



# Investigating statistical predictions with first graders in Greece

Anastasia Michalopoulou<sup>1\*</sup> , Sonia Kafoussi<sup>1</sup> 

<sup>1</sup>Department of Sciences of Preschool Education and Educational Design, University of the Aegean, Rhodes, GREECE

\*Corresponding Author: [psed19008@rhodes.aegean.gr](mailto:psed19008@rhodes.aegean.gr)

**Citation:** Michalopoulou, A., & Kafoussi, S. (2024). Investigating statistical predictions with first graders in Greece. *International Electronic Journal of Mathematics Education*, 19(2), em0771. <https://doi.org/10.29333/iejme/14247>

## ARTICLE INFO

Received: 30 Aug. 2023

Accepted: 05 Feb. 2024

## ABSTRACT

This paper argues that engaging students in informal statistical reasoning from early school years is essential for the development of statistical understanding. We investigated if and how children aged six-seven years old identified variation in a table of data and made predictions through the design of a teaching experiment. The classroom teaching experiment was comprised of four 45 minutes lessons addressing the understanding and interpretation of data sets. In order to describe students' informal predictive reasoning, we used the framework of "data lenses". More specifically, we analyzed the different types of answers the students produced as they engaged in predictive reasoning during an interview given before and after the teaching experiment. The participation of students in (classroom) and out-of-school (family) communities of practice was also taken into consideration. Our results demonstrate that the students benefited from their learning experience and developed data understanding.

**Keywords:** data lens, early childhood, parental involvement, statistical predictive reasoning

## INTRODUCTION

In recent years researchers' interest in supporting children's statistical education has been constantly increasing. The reasons why the teaching of statistics should be part of basic education have been repeatedly pointed out by various researchers (e.g., Ben-Zvi & Garfield, 2004; Franklin et al., 2007; Gal, 2002; Steen, 2001; Weiland, 2017). Among the basic reasons is the usefulness of statistics in everyday life (i.e., nutritional facts), its auxiliary role in other sciences (i.e., COVID-19 pandemic management, weather forecasts), the need for basic statistical knowledge that exists in many professions and its contribution to the development of critical reasoning (i.e., statistical arguments of political public discourse). Moreover, it is claimed that statistical literacy is the priority of statistical education, as it is fundamental for students' future life as citizens of modern society, where reality is shaped and informed by arguments based on data (Weiland, 2017). As a consequence, the main purpose of teaching statistics is to prepare students to interpret and critically evaluate information, draw conclusions, make predictions and make decisions under uncertain conditions in a variety of contexts. Garfield and Ben-Zvi (2007) have argued that statistics teaching needs to focus more on the unlocking of the stories in the data and making sense of the data through the investigation of patterns while noticing the unexpected.

According to English (2013), developing learning opportunities, where students are encouraged to be engaged in statistical reasoning with uncertain situations in the early school years will provide a valuable learning experience for formal statistical understanding in later years. Early engagement with informal statistical reasoning could lead to a solid statistical background for students' learning trajectory (Doerr et al., 2017). Towards this effort, researchers have investigated how students could meaningfully be engaged in informal statistical reasoning from early childhood (Aridor & Ben-Zvi, 2017; English, 2012, 2013; Lehrer & English, 2017; Makar, 2014, 2016; Makar & Rubin, 2017; Oslington et al., 2020; Papanistodemou & Meletiou-Mavrotheris, 2008). Echoing their findings, an informal perspective to statistical inference is crucial in the early years and informal prediction could be an effective access to inference (English, 2012; Makar, 2016; Meletiou-Mavrotheris & Papanistodemou, 2015). Young children are able to make informal inferences based on available data, with prediction being an important element of this process (Watson, 2007). However, there has been limited research on developing young children's informal predictive reasoning in statistical contexts (Oslington et al., 2020).

Moreover, nowadays research should consider that students participate in multiple local communities and broader groups in society (Yackel et al., 2011). The learning route of a child does not only depend by facts realizing in the school classroom, but it emerges as a result of the multiple ways of engaging with mathematics in the socio-cultural environments with which he/she interacts (Darragh, 2016). Especially, parents' participation in their children mathematics education is a critical issue, as many studies have highlighted that parental involvement could have positive influence on children's mathematical achievement as well

as on the formation of positive attitudes and beliefs towards mathematics, including mathematics homework (e.g., Boonk et al., 2018; Civil & Bernier, 2006; Crafter, 2012; Galindo & Sheldon, 2012; Knapp et al., 2017; Maloney et al., 2015; Quay & Pomeroy, 2022; Silinkas & Kikas, 2019; Van Voorhis et al., 2013; Zippert & Rittle-Johnson, 2020).

Based on the above, research in predictive reasoning of students in early childhood is relatively limited and lacks the consideration of parents' participation. In our study, we attempt to address the above gap. We investigated if and how children aged six-seven years old identified variation in a table of data and then made predictions, during interviews given before and after a teaching experiment. More specifically, we analyze and present the different types of answers related to data understanding the students developed, as they were engaged in predictive reasoning, using the theoretical framework of "data lenses" of Konold et al. (2015). Furthermore, our research explores students' answers in relation to parental involvement, considering if the students completed two homework assignments with the help of their parents during the teaching experiment.

## THEORETICAL BACKGROUND

### Insights on Statistical Reasoning

Statistical reasoning includes correct and incorrect ways that student reason about statistics (Garfield, 2002; Shaughnessy, 2007). At an introductory level, students should be given the learning opportunity to develop a data sense and engage in informal inference (Franklin et al., 2007). Nowadays, it is well accepted that inference is the essence of statistics, as it allows the creation of substantiated claims in the light of uncertainty when only some data are available (Makar, 2016). According to Makar and Rubin (2009) informal statistical inference is a claim with three key characteristics:

- (1) it goes beyond the data,
- (2) it utilizes data as evidence, and
- (3) it uses expressions of uncertainty.

Prediction could be an effective access to informal inference. Oslington et al. (2020) have mentioned that "prediction is an everyday statistical activity, where individuals draw upon past experiences and incomplete information to estimate, plan or draw conclusions" (p. 5). At an informal level, prediction and estimation are everyday experiences that can be used by non-specialists to access inferential processes (Paparistodemou & Meletiou-Mavrotheris, 2008). Predictions can be based on aspects of the problem context and children's understanding of the data presented.

Moreover, another critical issue when teaching statistics should be about finding appropriate situations through which students will have the opportunity to reason and reflect on the implications of their statistical decisions in real life (Mc Clain & Cobb, 2001). When students interpret data and make informal predictions in familiar contexts, they are given the learning opportunity to develop some of the basic ideas in statistics such as variability, range and aggregate properties of data sets (Makar, 2016). During the first school age, students can develop their statistical reasoning by understanding that data are not simple numbers, but information with reference to a specific situation, in other words numbers in context as well as by making predictions helping them to make decisions based on this information. According to Piaget's (1954) theory of cognitive development, children aged six-seven are in an operational stage, where thinking is based on concrete experiences and objects, so it is valuable to engage them in experientially real statistical activities. To this end, we designed tasks that were considered to be meaningful for a first grade student.

### Framework of the Study

In order to describe students' informal predictive reasoning, we used the framework "data lenses" from the work of Konold et al. (2015). The development of the framework was grounded on research reporting the difficulty the students face in perceiving the data set as an aggregate rather than as a "collection of points" (Cobb, 1999). Through their research on how students perceive and reason about aggregate, Konold et al. (2015) proposed four distinct perspectives (lenses) the students employ when reasoning about data. These lenses reveal different levels of students' understanding of data. More specifically, they identified four general perspectives in students' data understanding, namely:

- Interpreting data as pointers: students understand data as reminders of the larger event from which the data came.
- Interpreting data as case values: students take the individual data element as the perceptual unit and focus on the characteristics of individual cases.
- Interpreting data as classifiers: students understand cases with similar values as a unit.
- Interpreting data as an aggregate: students understand the data set as a whole.

The above descriptive framework has already been used by a small number of researchers who have focused on requiring students to predict unknown values from statistical data. For example, English (2012) has studied first-grade children's data modeling and predictions. Participants in her research were three classes of first grade children that engaged in a series of statistical activities using picture storybooks. In one of the activities, students engaged in making predictions. Students were presented with a table of data describing the number of different types of garbage found in a park during three consecutive days. Then students were asked to make predictions on the amount of garbage expected to be found on the fourth day. All groups recorded predictions of values within a small range of the values in the data set given. Her research has suggested that young children are capable of dealing with informal inference.

Makar (2016) investigated how young children (aged five to six years) engaged in informal statistical inference and data-based reasoning within an inquiry-based environment in a series of teaching experiments realized in two phases over a period of six months. In phase 1, the students were engaged in four activities involving inferring in stories, creating and working with patterns and recording. In phase 2, which lasted three lessons, the students made predictions about shoe sizes expected to be found in other classrooms, after drawing the data from their classroom themselves. She indicated six foundational skills of informal statistical inference:

- (1) articulating or predicting from personal observations,
- (2) recording data,
- (3) organizing data,
- (4) inventing methods,
- (5) working with aggregates, and
- (6) exposure to variability.

However, the participants were not “typical” students in that they were selected by the teacher as those who had strong communication and investigation skills and an above average performance in school. As a result, more research is needed with a more typical group of students.

Oslington et al. (2020) researched 46 third graders’ predictive reasoning strategies as they interpreted a table of temperature data in order to predict future monthly maximum temperatures. The students were withdrawn from class in groups and participated in a 90-min lesson. The students’ predictions, representations, written and verbal descriptions were also analyzed according to Konold et al.’s (2015) construct of data lenses. It was found that 54.0% of students made use of the given data to predict temperatures. Moreover, the students used all the lenses as they tried to make predictions. The most common lenses were found to be the case value lens and the pointer lens in their responses. However, participants were limited to a single school with a relatively high socioeconomic status.

Thus, recent research utilizes Konold et al.’s (2015) construct of data lenses, but further elaboration including which particular types of answers of informal prediction correspond to each data lens might be valuable for informing educators and researchers. In our study we categorized students’ answers to predictive reasoning tasks to the four data lenses and searched for further subcategories (types of answers).

## Research Questions

The present study was conducted in Greece. In our country, primary students receive mathematics homework frequently, which usually consists of exercises for repetition and practice of the content taught at school (Chaviaris & Kafoussi, 2009). Moreover, Greek parents believe that their role is to share responsibility with the teachers about their children learning mathematics (Kafoussi et al., 2020), thus adopting a “partnership-role focused construction for involvement” (Hoover-Dempsey & Jones, 1997). So, the participation of students in (classroom) and out-of-school (family) communities of practice was also taken into consideration in our study. As it has been stressed, schools need to take an active role in developing collaborative relationships among parents and teachers concerning school mathematics and learning at home is the most powerful type of cooperation between school and family concerning students’ performance in mathematics (Sheldon & Epstein, 2005). More specifically, learning at home concerns the provision of information and ideas to families about how to help students with homework.

In sum, this study extends the current research by considering the different types of answers corresponding to each data lens (Konold et al., 2015), as they emerged through the engagement of first graders in informal predictive reasoning in individual interviews before and after a teaching experiment. Moreover, it has taken into consideration the participation of the students in the family and school context, extending the research into including parental engagement in school mathematics through homework. Our research questions were the following:

1. Which types of answers emerge in each data lens as students engage in predictive reasoning prior to and after their participation in relevant activities?
2. Does the pattern of the data influence their predictions?
3. How is parents’ participation in school homework related to students’ data understanding?

## METHODS & PROCEDURES

### Context of the Study

In Greece, all students follow a national curriculum, and they have the same mathematics textbooks throughout the country. Statistics are not currently included in the mathematical content taught in the kindergarten school, the first and the second grades of primary school (ages four-eight years old). So, all students had no prior experience with statistical concepts.

Participants of the present research were 26 first-graders. The school was a typical public primary school situated in the suburbs of Athens, in Greece. The students came from middle-class families with most of their parents holding a university degree. The students of the research attended two clinical interviews and a classroom teaching experiment over a period of two weeks. The classroom teaching experiment was comprised of four 45 minutes of classroom lessons and took place near the end of the

**Table 1.** Phases of research

Phases of research	Tasks
Interview before teaching experiment	Task related to how students of a class spent their weekends (book reading, theater, playground, & bicycle) for three consecutive weekends. Students were invited to produce informal predictions for fourth weekend.
Teaching experiment: Introductory task	Task related to how students spend their school break (intermission). Students created data themselves by collecting & organizing data in a classroom survey.
Teaching experiment: 1 <sup>o</sup> task for school class	Students were given a table of data. Data presented how a certain set of students spent their break (playing football, playing chase, playing with cards, & not playing) for three consecutive days & students were asked to give their informal prediction for fourth day.
1 <sup>o</sup> homework	A similar task (activities during school break) with different data was given as homework.
Teaching experiment: 2 <sup>o</sup> task for school class	Task related to recycling in school (paper, plastic, & aluminum). Data set given described what material was recycled by a school class for three consecutive weeks & students were asked to produce predictions for fourth week.
2 <sup>o</sup> homework	A similar task (recycling materials) with different data was given as homework.
Interview after teaching experiment	Same task as interview before the teaching experiment was given.

school year (April-May 2022). It was considered that the students would have developed an early number sense by that time. At this point, we should note that, except of the classroom activities, two homework assignments were given to the students and their parents. So, we could discuss of a “classroom teaching experiment including homework”.

The students participated individually in two 15-minute task-based clinical interviews, which were conducted both prior to the teaching experiment and after the completion of the teaching experiment. They were withdrawn from the classroom for the interviews. The lessons and the interviews were audio-taped.

### Teaching Experiment

During the teaching experiment the students were engaged in two different tasks (see **Table 1**). In our research, we tried to use appropriate contexts in order to support them to develop their predictive reasoning. These tasks included elements of random variation without underlying causal forces, using “messy” data sources from classroom investigations (Oslington et al., 2020).

In the classroom, the students were initially presented with the task, they then worked in groups of four to six children (there were 6 groups in total in both classes) and in the end, they presented their predictions to the whole class. The groups were selected following the teachers’ advice in the two classes. The teachers were asked to create three groups of students for each class of first grade, ensuring that students with typical mathematical understanding would be represented both in the group with sophisticated mathematical understanding and in the group with low mathematical understanding. The researcher monitored students’ group activity and engaged in discussion during the group work. In **Table 1**, the phases and tasks of the research are presented.

The designing of the tasks supported inquiry-based learning and considered the importance of data collection, exploration of data and the interpretation of results, which were seriously dependent on context and at the introductory level involved limited formal mathematics as stated in GAISE report (Fraklin et al., 2007). The learning trajectory in the classroom began with young children’s investigations of the above tasks, progressing towards organizing, structuring, visualizing, and representing data. The learning trajectory developed into data modeling, involving the fundamental components of beginning inference, which included making predictions (Watson, 2006). At the end of the activity, the students had the opportunity to discuss the social issues emerging from the context of the tasks. Namely, they discussed social inclusion in the first task and environmental consciousness in the second task.

Moreover, the homework was designed by the researchers to engage parents at home in the learning of school mathematics. Parents were informed with handouts that they were expected to sit down with their child and record on the paper the exact words and phrasing of the student. In the introduction of the task a brief guideline on the completion of the task was given. The task was also accompanied by a questionnaire addressed to the family member helping the student. The questionnaire included questions asking the family member to share thoughts on his/ her experience such as “what was the most difficult part of the task?”, “share a student answer you found particularly interesting”. Overall, it was made clear to the parents involved that we were looking for the way students expressed their thinking rather than the right answer. The statistical tasks that were assigned as homework followed the tasks considered in the classroom. Both of the homework tasks had the same context as the school tasks with altered data values. Students were expected to produce prediction values for a given data set in cooperation with their parents. The present study is focused on the students’ interviews given individually before and after the classroom teaching experiment.

### Task of the Interview

The prediction task of the interview was designed to test students’ understanding of the range of the data given and their ability to use a number within the range of the data as their prediction. Each row of data values in the given table of data was selected so as to examine the extent to which students would consider:

- (1) the variability in the data values,
- (2) the presence of a prevailing data value,
- (3) the existence of a median value, and
- (4) the range of the data values (see **Table 2**).

The table of data values was accompanied by a pictogram of the data (see **Figure A1** in **Appendix A**).

**Table 2.** Task given to students during interview

Activity	First weekend	Second weekend	Third weekend	Next weekend
Playground	11	16	14	
Theater	1	1	1	
Book	5	3	4	
Bicycle	3	0	1	
Total	20	20	20	

Note. We asked 20 children attending first grade in nearby school how they spent their weekends. We wrote down on a table their activities for three consecutive weekends. Questions: (1) Can you predict how many children might play in playground next weekend? How did you think about it?; (2) Can you predict how many children might go to theater next weekend? How did you think about it?; (3) Can you predict how many children might read a book next weekend? How did you think about it?; & (4) Can you predict how many children might ride a bicycle next weekend? How did you think about it?

During the interviews prior to prediction questions, students were asked multiple questions testing their comprehension of the table of data (e.g., How many students read a book during the second weekend?). Most students were able to extract on their own the information needed from the data table and some students were given some help from the researcher. In the beginning of the interview, the students constructed on their own (and in some cases with the help of the researcher) a similar data table, while discussing the meaning of the data values in each row and column and in a similar pictogram. Furthermore, during the interviews in all four questions, we used the word “prediction” as in “Can you predict a number for the fourth week” together with equivalent expressions such as “Can you say a number for the fourth week”, so that by the middle half of the interview and during the teaching experiment it would be taken as shared (Stephan & Akyuz, 2022) that prediction entails using the available data to suggest a possible data value appearing in the future.

The students were presented with the same task both in the interview prior to the classroom lesson and in the interview following the classroom lesson. Interviews were conducted by the first author in an adjacent room to the classroom, away from other students. The task given during the interviews prior to and after the teaching session was the following (Table 2).

### Data Analysis

Our data derived from the students’ answers in the individual interviews before and after the teaching experiment. The data analysis was based on qualitative methods. Initially, the collected data were analyzed with the method of the content analysis (Bryman, 2016), concerning the four lenses of data understanding (Konold et al., 2015). Then, the data were subjected to thematic analysis (Boyatzis, 1998; Braun & Clarke, 2012), concerning the coding of the students’ answers in categories/types for each lens. More specifically, initially the students’ answers in the interviews before and after the teaching experiment were transcribed by the first author. Then all the answers of every student were categorized according to the four data lenses (Konold et al., 2015). Subsequently, all the answers categorized as data lens 1 from all students, corresponding to questions 1, 2, 3, and 4 from the interview, were gathered together and examined as to whether they could be grouped and labeled into different types of answers. This procedure was followed for each lens of data understanding. Each categorization in the four lenses and subsequent grouping of the answers into distinct type of answers was considered by both researchers and cross referenced for validity.

According to the thematic analysis, seven types of answers (random numbers, repeating previous prediction, operations using elements of the data, counting from the picture of the problem, personal experiences, noticing the numbers, and reporting of a story) were produced in the first data lens (data as pointers), three types of answers (isolation of a data value, isolation of a data value, and producing prediction by naming the next or previous number in the number line, finding the nearby tenth) in the second data lens (data as case values), four types of answers (prediction value within the range of data values, dominating value, prediction value near the range of data values, and median value) in data lens 3 (data as classifiers) and one type (taking into account the entire distribution of values) of answer in data lens 4 (data as an aggregate). For the creation of the types of answers the first step of analysis was for both authors to categorize independently students’ answers in the four data lenses. The initial categorization was cross referenced. Secondly the answers in each data lens were coded inductively to find types of answers within each data lens. Next, two authors met to discuss results of the first analysis, and they developed codes to encompass the types of answers that were evident across the whole data set. Types of answers, that is the results of the thematic analysis, will be presented and explored later, which includes students’ responses that explicate these types and highlight contrast between them.

Moreover, the students’ answers were also categorized into two groups, group A and group B. Group A consists of the answers of 18 students who during the teaching experiment completed the two homework tasks assigned with the help of their parents two times. Group B consists of the answers of eight students that did not fill out the homework assignments. Then, we analyzed if there were differences between the students’ answers of these two groups according to the four lenses as well as concerning the types of answers in each lens before and after the teaching experiment. This procedure was followed for each question of the interview (Q1, Q2, Q3, and Q4). For the presentation of the findings and the conclusions descriptive statistics (percentages and frequency distribution) were applied.

## RESULTS

In this section we present initially the different types of students’ answers that were produced in the four distinct data lenses before and after the teaching experiment with illustrative examples, using codes for the students (for example the code ST1 stands for the answer of student 1).

Then, for each lens, we present the types of answers formed in each of the four questions of the interview (Q1, Q2, Q3, and Q4) for group A and group B. Finally, we present the total number of answers in each level of data understanding.

### An Overview of Types of Answers That Emerged in Each Data Lens

#### Data lens 1

It was found that seven distinct types of answers emerged in data lens 1 (data as pointers). These types were named, as following:

1. Type of answer 1: Random numbers
2. Type of answer 2: Repeating previous prediction
3. Type of answer 3: Operations using elements of the data
4. Type of answer 4: Counting from the picture of the problem
5. Type of answer 5: Personal experiences
6. Type of answer 6: Noticing the numbers
7. Type of answer 7: Reporting of a story

Subsequently, we present an illustrative example of student answer for each of the seven types of answers.

1. In the type of answer 1 the students mentioned random numbers as their prediction value, without considering the numbers of the data. For example in question 1 (**Table 2**), where the student (ST24) was asked to predict how many children might play in playground the next weekend, when we know that 11 played the first weekend, 16 played the second weekend and 14 played the third weekend, the student replied “63” and added that: “I guessed it”.
2. In type 2 the students would repeat a number that had been a previous prediction value. An example of this type of answer in data lens 1 is the following answer of the student (ST23) in question 4. The student was asked to predict how many children might ride a bicycle next week when it is known that three rode on the first weekend, 0 on the second and one on the third. The student said: “two, because I want to put down two (he used two as a prediction value in all the previous questions)”.
3. Type 3 included answers that were based on the iteration of known sums or the sum of the data value. For example the prediction value of the student (ST17) in question 2, where she was producing a prediction of how many children might go to the theater next weekend, when it is known that 1 child went to the theater the first, the second and the third weekend was “3”, and she justified her answer: “three, because one and one and one is three”.
4. Type 4 included answers, where students counted images from the picture of the problem. In this type of answer an example is the answer of a student (ST18) in question 3, where the data values of previous weekends were five, three, four, and the student said: “five, because I counted the children in the picture (referring to the picture accompanying the task)”.
5. The type 5 of answers was formed by students’ answers that would recall a personal experience within the experiential environment of the given problem (for example the playground). An example of this type of answer is the answer of the student (ST22) that explains her prediction values in question 1 regarding how many children might play in the playground next weekend. The student provides the following answer: “one, because during weekends it is only me in the playground”.
6. In the answers categorized as type 6 in data lens 1 students would concentrate on picking up a number that was not already given as a data value. For example in question 1, when students are asked for a prediction value of children playing in the playground next weekend and the data values for previous weekends are 11, 16, and 14 a student (ST25) answered: “two, because we see all numbers (in the row in question) except from two”.
7. Type 7 corresponded to answers including a story, where students explain their answer with a story (for example saying they would not want children in the problem being crowded, so they would pick a small number). This particular answer was given by the following student (ST14) in question 3, regarding a prediction value for children that might read a book next weekend, when data values for previous weekends are five, three, and four: “I will put down one, because five and three is eight (and four), is 12, so I will write down one, because I do not want them to get crowded”.

#### Data lens 2

Data lens 2 included the answers that perceived data as case values. The answers of this data lens were further grouped into three types:

1. Type of answer 1: Isolation of a data value
2. Type of answer 2: Isolation of a data value and producing prediction by naming the next or previous number in the row
3. Type of answer 3: Finding nearby tenth

More specifically:

1. Type 1 included the answers that were based on singling out one number out of the data set and naming it as the prediction value. For example in question 1 students were asked to produce a prediction value on how many children might play in the playground next weekend, when the data values on previous weekends are 11, 16, and 14. One student (ST1) said: “16, because I saw it here (points to second weekend)”.



**Table 3.** Types of answers emerged in lens 1 (data as pointers) before teaching experiment

Type of answer-Question1	Type of answer-Question2	Type of answer-Question3	Type of answer-Question4
Type 1: Group A: n=1 & group B: n=0	Type 1: Group A: n=4 & group B: n=1	Type 1: Group A: n=1 & group B: n=2	Type 1: Group A: n=1 & group B: n=2
	Type 3: Group A: n=4 & group B: n=2	Type 3: Group A: n=6 & group B: n=1	Type 3: Group A: n=7 & group B: n=1
Type 5: Group A: n=2 & group B: n=0	Type 5: Group A: n=1 & group B: n=0		
Type 6: Group A: n=3 & group B: n=1		Type 7: Group A: n=1 & group B: n=0	

- The second type of answer was the one that singled out one of the data values and chose the next number in the row as the prediction value. An example of this type of answer in question 1 would be the answer of the student (ST4) that said: "15, from 14 (it is after 14)".
- The third type of answer was the one based on finding the nearby tenth. An example of this type of answer in question 1 is the answer of the student (ST20) that replied: "10, because 11, 16, and 14 are all above 10".

### Data lens 3

In the answers categorized as data lens 3, data is perceived as classifiers. Four types of answers were constituted in data lens 3. The types of answers were the following:

- Type of answer 1: Prediction value within the range of data values
  - Type of answer 2: Dominating value
  - Type of answer 3: Prediction value near the range of data values
  - Type of answer 4: Median value
- The type 1 of answers was formed by the students that proposed a prediction value within the range of data values. For example in question 1, where the data values of the previous weekends were 11, 16, and 14, a student (ST23) provided the following reasoning for her prediction value: "15, because it goes 11, 16, 14, and 15".
  - In type 2 the students observed the dominating value in the data set and suggested it as the prediction value. An example of this type is the answer of the student (ST4) that in question 2, where the data values of all previous weekends is one, said: "one, because the first weekend there is one child, the second weekend there is one child, the third one child, so the fourth it will be one".
  - Type 3 included answers that mentioned a trend in the data set and their prediction value was close to the range of the data set. The following student (ST10) replied for example in question 3, where the student gives a prediction value for the next weekend regarding the number of children that might read a book, when data values for previous weekends are five, three, and four that: "six, because it is five, four, five, and six. So, three, four, five, and six".
  - The last type of answers named the median of the data values as their prediction value. An example is the answer to question 3 (Table 2) given by the following student (ST25) was: "four, because five, three, and four. Before five is four and after three is four, so I thought it was four".

### Data lens 4

Finally, in lens 4, understanding data as an aggregate, only one type of answer emerged during the interviews in which students would consider the data set as a whole having as their concern to preserve it (named: considering the entire distribution of values). More particularly, the students would consider the numbers in the row and prediction column in the data table. It was recorded that answers of data lens 4 were documented only in the fourth question and only by students that in all previous questions gave answers of data lens 3. For example, we present the answer of a student (ST4) that in the previous questions gave the prediction values 15, one, and three, respectively and answered to the fourth question: "one, because 15 and one is 16, and three, 19 and one, 20".

### An Analysis of Types of Students' Answers in Each Question for Both Groups

The types of answers of group A and group B appear in tables, before and after the teaching experiment for each question in each data lens. In each type of answer the number of students' answers also appears (e.g., n=4).

### Data lens 1

Table 3 and Table 4 present the types of answers in the different questions in the first data lens before and after the teaching experiment, respectively.

As shown in Table 3 and Table 4, the types of answers in the different questions in the first lens were not the same before and after the teaching experiment. Before the teaching experiment, the type of answer "random numbers" (type 1) exists in all questions in group A. Furthermore, we noticed that the type of answer "operations using elements of the data" (type 3) was present in three of the four questions (Q2, Q3, and Q4). Interestingly, the types of answers included in the first lens were reduced in three of the four questions after the teaching experiment in group A. More specifically, in question 1 there is no type of answer of lens 1, and in question 2 there is only one type of answer ("random numbers").

**Table 4.** Types of answers emerged in lens 1 (data as pointers) after teaching experiment

Type of answer-Question1	Type of answer-Question2	Type of answer-Question3	Type of answer-Question4
	Type 1: Group A: n=1 & group B: n=0	Type 1: Group A: n=1 & group B: n=0	Type 1: Group A: n=1 & group B: n=0 Type 2: Group A: n=1 & group B: n=0
Type 3: Group A: n=0 & group B: n=1	Type 3: Group A: n=0 & group B: n=1		Type 3: Group A: n=1 & group B: n=0
Type 4: Group A: n=0 & group B: n=1		Type 4: Group A: n=1 & group B: n=1	Type 4: Group A: n=0 & group B: n=1

**Table 5.** Types of answers emerged in lens 2 (data as pointers) before teaching experiment

Type of answer-Question1	Type of answer-Question2	Type of answer-Question3	Type of answer-Question4
Type 1: Group A: n=1 & group B: n=1	Type 1: Group A: n=1 & group B: n=0	Type 1: Group A: n=2 & group B: n=0	Type 1: Group A: n=2 & group B: n=0
Type 2: Group A: n=6 & group B: n=3	Type 2: Group A: n=4 & group B: n=2	Type 2: Group A: n=2 & group B: n=3	Type 2: Group A: n=4 & group B: n=2
Type 3: Group A: n=2 & group B: n=1		Type 3: Group A: n=1 & group B: n=0	

**Table 6.** Types of answers emerged in lens 2 (data as pointers) after teaching experiment

Type of answer-Question1	Type of answer-Question2	Type of answer-Question3	Type of answer-Question4
Type 1: Group A: n=2 & group B: n=0		Type 1: Group A: n=1 & group B: n=0	Type 1: Group A: n=2 & group B: n=0
Type 2: Group A: n=2 & group B: n=2	Type 2: Group A: n=2 & group B: n=1	Type 2: Group A: n=1 & group B: n=1	Type 2: Group A: n=0 & group B: n=1

In the third question appeared also type 4, that is “counting from the picture of the problem”. In the fourth question appeared also an answer of type 2 (“repeating previous prediction”) and an answer of type 3 (“operations using elements of the data”).

In group B, before the teaching experiment, the type of answer “random numbers” (type 1) and the type of answer “operations using elements of the data” (type 3) appeared in three questions (Q2, Q3, and Q4). After the teaching experiment, type 4 (“counting from the picture of the problem”) appeared in three questions (Q1, Q3, and Q4). A viable interpretation of this practice might be that students recalled their experience of counting data images for the production of pictograms during the teaching experiment and applied this counting images experience during the interview. Moreover, there is one answer in question 1 and in question 2 for type 3.

In sum, the types of answers 1 and 3 were the most frequent in lens 1, that is data as pointers, for both groups before and after the teaching experiment. There were 15 answers of type 1 and 24 answers of type 3 for both groups.

### Data lens 2

The types of answers that emerged in the second lens of data in each question are presenting in **Table 5** and **Table 6**.

The answers grouped in the second lens of data understanding included two types before the teaching experiment for all questions: “isolation of a data value” (type 1) and “isolation of a data value and producing the prediction by naming the next number in the row” (type 2) in group A. In the first and third question one more type of answer emerged, that of type 3 (“the finding of the nearby tenth”). After the teaching experiment, the types of answers concerning lens 2 were reduced to the two types (type 1 and type 2) in group A. In the second question, the answers were only of type 2, whereas in question 4 the answers were only of type 1.

In group B, the answers corresponding to lens 2 before the teaching experiment included type 2 in all questions. In the first question, two more types of answers emerged (type 1 and type 3). The answers produced after the teaching experiment in group B formed only type 2 in all four questions.

Based on the above results, the type of answer 2 (“isolation of a data value and producing the prediction by naming the next number in the row”) was the most frequent in lens 2, adding the answers of both groups before and after the teaching experiment (n=36).

### Data lens 3

**Table 7** and **Table 8** present the types of answers that emerged in the third lens of data (data as classifiers).

In group A the answers produced before the teaching experiment at the third lens of data formed one type in each question. In question 1 and question 4 the answers formed type 1 (“prediction value within the range of data values”), in question 2 they formed type 2 (“dominating value”), and in question 3, type 3 (“a prediction value near the range of data values”). Furthermore, in questions 1, 2, and 4 the same type of answers was repeated in the interviews before and after the teaching experiment. In question 3 the answers given after the teaching experiment were grouped into two more types: type 1 and type 4 (“median value”).

In group B considering the types of answers of students that did not complete the homework we note that there were no answers categorized as interpreting data as classifiers (data lens 3) for question 1 before the teaching experiment. For the rest of



**Table 7.** Types of answers emerged in lens 3 (data as pointers) before teaching experiment

Type of answer-Question1	Type of answer-Question2	Type of answer-Question3	Type of answer-Question4
Type 1: Group A: n=2 & group B: n=0			Type 1: Group A: n=4 & group B: n=1
	Type 2: Group A: n=4 & group B: n=2		
		Type 3: Group A: n=4 & group B: n=0	

**Table 8.** Types of answers emerged in lens 3 (data as pointers) after teaching experiment

Type of answer-Question1	Type of answer-Question2	Type of answer-Question3	Type of answer-Question4
Type 1: Group A: n=14 & group B: n=3		Type 1: Group A: n=4 & group B: n=0	Type 1: Group A: n=8 & group B: n=3
	Type 2: Group A: n=15 & group B: n=6	Type 3: Group A: n=4 & group B: n=3	
		Type 4: Group A: n=5 & group B: n=0	

**Table 9.** Data lenses before & after teaching experiment in question 1

Lenses	1 <sup>st</sup> lens	2 <sup>nd</sup> lens	3 <sup>rd</sup> lens	4 <sup>th</sup> lens	No answer
Before	Group A: n=6 & group B: n=1	Group A: n=9 & group B: n=5	Group A: n=2 & group B: n=0	-	Group A: n=1 & group B: n=2
After	Group A: n=0 & group B: n=2	Group A: n=4 & group B: n=2	Group A: n=14 & group B: n=3	-	Group A: n=0 & group B: n=1

the questions the types of answers in group B were the same as group A. The answers for question 1 given after the teaching experiment are of type 1 (“the prediction value is within the range of data values”). The answers related to questions 2, 3, and 4 formed the same type of answer in both interviews. Specifically in question 2 the answers were of type 2 (“dominating value”). The answers to question 3 were of type 3 (“prediction value near the range of data values”) in the interviews given before the teaching experiment and of the type 1 (“prediction value within the range of data values”), type 3 (“prediction value near the range of data values”) and type 4 (“median value”) after the teaching experiment.

Comparing the types of answers that emerged in the perspective of interpreting data as classifiers between group A and group B, we note that the same types of answers were developed, with the exception of question 3. In group A, answers related to the question 3 can be divided into three types (type 1: “prediction value within the range of data values”, type 3: “prediction value near the range of data values”, type 4: “median value”). In contrast, in group B answers given in question 3 formed only type 3 (“prediction value near the range of data values”).

The most frequent types of answers in this lens for both groups were type 1 (“prediction value within range of data values”) and type 2 (“dominating value”). Total number of students producing answers of type 1 and type 2 were n=39 & n=27, respectively.

**Data lens 4**

Finally, in the fourth data lens in group A no student answer was corresponding to question 1, 2, and 3 given before and after the teaching experiment. In the answers given after the teaching experiment to question 4 one type of answer emerges, where some students considered the data set as a whole (n=5). The students giving this type of answer produced their prediction by summing the values of their previous predictions and calculating the difference from the value of the whole data set. The students interpreting data as an aggregate in question 4 were concerned with maintaining the total value of the data set stable and producing a prediction value within the range of the data values. We have to consider that maintaining the stability of the whole data set never occurred in the answers given to questions 1, 2, and 3. In group B there were no answers of the fourth perspective in all four questions. This fact possibly points to the conclusion that parental involvement in the homework produced by group A is maybe related to students producing answers of the fourth perspective in the interviews after the teaching experiment.

**Comparing Answers in Each Lens in Total for Both Groups**

We offer insight into the total number of answers in each level of data understanding, before and after the classroom lessons for each question in both groups. Group A consists of answers of 18 students and group B consists of the answers of eight students.

Concerning the first question, **Table 9** shows that most of the students’ answers 50.0% (n=9) of group A during the interview before the classroom lessons were classified in the second level of data understanding. Moreover, 33.0% (n=6) of the answers were of the first level of data understanding (data as pointers) and 11.0% (n=2) of the answers were characterized as the third level of data understanding (data as classifier). The majority of the answers after the classroom teaching experiment 78.0% (n=14), were of the third level of data understanding and a small percent of answers 22.0% (n=4) were categorized in the second level of data understanding. We observed that there were no answers in the post-classroom lessons interviews ranked in the first level of data understanding. In group B the results were almost the same, as we notice an increase in the number of answers corresponding to the third perspective (lens) (“interpreting data as classifier”) between the answers given before and after the teaching experiment. More specifically, the percentages of answers categorized as interpreting data as classifier for the first question after the teaching experiment was 37.5% (n=3), although there was no answer corresponding to this perspective before the teaching experiment.

**Table 10.** Data lenses before & after teaching experiment in question 2

Lenses	1 <sup>st</sup> lens	2 <sup>nd</sup> lens	3 <sup>rd</sup> lens	4 <sup>th</sup> lens	No answer
Before	Group A: n=9 & group B: n=3	Group A: n=5 & group B: n=2	Group A: n=4 & group B: n=2	-	Group A: n=0 & group B: n=1
After	Group A: n=1 & group B: n=1	Group A: n=2 & group B: n=1	Group A: n=15 & group B: n=6	-	-

**Table 11.** Data lenses before & after teaching experiment in question 3

Lenses	1 <sup>st</sup> lens	2 <sup>nd</sup> lens	3 <sup>rd</sup> lens	4 <sup>th</sup> lens	No answer
Before	Group A: n=8 & group B: n=3	Group A: n=5 & group B: n=3	Group A: n=4 & group B: n=1	-	Group A: n=1 & group B: n=1
After	Group A: n=2 & group B: n=2	Group A: n=2 & group B: n=1	Group A: n=13 & group B: n=3	-	Group A: n=1 & group B: n=2

**Table 12.** Data lenses before & after teaching experiment in question 3

Lenses	1 <sup>st</sup> lens	2 <sup>nd</sup> lens	3 <sup>rd</sup> lens	4 <sup>th</sup> lens	No answer
Before	Group A: n=8 & group B: n=3	Group A: n=6 & group B: n=2	Group A: n=4 & group B: n=1	-	Group A: n=0 & group B: n=2
After	Group A: n=3 & group B: n=2	Group A: n=2 & group B: n=2	Group A: n=8 & group B: n=3	Group A: n=5 & group B: n=0	Group A: n=0 & group B: n=1

**Table 10** presents a concise view of the grouping of answers in levels of data understanding for the second question. As shown in **Table 10** of question 2, concerning group A, 50.0% of the answers (n=9) before the classroom lessons were characterized as level 1 of data understanding, whereas 83.0% of the answers (n=15) given after the classroom lessons were associated with level 3 of data understanding. Level 2 of data understanding included 28.0% (n=5) of the answers before and 11.0% (n=2) of the answers after the lessons. Only 6.0% (n=1) of the answers were deemed to be in the first level of data understanding after the teaching experiment. Concerning group B, the percentage of answers tripled from 25.0% (n=2) before the teaching experiment to 75.0% (n=6) after the teaching experiment about the level 3 of data understanding.

**Table 11** presents the grouping of answers in levels of data understanding for the third question. Regarding question 3, the majority of the answers 44.0% (n=8) before the lessons were characterized as level 1 of data understanding, whereas the overwhelming majority of the answers 72.0% (n=13) given after the lessons were found to be of level 3 for group A. For group B, an increase in percentages occurred in question number 3, where answers before the teaching experiment were 12.5% and raised to 37.5% after the teaching experiment in the third level.

Finally, **Table 12** presents the grouping of answers in levels of data understanding for the fourth question. For question 4 the corresponding **Table 12** shows that before the teaching experiment, 44.0% (n=8) of the answers were of the first data lens, whereas after the teaching experiment, the same percentage was characterized as being of the third data lens for group A. After the teaching experiment, 17.0% (n=3) of the answers were in data lens 1 and 11.0% (n=2) in data lens 2. Concerning group B, there was also an increase concerning the third data lens (from 12.5% to 37.5%). However, answers in data lens 4 (data as an aggregate) were produced only in question 4 and only by the group of students that also completed the homework assigned after the teaching experiment. Specifically, the percentage of answers categorized as lens 4 was 28.0% (n=5) in group A and none in group B. This fact led us to believe that parental participation in school homework is related to students' data comprehension.

Summarizing, considering the answers of group A, we notice an increase in the number of answers corresponding to the third perspective (lens) (interpreting data as classifier) between the answers given before and after the teaching experiment. For example, in group A answers in question 1 classified as "interpreting data as classifier" increased from 11.0% before the teaching experiment to 78.0% after the teaching experiment and this finding concerns all four questions. Moreover, the engagement of students in classroom predictive reasoning activities influenced their data understanding, substantially decreasing the percentages of answers characterized as data lens 1 in all four questions in group A. For example, in question 1 from 33.0% before the teaching experiment, there were no answers corresponding to lens 1 after the teaching experiment. It was also of interest the fact that, although for question number 4 the percentage of answers characterized as interpreting data as an aggregate (fourth perspective) was zero before the teaching experiment, it was raised to 28.0% after the teaching experiment. Moreover, summarizing for group B, we should note that a decrease in percentages was also observed concerning the answers of data lens 1, with the exception of question 1, where there was actually an increase (from 12.5% before the teaching experiment to 25.0% after the teaching experiment). For the rest of the questions, there was a decrease in percentages concerning the answers of data lens 1 (for question 2: from 37.5% to 12.5% and for question 3 and question 4: from 37.5% to 25.0%).

## DISCUSSION & CONCLUDING REMARKS

The development of statistical reasoning is recommended by contemporary curricula in mathematics education (Bargagliotti et al., 2020; Franklin et al., 2007; National Council of Teachers of Mathematics [NCTM], 2000). It is vital that students in their early school years are given the learning opportunity to develop mathematical experiences in statistics. This study extended existing research in informal predictive reasoning in primary school by examining the types of answers that emerge when first-grade

students engage in informal predictive reasoning activities. More specifically, our present research extended Konold et al. (2015) framework into considering which particular type of answers emerge into each data lens.

According to our results, concerning the first research question, it was found that during the interviews there were seven distinct types of answers corresponding to lens 1 of data understanding. The most frequent answers were either type 1 (“random numbers”) or type 3 (“operation using elements of the data”). In lens 2 of data understanding 3 types of answers were found. The most common type of answer was type 2 (“isolation of a data value and producing the prediction value by naming the next number in the row”). Data lens 3 (understanding the data as classifiers) consisted of 4 types of answers. The most recurrent answers were either of type 1 (“prediction value within the range of data values”) or of type 2 (“the dominating value used as the prediction value”). Finally, in data lens 4 (understanding data as aggregate) there was one type of answer constituted. In this type of answer students would understand the entire distribution of values as the perceptual unit considering the variability within the given data and the preservation of the whole data set.

Observing the distribution of students’ answers in the different questions of the interview, we could conclude that the greater shift in lenses between the answers produced before and after the teaching experiment occurred in the first question. During the interviews before the teaching experiment, it was noted that the percentage of students producing answers of lens 3 was smaller in question number 1 in contrast to questions 2, 3, and 4. Maybe, before the teaching experiment students had difficulty considering the range of the data value. After the teaching experiment, students appear to consider the range of the data value so as to come up with a prediction value within the range of the data available. More specifically, in group A, there were no answers corresponding to the first lens, since all the answers had shifted to more sophisticated lenses. In group B, although there is no answer appearing to interpret data as classifier (third lens) in the first question before the teaching experiment, after the teaching experiment 37.5% of all answers produced were characterized as interpreting data as classifier. So, concerning the second research question, which explored how the pattern of the given data is related to the data lens of students’ answers, it was demonstrated that students have more difficulty producing a prediction when the given data values are sparse in a wide range instead of a small range.

It should be noted that, although data lens 4 is related to the aggregate, the shift of students’ answers to data lens 2 and data lens 3 is considered important with regard to the development of statistical reasoning. According to Konold et al. (2015), “it would be a serious mistake if educators saw as the major instructional goal to move students quickly past interpreting data as pointers, case values and classifiers” (p. 323). The overall shift of student answers to data lens 3 is considered to be an important indication that students participated in a meaningful learning experience related to the development of data understanding.

Regarding the third research question concerning the relation between students’ data understanding and parental participation, it was found that exclusively students that had completed the homework assignment (group A) were able to produce answers of data lens 4 (Michalopoulou & Kafoussi, 2024). This fact underlies the crucial role of the parents in forming the mathematical experience of the student in the early school years (e.g., Galindo & Sheldon, 2012; Sheldon & Epstein, 2005) and it supports the claim that the role of the parents is “crucial in shaping their children’s cognitive and affective relationship with mathematics” (Kafoussi et al., 2020). However, it is important to understand how the participation of students in communities such as the family affects their understanding of statistics. A detailed examination of the interactions during the homework should be considered. The findings of the present research will be further examined by the analysis of first-grade students’ answers at home.

In general, our results suggest that students participating benefited from their experience through the teaching experiment and developed their data understanding. In accordance with relevant research (English, 2012; Oslington et al., 2020) it was concluded that predictive reasoning constitutes an enriching experience for young learners to work with data and reason statistically. Our study provides an opportunity to overview the types of answers in each data lens that might emerge when students are engaged in predictive reasoning. It should be emphasized that it was limited to a single school with middle socio-economic status. Further research into predictive reasoning should be conducted focusing on ways to engage students in viewing data as an aggregate (Ben-Zvi et al., 2018). Moreover, as students participate simultaneously in different communities of practice, further research into how these communities (in particular school and family) are connected, would establish a meaningful way to understand students’ mathematical reasoning.

**Author contributions:** All authors have sufficiently contributed to the study and agreed with the results and conclusions.

**Funding:** No funding source is reported for this study.

**Ethical statement:** The authors stated that the study was approved by the Ethics Committee at University of the Aegean with Protocol Number: 200/20-01-2022 on 20 January 2022. Written informed consents were obtained from the participants.

**Declaration of interest:** No conflict of interest is declared by authors.

**Data sharing statement:** Data supporting the findings and conclusions are available upon request from the corresponding author.

## REFERENCES

- Aridor, K., & Ben-Zvi, D. (2017). The co-emergence of aggregate and modelling reasoning. *Statistics Education Research Journal*, 16(2), 38-63. <https://doi.org/10.52041/serj.v16i2.184>
- Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. (2020). *Pre-K-12 guidelines for assessment and instruction in statistics education II (GAISE II)*. American Statistical Association.

- Ben-Zvi, D., & Garfield, J. (2004). *The challenge of developing statistical literacy, reasoning and thinking*. Kluwer Academic Publishers. <https://doi.org/10.1007/1-4020-2278-6>
- Ben-Zvi, D., Gravemeijer, K., & Ainley, J. (2018). Design of statistics learning environments. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 473-502). Springer. [https://doi.org/10.1007/978-3-319-66195-7\\_16](https://doi.org/10.1007/978-3-319-66195-7_16)
- Boonk, L., Gijssels, H. J. M., Ritzen, H., & Brand-Gruwel, S. (2018). A review of the relationship between parental involvement indicators and academic achievement. *Educational Research Review*, 24, 10-30. <https://doi.org/10.1016/j.edurev.2018.02.001>
- Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. SAGE.
- Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper (Ed.), *APA handbook of research methods in psychology: Research designs* (pp. 57-91). American Psychological Association. <https://doi.org/10.1037/13620-004>
- Bryman, A. (2016). *Social research methods*. Oxford.
- Chaviaris, P., & Kafoussi, S. (2009). Epikinonia Skholiou-Ikoyenias Meso Phillon Ergasias sta Mathimatika: Mia Meleti Periptosis [Communication between school and family through worksheets in mathematics: a case study]. In F. Kalavasis, S. Kafoussi, M. Chionidou, C. Skoumpourdi, & G. Fesakis (Eds.), *Proceedings of the 3rd Hellenic Conference of the Greek Association Research in Mathematics Education* (pp. 181-191). GARME.
- Civil, M., & Bernier, E. (2006). Exploring mages of parental participation in mathematics education: Challenges and possibilities. *Mathematical Thinking & Learning*, 8(3), 309-330. [https://doi.org/10.1207/s15327833mtl0803\\_6](https://doi.org/10.1207/s15327833mtl0803_6)
- Cobb, P. (1999). Individual and collective mathematical learning: The case of statistical data analysis. *Mathematical Thinking and Learning*, 1, 5-44. [https://doi.org/10.1207/s15327833mtl0101\\_1](https://doi.org/10.1207/s15327833mtl0101_1)
- Crafter, S. (2012). Parental cultural models and resources for understanding mathematical achievement in culturally diverse school settings. *Educational Studies in Mathematics*, 81(1), 31-46. <https://doi.org/10.1007/s10649-011-9359-5>
- Darragh, L. (2016). Identity research in mathematics education. *Educational Studies in Mathematics*, 93(1), 19-33. <https://doi.org/10.1007/s10649-016-9696-5>
- Doerr, H., Delmas, R., & Makar, K. (2017). A modelling approach to the development of students' informal inferential reasoning. *Statistical Educational Research Journal*, 16(2), 86-115. <https://doi.org/10.52041/serj.v16i2.186>
- English, L. D. (2013). Reconceptualizing statistical learning in the early years. In L. English, & J. Mulligan (Eds.), *Reconceptualizing early mathematical learning* (pp.67-82). Springer. <https://doi.org/10.1007/978-94-007-6440-8>
- English, L. D. (2012). Data modeling with first-grade students. *Educational Studies in Mathematics*, 81, 15-30. <https://doi.org/10.1007/s10649-011-9377-3>
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-K-12 curriculum framework*. American Statistical Association.
- Gal, I. (2002). Adult's statistical literacy: Meaning, components, responsibilities. *International Statistical Review*, 70(1), 1-25. <https://doi.org/10.2307/1403713>
- Galindo, C., & Sheldon, S. (2012). School and home connections and children's kindergarten achievement gains: The mediating role of family involvement. *Early Childhood Research Quarterly*, 27(1), 90-103. <https://doi.org/10.1016/j.ecresq.2011.05.004>
- Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education*, 10(3). <https://doi.org/10.1080/10691898.2002.11910676>
- Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, 75(3), 372-396. <https://doi.org/10.1111/j.1751-5823.2007.00029.x>
- Hoover-Dempsey, K. V., & Jones, K. P. (1997). *Parental role construction and parental involvement in children's education* [Paper presentation]. The Annual Meeting of the American Educational Research Association.
- Kafoussi, S., Chaviaris, P., & Moutsios-Rentzos, A. (2020). Investigating parental influences on sixth graders' mathematical identity in Greece: A case study. *International Electronic Journal of Mathematics Education*, 15(2), em0572, <https://doi.org/10.29333/iejme/6279>
- Knapp, A., Landers, R., Liang, S., & Jefferson, V. (2017). We all as a family are graduating tonight: A case for mathematical knowledge for parental involvement. *Educational Studies in Mathematics*, 95(1), 79-95. <https://doi.org/10.1007/s10649-016-9741-4>
- Konold, C., Higgins, T., Russell S. J., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics*, 88, 305-325. <https://doi.org/10.1007/s10649-013-9529-8>
- Lehrer, R., & English, L. (2017). Introducing children to modelling variability. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *The international handbook of research in statistics education* (pp. 229-260). Springer. [https://doi.org/10.1007/978-3-319-66195-7\\_7](https://doi.org/10.1007/978-3-319-66195-7_7)
- Makar, K. (2014). Young children's explorations of average through informal inferential reasoning. *Educational Studies in Mathematics*, 86(1), 61-78. <https://doi.org/10.1007/s10649-013-9526-y>
- Makar, K. (2016). Developing young children's emergent inferential practices in statistics. *Mathematical Thinking and Learning*, 18(1), 1-24. <https://doi.org/10.1080/10986065.2016.1107820>
- Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, 8, 82-105. <https://doi.org/10.52041/serj.v8i1.457>

- Makar, K., & Rubin, A. (2017). Learning about statistical inference. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *The international handbook of research in statistics education* (pp. 261-294). Springer. [https://doi.org/10.1007/978-3-319-66195-7\\_8](https://doi.org/10.1007/978-3-319-66195-7_8)
- Maloney, E., Ramirez, G., Gunderson, E. A., Levine, S., & Beilock, S. (2015). Intergenerational effects of parents' math anxiety on children's math achievement and anxiety. *Psychological Science*, 26(9), 1480-1488. <https://doi.org/10.1177/0956797615592630>
- McClain, K., & Cobb, P. (2001). Supporting students' ability to reason about data. *Educational Studies in Mathematics*, 45, 103-129. <https://doi.org/10.1023/A:1013874514650>
- Meletiou-Mavrotheris, M. & Paparistodemou, E. (2015.) Developing students' reasoning about samples and sampling in the context of informal inferences. *Educational Studies in Mathematics*, 88, 385-404. <https://doi.org/10.1007/s10649-014-9551-5>
- Michalopoulou, A., & Kafoussi, S. (2024). Engaging first grade students in statistical predictive reasoning. *Quaderni di Ricerca in Didattica (Proceedings of CIEAEM74)* (in press).
- NCTM. (2000). *Principles and standards for school mathematics*. National Council of Teachers of Mathematics.
- Oslington, G., Mulligan, J., & Van Bergen, P. (2020). Third-graders' predictive reasoning strategies. *Educational Studies in Mathematics*, 104, 5-24. <https://doi.org/10.1007/s10649-020-09949-0>
- Paparistodemou, E., & Meletiou-Mavrotheris, M. (2008). Developing young students' informal inference skills in data analysis. *Statistics Education Research Journal*, 7(2), 83-106. <https://doi.org/10.52041/serj.v7i2.471>
- Piaget, J. (1954). *The construction of reality in the child*. Basic Books. <https://doi.org/10.1037/11168-000>
- Quaye, J., & Pomeroy, D. (2022). Social class inequalities in attitudes towards mathematics and achievement in mathematics cross generations: A quantitative Bourdieusian analysis. *Educational Studies in Mathematics*, 109(1), 155-175. <https://doi.org/10.1007/s10649-021-10078-5>
- Shaughnessy, J. M. (2007). *Research on statistics learning and reasoning*. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957-1010). National Council of Teachers of Mathematics.
- Sheldon, S., & Epstein, J. (2005). Involvement counts: Family and community partnerships and mathematics achievement. *The Journal of Educational Research*, 98(4), 196-207. <https://doi.org/10.3200/JOER.98.4.196-207>
- Silinkas, G., & Kikas, E. (2019). Parental involvement in math homework: Links to children's performance and motivation. *Scandinavian Journal of Educational Research*, 63(1), 17-37. <https://doi.org/10.1080/00313831.2017.1324901>
- Steen, L. (2001). *Mathematics and democracy: The case for quantitative literacy*. The National Council on Education and the Disciplines.
- Stephan, M., & Akyuz, D. (2022). Semiotics from a social constructivist perspective. *International Journal of Science and Mathematics Education*, 20, 1499-1519. <https://doi.org/10.1007/s10763-021-10212-y>
- Van Voorhis, F. L., Maier, M. F., Epstein, J. L., & Lloyd, C. M. (2013). *The impact of family involvement on the education of children ages 3 to 8. A focus on literacy and math achievement outcomes and social-emotional skills*. MDRC.
- Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. Lawrence Erlbaum.
- Watson, J. M. (2007). Inference as prediction. *Australian Mathematics Teacher*, 63(1), 6-11.
- Weiland, T. (2017). Problematizing statistical literacy: An intersection of critical and statistical literacies. *Educational Studies in Mathematics*, 96, 33-47. <https://doi.org/10.1007/s10649-017-9764-5>
- Yackel, E., Gravemeijer, K., & Sfard, A. (2011). A journey in mathematics education research. Insights from the work of Paul Cobb. *Mathematics Education Library*, 48, 167-177.
- Zippert, E., & Rittle-Johnson, B. (2020). The home math environment: More than numeracy. *Early Childhood Research Quarterly*, 50(3), 4-15. <https://doi.org/10.1016/j.ecresq.2018.07.009>



## APPENDIX A

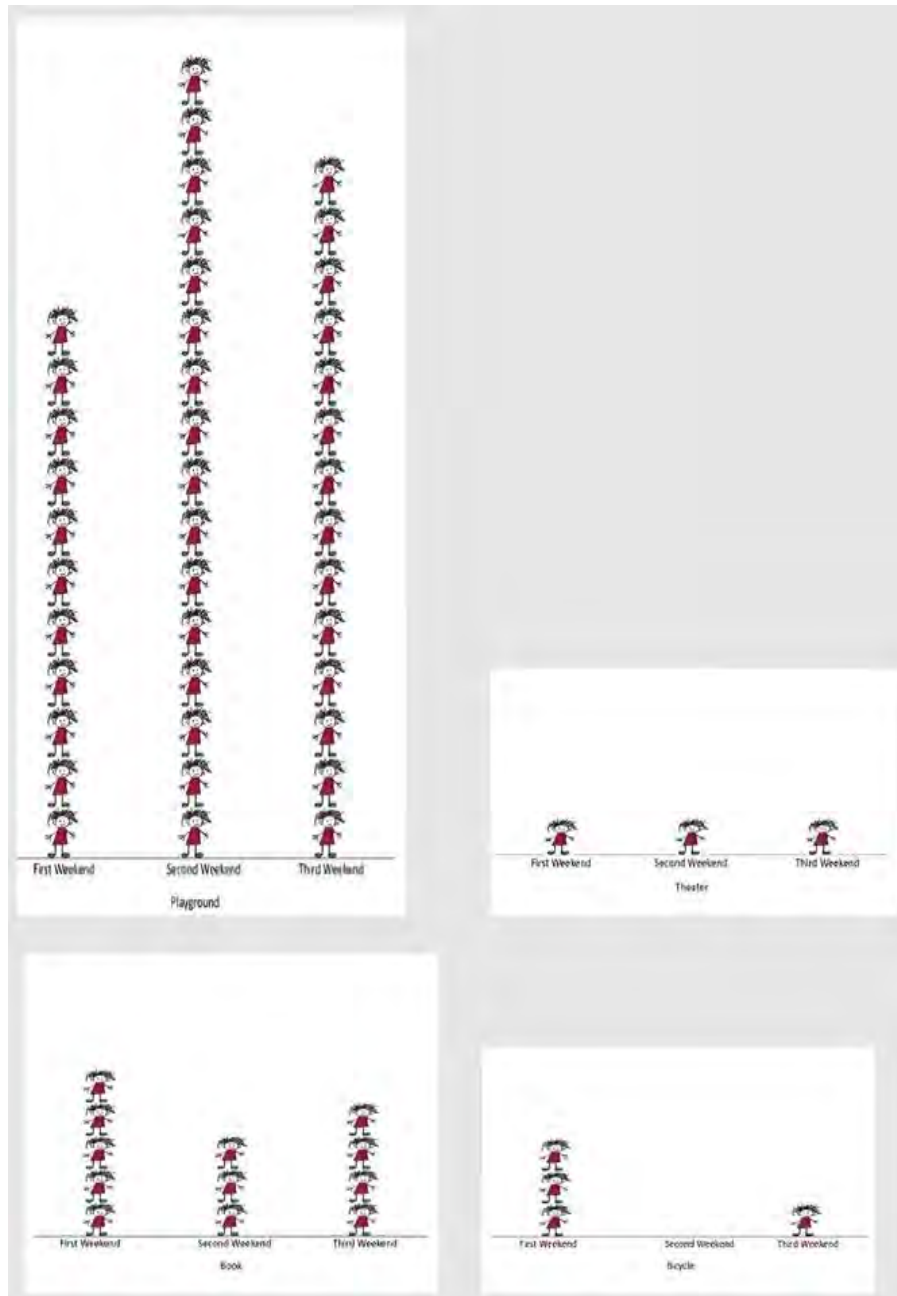


Figure A1. Pictogram of data (Source: Authors' own elaboration)