Osmangazi Journal of Educational Research Volume 10(1), Spring 2023

RESEARCH Open Access

Suggested Citation: Zora Öztürk L., & Saban Anapa, P. (2023). Examination of eighth grade students' statistical reasoning skills regarding pie chart. *Osmangazi Journal of Educational Research, 10*(1), 63-90.

Submitted: 27/02/2023 **Revised:** 09/08/2023 **Accepted:** 12/08/2023 **DOI:** 10.59409/ojer.1257451

Examination of Eighth Grade Students' Statistical Reasoning Skills Regarding Pie Chart

***Leyla Öztürk Zora, **Pınar Anapa Saban**

Abstract. The aim of the study is to examine in depth the eighth-grade students' levels of statistical reasoning by pie chart by using "The Middle School Student Statistical Thinking Model". The study used the case study design, which is a qualitative research method. The study group consists of three eight-grade students attending a public school in İstanbul, Turkey. The activities developed by the researchers, the clinical interviews based on activities and the researcher notes were used as the data collection tools. The results indicated that in the process of describing data, the students' statistical reasoning levels decrease from high academic achievement to low academic achievement. In addition, the sub-process with the lowest reasoning levels of the students is to determine the effectiveness of data display types that representing data. It was determined that the most significant differentiation between the reasoning levels of the students is in the process of analyzing and interpreting data. Students mostly had difficulties in the sub-process of making inferences about a data display. In line with the finding of the study, recommendations for future studies were presented.

Keywords. Statistical reasoning, statistics education, M3ST Model, reasoning process.

***(Responsible Author**) Mathematics Teacher. Ministry of National Education, İstanbul, Türkiye

e-mail: leyla.ozturk@marun.edu.tr

****** Prof. Dr. Eskisehir Osmangazi University, Faculty of Education, Eskisehir, Türkiye

e-mail: panapa@ogu.edu.tr

Note: This study is a part Leyla Öztürk Zora's master's thesis named "The Investigation of the Statistical Reasoning Levels of 8th Grade Students" completed in Eskişehir Osmangazi University in 2019.

In today's information society, statistical information emerges in a wide range of areas, such as population census, weather forecasts, election results, inflation rates, and fluctuations in the stock market. Individuals encounter these statistical data in newspapers, magazines, television, news websites, economic sections, or scientific articles, and they are required to think, interpret, and make inferences based on this data. In other words, reasoning on statistical information and data has become an indispensable part of daily life in contemporary conditions.

The fact that statistical information has an important place in real life and the necessity of being individuals who are productive, comprehend what they read, interpret the data they have, and make inferences has been demonstrated in studies conducted in the field of education (Gaise, 2016; Gal, 2002; Koparan, 2013; NCTM, 2000). For this reason, the importance of statistics and statistics education has started to be emphasized more in the mathematics curricula of most countries (CCSSI, 2010; MEB, 2018; NCTM, 2000; NGSS, 2013). Although there are ongoing questions about the nature of mathematics and how it should be taught, there is a strong consensus about the direction in which statistics education is heading at all grade levels (Greer, 2000). Because statistics is basically a process in which students collect data for problem situations, consider the reasons for data collection, how to organize the data, and the inferences that can be made (Cockcroft, 1982 as cited in Yıldız, 2022). However, studies indicate that students are unable to use statistical reasoning in a meaningful way because of the emphasis on calculation type studies in educational environments (Ben-Zvi and Garfield, 2004; McGatha, Cobb and McClain, 2002; Utts, 2003). Because statistical reasoning is not just about performing statistical calculations or defining concepts; it involves interpreting, reasoning, making inferences and generalizations based on data and graphical representations (Garfield and Ben-Zvi, 2008; Lovett, 2001; Mooney, 2002). In addition, conceptual understanding of important ideas such as graphical representations, measures of central tendency and distribution, and the relationship between variables are at the core of statistical reasoning (Kazak, 2015).

In the literature, there are various frameworks that have been put forward to examine the statistical reasoning processes of individuals at different levels of education and that deal with statistical reasoning in different aspects. For instance, Wild and Phannkuch (1999) focused on the thinking processes of higher education students in solving statistical problems. They defined a research cycle called the PPDAC (Problem, Plan, Data, Analysis, Conclusion) model for solving statistical problems in a real-life context. Chan, Ismail and Sumintono (2016) and Groth (2003) defined a framework to examine high school students' statistical reasoning processes. Chan et al.

(2016)'s statistical reasoning framework is defined as five levels of statistical reasoning and four key constructs which are describing data, organizing and reducing data, representing data, and analyzing and interpreting data. In addition to these four key constructs, Groth (2003) includes the process of "collecting data" in his framework. Mooney (2002) developed the Middle School Student Statistical Thinking (M3ST) framework to investigate middle school students' statistical reasoning processes. This model is based on the framework previously developed by Jones, Thorton, Langral, and Mooney (2000) for elementary school students.

In the study, Mooney's (2002) M3ST framework for middle school students was used. M3ST framework includes the four key statistical processes mentioned above, as well as sub-processes defined within each statistical process. "Describing data" process is the ability to read data in different visual displays; "organizing and reducing data" refers to arranging data using measures of center and spread; "representing data" is the capacity to construct different visual displays of the same data; and "analyzing and interpreting data" involve making inferences and predictions about statistical data (Mooney, 2002). Also, "sub-processes of describing data" are (i)to demonstrate awareness of the display features, (ii)to recognize the same data in different data displays, (iii)to evaluate the effectiveness of data displays in representing data, (iv)define units of data values; "sub-processes of organizing and reducing data" are (i)grouping or ordering data, (ii)describing data using measures of center, (iii)describing the spread of data; "sub-processes of representing data" are (i)to construct a data display for a given data set (ii)to complete a partially constructed data display (iii)to construct an alternate data display for data in a given data display; "sub-processes of analyzing and interpreting data" are to make (i) comparisons within data sets or data displays, (ii)comparisons between data sets or data displays (iii)inferences from a given data set or data display (Mooney, 2002). According to Mooney (2002), students' progress through four levels of reasoning in each of these processes: Level-1/Idiosyncratic, Level-2/Transitional, Level-3/Quantitative and Level-4/Analytical. In Level 1, students' reasoning is limited to subjective reasoning that is unrelated to the given data and often focuses on personal experiences or beliefs. In Level 2, students can show little awareness of the context, give partially correct answers but it is still not sufficient at this level. In Level 3, students can explain a problem mathematically and do not deal with irrelevant aspects of the problem. In Level 4, students can carry out all procedures without any error; they can fully read the data, make calculations and connections correctly as well as explain the aim of using different data displays, make transitions between them, draw meaningful conclusions, and generalize from the data.

In this study, the M3ST framework was used because the study group consisted of middle school students, and it provided the opportunity to examine the four key processes and sub-processes of statistical reasoning, just like the frameworks mentioned above. As a matter of fact, it is seen in the literature that students' statistical reasoning skills are also examined within the framework of these four key processes. In the describing data process, Oruç and Akgün (2010) emphasize that although 7th grade students are successful in one-dimensional graphs, they have difficulty in interpreting graphs in questions involving more than one graph; Koparan and Güven (2013) emphasize that 7th and 8th grade students can mostly read data in graphical representations at an analytical level; Tosun (2021) emphasizes that 8th grade students do not have the necessary reasoning skills for between-data and beyond-data reading levels. In the organizing and reducing data process, it was observed that students had the most difficulty in interpreting measures of central tendency and distribution (Çakmak and Durmuş, 2015; Koparan and Güven, 2013; McGatha et al., 2002) and tended to use the wrong data belonging to the graph (Koparan and Güven, 2013). In the representing data process, it is stated that students are successful in both reading and interpreting graphs, but they have more difficulty in constructing of data display (Gültekin, 2009; Tairab and Al-Nagbi, 2004), and they are inadequate in constructing an appropriate display representing the data set and evaluating the effectiveness of data display features based on the context in which the data is presented and the display features (Koparan and Güven, 2013). In the analyzing and interpreting data process, Tosun (2021) emphasizes that 8th grade students are successful in bar charts, Kaynar (2012) in line charts; Polat (2016) emphasizes that middle school students are successful in bar, line and pie charts, and Bursal and Yetiş (2020) in line, bar and pie charts, respectively.

In the literature, when the studies on statistical reasoning processes are evaluated together, it is seen that different results are obtained for the same statistical process (Bursal and Yetiş, 2020; Koparan and Güven, 2013; Oruç and Akgün, 2010; Polat, 2016; Tosun, 2021). This might be due to the context of the questions in the data collection instrument or the fact that statistical processes are considered as a whole, and sub-processes are not evaluated separately. Because in some studies conducted on the representing data process, it is stated that students are generally able to construct at least one of the bar, line or pie charts asked from them, but they cannot determine the most appropriate graph representing the data set (Gürbüz and Şahin, 2015; Hacısalihoğlu Karadeniz, 2016; Koparan and Güven, 2013; McGath et al., 2002; Özsevgeç and Yayla, 2014). Therefore, it can be said that it is important to examine students' reasoning in statistical processes based on each sub-process. In addition, the studies focusing on reasoning in four key statistical processes were limited to data collection instruments including bar and line graphs (Koparan and Güven, 2013; McGatha et al., 2002). However, in a few studies focusing on pie chart, it has been stated that students have difficulty in explaining graphical relationships and deciding on the most appropriate type of graph to represent data (Şahin, 2020) and that they cannot choose appropriate strategies to solve pie chart questions (Diezmann and Lowrie, 2009). In this context, it can be said that there is a need to examine students' statistical reasoning processes for pie chart in the context of statistical reasoning process and sub-processes. In line with these reasons, in this study, it was aimed to examine the statistical reasoning processes of eighth grade students towards the pie chart in depth within the framework of the M3ST model. In line with this purpose, answers to the following questions were sought:

1. What are the statistical reasoning levels of eighth grade students in the process and sub-processes of describing data in pie chart?
2. What are the statistical reasoning levels of eighth grade students in the process and sub-processes of organizing and reducing data in pie chart?
3. What are the statistical reasoning levels of eighth grade students in the process and sub-processes of representing data in pie chart?
4. What are the statistical reasoning levels of eighth grade students in the process and sub-processes of analyzing and interpreting data in pie chart?

Method

In this study, statistical reasoning processes of three eighth grade students were described individually. Each activity was handled separately for each student, but as a holistic situation. A case study is a research design in which a limited system, an individual, a group and a phenomenon are described and analyzed in depth (Merriam, 2013).

Regarding this study,

- It is related to the process because it is aimed to investigate the statistical reasoning levels of the students,
- It is descriptive in that it describes the decision-making and action processes of the students in-depth,
- It is inductive in that it deals with students' statistical reasoning levels based on statistical reasoning sub-processes,
- One of the researchers acted as a teacher, administrator, and researcher during the activity.

In addition, the research was limited to the eighth-grade acquisitions in the field of data processing learning in the mathematics curriculum. Data for these attainments were collected with more than one data collection tool. Based on the data obtained, students' statistical reasoning levels were examined in depth by focusing on "how" and "why" questions. With these aspects, the research is a case study.

Study Group

The study group of this study, which was determined using the purposive sampling method, consists of three eighth grade students in a public school in Istanbul in the 2018-2019 academic year. The reason why the study group consisted of eighth grade students is that all the learning outcomes addressed in the study are at this grade level and are associated with different subject areas. Besides, when the study group was selected, attention was given to ensure that the students were heterogeneous in terms of their academic achievement. Hence, it was aimed to obtain enriched data on how the reasoning processes of students at different academic levels differed in the same activity. The students in the study group have been studying in the same class since the 5th grade, and the researcher has been the course teacher at the school during this period. Therefore, students learned the topics with the same method and technique. During the teaching process, examples of the use of graphs in different disciplines were presented, and class discussions were held on which graph might have been preferred according to the context and why. In addition, both routine and non-routine problem situations were studied. Students were selected not only based on their performance in the course in which the topic was studied, but also based on their grade point average and the teacher's observations. In addition, the selection of students who have communication skills and are able to give appropriate answers to the research problem was taken into consideration. For the research ethics, the study group was created on a voluntary basis and consent forms were obtained from the students and parents. In addition, the real names of the participants were not used in the study.

The students in the study group have been studying in the same class since the 5th grade and their achievement levels are different from each other. The achievement levels of the students were classified according to their grade point averages in mathematics courses in the 5th, 6th, and 7th grades, as indicated in Table 1.

Table 1.

Student Code Names According to Achievement Levels

Student Code Names	Achievement Levels	Grade Point Averages
Zehra	High Achievement (HA)	98.86
İlayda	Medium Achievement (MA)	76.55
Damla	Low Achievement (LA)	48.50

Data Collection Tools

The data collection tools of this study consisted of the activities developed by the researchers, the clinical interviews based on the activities, and the researcher notes. In the process of preparing the activities, initially, a literature review on the subject of statistical reasoning was conducted. When the related studies are examined, it is suggested that (i) "how and why" questions be asked, (ii) the use of statistical concepts be justified, (iii) real data be included, and (iv) interdisciplinary contexts be used for the development and observation of statistical reasoning in students (delMas, 2002; Gaise, 2016; Savard & Manuel, 2016). Considering these results, three activities were prepared: “Area of Continents”, “Summer Olympics” and “School Canteen”. The questions in the activities were designed to allow the observation of the processes and sub-processes defined in the M3ST model (see Appendix-1). In addition, the activities were prepared in different disciplines, real data were used, and why and how questions were included to enable students to explain their reasoning.

The activities, for which expert opinion was taken, were applied to two students with high and low academic achievement, and the pilot implementation of the activities was carried out. During the pilot implementation process, no negative comments were received from the students about the activities. The questions in the activities and which sub-process they are related with are shown in Table 2.

Table 2.

Distribution of Questions in Activities for Pie Charts

Describing Data (DD)		
DD-1 Surface Area of Continents Question-1	DD-2 Surface Area of Continents Question-4	DD-3 Surface Area of Continents Question-2
Organization and Reduction Data (OR)		
OR-1 School Canteen Question-1	OR-2 Summer Olympics Question-2	OR-3 Summer Olympics Question-5
Representing Data (RD)		
RD-1 Summer Olympics	RD-2 Summer Olympics	RD-3 Summer Olympics

Question-3	Question-1	Question-6
Analyzing and Interpreting Data (AI)		
AI-1 Surface Area of Continents Que	AI-2 Summer Olympics Question-4	AI-3 Summer Olympics Question-7

*Sub-processes and abbreviations are explained under the heading of Analysis of Data.

Process

The research data were collected through three sessions held four weeks after the students learned the subject, based on the developed activities. Each session was held in the form of a one-on-one interview by making an appointment at the appropriate time intervals for the participants. In the research, only one activity was applied to three different students in each session.

Participants were informed that during the data collection process, the session would be recorded, but their faces would not be captured. It was clarified that their responses to the questions would only be used for this study and would not be evaluated as grades. As the researcher was also their teacher, participants did not feel uneasy and provided responses genuinely. In addition, audio recordings were taken during the interview. During the session, tools such as pencil, compass, miter, protractor, and ruler were provided. Immediately after the statistical reasoning activities for the pie charts were given to the students, necessary information was given, and an explanation was made as "you can start with any question you want". Just after the students answered the questions in the activity, a clinical interview was held for the relevant activity. In Table 3, the activity in each session, the duration of the interviews with the participants and the total duration of each session are given.

Table 3.

The Duration of Interviews with Participants

	ZEHRA(HA)	İLAYDA(MA)	DAMLA(LA)
SESSION-1 Activity: School Canteen	2' 11''	3' 16''	4' 32''
SESSION -2 Activity: Surface Area of Continents	8' 30''	16' 15''	16' 36''
SESSION -3 Activity: Summer Olympics	14' 41''	27' 21'	29' 15''

Data Analysis

Data analysis in qualitative research involves preparing and organizing the data for analysis, then coding the data and creating themes, and finally presenting the data (Creswell, 2018). However, this is a general process and there may be some differences according to the method of the research.

In this study, initially, transcripts of audio recordings obtained from student interviews were created. Subsequently, the transcribed data was classified into the levels of statistical reasoning described in the M3ST framework. Statistical reasoning processes and sub-processes in the defined framework are as follows:

The Sub-Processes of the Describing Data (DD) Process

DD-1: Awareness with data display feature

DD-2: Evaluating the effectiveness of data display in representing data

DD-3: Identification units of data values

The Sub-Processes of the Organizing and Reducing Data (OR) Process

OR-1: Grouping or ordering data

OR-2: Describing data using measures of center

OR-3: Describing the spread of data

The Sub-Processes of the Representing Data (RD) Process

RD-1: Constructing a data display for a given data set

RD-2: Completing a partially constructed data display

RD-3: Constructing an alternate data display for data in a given data display

The Sub-Processes of Analyzing and Interpreting Data (AI) Process

AI-1: Comparing within a data display

AI-2: Comparing between data display

AI-3: Inferring from data display

According to the defined framework, firstly, students' statistical reasoning levels were examined according to each sub-process and classified as Level-1/Idiosyncratic, Level-2/Transitional, Level-3/Quantitative and Level-4/Analytical. At this stage, the double-coding procedure defined by Miles and Huberman (2015) was used. The answers of the students were analyzed independently by two researchers according to the framework defined in the M3ST model. Independently from each other, two researchers determined the students' reasoning levels according to the definitions in the framework. While doing this process, the researchers determined for which description the student answers were more appropriate and assigned the given answer to the levels defined in the frame. During the determination of the levels, two different situations were encountered: (i)the frame includes a description suitable for the answer, (ii)descriptions in the framework are insufficient.

When the first of these two situations were encountered, the definition was thought to be appropriate, but when the second situation was encountered, the literature was searched again, and some changes were made in the existing definition and the missing parts were completed. Then, students' statistical reasoning levels according to the sub-processes were classified. In this context, the agreement in the evaluations of the researchers was calculated based on Miles and Huberman's (1994) percentage of agreement = [agreement / (agreement + disagreement)] x 100. As a result of this calculation, the percentage of agreement was found to be 84.8%. Reliability calculations over 70% are considered reliable for research (Miles and Huberman, 2015). Although a reliable percentage was obtained, the two researchers and the expert performing the data analysis came together again and discussed the points of disagreement until an agreement was reached.

Finally, student's statistical reasoning levels in each process were determined by calculating the mean value of the statistical reasoning levels in the sub-processes. Finally, student's statistical reasoning levels in each process were determined by calculating the mean value of the statistical reasoning levels in the sub-processes. Mean values that were halfway between two levels were rounded down to the lower level. Thus, a student receiving a mean value of 1.0 to 1.5 for a particular process would be coded as being a Level 1/idiosyncratic. A mean value of greater than 1.5 and less than or equal to 2.5 would be coded as a Level 2/transitional and so forth (Mooney, 2002).

Results

The results obtained in the study are presented in the order of sub-problems below.

The Results Related to the Describing Data Process and Sub-processes

The questions asked to the students for each sub-process of the describing data process and the responses given to the questions are presented in Table 4.

Table 4.

The Student Responses to the Questions Asked During the Describing Data Sub-Process

	Zehra (HA)	İlayda (MA)	Damla (LA)
	What information can be obtained from the graph of Area of Continents? Explain.		
DD-1	*At first glance, I can tell which continent takes up how much land in the world.* *Asia has the largest area (30%) and Australia the smallest (5%).* *The area of an average continent. However, since the data groups are not very close here, I have to comment by looking at the one in the middle.*	*In the distribution of the landmass, the Asian continent got the most share as a percentage.* *With a share of 5%, the smallest piece of land is Australia.* *If Europe and Australia unite, they will be equal to South America.*	*Most landmass is in Asia.* *Australia's landmass is the least.* *South America is more than Antarctica.* *Europe is more than Australia.*
	Can the data in the Area of the Continents pie graph be displayed with a different graph type? If so, what type of display would be more useful?		
DD-2	*Bar chart because it allows comparing the area of continents. Since the line chart is used to look at the increase and decrease of data, but the areas do not change constantly. If the goal is to compare data, a bar chart should be used, and if the purpose is to look at their distribution, a pie graph should be used.*	*It can be compared with a bar chart. ... The bar chart is used to compare data. Here we look at which continent is bigger and which continent is smaller. That's why a bar chart is appropriate.*	*It should be compared with the line chart because the increases and decreases between them are more pronounced.* *…* *I can't make the pie graph. It is difficult for me, so it should not be shown with a circle. The line is easier to interpret.*
	Which two continents have the surface area equal to half the area of the Earth?		
DD-3	*Considering that the whole of the Earth, that is, the land on the Earth, constitutes 100% of the surface area, I thought which two continents I should take to make 50%. The Asian continent has 20% land and Africa 30%, and since the sum of the two is 50%, they have half of the Earth's landmass.*	*If I add them all together, it's 100%. Accordingly, I need to think about which continents half, that is, 50%, is equal to the sum of. When I look at the graph, the sum of the surface areas of the African and Asian continents is 50%, that is, it is equal to half the surface area of the Earth.*	*How do I compare it to the Earth, I don't know the size of the Earth. ... How will I find it now? I have to choose two continents, Africa and Asia have the largest numbers then let these two be.*

Considering the student responses presented in Table 4, the statistical reasoning levels of the students in each sub-process of the describing data process were examined in Table 5.

Table 5.

The Reasoning Levels of the Students in the Sub-Processes of Describing Data

	Zehra (HA)	**İlayda (MA)**	**Damla (LA)**
	Level-4 / Analytical	Level-3 / Quantitative	Level-2 / Transitional
DD-1	Reaches information read directly from the graph. Reaches information that is not read directly from the graph (such as mean value or median). Associates variables with each other. Explains quantitatively.	Reaches the information read directly from the graph. Associates variables with each other. Explains datasets quantitatively.	Reaches the information read directly from the based on the graph's visual characteristics. Does not associate variables with each other. Does not use quantitative values when describing datasets.
	Level-4 / Analytical	Level-2 / Quantitative	Level-1 / Transitional
DD-2	Determines the graphs that are appropriate and not appropriate for displaying the dataset. Evaluates appropriate graphs with both the property of the graph and the property of the dataset.	Accurately determines the graphs that are appropriate for displaying the dataset. Decides the appropriate graph by only looking at the characteristic of the graph.	Cannot determine the graph appropriate for the display of dataset.
	Level-4 / Analytical	Level-4 / Analytical	Level-1 / Transitional
DD-3	Reads labels. Explains quantitatively. Recognizes general data unit (%).	Reads labels. Explains quantitatively. Recognizes general data unit (%)	Shows no awareness of data units.

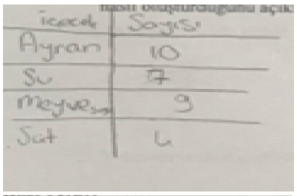
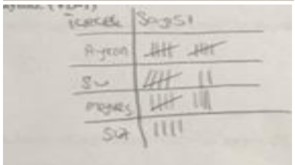
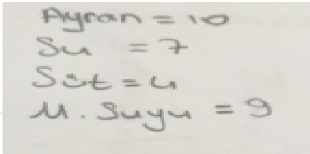
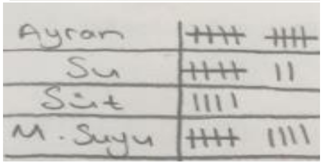
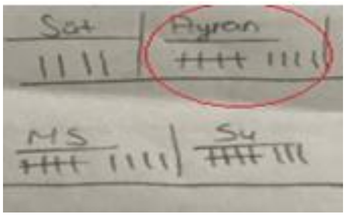
In Table 5, it is seen that Zehra (HA)'s statistical reasoning level is Level-4/Analytical in all sub-processes of the describing data process. İlayda (MA) is Level-4/Analytical only in DD-3. Damla (LA)'s statistical reasoning level is Level-2/Transitional in DD-1, Level-1/ Idiosyncratic in DD-2 and DD-3. In addition, in the description data process related to the pie chart, it is seen that the differentiation between the statistical reasoning levels is the most in DD-2. Within describing data process when the mean value of students' reasoning levels in sub-processes is calculated, it can be said that the statistical reasoning levels of Zehra, Dilara and Damla are Level-4/Analytical, Level-3/Quantitative and Level-1/Idiosyncratic, respectively.

The Results Related to the Organizing and Reducing Data Process and Sub-processes

The questions asked to the students for each sub-process of the organizing and reducing data and the responses given to these questions are presented in Table 6.

Tablo 6.

The Student Responses to the Questions in the Organizing and Reducing Data Sub-Process

	Zehra (HA)	**İlayda (MA)**	**Damla (LA)**
	Arrange the data given in the table with a tally table and a frequency table. (School Canteen)		
OR-1	*First, I determine what types of drinks there are. First, I will determine the frequency table, then the tally table.*  	*I remember the tally table but not the frequency table. First, I'll identify the types of drinks and how many of each, and then draw the tally table.*  	*I remember the tally table but not the frequency table.* (There are missing data in Damla's tally table) 
	What is the average number of athletes participating in the Rio Olympics? Explain what your result means. Prerequisite: Determining the number of people by using a pie chart. (Summer Olympics)		
OR-2	*To find the average of athletes, I find how many people participated in each branch and divide it by the number of sports branches.* *... I don't know wrestling and shooting in the graph. We know how many degrees of slices they have in their pie graph, so I can easily find these branches by proportions.* *... Having an average of 12 means that an average of 12 athletes from a sport branch participated.*	*To find the average number of athletes, I first need to find the number of athletes participating in each branch. ... First, I will find the number of athletes in shooting and wrestling by using the central angles given in the pie graph. Then I will divide the total number of athletes by 5.* *... I don't know what it means to have an average of 12. That's how it was calculated.*	*I find the total number of athletes and divide it by the number of sports branches available.* *To find the info not provided, I divide 360^0 by each central angle.* *... The average was 11. All summed and divided gives 11.*
	Find the range in the number of athletes participating in the 2016 Rio Olympics and the 2012 London Olympics. Explain what your result means. (Summer Olympics)		
OR-3	*The range is found by subtracting the smallest number from the biggest number in a dataset.* *The range of Rio is 26, and London Olympics is 24. A smaller range means that the number of athletes participating is closer to each other. ... In London, the number of athletes in each sport branch is closer.*	*The range is the difference between the largest number and the smallest number in a dataset.* *...The range is 24 at the 2012 London Olympics, and 26 at the 2016 Rio Olympics.* *The range is greater in 2016.*	*The range is the subtraction of the largest value and the smallest value in the data set. Therefore,* ***the range of London is 26, and the range of Rio is 24.*** *...* *What does such range mean? Well, one is bigger than the other.*

Considering the student responses presented in Table 6, the statistical reasoning levels and indicators of the students in each sub-process of the organizing and reducing data process are given in Table 7.

Table 7.

The Reasoning Levels in Organizing and Reducing Data Sub-Processes

	Zehra (HA)	**İlayda (MA)**	**Damla (LA)**
	Level-3 / Quantitative	Level-3 / Quantitative	Level-1 / Idiosyncratic
OR-1	Grouping the data in multiple ways Not being aware of the intended use of tables	Grouping the data in a single way Not being aware of the intended use of tables	Grouping the data such that it does not represent the data set
	Level-4 / Analytical	Level-3 / Quantitative	Level-2 / Transitional
OR-2	Proportional thinking Explaining the arithmetic mean both operationally and conceptually	Proportional thinking Explaining the arithmetic mean only operationally	No proportional thinking Explaining the arithmetic mean only operationally
	Level-4 / Analytical	Level-3 / Quantitative	Level-2 / Transitional
OR-3	Explains and calculates the range both operationally and conceptually.	Explains and calculates the range only in operational terms.	Explains the range operationally but cannot calculate it. (misreading the chart)

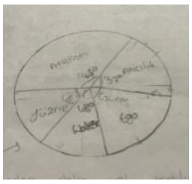
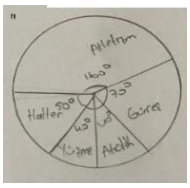
In Table 7, it is seen that the statistical reasoning level of İlayda (MA) in all sub-processes of the organizing and reducing data process is Level-3/Quantitative. While Zehra (HA)'s statistical reasoning levels were Level-3/Quantitative in OR-1, Level-4/Analytical in OR-2 and OR-3. Damla (LA) is in Level-1/Idiosyncratic in OR-1, while in Level-2/Transitional in OR-2 and OR-3. Within organizing and reducing data process when the mean value of students' reasoning levels in sub - processes is calculated, it can be said that the statistical reasoning levels of Zehra, Dilara and Damla are Level-4/Analytical, Level-3/Quantitative and Level-2/Transitional, respectively.

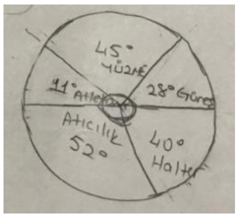
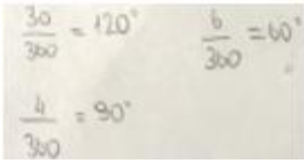
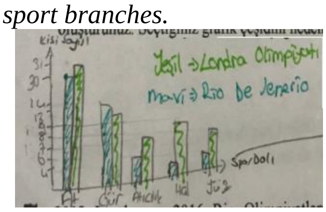
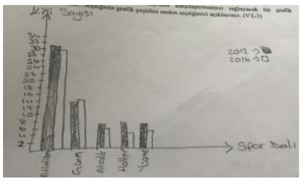
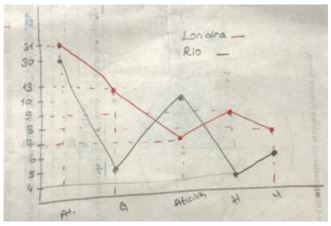
The Results Related to the Representing Data Process and Sub-processes

The questions asked to the students for each sub-process of the representing data process and the responses given to the questions are presented in Table 8.

Table 8.

The Student Responses to Questions Asked During the Representing Data Sub-Process

<table>
<tr><th></th><th>Zehra (HA)</th><th>İlayda (MA)</th><th>Damla (LA)</th></tr>
<tr><td colspan="4">Construct a graph showing the distribution of data shown in the bar chart. (Summer Olympics)</td></tr>
<tr><td>RD-1</td><td>Pie chart because when it comes to the graph showing the distribution in the questions, there is always a pie chart in the options. ... I do the proportion. But the central angles are not integers so let me take the integer part…But the sum of the angles is not 360°. I better round the numbers (to the ones digit).
</td><td>I will form a pie graph since it says distribution of data. Line is doesn't fit anyway, because the line chart was used to look up increase and decrease.... The result is not an integer, I will take an approximation of the numbers (in the tens digit).
</td><td>Since it says distribution, I'll draw a circle. Because in the previous question, he said the distribution of athletes and drew a circle.

I divide each number in the bar graph by 360°.

</td></tr>
<tr><td colspan="4">Complete the partial pie chart according to the information given in the table. (Summer Olympics)</td></tr>
<tr><td>RD-2</td><td>The circle segments of the two sports branches are shown in the graph. We need to place the data in the table in the rest of the circle. So, I'm going to add up the angle of wrestling and shooting, subtract it from 360°, and set up a proportion.
Completed the chart correctly.</td><td>I must sum all the data in the table, make a ratio of 360° and find the angle of each sport. (When forming the chart) I miscalculated because it has already filled two slices in the pie chart. Here I should subtract the sum of these two slices from 360° and since will place the remaining part in the table.
Completed the chart correctly.</td><td>When the circumference of the circle is 360°, I will divide the number of people given in the table by the circumference of the circle.

</td></tr>
<tr><td colspan="4">Create a graph to compare the number of athletes in both Olympics. (Summer Olympics)</td></tr>
<tr><td>RD-3</td><td>Line graph doesn't fit because the number of athletes did not change constantly. Also, I can't show two data in the same pie chart. But I can show both data groups in one bar graph and compare the number of athletes more easily according to the sport branches.

</td><td>I would use a bar chart to compare the two. It would be difficult for me to show it in a pie graph, the line chart is also suitable for looking up and down, so it's a bar chart.

</td><td>Line chart because increases and decreases are more visible.

</td></tr>
</table>

Considering the student responses presented in Table 8, the statistical reasoning levels and indicators of the students in the data display sub-processes are given in Table 9.

Table 9.

The Reasoning Levels of Students in Representing Data Sub-Processes

	Zehra (HA)	İlayda (MA)	Damla (LA)
RD-1	Level-3 / Quantitative Forms a pie graph even if the dataset is not an integer multiple of 360°. Proportional thinking The heading of the graph is missing.	Level-3 / Quantitative Forms a pie graph even if the dataset is not an integer multiple of 360°. Proportional thinking The heading of the graph is missing.	Level-1 / Idiosyncratic Not being able to transform bar chart into pie graph. No proportional thinking. The heading of the graph is missing.
RD-2	Level-4 / Analytical Complete the chart to represent the dataset. Proportional thinking	Level-4 / Analytical Complete the chart to represent the dataset. Proportional thinking	Level-1 / Idiosyncratic Completing the chart incorrectly and not representing the dataset. No proportional thinking.
RD-3	Level-3 / Quantitative Selecting and forming the appropriate graphical display, considering the features of more than one graphical display and the dataset. The heading of the graph is missing.	Level-3 / Quantitative Selecting and forming the appropriate graphical display by looking only at the property of the graphical displays. The heading of the graph is missing.	Level-1 / Idiosyncratic Selecting and forming inappropriate graphical display. In the generated chart, the chart does not represent the dataset.

When Table 9 is examined, it is seen that the statistical reasoning level in all sub-processes of the representing data process of Damla (LA) is at Level-1/Idiosyncratic. Zehra (HA) and İlayda (MA) are at Level-4/Analytical reasoning levels in the RD-2 sub-process, while at Level-3/Quantitative in RD-1 and RD-3. Within representing data process when the mean value of students' reasoning levels in sub -processes is calculated, it can be said that the statistical reasoning levels of Zehra, Dilara and Damla are Level-3/Quantitative, Level-3/Quantitative and Level-1/Idiosyncratic, respectively.

The Findings Related to Analyzing and Interpreting Data Process and Sub-processes

The questions asked to the students for each sub-process of the analyzing and interpreting data process and the responses given to the questions are presented in Table 10.

Table 10.

The Student Responses to Questions Asked During Analyzing and Interpreting Data Process

	Zehra (HA)	İlayda (MA)	Damla (LA)
	«Area of Continents» If the area of the continent of Europe is approximately 10 million square kilometers, approximately how many square kilometers would the surface area of the Earth be?		
AI-1	*… If 7% of it is 10 million km^2, I need to find 100% of it. … I wish you had given 4% or something. …I will round to the ones digit, ignoring the decimal point. The surface area of the Earth is one billion four hundred and twenty-eight million five hundred seventy-one thousand four hundred and twenty-eight km^2*	*Europe covers 7% of the Earth and 7% of it is 10 million, then 70% would be 100 million. But how much is 30%? 28% make 40 million. So, I found 98%, but I don't know how to find the remaining 2%.* *I will say an estimation, it will definitely be more than 140 km^2, so it is about 140.2 km^2.*	*How will I do it now? Am I going to write this (7%) as a fraction?* *…* *I don't know, there are 7 continents. So, it can be 700 million km^2.*
	«Summer Olympics» Compare the distribution of players in the Rio and London Olimpics.		
AI-2	*… If the total number of players participating in the Olympics had been the same (it is not the same) it would have been more accurate to compare them on the bar graph, but they did not participate, so I will make the comparison according to the pie graph.* *.. For example, in the London Olympics, there are 31 more people in athletics, but it is shown with a 164° circle. In Rio, it is 30 people, but the circle zone is 180°. That is, the number of people and their ratio are different.*	*…since it asks to compare according to their distribution, I will compare by looking at the pie graphs.* *… 160 people participated in athletics in London. In fact, more people participated in athletics at the Rio Olympics, but more people participated in other sports in London.* *…I looked at their people, I looked at their degrees, so the number of people in athletics is higher in Rio and in other sports branches in London.*	*I will compare the data given in the bar and pie graph. I add up all the numbers.* *…* *There are 55 people in the Rio Olympics and 68 people in the London Olympics, the difference between them is 13 people.*
	«Summer Olympics» The numbers and distribution of athletes participating in the 2012 London and 2016 Rio Olympics are given in the graphics above. What *can be* ***the total number of athletes who will participate in the five branches (athletics, wrestling, shooting, weightlifting and swimming) in the summer Olympics to be held in Tokyo in 2020?***		
AI-3	*It could be 64. I took the average of the two. One has 60 participants and the other 68 participants. I think the middle of them might be 64.*	*In fact, fewer people participated in the Rio Olympics than in London. Therefore, it may decrease further in 2020. In other words, 8 people decreased in 4 years. In the 2020 Olympics, 8 people may decrease. Considering that there is a pattern, 52 people can participate.* *…* *I can't say anything about sports branches.*	*...I don't know, there is much time to 2020, anything can happen.*

Regarding the student responses in Table 10, the statistical reasoning levels, and indicators of the students in the analyzing and interpreting data sub-processes are explained in Table 11.

Table 11.

The Reasoning Levels in Analyzing and Interpreting Data Sub-Processes

	Zehra (HA)	İlayda (MA)	Damla (LA)
	Level-4 / Analytical	Level-2 / Transitional	Level-1 / Idiosyncratic
AI-1	Comparing the part with the whole Proportional thinking	Ability to think proportionally when the whole is an exact multiple of the part Partially correct comparison	Inability to relate the part to the whole Makes inaccurate comparisons
	Level-4 / Analytical	Level-1 / Idiosyncratic	Level-1 / Idiosyncratic
AI-2	Makes accurate comparisons between multiple data displays	When comparing multiple pie graphs, she thinks that the central angles represent the number of people. Makes inaccurate comparisons	Selecting the wrong charts when comparing between data displays Makes inaccurate comparisons
	Level-1 / Idiosyncratic	Level-2 / Transitional	Level-1 / Idiosyncratic
AI-3	Non-data-based prediction in graphical displays.	Ability to estimate partially correct based on data	Not attempting to estimate

When Table 11 is examined, it is seen that Damla (LA) uses Level-1/Idiosycratic reasoning in all sub-processes of the analyzing and interpreting data process. While İlayda (MA)'s statistical reasoning levels are at Level-2/Transitional in AI-1 and AI-3 sub-processes, she is at Level-1/ Idiosycratic in AI-2. It is noteworthy that Zehra (HA) is at Level-4/Analytical reasoning in AI-1 and AI-2 sub-processes, while she is at Level-1/ Idiosycratic reasoning level in AI-3. Within representing data process when the mean value of students' reasoning levels in sub -processes is calculated, it can be said that the statistical reasoning levels of Zehra, Dilara and Damla are Level-3/Quantitative, Level-2/Transitional and Level-1/Idiosyncratic, respectively.

Discussion and Conclusion

In this study, when the responses to the sub-processes of describing data were analyzed, it was seen that the high academic achievement student is at Level 4/Analytical reasoning in all sub-processes, while the reasoning levels of other students differed. In particular, it was observed that students with low academic achievement generally are at Level-1/ Idiosyncratic reasoning. When the studies conducted in this field were examined, it was seen that seventh and eighth grade students generally reasoned at the Level-4/Analytical in describing data (Koparan and Güven, 2013; Mooney, 2002; Tosun, 2021). Therefore, it can be said that the results of this study differ from the related studies. However, in the detailed analysis, it was seen that the study group of Tosun (2021) consisted of students with high academic achievement, whereas in the studies of Koparan and Güven (2023) and Mooney (2002), no questions were asked about the pie chart. Therefore, in this study, students might have struggled to read the data in the pie chart and recognized the parts of the data values.

When the sub-processes of describing data, awareness of the data display feature (DD-1) and identification units of data values (DD-3) were considered together, it was observed that students who were able to define data value units interpreted graphs by focusing on numerical quantities rather than the visual features of the graph. Similarly, Pfannkuch and Wild (2004) and Mooney (2002) stated that students focused on the visual features of the graph rather than numerical data. In this regard, it can be said that the awareness of the data units in the process of describing data positively affects the level of awareness about the display features of the graph. In the process of describing data, the sub-process in which the students' reasoning levels are the lowest is the evaluating the effectiveness of data display in representing data (DD-2). As a matter of fact, this result shows parallelism with studies in the literature emphasizing that students are insufficient in evaluating the effectiveness of data display features (Koparan and Güven, 2013; Mooney, 2002).

The results showed that students with high, medium, and low academic achievement reasoned at Level-4/Analytical, Level-3/Quantitative and Level 2/Transitional, respectively, when all sub-processes of organizing and reducing were evaluated together. It is stated in the literature that most middle school students reason at Level-1/Idiosyncratic (Koparan and Güven, 2013); Level 2/Transitional and Level-3/Quantitative (Mooney, 2002). Although the related studies examined students' reasoning about the organizing and reducing data in table, line, and bar graphs, it is noticeable that there were no Level-4/Analytical reasoning students. As a matter of fact, many studies in literature emphasize that students have more difficulty in pie charts than other graphs (Bursal and Yetiş, 2020; Çakmak and Durmuş, 2015; Kaynar, 2012; Polat, 2016). In the detailed analysis conducted with this perspective, it was seen that the study groups of both studies consisted of students at different grade levels. When the findings of the studies are analyzed, it is seen that almost half of the 8th grade students in Koparan and Güven's (2013) study reasoned at Level-4/Analytical, while students in Mooney's (2002) study reasoned at Level-4/Analytical in some sub-processes. Based on these results, it can be said that the type of graphic representation of the data could not be related to the level of statistical reasoning in the organizing and reducing data process.

In the sub-process of grouping and ordering data (OR-1), which is one of the sub-processes of the organizing and reducing data process, it has been determined that although the students can form the tables formally, they are not aware of the purpose for which the tables are used. Similarly, Hacısalihoğlu Karadeniz (2016) and Selamet (2014) in their studies aiming to reveal the perceptions of fifth grade students about data processing, stated that almost all the students had difficulties in forming a frequency table and summarizing the data shown with a frequency table. In the sub-processes of describing data using measures of center (OR-2) and describing the spread of data (OR-3), all the students could not explain conceptually, although they could explain how the arithmetic mean and range could be calculated operationally. For this reason, students with high and medium academic achievement were included in Level-3/Quantitative in OR-2 and OR-3 sub-processes. The student with a low level of academic success, on the other hand, could not read the data in the pie chart, so she calculated the central tendency and distribution measures incorrectly. For this reason, the student took place in Level-2/Transitional in OR-2 and OR-3. In this context, it can be said that the difficulties experienced in the process of identification of data negatively affect the process of organizing the data. The results obtained from the OR-2 and OR-3 sub-processes are similar to the studies in the literature that emphasize that while students are more successful in routine problems based on formulas related to central tendency and distribution calculations, deficiencies arise in questions about where and for what purpose these calculations should be used (Cai, Moyer and Grochowski, 1999; Çakmak and Durmuş, 2015; Gal, Rothschild and Wagner, 1989; Koparan and Güven, 2013; McGatha et al., 2002; Mokros and Russel, 1995; Strauss and Bichler, 1998; Toluk Uçar and Akdoğan, 2009; Watson and Moritz, 2000).

In all sub-processes of the representing data process, it was determined that the students with medium and high academic achievement have Level-3/Quantitative reasoning, while students with low academic achievement have Level-1/Idiosyncratic reasoning. It is thought that the reason for this difference in the statistical reasoning levels of the students is due to the lack of knowledge about the concepts of ratio, proportion, percentage, and angle. In literature, there are studies emphasizing that angles, percentages, and proportions are necessary preliminary learning in the process of constructing a pie chart (Çakmak & Durmuş, 2015; Savard & Manuel, 2016; Schield, 2001; Schield, 2006). Thus, the concepts belonged to mathematics (angles, percentage, circle), but were used for creating statistical displays, in this case a pie chart. This finding suggests that making pie charts cannot be done in the statistical context alone.

In the sub-processes of data representation, students with high and medium academic achievement reason at Level-3/Quantitative reasoning in constructing a data display for a given data set (RD-1) and constructing an alternate data display for a given data display (RD-3). On the other hand, students with low academic achievement have Level-1/Idiosyncratic in all sub-processes of data representation. It is stated that students reason at Level 2/Transitional by Mooney (2002); at Level-1/Idiosyncratic by Koparan and Güven (2013). Although these results are similar for students with low academic achievement, they differ for other students. Students with high and medium academic achievement are aware of data representation features and which graph type is used for which purpose. However, all three students did not put titles on the graphs they created. This differentiation may have resulted from the difference in the study group. As a matter of fact, in related studies, it has been stated that 6th and 7th grade students especially have difficulty in creating graphical representations and determining their effectiveness (Koparan & Güven, 2013; Mooney, 2002). In addition, it can be said that students with low academic achievement have more difficulty in constructing circle graphs than others (Kaynar, 2012).

Within the statistical reasoning sub-processes of the M3ST model in the research, the most significant difference between the reasoning levels of the students was experienced during the analysis and interpretation of the data. In the sub-processes of making comparisons between (AI-1) and (AI-2) data display of the process of analyzing and interpreting the data, the high-achieving student was able to use proportional reasoning skills both when making a part-whole comparison in a pie chart and when comparing different pie charts. Therefore, the reasoning level was determined as Level-4/Analytical. The medium-achieving student, on the other hand, makes Level-2/Transitional reasoning when comparing the data in a pie chart (AI-1) within himself, as she can use proportional reasoning if the whole is a solid multiple of the part. In addition, she developed the perception that the central angle represents the number of people in the graph while making comparisons (AI-2) between multiple pie charts, and therefore, statistical reasoning in AI-2 was determined as Level-1/Idiosyncratic. The student with low academic achievement, on the other hand, made erroneous comparisons both within a pie chart (AI-1) and between more than one pie charts (AI-2) and could not use proportional reasoning skills. Therefore, statistical reasoning in AI-1 and AI-2 was determined as Level-1/Idiosyncratic. Making inferences in data display (AI-3) was determined as the sub-process that students had the most difficulty regardless of their academic achievement levels. In AI-3, only moderately successful students could partially estimate based on the data in the pie chart, while high and low achieving students made wrong estimations or did not even attempt to estimate. Therefore, in AI-3, students' statistical reasoning levels varied as Level-1/Idiosyncratic and Level-2/Transitional. This result shows parallelism with the results of the studies in the literature emphasizing that students have difficulty in making inferences and estimations from a data display (Jones et al., 2000; Koparan and Güven, 2014; Mooney, 2002). It can be thought that the reason for this difficulty is that the students have not encountered tasks based on inference and interpretation before. As a matter of fact, Yanık, Özdemir and Eryılmaz Çevirgen, (2017) examined the student textbooks in the context of statistical problems and determined that the tasks requiring inference were limited in the 5th, 6th and 7th grades, and that there were no tasks at the 8th grade level. Tasks requiring forward-looking estimations were not encountered at any grade level Jones et al. (2015) stated that when the textbooks focus heavily on operational skills, students will develop limited skills in problem solving, interpretation and estimation skills. For this reason, it is thought that focusing on interpretation and estimation skills in addition to procedural skills in textbooks and learning environments will positively affect students' reasoning skills.

Recommendations

Since this study is limited to three sub-processes of the data identification process, it is recommended to conduct a study that addresses all the sub-processes of describing data. In addition, in the process of describing data, it was determined that there were differences in the reasoning levels of the students in the national and international literature. The reasons for this situation can be investigated in future studies.

In the field of data processing learning, students' reasoning levels can be examined at different grade levels other than 8^{th} grade and with larger samples.

Also in future studies, students' reasoning levels can be examined according to different statistical reasoning models other than M3ST model.

About Authors

First Author: Leyla Öztürk Zora is a math teacher at Ministry of Education. She got master's degree at Eskişehir Osmangazi University. She is currently PHD student at Marmara University. She mainly works in the fields of Mathematics Education.

Second Author: Pınar Anapa Saban is a member of Eskişehir Osmangazi University. She works at the Faculty of Education. She is currently working at the Mathematics and Science Education Department. She completed her doctorate at Eskişehir Osmangazi University and her subject is on Mathematics. She mainly works in the fields of Mathematics Education.

Conflict of Interest

It has been reported by the authors that there is no conflict of interest.

Funding

No funding was received.

Ethical Standards

We have carried out the research within the framework of the Helsinki Declaration. The participants are volunteers.

ORCID

Leyla Öztürk Zora https://orcid.org/0000-0003-1595-5312

Pınar Anapa Saban https://orcid.org/0000-0003-3613-7001

References

Ben-Zvi, D., & Garfield, J. (2004). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. *The challenge of developing statistical literacy, reasoning, and thinking, 66*, 3-15.

Bursal, M., & Yetiş, S. (2020). Middle school students' graph skills and affective states about graphs. *International Journal of Research in Education and Science (IJRES), 6*(4), 692-704.

Cai, J., Moyer, J. C., & Grochowski, J. N. (1999). Making the mean meaningful: An instructional study. *Research in Middle Level Education, 22*(4), 1-24. doi:10.1080/10848959.1999.11670153

Çakmak, Z. T., & Durmuş, S. (2015). Determining the concepts and subjects in the area of learning statistics and probability that 6-8th grade math students have difficulties. *Abant Izzet Baysal University Journal of Faculty of Education, 15*(2), 27-58.

Creswell, J. W. (2013). *Qualitative research methods.* (M. Bütün & S.B. Demir, Trans.). İstanbul: Siyasal Kitapevi.

Chan, S. W., Ismail, Z., & Sumintono, B. (2016). A framework for assessing high school students' statistical reasoning. *Plos one, 11*(11), 1-32. doi: 0.1371/journal.pone.0163846

Common Core State Standards Initiative (2010), "*Common core state standards for mathematics* (CCSSI report)" Available at https://learning.ccsso.org/wp-content/uploads/2022/11/ADA-Compliant-Math-Standards.pdf.

delMas, R. C. (2002). Statistical literacy, reasoning, and learning: A commentary. *Journal of Statistics Education, 10*(3), 1-11. doi: 0.1080/10691898.2002.11910674

Diezmann, C. M., Lowrie, T. (2009). The role of fluency in a mathematics item with an embedded graphic: interpreting a pie chart. *ZDM Mathematics Education, 41*, 651-662. doi: 0.1007/s11858-009-0212-6

GAISE. College Report ASA Revision Committee. (2016). "*Guidelines for assessment and instruction in statistics education college report 2016*" Available at http://www.amstat.org/education/gaise.

Gal, I. (2002). Adults' statistical literacy: meanings, components, responsibilities. *International Statistical Review, 70*(1), 1–52. Doi: 10.1111/j.1751-5823.2002.tb00336.x

Gal, I., Rothschild, K., & Wagner, D. A. (1989). *Which group is better? The development of statistical reasoning in elementary school children*. Paper presented at meeting of the Society for Research in Child Development, Kansas City.

Garfield, J.B., & Ben-Zvi, D. (2008). *Creating statistical reasoning learning environment*. In J.B. Garfield, and D. Ben-Zvi (Ed.), *Developing students' statistical reasoning: Connecting research and teaching practice* (pp.45-63). Springer.

Greer, B. (2000). Statistical thinking and learning. *Mathematical thinking and Learning, 2*(1-2), 1-9.

Groth, R.E. (2003). *Development of a high school statistical thinking framework* (3087867). [Doctoral dissertation, Illinois State University]. ProQuest Dissertations and Theses Global.

Gültekin, C. (2009). *Examining 9th grade students' abilities on drawing, reading and interpreting of graphs about solutions and their properties* (Master's Thesis). Balıkesir University, Balıkesir.

Gürbüz, R., & Şahin, S. (2015). 8th grade students' skills in translating among multiple representations. *Kastamonu Education Journal, 23*(4), 1869-1888.

Hacısalihoğlu Karadeniz M. (2016). Determining the data processing learning domain attainments of 5th grade students. *Mediterrean Journal of Humanities, 6*(1), 221-236.

Jones, D.L., Brown, M., Dunkle, A., Hixon, L., Yoder, N., & Silbernick, Z. (2015). The statistical content of elementary school mathematics textbooks. *Journal of Statistics Education, 23*(3), 1-22.

Jones, G., Thornton, C., Langrall, C., Mooney, E., Perry, B., & Putt, I. (2000). A framework for characterizing children's statistical thinking. *Mathematical Thinking and Learning, 2*(4), 269-307.

Kaynar, Y. (2012). *The investigation of the "statistics" dimension of the new middle school mathematics curriculum* (Master's Thesis). Afyon Kocatepe University, Afyon.

Kazak, S. (2015). *Matematik eğitiminde teoriler*. In E. Bingölbali, S. Arslan ve İ.Ö. Zembat (Ed) *İstatistiksel akıl yürütme üzerine teorik çerçeveler* (pp. 202-213). Ankara: Pegem Akademi Yayınevi.

Koparan, T. (2013). Examination of statistical thinking models. *Elementary Education Online, 12*(3), 730-739.

Koparan, T., & Güven, B. (2013). A study on the differentiation levels of middle school students' statistical thinking. *Elementary Education Online, 12*(1), 158-178.

Lovett, M. (2001). *A collaborative convergence on studying reasoning processes: A case study in statistics*. In S. M. Carver & D. Klahr (Ed), *Cognition and instruction: Twenty-five years of progress* (pp. 347-384). Hillsdale, NJ: Erlbaum.

McGatha, M., Cobb, P., & McClain, K. (2002). An analysis of student's initial statistical understandings: Developing a conjectured learning trajectory. *The Journal of Mathematical Behavior, 21*(3), 339-355.

MEB. (2018). *Mathematic curriculum and guide K-8*. Ankara: Devlet Kitapları Müdürlüğü.

Merriam, S. B. (2013). *Qualitative Research: A guide to design and implementation* (3rd ed.). A Wiley Imprint.

Miles, M. B., & Huberman, A. M. (2015). *Early steps in analysis* (A. Ersoy, Trans.). In S. Akbaba Altun, and A. Ersoy (Trans Ed.), *Qualitative Data Analysis*. Ankara: Pegem Akademi

Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. Sage.

Mooney, E. S. (2002). A framework for characterizing middle school students' statistical thinking. *Mathematical Thinking and Learning, 4*(1), 23–63.

National Council of Teachers of Mathematics [NCTM]. (2000). *Principles and standards for school mathematics*. Reston, VA: Author

Next Generation Science Standards [NGSS]. (2013). *Next generation science standards: For states, by states.* Washington, DC: The National Academies Press.

Oruç, Ş., & Akgün, İ. H. (2010). The acquisition level of graphic reading skills of elementary social studies 7th grade students. *International Journal of Eurasia Social Sciences, 1*(1), 51-58.

Özsevgeç, T., & Yayla, G. (2014). The examination of elementary school students' graphic skills: construction and interpretation of line graphs. *Kastamonu Educational Journal, 23*(3), 1381-1400.

Pfannkuch, M., & Wild, C. (2004). *Towards an understanding of statistical thinking*. In D. Ben-Zvi, & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 17-46). Netherlands: Kluwer Academic Publishers.

Polat, F. (2016). *The skills of reading graphics used in science lessons and the visions of secondary school students towards graphics* (Master's Thesis). Cumhuriyet University, Sivas.

Savard, A., & Manuel, D. (2016). Teaching statistics: Creating an intersection for intra and interdisciplinarity. *Statistics Education Research Journal, 15*(2), 239-256.

Şahin, S. (2020). *Graphic literacy skills and challenges to secondary scholl students: pie graph example* (Doctoral Dissertation). Anadolu University, Eskişehir.

Schield, M. (2001). *Statistical literacy: Reading tables of rates and percentages.* In Proceedings of the Annual Meeting of the American Statistical Association.

Schield, M. (2006). *Statistical literacy survey analysis: reading graphs and tables of rates and percentages*. Paper presented at meeting of the International Conference on Teaching Statistics (ICOTS-7) Conference, Salvador, Brazil.

Selamet, C. S. (2014). *The examination of reading and interpretation achievement levels of 5th grade students about frequency table and graphs* (Master's Thesis). Afyon Kocatepe University, Afyon.

Tairab, H. H., & Al-Naqbi, A. K. (2004). How do secondary school science students interpret and construct scientific graphs? *Journal of Biological Education, 38*(3), 127-132.

Toluk Uçar. Z., & Akdoğan, N. E. (2009). Middle school students' understanding of average. *İlköğretim Online*, *8*(2), 391-400.

Tosun, T. (2021). *Examining of graph sense and graph construction competences of secondary school students* (Master's Thesis). Adnan Menderes University, Aydın.

Utts, J. (2003). What educated citizens should know about statistics and probability. *The American Statistician, 57*(2), 74-79.

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

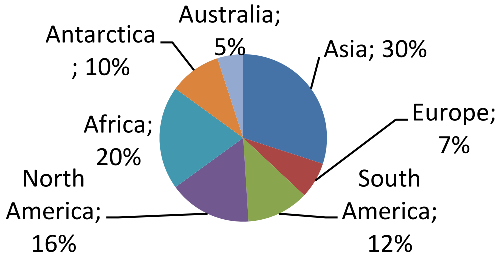
Yanık, H. B., Özdemir, G., & Eryılmaz Çevirgen, A. (2017). Investigating data processing related tasks in middle school mathematics textbooks. *Inonu University Journal of the Faculty of Education, 18*(2), 45-61. doi: 10.17679/inuefd.323407

Yıldız, S. G. (2022). "Matematik eğitiminde istatistiksel düşünmenin yeri ve önemi", In *Eğitim Bilimleri Alanında Uluslararası Araştırmalar XIV*, (103-114), Eğitim Yayınevi.

Appendix

AREA OF THE CONTINENTS

Distribution of Land Pieces in the World



1. What information do you get from the graph above? Please explain. (DD-1)

2. Which two continents have half the surface area of Earth? (DD-3)

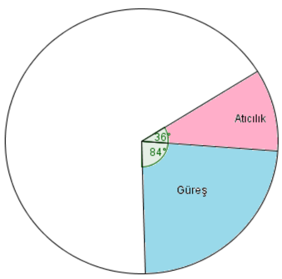
3. If the area of the continent of Europe is approximately 10 million square kilometers, approximately how many square kilometers would the surface of the earth be? (AI-1)

4. Is there a different type of chart that represents the dataset given in a pie chart? Which chart type do you think would be more useful to show the data above? Please explain with reasons. (DD-2)

SUMMER OLYMPİCS

The pie chart below shows the distribution of athletes participating in the 2016 Rio De Janeiro Summer Olympics from five different branches. However, only the distribution of athletes participating in wrestling and shooting branches is given in the table. The number of athletes from athletics, weightlifting and swimming sports branches are given in the table.

Number of Athletes from Turkey Participating in the 2016 Rio Olympics



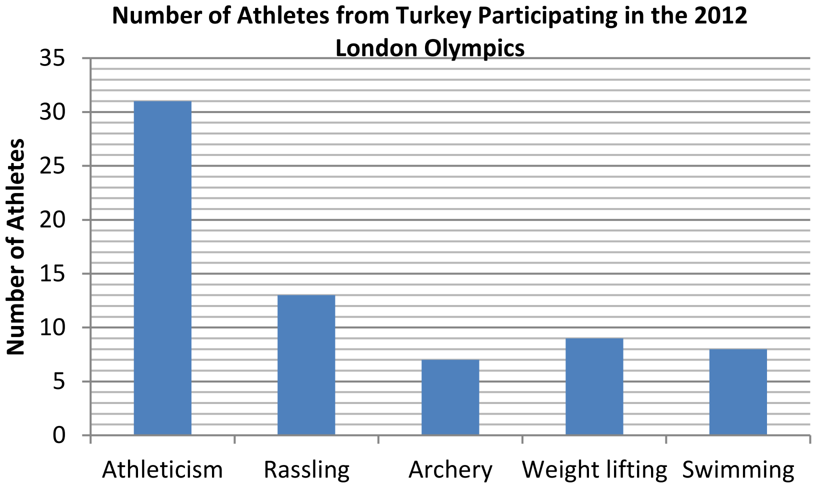
Number of Athletes from Turkey Participating in the 2016 Rio Olympics	
Spor Dalı	**Kişi Sayısı**
Atletizm	**30**
Halter	**4**
Yüzme	**6**

Cited: **(http://www.olimpiyatkomitesi.org.tr)**

1. Complete a partially constructed pie chart above based on the table. Explain how you completed the graph. (RD-2)

2. What is the average number of athletes participating in the 2016 Rio Olympics from Turkey? Please explain. (OR-2)

3. The bar chart below shows the number of athletes from Turkey participating in the same sport at the 2012 London Summer Olympics. Create a chart that shows the distribution of the number of athletes in a bar chart. Please explain. (RD-1)



4. Compare the 2012 London Olympics and the 2016 Rio Olympics according to the distribution of athletes? Please explain. (AI-2)

5. How many athletes from Turkey participated in the 2016 Rio Olympics and 2012 London Olympics? Explain how you calculated the range and compare results focusing on the meaning of the range. (OR-3)

6. Create a table to compare the number of athletes in both Olympics. Explain why you chose the chart type you chose. (RD-3)

7. The numbers and distribution of the athletes participating in the 2012 London and 2016 Rio Olympics are given in the graphics above. Guess the total number of athletes participating in the summer Olympics to be held in Tokyo in 2020 in five branches (athleticism, rassling, archery, weightlifting, swimming)? (AI-3)