in the hat?	Participants are invited to answer the questions proposed in the story book.

3.2. DATA COLLECTION AND ANALYSIS

Data collection relied on videos from four individual one-hour sessions, one week apart during one month with each participant (eight sessions in total) and the participants' written productions in the problem activities. The sessions were held on-line and were recorded using the functions of the platforms used (Zoom or Google Meet). The videos were transcribed word by word in Spanish and later translated into English to facilitate the analysis. The focus was on the participants' articulation of their statistical reasoning in particular stages of the sessions and how the context played a crucial role in the process. The analysis is illustrated with episodes of interest (Powell et al., 2003) from the data gathered. For each problem activity, instances were tracked in which participants (1) used context to understand the situation, offered explanations, provided alternative interpretations, or suggested conjectures, and (2) displayed early statistical reasoning. This scheme was used as a basis for coding the data. The transcripts for each session were reviewed and annotated. Each session was deconstructed into episodes that reflected the role of context or early statistical reasoning. Each episode was coded and when appropriate, descriptors were assigned to further characterize the episode. The episodes were coded and examined for patterns and trends using constant comparative strategies (Strauss & Corbin, 1990). Those episodes that offered richness to better explain the issues of interest were the ones selected to illustrate the analysis.

4. ANALYSIS

The results use data from the sessions with the participants in the attempt to investigate the role of context in supporting early statistical reasoning in the data modeling process. The focus is on the informal component of statistical reasoning since at seven years there was no formal instruction. The analysis is illustrated with two episodes that make evident the role of context and the statistical reasoning: in the data modeling and in the informal inference. The episodes were selected purposefully because they showed events that are worth studying in depth due to their richness and complexity (as described in Powell et al., 2003). The episodes are not chosen because they are representative of what happened in the sessions with the participants, but because of their explanatory power to support the arguments about the role of context in young children's statistical reasoning in the data modeling process. The exchanges are numbered in order of presentation in the paper rather than order of occurrence in the transcripts.

4.1. EPISODE 1 - CONTEXT SHAPED THE DATA MODELING ACTIVITY

Feliciano's Toys was the context for all the proposed problem activities in some sessions. In that story book, Feliciano had made a mess with his toys and the participants were invited to help him think of strategies to organize them. One of the activities asked the participants to organize twelve items of different nature (a piece of *Lego*™, a teddy bear, a ball, a book, a jump rope, a piece of a puzzle, an airplane toy, a group of color pencils, a kite, some blank paper, some watercolors, scissors) from a mountain of disorder located in a corner of Feliciano's room. The participants were given three baskets to propose attributes and apply them in the organization and classification of the objects. Here is a fragment of the exchange.

1	Researcher	(05:58) Feliciano has to organize these items into groups that share the same characteristic. For that purpose, he has three baskets. How would you help him?
2	Participant	(08:10) The paper, the colors pencils and the book need to be kept in the same basket because those are study items (10:00) Wait a moment, paper and color pencils are not the only study items, scissors and watercolors too (11:45) I think I have an idea. I think I will have a basket left over in case more toys come. I mean, in the second box, I'm going to put the remaining of the toys: the Teddy bear, the Lego, the airplane, what else, the puzzle, the kite, and the jump rope. And I have a left-over basket in case more toys come so that he has a place to store them. (May 29, 2021)

In the exchange, it is evident that the participant classified the twelve items according to the attribute he developed. The attribute was whether the item was a toy or a study item. Then, the researcher asked the participant to think about another strategy to organize those twelve items and he modified the earlier strategy to include a more refine one.

3	Participant	(15:20) We could get the kite from here and the jump rope because those are outside toys [he places them in another basket]. Inside toys could be put in one basket and outside toys in another [he forms tree baskets: one with the study items, one with the inside toys and the last one with the outside toys]. But I think that is not a good organization because if more toys come for birthdays, Christmas or special events, there will be toys with no room. It is not very efficient strategy because there will not be space left (May 29, 2021).

The participant found two different ways to classify the objects given in the three baskets. The first strategy included three categories: study items, toys, and toys that will come. The second strategy also included three categories: study items, inside toys, and outside toys. The familiarity with the context helped the participant to identify different attributes to apply to his data modeling structure. With his insistence to keep one basket left empty, "I think I will have a basket left over in case more toys come"

(Exchange 2), it is clear that he applied his personal experience of getting toys as presents in special events. “If more toys come for birthdays, Christmas or special events, there will be toys with no room” (Exchange 3). The researcher questioned, “Could you think about another strategy to organize these toys?” which led the participant to think about a more refined attribute.

4.2. EPISODE 2: INFORMAL INFERENTIAL REASONING

After reading *Juanita’s Birthday*, the participants were asked to answer the following question: Did teeth fall out start at the same age for everybody in the party? To answer the question, one of the participants decided to organize the information from the party guests into four groups according to the age at which the baby teeth fall out started. The visual representation was a graphical abstraction where the name of the guest and the age at which the first tooth fell out (see Figure 1) were considered as the most relevant characteristics of the context. That representation functioned not only to summarize the information from the party guests, but also as support for later informal inference.

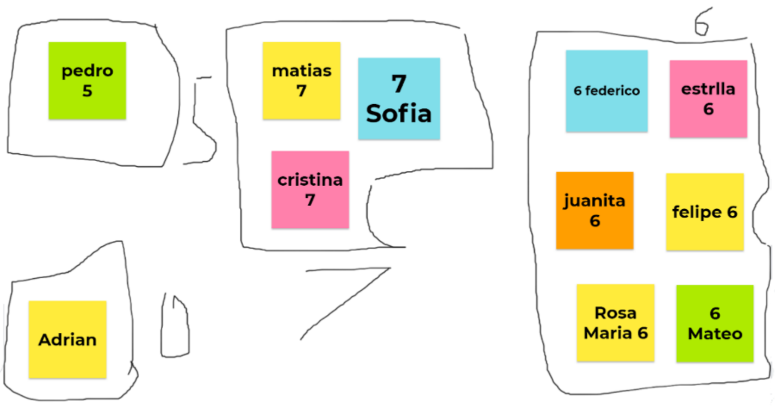

Figure 1. Visual representation of the participant using the Jam board from Google Meet

Once the participant created his graphical representation and explained it, the researcher showed him an image with a different visual organization (see Figure 2) that another child created. The aim was for the participant to comment on the representation.

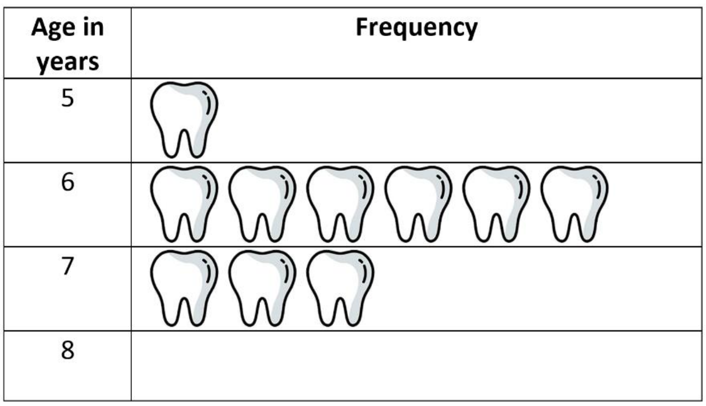

Figure 2. Visual representation to contrast with the participant’s representation

4	Researcher	(19:00) This image was the strategy another child used to organize the information (see Figure 2). What do you think about it?
5	Participant	(19:19) It was something very similar to mine […] (19:30) he gave it a number, he put a box on it. Except, he only put a tooth. (June 5, 2021)

The participant recognized the similarity of the two representations to summarize the information. However, his representation was more detailed than the one presented by the researcher. The participant's representation included two variables: age when the baby teeth fall out started and name of the party guest. The expression, "Except, he only put a tooth" (Exchange 5), seems to be a call of attention to refer that the visual organization of the data under discussion failed to identify an important variable in the information of the party guest. Using the visual representation was a crucial strategy in the data modeling process. The participant identified attributes, structured them, and used the organization to communicate a strategy. The organization of the information further helped the participant to support some informal inference, as illustrated in the following exchanges.

6	Researcher	(16:00) If we are going to answer that question from the beginning: Did teeth fall out start at the same age for everybody in the party? With this information, what could you answer?
7	Participant	(16:10) Yes and no.
8	Researcher	(16:14) Can you explain that a little bit more?
9	Participant	(16:22) If we look within the small groups [could be yes], but if they are talking about everyone, they [the first tooth] fall out at different ages
10	Researcher	(16:40) When do you think Adrián's [a five-year-old character in the story book] first tooth will fall out? When is that tooth most likely to fall out?
11	Participant	(17:14) It could be at six.
12	Researcher	(17:17) Why do you think at six? How sure are you that it could be at six?
13	Participant	(17:20) Because there are more children whose tooth fell out at six
14	Researcher	(19:47) If a new friend comes to the party, when do you think their teeth may have started falling out?
15	Participant	(19:58) Maybe at six years old
16	Researcher	(21:05) Why?
17	Participant	(21:03) Because there are more children with teeth that fell out at six
18	Researcher	(21:07) With the information you have, is it possible to think that some child could have started their teeth fall out at eight years of age?
19	Participant	(21:28) Not much.
20	Researcher	(21:30) Tell me a little more why do you think that?
21	Participant	(21: 50) Because there is no one who lost a tooth [the first one] at the age of eight and there are no cases.
22	Researcher	(21:58) If a new six-year-old friend comes to the party, could you predict how many teeth he/she may have lost so far?
23	Participant	(22:38) Maybe a tooth.
24	Researcher	(22:40) How can you suspect that?
25	Participant	(22:48) Wait a minute, but I do have two fallen out teeth. Let's see.
		[Session 3, video timestamp 16:00–22:48, June 5, 2021]

In this exchange, the participant used his visual representation to talk about the composition of the tooth fall out of the party guests. When he referred to the whole group he said, "If we are talking about everyone, they [the first tooth] fall out at different ages" (Exchange 9). He referred to the aggregate, not an individual, and considered the data variation within the whole data set. Later, the researcher asked some questions that required some prediction, and the participant went back and used his graphical representation to establish the prediction. The researcher asked the participant to predict for (a) a five-year-old guest in the party who's baby teeth had not started his falling out (Adrian), (b) a new guest whose age was unknown, (c) a six-year-old guest, and (d) the starting of baby teeth fall out for an eight-year-old child.

For the first prediction, the participant said, "It could be at six" (Exchange 11) and justified his answer based on the graphical representation and stated, "Because there are more children whose tooth fell out at six" (Exchanges 13 and 17). For the new guest that joined the party, that person's age was unknown, so the participant went back to his graphical representation and stated, "Maybe at six years old" (Exchange 15). The researcher asked for the possibility of the teeth starting to fall out at eight years old and the participant answered, "Not much" (Exchange 19), "Because there is no one who lost a tooth at the age of eight and there are no cases" (Exchange 21). In the predictions, the participant displayed a degree of uncertainty in drawing conclusions that were based on data.

In the prediction for the number of teeth already fallen out of a six-year-old friend that comes to the party, the participant initially answered, “Maybe a tooth” (Exchange 23), but after the researcher asked for justification, he hesitated and put his personal experience into the situation to express some doubt, “Wait a minute, but I do have two fallen out teeth. Let’s see” (Exchange 25). His personal experience functioned to contrast the information from the graphical representation and as a strategy to assess the conclusions given from the model. It was also a way of going back and forth between the context and the data.

5. DISCUSSION

This paper examined the role of contexts from story books in supporting early statistical reasoning in the data modeling process. The participants responses to the problem activities were explored in terms of how they made sense of data, identified attributes in data, created representations, organized and structured the data, considered variation, assessed the model, and made informal statistical inferences.

The two episodes presented illustrate the participants responses to the proposed problem activities and make evident the possibilities for learning. The findings show that participants made sense of the data from the story books, managed the data, experienced variations, made predictions and generated their own models to make sense of the data. These findings coincide with those from English (2018), who also highlighted the learning affordances generated throughout the problem activities proposed.

The participant in the first episode developed a well-defined classification strategy and, after being questioned by the researcher, established a more refined attribute in the data modeling activity. He created, analyzed, and revised his classification models. These results coincide with those from English (2010, 2012) who showed that students developed a classification strategy, but after talking to others, they refined their classification. The results, however, differed from Leavy and Hourigan’s (2018) research who reported that the objects under classification did not fix clearly in the categories the children came out with. Most of the attributes from the participants in Leavy and Hourigan’s (2018) study were perceptual (e.g., color, size, presence of wings), less taxonomical and less visible (e.g., diet). This may be explained because participants in that study were younger (5–6-year-old children) than the ones in the present study. Another aspect to consider in the discussion is the interaction with the adult, which was essential for the participants to go beyond the data collected. As Vygotsky (1986) suggested, learners move beyond their level of understanding through the help of more knowledgeable others. The expertise in the topic also helps to develop reasonable classifications. Toys is a very common topic for children, more so than a zoo design that inspired Leavy and Hourigan’s (2018) work. In this regard, Langrall et al. (2006) showed that discussion instances increase in familiar contexts. This contrast suggests that the knowledge of the topic plays an important role in the data modeling process.

The results suggest that the contexts used in this research were determinant in helping the participants to develop their data modeling strategies. Culture had an important influence in participants’ decisions. Toys were common gifts that the participants received in special occasions and the loss of the baby teeth was a natural process that both had gone through. In the two episodes presented, the participants were engaged in the contexts proposed by the story books and in solving the problem activities that took them throughout data modeling process. Early statistical reasoning emerged through the discussions in the sessions within familiar situations. The story books’ contexts gave the participants the confidence to talk about situations that were familiar for them and reason statistically from an informal point of view in the data modeling process.

The informal statistical reasoning that took place was similar to that found in later years with more formal statistical instruction. The contexts given by the children’s story books was content-rich and complex, allowing for multiple aspects of statistical reasoning in which participants developed strategies to identify attributes of data, use such attributes to organize, structure, and visualize data, as well making informal inferences. Tools, strategies, and concepts were blended with meaningful contexts, and not in isolation, to solve problem activities and to develop holistic knowledge (Bakker & Derry, 2011; Makar, 2018). These findings suggest that children’s story books and corresponding problem activities appear to play an important role in the way children perceive the tasks and, therefore, could influence how they engage in them and the reasoning they stimulate.

Similar findings were reported by Makar (2018), who suggested that within rich context, young children have the potential to integrate concepts, relationships, tools, and structures that had analogies in the context of reference to purposefully solve a statistical question. Kinnear (2018) reached similar conclusions, stating that the context in which the activities are presented influences the way data are approached, engaged, analyzed and interpreted, and consequently context influences the statistical reasoning and knowledge development that takes place. Langrall et al. (2006) also concluded that context has an important role in children's task perception and can influence their engagement and reasoning. Likewise, the study of Ben-Zvi and Aridor-Berger (2016) found that children go back and forth from data world to the context world to develop their reasoning and to better express their statistical ideas.

6. CONCLUSIONS

The study reported in this paper revealed that contexts from children's story books used in data modeling activities (1) offer meaningful scenarios for participants to develop attributes, structure them and apply them to data sets in classification processes, and (2) function as tools to promote early statistical reasoning in which participants go back and forth between the data and the context to make claims and predictions that have the form of informal inferences. The results revealed that participants genuinely engaged in solving a series of problem activities that stimulated early statistical reasoning in familiar and meaningful contexts. This familiarity and significance provide clues to suggest that both the design of teaching activities and curricular materials should be thought of from spaces that are attractive to students and in which there is an affective bond.

The findings from this study provide direction for future research. There is potential for further studies to examine additional age groups in a variety of educational settings. To find more robust evidence, this study needs to be done with a larger number of participants from different backgrounds. The contexts from the story books reported in this study also have interesting possibilities for further research. Although the results did not report on this specific point, each story book stimulated different skills in the early statistical reasoning. It would be interesting to explore this issue deeply. This study also showed that early childhood learners developed affective bonds with the problem activities inspired by the story book narratives. In this sense, a practical implication for both research and teaching would be to consider story books in the design of data modeling problem activities with early childhood participants.

ACKNOWLEDGEMENTS

Thanks to the Universidad de Antioquia research committee-CODI and the participants who generously accepted the invitation to read story books and talk about them. Special thanks to the reviewers who provided feedback via the blind review process and to the editors for their helpful comments to improve this paper.

REFERENCES

Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning, 13*(1–2), 5–26. https://doi.org/10.1080/10986065.2011.538293

Ben-Zvi, D., & Aridor-Berger, K. (2016). Children's wonder how to wonder between data and context. In D. Ben-Zvi & K. Makar (Eds.), *The teaching and learning of statistics*. Springer. https://doi.org/10.1007/978-3-319-23470-0_3

Campos, C. R. (2016). *Towards critical statistics education: Theory and practice*. Lambert Academic Publishing.

Campos, C. R., Jacobini, O. R., & Wodewotzki, M. L. (2013). *Educação Estatística: teoria e prática em ambientes de modelagem matemática* (2nd ed.). *[Statistical Education: Theory and practice in mathematical modeling environments]*. Autêntica Editora.

Creswell, J. W., & Creswell, J. D. (2018). *Research design qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Elia, I., Georgiou, A., & van den Heuvel-Panhuizen, M. (2010). The role of pictures in picture books on children's cognitive engagement with mathematics. *European Early Childhood Research Journal*, *18*(3), 275–297. https://doi.org/10.1080/1350293X.2010.500054

English, L. D. (2010). Young children's early modelling with data. *Mathematics Education Research Journal*, *22*, (2) 24–47. https://doi.org/10.1007/BF03217564

English, L. D. (2012). Data modelling with first-grade students. *Educational Studies in Mathematics*, *81*, 15–30. https://doi.org/10.1007/s10649-011-9377-3

English, L. D. (2018). Young children's statistical literacy in modelling with data and chance. In A. Leavy, M. Meletiou-Mavrotheris & E. Paparistodemou (Eds.), *Statistics in early childhood and primary education* (pp. 295–313). Springer. https://doi.org/10.1007/978-981-13-1044-7_17

Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education*, *(10)*3, 1–12. https://doi.org/10.1080/10691898.2002.11910676

Garfield, J. B., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, *75*(3), 372–296. https://doi.org/10.1111/j.1751-5823.2007.00029.x

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer.

Gómez-Blancarte, A. L., Chávez Aguilar, R. D., & Rocha Chávez, R. (2021). A survey of the teaching of statistical literacy, reasoning and thinking: Teachers' classroom practice in Mexican high school education. *Statistics Education Research Journal*, *20*(2), Article 13. https://doi.org/10.52041/serj.v20i2.397

Kinnear, V. (2018). Initiating interest in statistical problems: The role of picture story books. In A. Leavy, M. Meletiou-Mavrotheris & E. Paparistodemou (Eds.), *Statistics in early childhood and primary education* (pp. 183–199). Springer. https://doi.org/10.1007/978-981-13-1044-7_11

Konold, C., Finzer, W., & Kreetong, K. (2017). Modeling as a core component of structuring data. *Statistics Education Research Journal*, *16*(2), 191–212. https://doi.org/10.52041/serj.v16i2.190

Langrall, C. W., Mooney, E. S., & Nisbet, S. (2006). The interplay between students' statistical knowledge and context knowledge in analysing data. In A. Rossman & B. Chance (Eds.), *Working co-operatively in statistics education.* Proceedings of the 7th International Conference on Teaching Statistics (ICOTS7), Salvador, Brazil. International Statistics Association. https://iase-web.org/documents/papers/icots7/2A3_LANG.pdf?1402524964

Leavy, A., & Hourigan, M. (2018). The role of perceptual similarity, context, and situation when selecting attributes: considerations made by 5–6-year-olds in data modeling environments. *Educational Studies in Mathematics, 97*, 163–183 https://doi.org/10.1007/s10649-017-9791-2

Lehrer, R., & Schauble, L. (2005). Developing modeling and argument in the elementary grades. In T. Romberg, T. Carpenter & F. Dremock (Eds.), *Understanding mathematics and science matters* (pp. 29–53). Lawrence Erlbaum.

Lehrer, R., & Schauble, L. (2006). Cultivating model-based reasoning in science education. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 371–386). Cambridge University Press.

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). Taylor & Francis.

Makar, K. (2018). Theorising links between context and structure to introduce powerful statistical ideas in the early years. In A. Leavy, M. Meletiou-Mavrotheris & E. Paparistodemou (Eds.), *Statistics in early childhood and primary education* (pp. 3–20). Springer. https://doi.org/10.1007/978-981-13-1044-7_1

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, *8*(1), 82–105. https://doi.org/10.52041/serj.v8i1.457

Moore, D. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95–137). National Academy Press.

Pfannkuch, M. (2011). The role of context in developing informal statistical inferential reasoning: A classroom study. *Mathematical Thinking and Learning*, *13*(1&2), 27–46. https://doi.org/10.1080/10986065.2011.538302

Pfannkuch, M., Budgett, S., Fewster, R., Fitch, M., Pattenwise, S., Wild, C., & Ziedins, I. (2016). Probability modeling and thinking: What can we learn from practice? *Statistics Education Research Journal*, *15*(2), 11–37. https://doi.org/10.52041/serj.v15i2.238

Powell, A. B., Francisco, J. M., & Maher, C. A. (2003). An analytical model for studying the development of learners' mathematical ideas and reasoning using videotape data. *Journal of Mathematical Behavior, 22*(4), 405–435. https://doi.org/10.1016/j.jmathb.2003.09.002

Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal*, *7*(2), 5–19. https://doi.org/10.52041/serj.v7i2.467

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Ground theory procedures and techniques*. SAGE Publications.

van den Heuvel-Panhuizen, M., & van den Boogaard, S. (2008). Picture books as an impetus for kindergartners' mathematical thinking. *Mathematical Thinking and Learning*, *10*(4), 341–373. https://doi.org/10.1080/10986060802425539

Vygotsky, L. (1986). *Thought and language* (Translation A. Kozulin). The Massachusetts Institute of Technology Press.

Zapata-Cardona, L. (2018). Students' construction and use of statistical models: a socio-critical perspective. *ZDM Mathematics Education*, *50*(7), 1213–1222. https://doi.org/10.1007/s11858-018-0967-8

LUCÍA ZAPATA-CARDONA
Universidad de Antioquia
Calle 67 No. 53–108, Medellín
Colombia, 050034
ORCID https://orcid.org/0000-0003-4266-5273