SEVENTH GRADE STUDENTS' POSSIBLE SOLUTION STRATEGIES, ERRORS AND MISINTERPRETATIONS REGARDING THE CONCEPTS OF MEAN, MEDIAN AND MODE GIVEN IN BAR GRAPH REPRESENTATIONS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

DİDEM ENİSOĞLU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
THE DEPARTMENT OF ELEMENTARY SCIENCE AND MATHEMATICS EDUCATION

NOVEMBER 2014

Approval of the Graduate School of Social Sciences

_____

Prof. Dr. Meliha ALTUNIŞIK
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

_____

Prof. Dr. Ceren ÖZTEKİN
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

_____

Assoc. Prof. Dr. Mine IŞIKSAL-BOSTAN
Supervisor

Examining Committee Members

Assoc. Prof. Dr. Erdinç ÇAKIROĞLU (METU, ELE) _____

Assoc. Prof. Dr. Mine IŞIKSAL-BOSTAN (METU, ELE) _____

Assist. Prof. Dr. İ. Elif YETKİN-ÖZDEMİR (H. UN., ELE) _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Didem ENİSOĞLU

Signature :

ABSTRACT

SEVENTH GRADE STUDENTS' POSSIBLE SOLUTION STRATEGIES, ERRORS AND MISINTERPRETATIONS REGARDING THE CONCEPTS OF MEAN, MEDIAN AND MODE GIVEN IN BAR GRAPH REPRESENTATIONS

Enisoğlu, Didem
M.S., Department of Elementary Science and Mathematics Education
Supervisor: Assoc. Prof. Mine IŞIKSAL-BOSTAN

November 2014, 195 pages

The aim of the present study was to investigate the solution strategies used by seventh grade students and their possible errors while solving questions regarding the concepts of mean, median and mode given in bar graph representations. Students' possible misinterpretation regarding the concepts of mean, median and mode given in bar graph representations was also examined.

Partipants were 233 seventh grade students from two public middle schools in Gelibolu district of Çanakkale. Data were collected via Statistics Achievement Questionnaire (SAQ) during 2013-2014 spring semester. To address the research questions, students' possible solution strategies, errors and misinterpretations were identified through item based in-dept analysis.

The results of the study indicated that students' used three different solution strategies to solve questions regarding each concept. On the other hand, analysis revealed that students made errors regarding each concept. For the mean concept, identified errors were operational error, finding total number, incorrect usage of averaging algorithm, accepted close values, looking pattern, turned to find smaller or larger numbers, not meeting all requirements of a problem, forming incomplete data set, and incorrect reading of values from graph. Furthermore, for the median concept, identified errors were operational error, wrong decision on unordered data set, wrong decision on ordered data set, wrong interpretation of graph, forming incomplete data set, incorrect reading of values from graph and incorrect largest and/or smallest data values. In addition to these, students' errors in the questions regarding the mode concept were inappropriate usage of averaging algorithm, forming incomplete data set and incorrect largest and/or smallest data values. Besides, students have several misinterpretations regarding the concepts of mean, median and mode.

Key Words: Mean, Median, Mode, Solution Strategies, Errors

ÖZ

YEDİNCİ SINIF ÖĞRENCİLERİNİN SÜTUN GRAFİĞİ GÖSTERİMİNDE VERİLEN ARİTMETİK ORTALAMA, ORTANCA VE TEPE DEĞER KAVRAMLARI İLE İLGİLİ PROBLEMLERİ ÇÖZERKEN KULLANDIKLARI OLASI ÇÖZÜM STRATEJİLERİ,YAPTIKLARI HATALAR VE YANLIŞ YORUMLAMALARI

Enisoğlu, Didem
Yüksek lisans, İlköğretim Fen ve Matematik Alanları Eğitimi
Tez Danışmanı: Doç. Dr. Mine IŞIKSAL-BOSTAN

Kasım 2014, 195 sayfa

Bu çalışmanın amacı, yedinci sınıf öğrencilerinin sütun grafiği gösteriminde verilen aritmetik ortalama, ortanca ve tepe değer kavramları ile ilgili problemleri çözerken kullandıkları olası çözüm stratejilerini, yaptıkları hataları ve yanlış yorumlamaları incelemekti.

Çalışmaya Çanakkale'nin Gelibolu ilçesinden 233 yedinci sınıf devlet ortaokulu öğrencisi katılmıştır. Veriler, İstatistiksel Başarı Testi aracılığıyla 2013-2014 öğretim yılı bahar döneminde toplanmıştır. Verilerin analizinde, öğrencilerin soruları çözerken kullandıkları çözüm stratejilerini, yaptıkları hataları ve yanlış yorumları belirlemek için her soruya verilen cevaplar derinlemesine incelenmiştir.

Çalışmaların sonuçları göstermiştir ki öğrenciler her kavramla ilgili soruları çözerken üç farklı strateji kullanmıştır. Diğer yandan, veri analizi göstermiştir ki öğrenciler bu kavramlarla ilgili soruları çözerken de hata yapmıştır. Aritmetik ortalama kavramı için belirlenen hatalar; işlem hatası, toplam sayıyı bulma, ortalama algoritmasını yanlış kullanma, daha küçük veya daha büyük sayılara yönelme, problemin bütün gerekliliklerini yerine getirmeme, eksik veri grubu oluşturma ve grafikten değerleri yanlış okumadır. Bunlara ek olarak, ortanca kavramı ile ilgili sorularda belirlenen hatalar ise şunlardır: işlem hatası, sıralanmamış veri grubu üzerinden yanlış karar, sıralanmış veri grubu üzerinden yanlış karar, grafiği yanlış yorumlama, eksik veri grubu oluşturma, grafikten değerleri yanlış okuma ve en büyük ve/veya en küçük değeri yanlış tespit etmedir. Bunlara ek olarak, tepe değer kavramı ile ilgili sorularda yapılan hatalar ise şunlardır: Ortalama algoritmasını yanlış yerde kullanma, eksik veri grubu oluşturma ile en büyük ve/veya en küçük değeri yanlış tespit etmedir. Ayrıca, öğrenciler aritmetik ortalama, ortanca ve tepe değer kavramları ile ilgili yanlış yorumlar da yapmıştır.

Anahtar Kelimeler: Aritmetik Ortalama, Ortanca, Tepe Değer, Çözüm Stratejileri, Hatalar

To My Husband and My Family

ACKNOWLEDGMENTS

I would like to express my deepest appreciation and gratitude to my supervisor Assoc. Prof. Mine IŞIKSAL-BOSTAN for her invaluable assistance, guidance, encouragement and support throughout this process. She encouraged me to maintain the research whenever I was in the complicated processes of the thesis. Her insightful comments and suggestions have shaped my thesis work.

I sincerely thank to my committee members Prof. Dr. Erdinç ÇAKIROĞLU, Assist. Prof. Dr. İ. Elif YETKİN-ÖZDEMİR for their invaluable contributions for my study.

A special gratitude to Assist. Prof. Dr. Sibel KAZAK and Assist. Prof. Dr. İ. Elif YETKİN-ÖZDEMİR for her suggestions, feedbacks and comments for my study.

I would also like to thank all members of my family. They always believe and encourage me throughout the thesis process. My mother Şengül KARAMAN, my father Nihat KARAMAN and my sisters Gökçe KARAMAN, Beyza KARAMAN and İrem KARAMAN, thank you all.

I really appreciate and thank to my husband, İlter ENİSOĞLU for his love, everlasting patient and support. He always motivated me when I felt tired and depressed throughout the thesis process. I extend my sincere thanks to ENİSOĞLUs family for their support.

I would like to express my special thanks to Gizem SEVİM ATAYEV and Ayşegül ÇABUK for their contributions and support during this process. They have been special for me in my life. I would also like to thank Rukiye AYAN and Seçil YEMEN KARPUZCU for their help and support for this study.

Finally, I would like to thank TUBİTAK for the scholarship which helped me to pursue my master study.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATION

MoNE	Ministry of National Education
METU	Middle East Technical University
NCTM	National Council of Teachers of Mathematics
SAQ	Statistics Achievement Questionnaire

CHAPTER 1

INTRODUCTION

In recent years, the emphasis laid upon statistics in the mathematics curriculum has increased by means of international movements (Jacobbe & Carvalho, 2011). One of the reasons underlying this increase in emphasis may be that statistics is an essential part of daily life. More specifically, statistics is essential to effectively participate in political debates on education, equity, health and environment in order to analyze and interpret data and results of research on such issues accurately (Konold & Higgins, 2003). Another reason may be that statistics establishes links between other areas, such as history, science and geography (Konold & Higgins, 2003).

With the increasing importance attached to statistics in many fields, NCTM added statistics to the curricula of K-12 mathematics. Moreover, *The Principles and Standards for School Mathematics* (NCTM, 2000) set out of 5 content areas, one of which is Data Analysis and Probability (Pratt, 2005). In other words, the national organization of mathematics teachers in the United States accepted geometry, number sense, measurement, algebra and statistics as the foundations of statistics (Shaughnessy, 1992). Similarly, in Turkey, with the changes made in the mathematics program in 1990 and 1992, numerous concepts related to probability and statistics were added, and the new textbooks included basic concepts, such as measures of central tendency, measures of spread and probabilistic events (Bulut, Ekici, & Iseri, 1999). In addition, the elementary mathematics curriculum in Turkey included real life contexts for the purpose of making interpretations and making decisions by resorting to statistical knowledge (MoNE, 2005). Furthermore, the last revised mathematics curriculum has also included statistics

in each level of the middle school mathematics curriculum in Turkey (MoNE, 2013). In statistics, data analysis is crucial; it includes asking questions, collecting data, analyzing data, and forming and communicating conclusions (Friel & Bright, 1998). The present research study focused on analyzing data, and forming and communicating conclusions stages of data analysis. The analyzing data stage includes choosing appropriate numerical and graphical representation methods, and using these methods to analyze collected data (Franklin, Mewborn, Moreno, Peck, Perry, & Scheaffer, 2005). One of the frequently used numerical representations in data analysis is measures of central tendencies, namely the mean, the median and the mode (Van de Walle, 2013). Furthermore, some of the graphical representations are bar graphs, pictographs, circle graphs and line graphs (Van de Walle, 2013). However, at that point, it is important to note that no one method of numerical and graphical representation is more useful than the other. Each method of representation is appropriate as long as it is used in parallel to the purpose of the collected data (Konold & Higgins, 2003). On the other hand, the forming and communicating conclusions stage includes interpreting the analysis and relating the interpretation to the research question of a study (Franklin et al., 2005). Therefore, in order to make sense of collected data, the analyzing data and forming and communicating conclusions stages are essential parts of data analysis. Hence, the mean, the median and the mode are commonly used numerical representations in analyzing data stage of data analysis and interpretation of these concepts is important to communicate on conclusions (Shaughnessy, 1992; Van de Walle, 2013).

With increasing emphasis laid upon data analysis in statistics as an outcome of curriculum reform efforts, studies regarding students' understanding of statistical concepts have become a topic of interest. However, different research study results indicated that students' make different types of errors, encounter various difficulties and hold misconceptions related to basic statistical concepts such as the mean, the median and the mode (Bright & Friel, 1998; Cai, 2000; Pollatsek, Lima & Well, 1981; Smith, diSessa, & Roschelle, 1993; Strauss & Bichler, 1988). For example, students face difficulties while calculating the weighted mean (Pollatsek, Lima & Well, 1981) and they made errors while using computational algorithm of the mean to find a missing value of a data set when mean of the data set was given (Cai, 1998). Besides, according

to Cai (2000), the difficulties students face regarding the averaging algorithm derive from students' lack of conceptual knowledge rather than their lack of procedural knowledge (Cai, 2000). Therefore, in order to obtain information regarding students' procedural and conceptual knowledge regarding the concepts of mean, median and mode, their errors while solving questions related to the concepts should be investigated. On the other hand, elementary students tend to have misconceptions on the interpretation of collected data (Capraro, Kulm & Capraro, 2005; Tarr & Shaughnessy, 2007). Moreover, they made errors while interpreting graphs (Bright & Friel, 1998). Hence, researchers agreed on the fact that students can make errors while calculating the mean, the median and the mode concepts and they can also make wrong interpretations while interpreting these concepts. In this regard, focusing on students' errors regarding the concepts of mean, median and mode while solving questions given in graph representations would be valuable to investigate to gain information about students' procedural and conceptual understanding of the three average concepts.

In accessible literature, although there are several studies that examine the problem solution strategies used by students (Becker, 1992; Cai 1995; 1998), there is a limited number of studies on the solution strategies used by students while solving questions regarding statistical concepts (Cai, 2000). Identifying students' possible solution strategies is important since if a student have conceptual understanding regarding the averaging algorithm of the mean, the student can correctly and flexibly apply the averaging algorithm to solve questions regarding the concept (Cai, 1998; Watson & Moritz, 2000). However, the number of studies in Turkey are not sufficient regarding students' solution strategies while solving questions regarding the mean, the median and the mode as well as errors and interpretations regarding these concepts (Uçar & Akdoğan, 2009). Focusing on students' solution strategy preferences while trying to solve questions related to these concepts can reveal important information about their understandings of these concepts (Cai, 1998); therefore, one of the purposes of the present study is the solution strategies utilized by students while solving questions regarding the concepts of mean, median and mode.

As it is mentioned above, it is known that middle school students have errors and difficulties regarding the concepts of mean, median and mode (Bright & Friel, 1998; Cai, 2000; Smith, diSessa, & Roschelle, 1993; Strauss & Bichler, 1988; Watson & Moritz, 2000) and identifying students' solution strategies is essential to gain information regarding conceptual understanding about the concepts (Cai, 1998; Watson & Moritz, 2000). Therefore, the aim of the present study was to investigate possible solution strategies used by seventh grade students and their errors while solving questions regarding the concepts of mean, median and mode given in bar graph representations. Lastly, the students' misinterpretations regarding the concepts of mean, median and mode given in bar graph representations were investigated.

1.1 Purpose of the Study

The first purpose of the present study was to investigate the solution strategies used by seventh grade students while solving questions regarding the concepts of mean, median and mode given in bar graph representations. The other purposes of this study were to examine seventh grade students' possible errors and misinterpretations regarding the concepts of mean, median and mode given in bar graph representations. In order to achieve the purposes of this study, three research questions were posed:

1) What are the possible solution strategies used by middle school seventh grade students while solving questions regarding the concepts of mean, median and mode given in bar graph representations?

2) What are the possible errors made by middle school seventh grade students regarding the concepts of mean, median and mode given in bar graph representations?

3) What are the possible misinterpretations made by middle school seventh grade students regarding the concepts of mean, median and mode given in bar graph representations?

1.2 Significance of the Study

In more recent years, the importance of the mean, the median and the mode in the mathematics curriculum has been accepted by many countries by emphasizing understanding and usage of these concepts in suitable contexts (Watson & Moritz, 2009). Some of these countries are the United States (National Council of Teachers of Mathematics [NCTM], 1989), Australia (Australian Education Council [AEC], 1991, 1994), and England and Wales (Department for Education [DFE], 1995) (Watson & Moritz, 2009). Similar to the other curriculum efforts, the mathematics curriculum in Turkey has included real life context to interpret statistics and make decisions based on the base of students' statistics knowledge (MoNE, 2005). Besides, the new Turkish curriculum indicated the importance of learning statistics by placing statistics instruction into each level of middle school mathematics education (MoNE, 2013).

Students could use different solution strategies while solving questions regarding the mean concept such as levelling and using average formula (Cai, 2000). The solution strategies used by students to solve given questions regarding averages shed light upon their procedural and conceptual knowledge regarding the concepts of mean, median and mode. More specifically, procedural knowledge regarding the concepts of mean, median and mode is knowing procedures, rules and algorithms of these averages and familiarity with symbols (Hiebert & Lefevre, 1986). Furthermore, as an example for conceptual knowledge regarding the average concepts is deciding appropriate numerical and graphical representation according to purpose of the collected data (Baroody, Feil & Johnson, 2007). Thus, focusing on students' solution strategies also give information regarding students' understandings of average concepts (Mokros & Russell, 1995; Watson & Moritz, 2009). In other words, identifying students' possible solution strategies is important since if a student have conceptual understanding regarding the average concepts, the student can correctly and flexibly apply procedures, rules and algorithms of these concepts to solve questions regarding the concepts (Cai, 1998). More specifically, center of balance and fair-share understandings, which were defined as two understandings of the mean by past researchers (Van de Walle, 2013), can be used to solve problems regarding the mean if a student have conceptual understanding of the

concept. Thus, solution strategies can also give information regarding these understandings of middle school students. Therefore, in order to gain information about middle school students' understandings regarding the concepts of mean, median and mode, the present study is significant.

On the other hand, analyzing students' solution strategies can also reveal information regarding students' errors. Several studies revealed that the concepts of average are not concepts that are easily comprehensible (Bright & Friel, 1998; Cai, 2000; Smith, diSessa, & Roschelle, 1993; Strauss & Bichler, 1988). As mentioned by Garcia and Garrett (2008):

> It is not so easy for students to understand basic notions associated with this concept [arithmetic mean]. These difficulties are related to different aspects: understanding of the algorithm, understanding of the concept and its properties, use of representations and language, ability to put forward arguments, among others (p.50).

Thus, it seems that there is an agreement in the literature that students experience difficulties and make errors regarding averages in statistics (Garcia & Garrett, 2008). Therefore, this research is important from the aspects of identifying seventh grade middle school students errors related to the concepts of mean, median and mode.

Interpretation of the concepts of mean, median and mode is important to communicate on conclusions of a research (Shaughnessy, 1992). However, while interpreting these concepts, students tend to have misconceptions (Mathew & Clark, 2007; Tarr & Shaughnessy, 2007). To illustrate, the mean was accepted as a process not measures of central tendency (Mathew & Clark, 2007). Therefore, analyzing how Turkish middle school students interpret the average concepts is essential to gain information about their misinterpretations regarding the concepts of mean, median and mode.

In addition, it is known that since knowing statistical concepts affect students' life both practically and educationally (Konold & Higgins, 2003), their understanding regarding these concepts is important. Furthermore, the questions in the questionnaire used in the

present study were asked through bar graphs since the graph type is generally encountered in daily life contexts, such as in newspapers and in analysis of results, and they are used in some of the questions related to concepts of average in seventh grade mathematics textbooks in Turkey. However, although there are some research studies having investigated students' errors and their solution strategies regarding the concepts of average (Bright & Friel, 1998; Cai, 2000; Pollatsek, Lima & Well, 1981, Smith, diSessa, & Roschelle, 1993), there is still limited research regarding this topic (Cai, 2000; Uçar & Akdoğan, 2009). Similarly, studies in Turkey investigating the possible errors seventh grade students make regarding the concepts of mean, median and mode and the solution strategies they use to solve questions related to these concepts remain insufficient in the accessible literature. Therefore, it is important to investigate not only the solution strategies students resort to while solving questions regarding the mean, the median and the mode through bar graphs, but also the errors and misinterpretations they make regarding these concepts. In addition to contributing to related literature, this study will provide guidance and, thus, be of significance to, both teachers and mathematics curriculum developers.

1.3 Definitions of the Important Terms

The research questions include some important terms. Constitutive definitions of important terms are presented below:

Mean: "The mean is average of all the scores in a distribution. It is determined by adding up all of the scores and then dividing this sum by the total number of scores" (Fraenkel & Wallen, 2006, p. 197). The mean is one of the describers of collected data (Fraenkel & Wallen, 2006).

Median: The median is defined as "the point below and above which 50 percent of the scores in a distribution fall−in short, the midpoint" (Fraenkel & Wallen, 2006, p.196). It is one of the describers of collected data. The median of a distribution is affected by the number of terms (Fraenkel & Wallen, 2006).

Mode: The mode is another method of calculating the average of values in statistics and is defined as "the most frequent score in a distribution" (Fraenkel & Wallen, 2006, p.196). Some distributions may have two modes that are named as *bimodal distribution* (Fraenkel & Wallen, 2006).

Graph: Graph refers to "information transmitted by position of point, line or area on a two-dimensional surface" (Fry, 1984, p.5).

In this study, especially bar graphs are used. According to Oxford Dictionary (2014) the bar graph refers to "a diagram in which the numerical values of variables are represented by the height or length of lines or rectangles of equal width."

***Error*:** "An error is a mistake, slip, blunder, or inaccuracy, and a deviation from accuracy" (Luneta & Makonye, 2010, p.36).

In this study, the term 'errors' was used for errors originating from algorithmically based errors due to operational errors, errors based on formal knowledge due to limited conceptions or inadequate knowledge regarding properties of the mean, the median and the mode concepts, and other errors that occurred due to neither algorithmic nor formal knowledge inadequacy. Moreover, the term 'errors' was used to identify incorrect answers but they might derive from students' lack of knowledge or their misconceptions.

1.4 Organization of the Study

In this chapter, the purpose of the study, the significance of the study and the definitions of important terms have been explained. The second chapter is devoted to the literature review, which presents definitions of important terms, such as error, mean, median and mode. Additionally, related studies on students' understandings on averages are mentioned. The third chapter includes information about the design of the study, the population and sample, data collection instruments, validity and reliability of the instruments, data collection procedure, analysis of data, assumptions and limitations. The fourth chapter presents findings of the study with respect to the solution strategies used by seventh grade students in the questions of the Statistics Achievement Questionnaire (SAQ), and the errors and misinterpretations they made regarding the

mean, median and mode concepts given in bar graph representations. The last chapter involves the discussion and implications of the study and presents recommendations for further studies.

CHAPTER 2

LITERATURE REVIEW

In this chapter, review of related literature is presented. Based on the purposes of the study, the chapter consists of six sections: conceptual and procedural knowledge regarding the concepts of mean, median and mode, definitions of error, definitions of the various concepts of average, related studies on understandings of the mean, the median and the mode, related studies on graphs, and studies conducted in Turkey on the concepts of average and graphing. At the end of the chapter, a summary of these six sections is presented.

2.1 Conceptual and Procedural Knowledge Regarding the Concepts of Mean, Median and Mode

Conceptual and procedural knowledge are defined as two types of knowledge that shape mathematical understanding (Schneider & Stern, 2010). The first type is conceptual knowledge, which involves deeper understanding through building relationship between mathematical ideas (Van de Walle, 2013). The second type is procedural knowledge, which focuses on skills and step-by-step procedures without understanding the mathematical ideas behind the procedures (Hope, 2006). However, in order to gain mathematical competence, establishing connection between procedural and conceptual knowledge is necessary (Bisanz & Lefevre, 1992; Van de Walle, 2013). In the following parts, more information regarding conceptual and procedural knowledge will be given.

2.1.1 Conceptual Knowledge

Hiebert and Lefevre (1986) defined conceptual knowledge as follows:

> Conceptual knowledge is characterized most clearly as knowledge that is rich in relationships. It can be thought of as a connected web of knowledge, a network in which the linking relationships are as prominent as the discrete pieces of information. Relationships pervade the individual facts and propositions so that all pieces of information are linked to some network (p. 3).

In the above description, Hiebert and Lefevre (1986) maintain that the quality of conceptual understanding is related to forming new cognitive networks and establishing relationships between these new networks and the existing schemas. Similarly, according to Baroody, Feil and Johnson (2007), new knowledge is learned conceptually if it is assimilated into the existing knowledge. Examples regarding the concepts of mean, median and mode to clarify the meaning of conceptual knowledge could be useful. For instance, the three measures of central tendency were used to find the typical or central value in a data set, and for some data sets, one of the measures was more suitable than the others to summarize them (Groth & Bergner, 2006).

2.1.2 Procedural Knowledge

Procedural knowledge "...is made up of two distinct parts. One part is composed of the formal language, or symbol representation system, of mathematics. The other part consists of the algorithms, or rules, for completing mathematical tasks" (Hiebert & Lefevre, 1986, p.5). The first part “includes a familiarity with the symbols used to represent mathematical ideas and an awareness of the syntactic rules for writing symbols in an acceptable form” (Hiebert & Lefevre, 1986, p. 6). On the other hand, the second type "consists of rules, algorithms, or procedures used to solve mathematical tasks” (Hiebert & Lefevre, 1986, p. 6). An example from statistics to explain the difference between the two types of procedural knowledge could be given. Knowing the meaning

of the symbol X to denote the mean of a set of data is an example for the first type of procedural knowledge because it is syntax. On the other hand, knowing the "add-and-divide" algorithm to find the mean is an example for the second type of procedural knowledge (Groth & Bergner, 2006).

In the previous section, definitions of conceptual and procedural knowledge were presented because one of the purposes of the study was to investigate the students' errors regarding the interpretation of the mean, the median and the mode given in bar graph representations based on their conceptual and procedural knowledge regarding the concepts. In accordance with the purpose of the study, in the following section, definition of error is given.

2.2 Error/Mistake

As mentioned before, one of the purposes of the study was to investigate students' possible errors regarding the concepts of mean, median and mode, the definition of error was presented in the section.

Within there lated literature, two different definitions of ‘error’ can be found. For example, Koshy (2000) defined error as “a wrong idea or wrong action that often is the result of a misconception, but not always so” (p.172). He claimed that an error can sometimes be the cause of a misconception based on the relationship between the mistake and the misconception. In a further study, Luneta and Makonye (2010) defined error as “… a mistake, slip, blunder or inaccuracy and a deviation from accuracy” (p.35). Additionally, Green, Piel and Flowers (2008) defined three types of errors, which are unsystematic, systematic or random. Unsystematic errors are unintended, non-repeated and trivial incorrect answers. This error type can be corrected by learners when it is realized by them (Khazanov, 2008). As opposed to unsystematic errors, systematic errors are repeated wrong responses orderly recreated across time and space. Moreover, systematic errors cannot be realized by learners and they are based on symptoms of a faulty hypothesis referred to as misconceptions (Green, et. al, 2008; Nesher 1987; Riccomini 2005). Lastly, random errors may or may not be repeated. Similar to Green et al. (2008), Cox (1975) categorized errors into two groups, systematic errors and random

errors. Systematic errors are defined as repeated wrong answers that are detected in a specific algorithmic calculation. On the contrary, random errors do not give any evidence related to the repeated wrong thinking process.

Another researcher, Radatz (1979) studied mathematical mistakes and he defined mistake as follows:

> First, mistakes in the learning of mathematics are not simply the absence of correct answers or the result of unfortunate accidents. They are the consequence of definite processes whose nature must be discovered. Second, it seems to be possible to analyze the nature and the underlying causes of mistakes in terms of the individual's information processing mechanisms. (p. 170).

In addition to the above definition, he stated that while learning mathematics, mistakes are not lack of correct answers or the result of unfortunate occurrence. He also added that causes of mistakes can be analyzed by investigating students' information processing mechanisms. Thus, he focused on causes of students' mistakes in mathematics education and he identified causes of mistakes in five categories: language difficulties, getting spatial information difficulties, deficiency in prerequisite skills, facts and concepts, incorrect associations, and applying unrelated strategies and rules. In addition to this study, Radatz (1980) studied students' mistakes in mathematical learning process. Similar to his previous study (Radatz, 1979), he claimed that the causes of mistakes are not carelessness, unsafety or unique circumstances in a situation; in fact, the causes are previous experiences of students' in their classrooms. Rather, he stated the features of student mistakes based on the research results as follows, claiming that student mistakes:

- are causally determined, and very often systematic;
- are persistent and will last for several school years, unless the teacher intervenes pedagogically;
- can be analyzed and described as error techniques;

- can be derived, as to their causes, from certain difficulties experienced by students while receiving and processing information in the mathematical learning process, or from effects of the interaction of the variables acting on mathematical education (teacher, curriculum, student, academic environment, etc.) (p.16).

Thus, mistakes of students are generally unsystematic, persistent, appropriate to analyze and affected by other variables, such as teacher, curriculum and academic environment.

After defining term of error in this study, errors made by students regarding the concepts of mean, median and mode will be focused. In the following section, based on the purposes of this study, the meanings of the concepts of averages, which are the mean, the median and the mode, are presented.

2.3 Concepts of Averages

Concepts of averages are highly influential in statistics because these concepts are used to describe a representative value that summarizes information about large data set **(**Shaughnessy, 1992**).** The mean, median and mode are three types of averages among many average types (Van de Walle, 2013). In order to give more information regarding the average concepts, the section is separated into three parts, namely the mean, the median and the mode.

2.3.1 The Mean

The mean is a very strong and powerful concept in statistics and it sheds light on students' basic understanding of summary statistics (Shaughnessy, 1992). Although the mean is one of the statistical concepts and defined independently of mathematical concepts, the process of calculating the means requires mathematics extensively. From this perspective, the mean is defined as "a mathematical construction that represents certain relationship in the data" (Russell & Mokros, 1996, p.362). Thus, there is a relationship between the statistic concept and mathematics. To illustrate, to calculate the mean, all the values in a data set are added and then divided by the number of values in the data set (Van de Walle, 2013).

On the other hand, most of the students who compute the mean correctly do not understand the meaning of the value and what the value represents (Konold & Higgins, 2003). Although the computational algorithm of the mean is easy to apply, developing conceptual understanding of this concept is difficult (Konold & Higgins, 2003). Related to understandings of the mean, past researchers defined two models, namely *center of balance* and *fair-share* (Van de Walle, 2013). The understandings also combined the mean concept's mathematical and statistical aspects. Firstly, the *center of balance* accepts the mean as point of balance of a data set (Hardiman, Well & Pollatsek, 1984). This process is presented in Figure 2.1. Secondly, the *fair-share* accepts the mean as equal distribution of a data set (Mokros & Russell, 1995). An illustration of this approach is given in Figure 2.2.

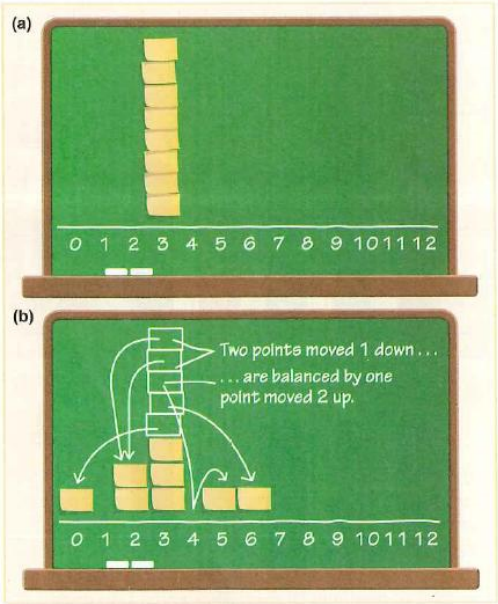


Figure 2.1 The Mean as a Point of Balance (Van de Walle, 2013, p.448)

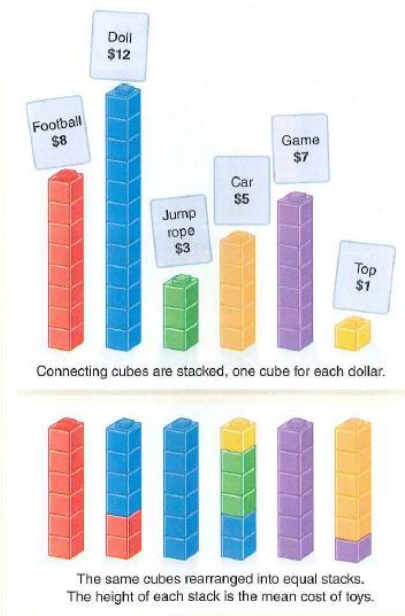


Figure 2.2 The Mean as a Point of Balance (Van de Walle, 2013, p.448)

It is claimed that fair-share and center of balance understandings provide a cognitive bridge between the mean as a statistical concept of representativeness and its computational algorithm (Marnich, 2008).

Additionally, Strauss and Bichler (1988) mention seven properties of the mean concept:

> The average is located between the extreme values, the sum of the deviations from the average is zero, the average is influenced by values other than the average, the average does not necessarily equal one of the values that was summed, the average can be a fraction that has no counterpart in physical reality, when one calculates the average, a value of zero, if it appears, must be taken into account, the average value is representative of the values that were averaged (Strauss &Bichler, 1988, p.66).

In order to understand these seven properties completely, conceptual understanding of the mean is required. Furthermore, the properties are parts of three aspects regarding the mean concept, which are statistical, abstract and representative. The first three of these properties are related to the statistical aspect, the second three are related to the abstract aspect and the last one is related to the representative aspect of the mean.

In the next section, definition and properties of the median will be presented.

2.3.2 The Median

Another concept of average is the median, which is defined as "the point below and above which 50 percent of the scores in a distribution fall−in short, the midpoint" (Fraenkel & Wallen, 2006, p.196). The median of a distribution that includes an odd number of values is the middle score when the values are listed in order. To illustrate, the median of the distribution 2, 4, 6, 8, and 10 is 6. On the other hand, the median of a distribution that includes an even number of values is the midpoint of the two middle scores when the values are listed in order. For example, in the distribution 1, 3, 4, 6, 7, and 8, the median is 5. Thus, the median could be a different value than the actual values in the distribution being represented (Fraenkel & Wallen, 2006).

According to most researchers and curriculum developers, the median should not be introduced to students before the students reach elementary grade level (Russell, 2006). Another point worth mentioning is that the median is not influenced by extreme values in a data set (Van de Walle, 2013). Thus, if a data set has extreme values, the median is the most appropriate average type to summarize the data set (Fraenkel & Wallen, 2006).

In the next section, definition and properties of the mode will be presented.

2.3.3 The Mode

The mode is another average type to summarize a data set, which is defined as "the most frequent score in a distribution" (Fraenkel & Wallen, 2006, p.196). For example, the mode of the distribution, 20, 11, 12, 12, 14, 11, and 12 is 12. However, some distributions may have two modes that are named as *bimodal distribution* (Fraenkel & Wallen, 2006). To illustrate, the mode of the distribution 5, 8 ,5 ,6, 7, 8, 9, and 10 is 5 and 8. On the other hand, the mode is only one type of average to summarize and describe qualitative data, such as favorite games (Russell, 2006). Thus, if a data set is qualitative, the mode is the appropriate average type to represent the data set.

In the previous section definitions of the concepts of averages have been presented. Since the present study aimed to investigate the solution strategies used by students while solving questions which included the mean, the median and the mode, and the errors they made while interpreting the concepts, related studies about students' understandings will be discussed in the next section.

2.4 Studies on Students' Understandings of Mean, Median and Mode

In recent years, most of the research studies on statistics education have been conducted to examine individual summary statistics such as the mean, the median and the mode (Leavy, 2004) because the statistics are important not only for statisticians but also for educated citizens (NCTM, 1989). Although many students can calculate the mean, the median and the mode, they have problems in using the values to analyze a given data (Suco, 2012). In this section of the literature review, some of the studies on the mean, the median and the mode are presented.

As mentioned in the previous section, there is a distinction between procedural knowledge, which includes computational algorithms, and conceptual knowledge of mathematical concepts in mathematics education (Strass & Bichler, 1988). It is stated that although the computational algorithm of the mean concept is easily understood, precollege and college students have problems in understanding the properties of the mean related to the computational algorithm (Pollatsek, Lima, & Well, 1981; Strauss & Bichler, 1988). For instance, a study by Strauss and Bichler (1988) concentrated on the computational properties of the mean. The study was conducted with twenty children of 8-14 ages. The purpose of the study was to test children's understanding of arithmetic average by investigating their understanding of its computational properties. To illustrate, one of the tested properties was whether the children realized that the arithmetic average has to be placed between the extreme values in a data set. Another one was whether the students realized that the average is affected by values except the average. They also tested whether the students realized that the average may not be equal to one of the values in the data set and whether the children were taking into account zero in a data set while calculating the average of the data set. The results of the study indicated that most of the students were not aware that zero as a data value

influenced the mean. On the other hand, the students were totally aware that the mean had to be between extreme values and the mean was affected by specific values in a data set and that the mean may not be equal to any one of the values in a data set. Similarly, Leon and Zawojewsky (1990) examined students' understanding of computational algorithm of the mean by testing the same properties as the previous study by Strauss and Bichler (1988). The participants of the study were from three educational levels: primary, middle and high. The results of the study indicated that the primary education students experienced more problems while solving questions which included four properties of the mean, but students' performance in solving the problems improved with age.

Additionally, other studies regarding procedural knowledge of the mean were conducted to investigate students' difficulties on weighted mean problems. Pollatsek, Lima and Well (1981) designed a study to examine how 37 undergraduate students calculate weighted mean while solving given problems. It was found that the students did not calculate weighted means of the problem situations correctly. The study also revealed that when the students were asked to find the mean of two unequal sized samples by using given means of the samples, they calculated the weighted mean of the sample as if the samples were equal sized. Pollatsek et al. (1981) stated that lack of conceptual understanding of the mean could be the reason for this result. In a similar way, Mevarech (1983) carried out a study to investigate students' calculation procedure of the weighted mean. According to the results of the study, the college students had a wrong assumption that a set of arithmetic averages with simple mean calculation procedure formed a mathematical group that satisfies four axioms, namely associativity, identity, closure and inverse. She stated that the misconception was so deeply rooted in the students' knowledge base that more advanced statistic course was not sufficient to overcome the misconception. However, most of the students formed appropriate schema of statistical concepts by means of diagnostic activities with feedback from an expert. Furthermore, the results of the study conducted by Watson and Moritz (1999) indicated that students were gaining competence to calculate the weighted mean by means of instruction with real life contexts.

On the other hand, several studies have been conducted to investigate students' conceptual understanding of averages. In one of the studies, Mokros and Russell (1995) investigated grades 4, 6 and 8 students' conceptions of averages by examining their preferences while proceeding with averages. Participants of the study consisted of 21 students, and data were collected by an instrument that included seven open ended problems. Four of the problems were described in their report. Two of these four problems were named as Allowance Construction Problem and Potato Chips Problem. In the problems, the students were expected to find a possible data set for the given mean by working backward. For example, in the Potato Chips problem, the mean of the data set was given as $1.35. Then, students were asked to find a possible data set which had that mean without using the actual mean as a data value. At this step, the students were supposed to work backward to find the data set. The Allowance problem was different from the Potato Chips problem in that the Allowance problem included data points on a bar graph and asked weighted mean of the data set. The remaining two problems required the interpretation of a given data set and calculation of weighted means. Based on the results of the study, five different constructions of representativeness by students were identified: average as mode, average as algorithm (arithmetic mean), average as reasonable, average as midpoint and average as mathematical point of balance. Another result of the study was that the students who focused on mode as average encountered problems while working backwards from the mean to construct a possible data set without using the mean as a data value because they were focusing on individual data values instead of the whole data set. Moreover, it was found that the students who focused on average as algorithm wanted to apply algorithm directly to find the mean of a given problem. The students did not take into account whether their answer made sense in reality because they over-trusted the algorithm. Similarly, a study by Cai (1995) found that although most of the students easily calculated the mean of a given data set, they had problems while working backwards to find missing values in a data set by using the given mean.

In addition to theoretical findings of Mokros and Russell's work, Konold and Pollatsek (2002) added new terms to the literature on the average concept. They identified four conceptual terms for the mean concept, namely typical value, fair share, data reducer

and signal amid noise. They claimed that the mean as "signal amid noise" is the most important one from a statistical perspective because the conception is the most useful while comparing two different data sets. Moreover, they indicated that mean as typical value and fair share are not useful while comparing two data sets and mean as data reducer and typical value are helpful for decision making in statistics. The researchers suggested that mean should be introduced as fair share and typical value because the conceptions can be built on students' major intuitions. Moreover, mean concept should be introduced to students in a context that include comparison of two data sets. Similarly, Watson and Moritz (1999) suggested that concepts of average should be introduced with activities that involve comparison of two equal sized data sets.

With regard to the effect of conceptual knowledge on procedural knowledge, Cai (1998) designed a study to examine US students' conceptual understanding of mean by investigating their solutions of a given contextualized average problem that required computational algorithm of mean. The study results revealed that students used different representations, which were algebraic, arithmetic, verbal and pictorial representations and solution strategies to solve the problem. In fact, the students' mathematical problem solving abilities affect their selection of solution strategies and representations. Moreover, students using arithmetic and algebraic representations experienced more problems than students using pictorial and verbal representations to solve the average problem. In addition, the students used three different solution strategies to solve the given average problem, that is *levelling, using average formula* and *guess and check*. Furthermore, five different types of errors were identified based on the findings of the study. The error types were *minor errors, violation of stopping rule, incorrect use of computational algorithm, unjustified symbol manipulation* and *others*. More specifically, a student who made the incorrect use of computational algorithm error tried to apply computational algorithm of the mean but the application was incorrect. Furthermore, a student who made the minor error applied correct procedure to solve a problem regarding the mean but the student made a minor calculation error. The results of the study also indicated that half of the students made mistakes while applying the computational algorithm to solve the problem although they knew the algorithm. The

study results also indicated that the average concept should be taught both as a statistical idea and computational algorithm.

In a further study, Cai (2000) conducted a cross national study to explore Chinese and US students' understanding of averaging computational algorithm and their representations while solving questions regarding the arithmetic average. Data were collected by an instrument that included two open-ended problems involving arithmetic average. One of the open-ended problems gave data set on a pictograph. It was found that the number of Chinese students who solved the problems was more than the number of US students who solved them. He also stated that the difficulties were caused due to their lack of conceptual knowledge of the averaging algorithm. Furthermore, it was found that Chinese students had a higher tendency to use algebraic representations, while US students had a higher tendency to use verbal and pictorial representations to solve the problems. Besides, while solving the problems, the students used three different strategies, namely levelling, using average formula and guess-and-check. Furthermore, in addition to observed errors in the study by Cai (1998), two different errors was observed, namely total number of hats sold in four weeks and looking a pattern for the number of hats sold. Also, Cai (1998) included same question with Cai (2000), which gave a data set on a pictograph.

Additionally, a longitudinal study of students in Grades 3 to 9 was conducted by Watson and Moritz (2009) to investigate the development of their concepts of average. These researchers developed an instrument by examining other studies on the same issue and the instrument included some questions displaying similarities with those used in the study of Mokros and Russell (1995). In addition, the instrument included different questions that measured students' interpretation of average in a real life situation, such as "Have you heard of the word average before? Where? What does it mean?" In this study, 94 interviews were conducted with students from Grades 3 to 9, and follow-up interviews were conducted with 22 of 94 students after three years and 21 of the remaining students after four years. Results of the study indicated that students' ability to apply the calculation rule of arithmetic average increased with their grade level. On the other hand, the students' interpretation ability of arithmetic average demonstrated limited

development with grade. Similarly, the present study investigated students' procedural and conceptual knowledge through identifying their solution strategies, errors and wrong interpretations.

Additionally, Mathew and Clark (2007) carried out a study which aimed to examine students' development of knowledge of some main statistical concepts. The participants of the study were eight students who were successful first grade students in a college. It was found that students had process conception about the mean. Process conception is being able to base answers on computational formula of the 'the mean' when asked why "the mean" is used. The students accepted the mean as merely a process, not as a measure of central tendency.

According to Keneddy, Tips and Johnson (2008), some students have difficulties while trying to find the mean and median values of large data sets. Students think that the mean and the median are always close to each other. However, this situation is true when the data are not clustered at any extreme since when the data are clustered, the mean will come close to extreme values of the data set.

Another confusion related to statistical concepts is that many students use the terms 'mean' and 'average' interchangeably (Brown & Silver, 1989). Average includes all measures of central tendency; that is, mean, median and mode. Students also tend to use mean if the best representative of a data set is asked. They generally do not consider the median and the mode as an appropriate representative of the data set. The misconception can be overcome through counter-examples which shows that the median or the mode can be the best representative of a data set (Seftlage, 2007).

In this section, studies regarding students' understandings of the mean, the median and the mode have been presented. In the following section, in accordance with the purpose of the study, students' errors and difficulties in graphs will be presented since the present study investigated seventh grade students solution strategies, errors and misinterpretations given in bar graph representations.

2.5 Studies on Errors and Difficulties in Graphs

Studies on graphing is essential because graph comprehension is necessary not only in students' educational life but also in their daily life (Friel et al., 2001). Moreover, related literature indicates that students have difficulties and misconceptions regarding graph concepts (Tortop, 2011). Middle grade students tend to have misconceptions on interpretation of given graphs and selecting convenient graph type to represent data. On the other hand, secondary students tend to have misconceptions regarding measures of central tendency and interpretations of the values. The misconceptions are also observed in undergraduate statistics students (Capraro, Kulm, Hammer, & Capraro, 2002; Conners, McCown & Roskos-Ewoldsen, 1998). In addition to these, there were some studies in accessible literature, which investigated students' interpretation of the concepts of mean by means of graph (Cai, 1998; 2000; 2002; Mokros & Russell, 1995).

Additionally, some studies have indicated that sample size of data have an effect on students' representation of the data. For example, Nisbet (2002a) designed a study to investigate the effect of sample size on organization of data. The study was conducted with 10 seventh grade students. The instrument used in the study included two numerical sets of data; the first one of these included 10 pieces of data and the second one included 30 pieces of data. The results of the study indicated that students could easily represent smaller data sets than larger data sets. Similar results were found in another study by Nisbet (2002b). The participants of the study consisted of 20 students from grades 9 through 11. In a similar way, the instrument used in the study included two numerical data sets. One of them included 10 pieces of data and the other one included 30 pieces of data. Results of the study were consistent with those reported in his earlier study (Nisbet, 2002a). Thus, independent of their grade level, most of the elementary and high school students could easily represent smaller data sets.

On the other hand, there were also studies which indicated that middle grade students tended to have misconceptions concerning interpreting given data and selecting the convenient type of graph for the given data (Capraro, Kulm, Hammer, and Capraro, 2002). To illustrate, Capraro, Kulm and Capraro (2005) carried out a study with 134 sixth grade students to investigate the students' misconceptions and naïve conceptions

on interpretation and graphical representation of data. Instrument of the study consisted of an open-ended problem named as Vet Club. The problem was formed by asking 14 students' number of pets and their type of pets and the problem required the construction of a graph to help find the typical number of pets in the data set. After students constructed graphs, their explanations were asked in an interview to identify the causes of their choices of graph type and the typical number. The collected data of the study indicated that 46% of the sample had constructed bar graphs, 35% of them had constructed pie graphs and the others had constructed line graphs. According to the results of the study, while constructing a bar graph, the students indicated zero, which was a categorical variable in the given data, on the x axis of the graph not on the origin. Their explanation of the situation was that any number cannot be identified without some representation. Moreover, the students tended to draw histograms instead of bar graphs. In fact, while the students intended to draw bar graphs, they were drawing histograms. Thus, it was observed that the students could not distinguish between bar graphs and histograms. Furthermore, most of the students who created a pie graph to represent data omitted the zero value while determining the percentage of each piece in the pie graph because they believed that zero could not be accepted as a categorical variable. Additionally, students could easily provide correct answers to questions based on graphs if the information was clearly demonstrated. When the typical number of pets was asked in the problem, some of the students did not interpret their graphs to answer that question. Similarly, Bright and Friel (1998) found that students made errors while constructing and interpreting their graphs and some of them could realize their errors. In addition to these, some students experienced problems while translating from a representation to another. Results of TIMSS indicated that when a data set was given on a table, 30% of third grade students and 55% of fourth grade students correctly represented the given data set on a bar graph (Beaton, Mullis, Martine, Gonzalez, Kelly, & Smith, 1996).

It was observed that some students sometimes make errors and hold misconceptions on interpretation of graphs due to cognitive obstacles. They generally did not know the reason for selecting a certain type of graph to represent a graph (Friel et al., 2001).For example, students create bar graphs without knowing the underlying reason for selecting

the graph type. In addition, students tend to find patterns in graphs (Piereira-Mendoza & Mellor, 1991) and students' language and reading ability affect their interpretation of graphs (Friel & Bright, 1996, as cited in Tiefenbruck, 2007).

In addition to studies related to the construction and interpretation abilities of students, some studies were conducted to investigate students' reading and interpretation abilities of graphs (Carr & Begg, 1994; Pereira-Mondeza & Mellor, 1991;Watson & Moritz, 2001). For instance, Pereira-Mendoza and Mellor (1991) conducted a study with 121 fourth grade and 127 sixth grade students to examine the students' ability in reading and interpretating bar graphs. It was found that the students had few problems regarding reading data from graphs. However, most of them had problems while interpreting data from graphs. Furthermore, the study results proposed that there were four general error types based on graphing, namely *topic, scale, data arrangement* and *the fact,* which were not demonstrated on the graph. In addition, students tended to find patterns by graphs and they stated that they could not answer questions if the answer was not on the graphical representations. For instance, if a question required making predictions about a future situation (the year is not on the graph), students say that "the year is not on the graph, so I cannot answer the question".

On the other hand, there were studies which investigated students' understandings regarding the mean concept by means of problems which were given on graphs (Cai, 1998; 2000; 2002; Mokros & Russell, 1995). For example, a study by Mokros and Russell (1995) investigated students' understandings of the average concept in statistics. As previously mentioned, instrument of the study included a problem which was named as the Allowance problem. The Allowance problem included data points on a bar graph and asked weighted mean of the data set. Participants of the study decided the mode of given data set based on the highest bar of the given bar graph representation. Furthermore, the study of Cai (1998) investigated students' conceptual understanding regarding the mean by analyzing their solution strategies and errors while solving a question. Data set of the question was given on a pictograph and a missing value of the data set was asked when mean of the data set was also given. As previously mentioned, students used leveling, average formula and guess and check solution strategies to solve

asked question. More specifically, students who used the leveling solution strategies benefited from the pictograph representation of asked question in the study. Thus, students might use different solution strategies while solving questions regarding the averages when the questions were given in graph representations.

One of the purposes of the present study was to investigate the solution strategies used by seventh grade students while solving questions regarding the mean, median and mode concepts given in bar graph representations. The other purposes of the present study were to investigate the possible errors and misinterpretations seventh grade students' would make while solving questions regarding the concepts of mean, median and mode given in bar graph representations. In accordance with the purposes, interpretation of graphs were important since purposes of the present study included interpretation of the mean, the median and the mode concepts given in bar graph representations. As mentioned above, in related literature, some studies were used graphs while investigating students' understandings of the mean concept since giving data set a set on a graph might provide different interpretations regarding the concepts of mean, median and mode (Cai, 1998; 2000; Mokros & Russell, 1995).

2.6 Studies in Turkey on Averages and Graphing

Although the importance of statistics in education and daily life situations has been emphasized in the national curriculum of Turkey (MoNE, 2013), there were few studies related to measures of central tendency and graphs in accessible literature of Turkey (Uçar & Akdoğan, 2009). More specifically, studies regarding statistical concepts investigating students' understanding of averages are also limited. However, there are research studies on graphs, which investigated students' interpretation abilities of graphs, their typical errors and misconceptions on graphing, and developing measuring tools for assessing graphing ability. In addition to these studies, measuring tools for assessing graphing ability of students were developed in Turkey (Taşar, İngeç & Güneş, 2009; Temiz & Tan, 2009b).

Uçar and Akdoğan (2009) conducted a study with 18 students from Grades 6 to 8 and 6 students were selected from each level to examine the effect of grade level on students'

understanding of average. Semi-structured interviews which included five questions were conducted to collect data. According to the results of the study, it was observed that average was understood as arithmetic mean and not accepted as a representative value for the given data set. Moreover, there was a positive correlation between students' understanding of average and students' grade level since 2 students from Grade 6 and4 students from Grade 8 accepted the mean as a summary statistics.

Another study was conducted by Tortop (2011) to examine the effect of instruction on seventh grade students' typical errors and misconceptions on graphs through questionnaire and interviews. The questionnaire was formed by the researcher. The instrument was administered to 71 seventh grade students and a teacher before and after instruction in Afyonkarahisar. After the questionnaire, semi-structured interviews were conducted with 8 students and the teacher. The results of the study showed that the teacher could not completely realize and prevent students' errors and misconceptions during instruction. Moreover, data from interview with the teacher indicated that she also had limited knowledge about students' possible errors and misconceptions on graphs, and data from observations in the instruction indicated that the teacher did not effectively plan her lesson based on the curriculum objectives of the concept. In addition to these results, the researcher proposed that in-service teachers should be educated on students' typical errors and misconceptions, and that effective planning should be made by taking into account students' possible errors and misconceptions.

2.7 Summary of the Literature Review

In this chapter, the literature review related to the purposes of the study was presented. In accordance with purposes of the present study, first of all, definitions of conceptual and procedural knowledge were stated. Then, definitions of error and averages were included. In addition, studies on students' understandings of the mean, the median and the mode, and studies on student errors and difficulties from related literature were reviewed. Lastly, the studies in Turkey on the averages and graphs were presented.

Results of related studies regarding the mean, the median and the mode showed that although computational algorithm of the mean was easily comprehensible, many pre-

college and college students had problems regarding properties of the algorithm (Pollatsek et. al., 1981; Strauss & Bichler, 1988). Furthermore, there were some other studies which indicated that students have difficulties on weighted mean problems (Pollatsek et. al., 1981; Mevarech, 1983). It is reported that the difficulties students face regarding the averaging algorithm derive from students' lack of conceptual knowledge rather than their lack of procedural knowledge (Cai, 2000). On the other hand, identifying students' possible solution strategies while solving questions regarding the mean is important since if a student have conceptual understanding regarding the averaging algorithm of the mean, the student can correctly and flexibly apply the averaging algorithm to solve questions regarding the concept (Cai, 1998). However, there is a limited number of studies on the solution strategies used by students while solving questions regarding statistical concepts (Cai, 2000). Therefore, in order to identify students' conceptual and procedural understanding of the mean, the median and the mode concepts, study on their solution strategies and errors while solving questions regarding these concepts is necessary. Besides, in order to have information regarding effect of graphical representations on interpretation of the average concepts, it is worth to study on interpretation of the concepts of mean, median and mode given in graph representations. Hence, it was investigated in the present study students' interpretations of the mean, the median and the mode concepts given in bar graph representations.

Additionally, literature review showed that there are few studies related to averages and graphing in Turkey. Results of a study by Uçar and Akdoğan (2009) showed that averages was understood as arithmetic mean and not accepted as representative of a given data by elementary students. Moreover, Uçar and Akdoğan (2009) also stated that there is necessity to study on students' solution strategies while solving questions regarding the mean, the median and the mode concepts. Furthermore, graph interpretation was also important part of statistics in Turkish mathematics curriculum. Also, graphical representations might affect students' interpretations of average concepts. Thus, when the accessible related literature in Turkey is investigated, it can be seen that there is a limited number of studies with similar purposes, so there is need to conduct further studies on the topic. Hence, the purpose of the present study was to investigate the solution strategies used by seventh grade students and their possible

errors while solving questions regarding the concepts of mean, median and mode given in bar graph representations. In addition, students' possible misinterpretations regarding the concepts of mean, median and mode given in bar graph representations was investigated.

CHAPTER 3

METHOD

The aim of this chapter is to explain the methodology used in the study. Hence, this chapter is devoted to providing information about the research design, population and sample, data collection instruments, validity and reliability of the instruments, data collection procedure, analysis of data, assumptions and limitations. At the end of the chapter, the internal and external validity of the study is presented.

3.1 Design of the Study

The aim of the present study was to investigate the solution strategies used by seventh grade students and their possible errors while solving questions regarding the concepts of mean, median and mode given in bar graph representations. In addition, the other purpose of the study was to investigate seventh grade students possible misinterpretations regarding the concepts of mean, median and mode given in bar graph representations.

In order to reach these purposes, the survey research design was preferred in this study because surveys are conducted to define some aspects and characteristics of a population or a sample (Fraenkel & Wallen, 2006). More specifically, the cross-sectional survey design was used in the study since the cross-sectional survey requires collecting data at one point of time from a selected sample from a predetermined population to describe the characteristics of the population (Fraenkel & Wallen, 2006). The collected data were analyzed through item based in-depth analysis to identify solution strategies and errors,

and descriptive statistics were computed since the research questions of the study aimed to identify the frequencies of the students' solution strategies and errors.

3.2 Participants of the Study

The target population of this study was all seventh grade students in Çanakkale. Moreover, the accessible population of the study was all seventh grade students in the Gelibolu district of Çanakkale. The convenience sampling method was used in this study since convenience sample is a class of readily available individuals (Fraenkel & Wallen, 2006). The sampling procedure was selected since the researcher was a middle school mathematics teacher at a public middle school in the Gelibolu district. Thus, the study was conducted at this school and at another central school in the district. Accordingly, the Statistics Achievement Questionnaire was conducted to 233 students in the Gelibolu district, Çanakkale. All seventh grade students of these 2 public middle schools in Gelibolu district completed the Statistics Achievement Questionnaire. One of these schools had three classes and the other had five classes. The selected seventh grade students had completed each objective which were intended to examine in the study. Demographic information of the participants, such as their class, age and gender was asked for while collecting data. In general, the number of boys was more than the number of girls in the sample, and the average age of the participants was 11.97. Details of the demographics are presented in Table 3.1 below.

Table 3.1 Participants' Demographic Information

Classes	Sample size(n)	Age (years)	Gender	
		Average	Boys	Girls
7-A	32	11.88	17 (53.1%)	15(46.9%)
7-B	33	11.94	15 (45.5%)	18 (54.5%)
7-C	32	11.84	18 (56.3%)	14 (43.7%)
7-A	29	12.03	17 (58.6%)	12 (41.4%)
7-B	27	11.96	16 (59.3%)	11 (40.7%)
7-C	27	12.11	14 (51.9%)	13 (48.1%)
7-D	25	12.08	15 (60.0%)	10 (40.0%)
7-E	28	11.89	15 (53.6%)	13 (46.4%)
Total (N)	233	11.97	127 (54.5%)	106 (45.5%)

3.3 Data Sources

This study investigated the solution strategies used by seventh grade students while solving questions related to the concepts of mean, median and mode given in bar graph representations. Moreover, the study also aimed to examine possible errors and misinterpretations made by seventh grade students related to the concepts of mean, median and mode given in bar graph representations. Data for the study were gathered through the Statistics Achievement Questionnaire (SAQ).

3.3.1 Statistics Achievement Questionnaire

In the present study, data were collected through an instrument that was developed by the researcher. The instrument was an achievement test. The test was prepared in three phases. Firstly, the objectives of seventh grade Turkish National Middle School Mathematics Education Curriculum related to bar graphs and the concepts of mean, median and mode were identified. Furthermore, the objectives of fifth and sixth grade Turkish National Middle School Mathematics Education Curriculum were investigated in order to define students' previous knowledge related to the mean, median, mode

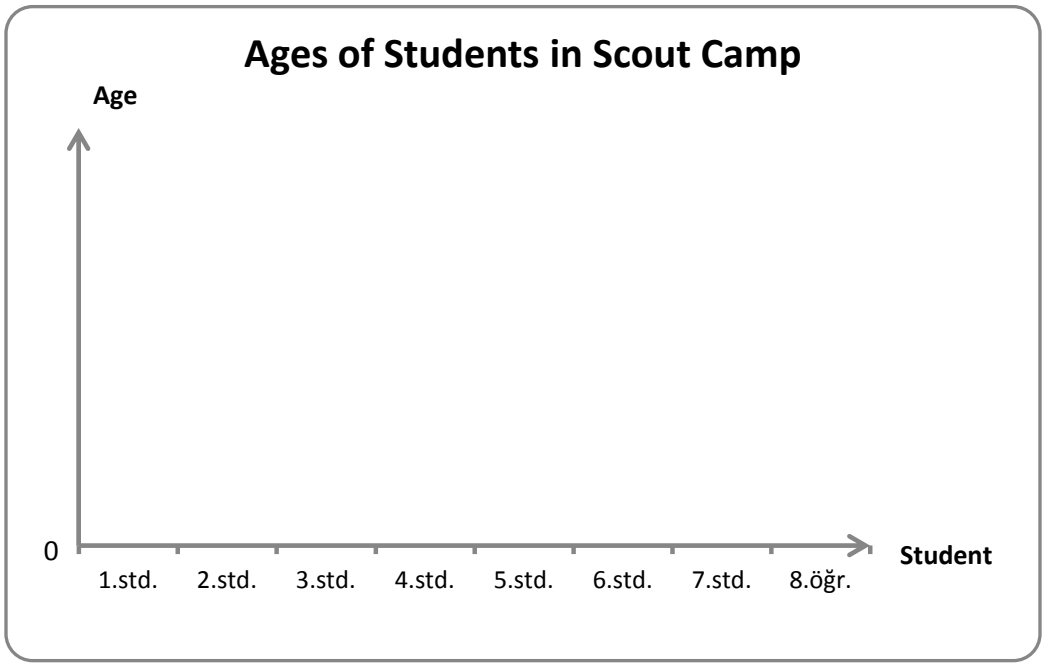
concepts and bar graphs. Secondly, related literature was reviewed before the achievement test was prepared. Based on the literature review some questions were adapted from literature. Lastly, the researcher prepared additional items based on the objectives mentioned above. While the researcher was developing the items of SAQ, each question was matched with the objectives in order to ensure that there was at least one item measuring each objective from the selected objectives of the seventh grade mathematics curriculum.

The Statistics Achievement Questionnaire (SAQ) consisted of 6 open-ended questions. In this test, questions 1-a, 1-b and 2-b were adapted from literature. The remaining questions were developed by the researcher to investigate the research questions of the study. Detailed information related to each question of SAQ are given below.

Question 1, which was adapted from study of Mokros and Russell (2005) consists of two parts. In the question, students' understanding of the relationship between data and average were investigated. In the first part of the question, the average and sample size of a data set were given and a possible data set for the given information was required. To state differently, the purpose of the question was to investigate students' procedural and conceptual knowledge regarding the averages by constructing a data set for the identified average. In addition, students were expected to draw a bar graph according to their constructed data set. The first part of the 1st question is given in Figure 3.1 below.

Question 1:

a) In a scout camp, there are 8 students whose ages are different from each other. The average age of the students is 15. Based on the information, what could their ages be? Draw a bar graph in order to show a possible data set of the ages of the students in the camp. (Note: None of age of the students in the camp is not 15.)



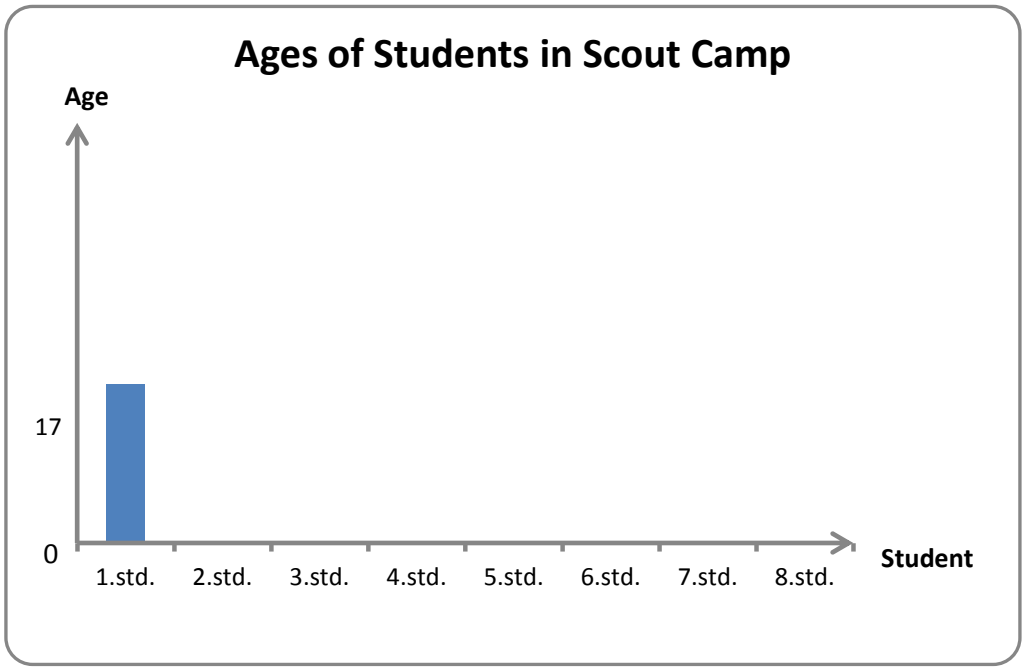
How did you give your answer? Explain.

Figure 3.1 The first part of question 1 of the SAQ

In the second part of the 1st question, average of a data set was given; additionally, a value from the data set was also given. Thus, while constructing a possible data set for the average, students had to take into account the given value. This part evaluated students' understanding of the relationship between the average concepts and data through construction of a possible data set for the given average by taking into account the given value from the data set. In addition, the part also required students to draw a bar graph according to their constructed data set. The second part of the 1st question is given in Figure 3.2 below:

Question 1:

b) In a scout camp, the average age of 8 students is 15 and if the age of a student is 17, what could be the ages of the other students so that the average will still turnout to be 15? Draw a bar graph to show a possible data set of the students' ages in the camp.(Note: None of age of the students in the camp is not 15.)



How did you give your answer? Explain.

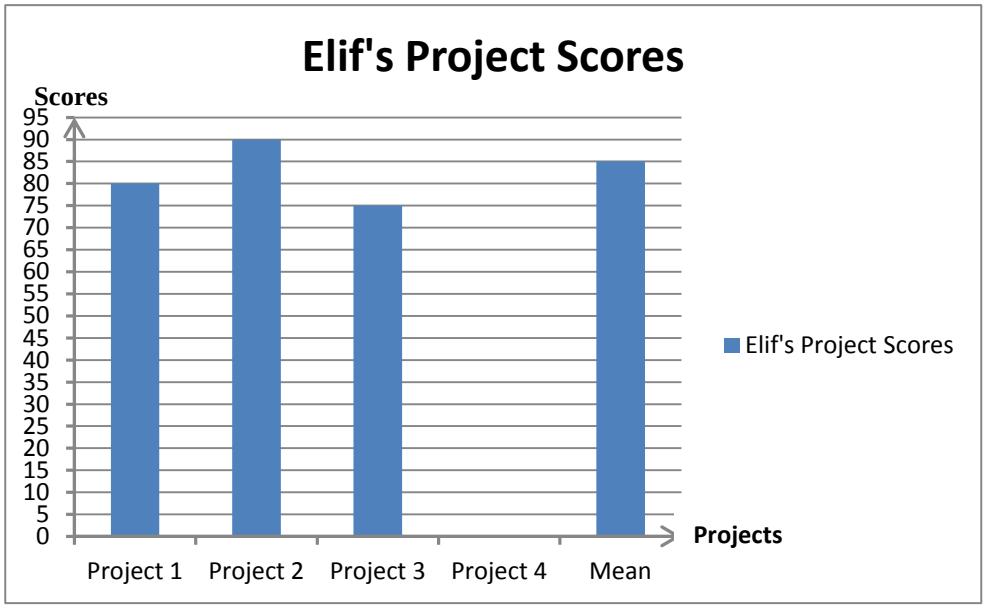
.....

Figure 3.2 The second part of question 1 in the SAQ

Question 2 in the SAQ consists of four parts. The second part of the question was adapted from Connected Mathematics 2/Grade 6 (Lappan, Fey, Fitzgerald, Friel, & Philips, 2006). The remaining parts of the question were developed by the researcher. In general, the three scores of four projects and the arithmetic average of all projects were given on the bar graph, as seen in Figure 3.3, and the question asked for the remaining score of the four projects. The first part evaluated students' reading ability of bar graphs. In the second part, the arithmetic average and three out of four values from the data set were given. This part evaluated students' procedural knowledge related to the calculation

of the arithmetic average. In the third part, a statement was given and students were expected to decide whether the given statement was true or not. In the last part, students' interpretation of the arithmetic average was investigated.

Question 2:



The bar graph shows Elif's scores on three of these projects. In the graph, a bar that shows Elif's mean score for all four projects.

a) Fill the below table according to given graph.

1. project score	
2. project score	
3. project score	
Mean of the project scores	

b)Find Elif's score on Project 4. Explain your reasoning.

c) Elif claims that If I had 20 points higher than 3rd project score, mean of my project scores would increase 4 points. Do you agree with Elif? Explain your reasoning.

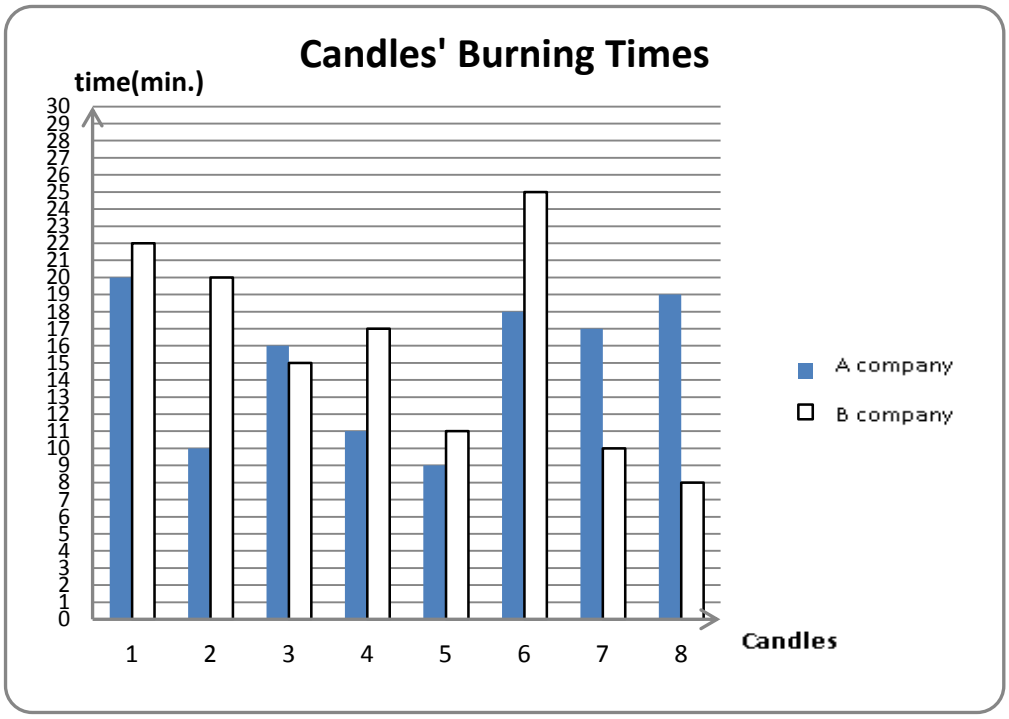
d) If we did not know any of the Elif's project scores and we knew only mean of the projects' scores; how could we comment on Elif's project scores?

Figure 3.3 The second question of the SAQ

Question 3 in the SAQ, developed by the researcher, consists of three parts. In the first part, which is presented in Figure 3.4, students' abilities to calculate the mean through the obtained data set from the given bar graph were examined. Similarly, the second part

evaluated students' abilities to find the median through the same data set. The third part investigated students' interpretation of averages based on the given statement. In addition, the last part aimed to reveal information about students' conceptual understanding regarding the mean and the median concepts since they were required to justify their reasoning. Each part of the 3rd question is given in Figure 3.4 below:

Question 3:The burning durations of the candles produced by companies A and B are wanted to be compared. In order to compare the burning durations of the candles of Company A and Company B, the same size 8 candles from each company were chosen. Then, the candles were burned and the number of minutes that each candles burned were recorded.



a) Determine the mean of burning times of candles for each company.

<u>Mean</u>

A Company ….....

B Company ……

b) Determine the median of burning times of candles for each company.

<u>Median</u>

A Company ….....

B Company ……

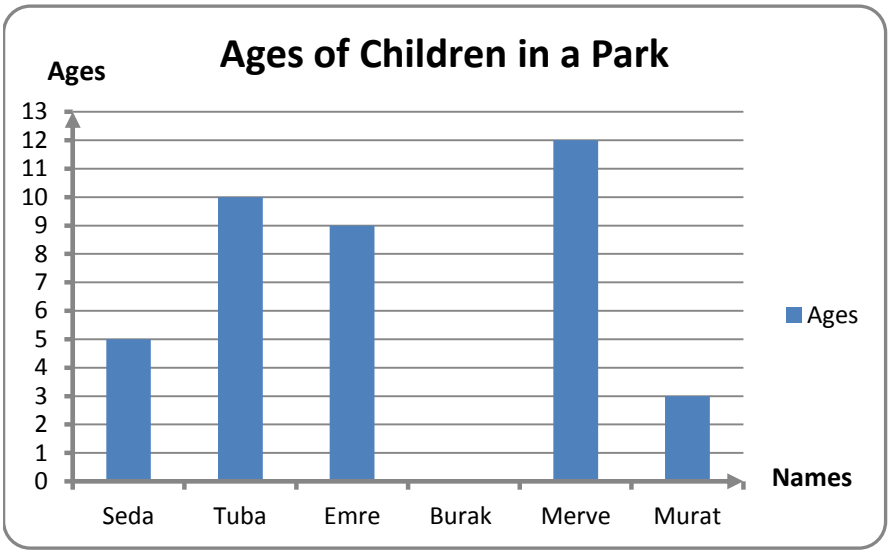
c) Ahmet claims that candles of A Company have longer burning times than candles of B Company. Do you agree with Ahmet?.....

- How did you decide whether Ahmet is right or not? Explain your reasoning.

Figure 3.4 The third question of the SAQ

Question 4 in the SAQ, generated by the researcher to investigate students' conceptual understanding of the mode and the average concepts, consists of three parts. Additionally, the question examined students' reading and interpretation abilities of bar graphs. In the first part, students' understanding of the mode was examined. The second part measured students' procedural knowledge related to the average concept. Moreover, in this part, it was also aimed to reveal information regarding students' understanding of the average concept since it was predicted that when the average is asked, most of the students understood the term arithmetic average. However, the term " average" includes the arithmetic average, the median and the mode concepts (Van de Walle, 2013). In the third part, it was examined how students interpret average of a data set when a value was subtracted from the data set. Each part of 4th question is given in Figure 3.5 below:

Question 4:



The above bar graph shows the ages of 5 children out of 6 children in a park. Answer the following questions according to the above graph.

a) In this park, if the number of nine aged-children is higher than the others, how old is Burak? Explain your reasoning?

b) What is the average of the ages of the children in the park?

c) Seda claims that if Murat leaves the park, the average age of the children will decrease. Do you agree with Seda? .....

- How did you decide whether Seda is right or not? Explain your reasoning.

Figure 3.5 The forth question of the SAQ

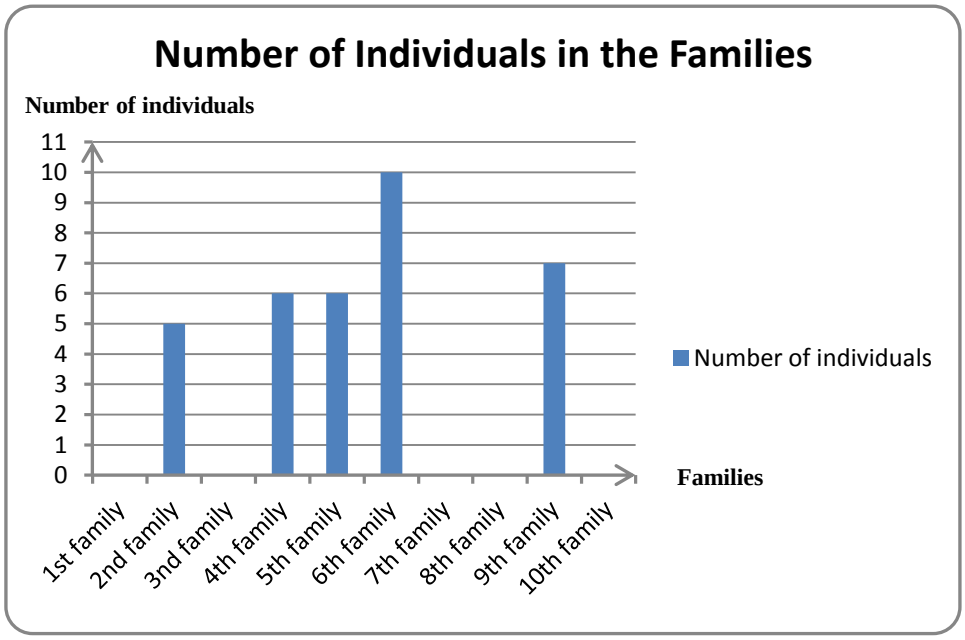
Question 5 examined students' interpretation abilities regarding the median and the mode by constructing a possible data set for the given information. Additionally, the question assessed students' reading and interpreting abilities of bar graphs. In the question, the mode, the median and the range of a data set were given. In addition, as shown in Figure 3.6, some values of the data set were given on a bar graph and some values were missing. The question required students to find a possible data set by using the given information and to fill the missing parts of the bar graph. Question 5 was also developed by the researcher.

Question 5:

There are 10 students in a class. Use the information about the number of individuals in the students' families below.

- Mode: 6 individuals
- Median: 6 individuals
- In the data set, there are families having at least 3, at most 10 individuals.

Find a possible data set by filling the missing parts of the bar graph below by using the above information and the given values in the bar graph.



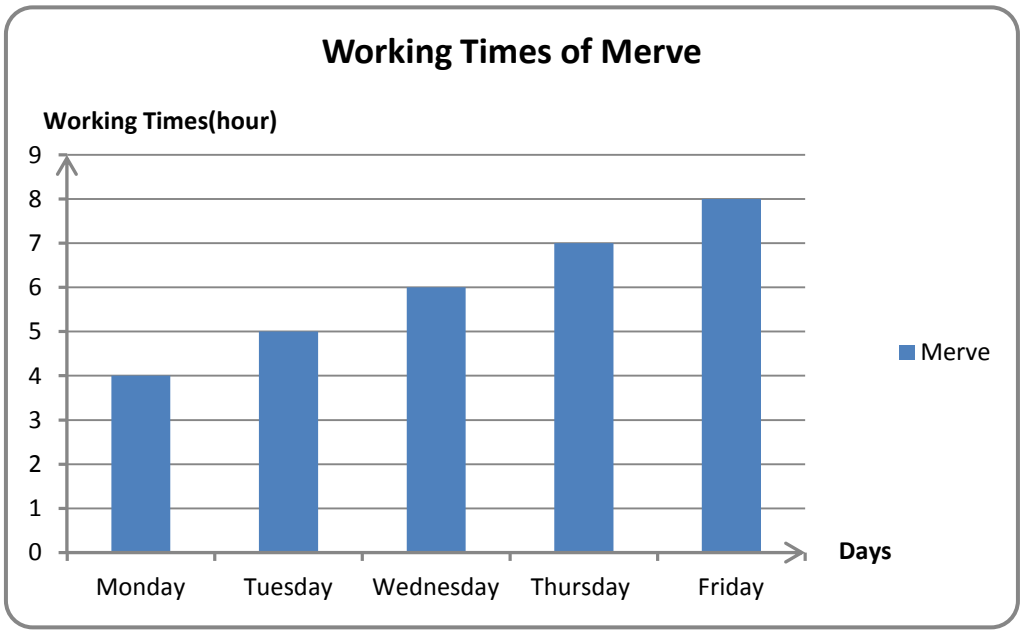
- How did you construct your possible data set? Explain.

Figure 3.6 The fifth question of the SAQ

In question 6, the students were asked if the working days of two employees and the arithmetic average of their working times were the same, what the working times of the other employee were, and whose working times was not given. In order to answer the question, a possible data set would be constructed by taking into account the given average and the bar graph. The purpose of the question was to investigate students' understanding of the relation between the arithmetic average and a data set. Additionally, students' ability to calculate and interpret arithmetic average was examined. The 6th question is given in Figure 3.7 below:

Question 6:

Merve and Seda are two employees in the same workplace. The working durations of Merve is given in the graph below.



If the working days of Seda and Merve and the arithmetic average of their working durations are the same for a week, what could be the working durations of Seda? Show by drawing a bar graph in the area below. (Note: Seda and Merve have different working durations on the same days.

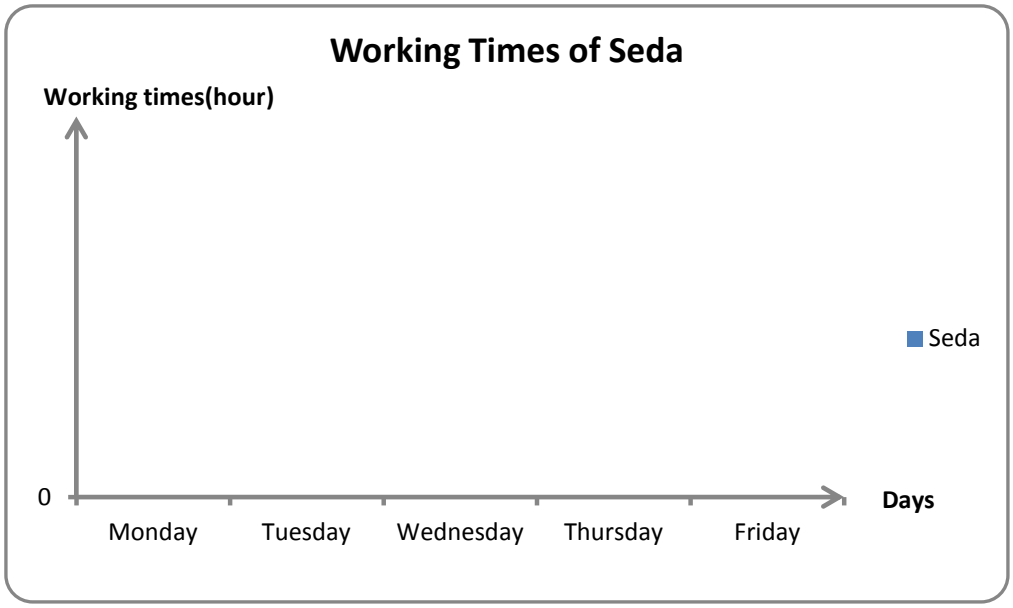


Figure 3.7 The sixth question of the SAQ

To sum up, totally 6 questions were asked in the SAQ. Details regarding the pilot study of the SAQ are explained in the following part of the chapter.

3.4 Pilot Study

The pilot study was implemented to determine the appropriate testing time duration for the implementation of SAQ, to adjust the difficulty level of the questions, to control the comprehensiveness of each question and to check the validity and reliability of SAQ.

The pilot study of SAQ was conducted by the researcher in a middle school at the Gelibolu district, Çanakkale during the 2013-2014 fall semester. Seventy two eighth grade middle school students who had reached the identified objectives of the study in their previous semester took the test including six questions with sixteen sub questions In the pilot study, the students were given forty five minutes to answer all the questions. However, it was noticed that they needed more time than fifty five minutes to complete the test. Therefore, the students were given sixty minutes in the actual study. Additionally, the pilot study results indicated that names of bars in given bar graph in question 3 of the SAQ were incorrect. The last form of SAQ is presented in Appendix C.

3.5 Validity and Reliability of the Instrument

Validity is "the appropriateness, correctness, meaningfulness, and usefulness of the specific inferences researchers make based on the data they collect" (Fraenkel & Wallen, 2006, p.151). In order to ensure validity of the instrument, content related evidence was provided. Before the pilot study was conducted, the SAQ was examined by three experts from the Elementary Mathematics Education Department of different universities to provide content related evidence of validity of the instrument. The experts checked the questions of the instrument based on the table of specification in terms of appropriateness of each question and their objectives. The experts checked whether or not a question met the intended objectives. Moreover, the experts evaluated the appropriateness of the questions in terms of comprehensibility for seventh grade students and level of difficulty. In addition, an Middle School Turkish Teacher checked the language of the items and two graduate students in Elementary Science and Math

Education at METU also gave opinion related to the comprehensibility and appropriateness of the questions for seventh grade students. The table of specification of the SAQ questions is presented in Table 3.2 below:

Table 3.2 Table of Specification for the SAQ Items

	Objectives
Mean (Arithmetic average)	Students are able to calculate the mean of data. 1a, 1b, 2b, 2c, 3a, 4b, 6
	Students are able to interpret the mean of data. 1a, 1b, 2d, 3c, 4c, 6
Median	Students are able to calculate the median of data. 3b, 4b
	Students are able to interpret the median of data. 3c, 5
Mode	Students are able to calculate the mode of data. 4a, 4b
	Students are able to interpret the mode of data. 4a, 5

Reliability is "the consistency of the scores obtained–how consistent they are for each individual from one administration of an instrument to another and from one set of items to another" (Fraenkel & Wallen, 2006, p.157). In order to check the reliability of the SAQ questions, the inter-rater reliability was computed since the questions are open-ended and they were scored according to the rubric. The responses of 35 seventh grade students' answers were scored by a graduate student in mathematics education as a second rater. A correlation of 98% was found between the two scorings.

3.6 Data Collection Procedure

The purpose of this study was to investigate the solution strategies used by seventh grade students in questions regarding the concepts of mean, median and mode given in bar graph representations. Another goal of the study was to examine the students' errors and misinterpretations regarding the concepts of mean, median and mode given in bar graph representations. At the beginning of the data collection procedure, the official permissions were received from the Middle East Technical University Human Subjects Ethics Committee (see Appendix A). Then, the necessary permissions were obtained from the Ministry of National Education (see Appendix B) for the administration of SAQ in the identified public schools. Following these permissions, two mathematic teachers, who were the actual teachers of the participants of this study, were informed about the purpose and the procedures of this study.

Subsequently, SAQ was administered to 233 middle school seventh grade students during the spring semester of the 2013-2014 academic year. The data were collected in this time period because the teachers had just completed teaching the statistics concepts. Data from both schools were collected at the same time. Thus, SAQ was administered not only by mathematics teachers but also by the teachers from different branches. Before the administration process, teachers were informed about purpose of this study and procedures of this administration process. More specifically, in one of the schools, one mathematics teacher, one Turkish teacher and the researcher administered the instrument at the same time. Furthermore, in the other school, the SAQ was administrated by a mathematic, a science, an English, a religion and an ethics teacher, and the researcher at the same time. To reach the purpose of the study, before the administration of SAQ in each classroom, the researcher emphasized to the participants that it was of utmost importance to provide explanations to the methods of solution used to answer each question. Totally, 60 minutes was given to the students for the completion of the Statistics Achievement Questionnaire. The instrument was implemented in students' own classrooms since their classrooms had similar conditions. The time schedule for the data collection procedure is presented in Table 3.3 below.

Table 3.3 Time Schedule for the Data Collection Procedure

Date	*Events*
January - November 2013	Development of the instrument of the study
December 2013	Pilot study-Revision of the instrument
March 2014	Implementation of the instrument
April- March 2014	Analysis of the data

3.7 Analysis of Data

In order to answer the research questions of the study, item based analysis was conducted. More specifically, the rubric was developed by the researcher to identify the solution strategies used and the errors made by the participants. The six items of SAQ were scored using the prepared scoring system (0 to 4 points). The rubric, which is presented below, was prepared by taking into account the participants' responses for each item. Since some of the items required finding an average type and some of them required the construction of a possible data set based on the given averages, a single rubric could not be used to analyze all the items of SAQ. However, there were general groups in this rubric. Firstly, four points was given to correct solutions with acceptable solution strategies; secondly, three points was given to partially correct answers; thirdly, two points was given to correct answers without explanation; one point was given to incorrect answers with missing or an unacceptable solution strategy and lastly, no point was given when an item was left blank or an irrelevant answer was given. Details of the rubric are presented in Table 3.4.

Table 3.4 Scoring Rubric for Open-Ended Questions

Codes	Answer Types
0	• Blank • Irrelevant answers • No interpretation
1	• Incorrect answer with missing or unacceptable solution strategy • Constructed an inappropriate data set with missing or an unacceptable solution strategy with erroneous or correct drawing • Incorrect answers due to operational error • Misinterpretation
2	• Correct answer without explanation • Constructed appropriate data set without explanation with erroneous/correct drawing
3	• Partially correct answer with an acceptable solution strategy with erroneous/correct drawing
4	• Correct answer with acceptable solution strategy • Constructed appropriate data set with an acceptable solution strategy with erroneous/correct drawing • Correct interpretation

In order to answer the research questions of the study, participants' correct answers were analyzed in the second part of the analysis. More specifically, correct answers with acceptable solution strategies, which were coded as 4, were analyzed in order to determine the solution strategies used by participants while solving the questions regarding the mean, the median and the mode given in bar graph representations. However, partially correct answers with acceptable solution strategies, which were coded as 3, were not analyzed to identify used solution strategies to solve questions of the SAQ since they were not accepted as totally correct answer. Before identifying students' solution strategies, frequencies of students who gave correct answer with acceptable solution strategies for item 1-a, 1-b, 2-ab, 2-c, 3-a, 3-b, 4-a, 4-b, 5 and 6 were determined based on analysis of 233 students' answers and are presented in Table 3.5 as follows:

Table 3.5 Frequencies of students' correct answers with acceptable solution strategy

Items	1-a	1-b	2-ab	2-c	3-a	3-b	4-a	4-b	5	6
Frequencies of correct answers	74	63	88	75	94	53	128	107	28	65

Then, their correct answers were analyzed to identify students' solution strategies regarding the concepts of mean, median and mode given in bar graph representations. For mean concept, using balance model, using average formula and guess and check solution strategies were identified based on the analysis of results of the present study. For median concept using numerical procedures, depending on graph representations and guess and check solution strategies were identified. Lastly, using numerical procedures, depending on graph representations and guess and check solution strategies were identified as solution strategies to solve the questions regarding the mode.

In the third part of analysis of present study, participants' incorrect answers with unacceptable solution strategies and their misinterpretations, which were coded as 1, were analyzed in order to identify 233 participants' errors and misinterpretations regarding the concepts of mean, median and mode given in bar graph representations,. Before identifying students' errors and misinterpretations, frequencies of students who gave incorrect answers with unacceptable solution strategies for items 1-a, 1-b, 2-ab, 2-c, 3-a, 3-b, 4-a, 4-b, 5 and 6 and made misinterpretations for items 2-d, 3-c and 4-c were determined based on analysis of students answers and are given in Table 3.6 as follows:

Table 3.6 Frequencies of students' incorrect answers for each item

Items	1-a	1-b	2-ab	2-c	2-d	3-a	3-b	3-c	4-a	4-b	4-c	5	6
Frequencies of incorrect answers	71	49	78	49	48	62	94	30	21	59	80	76	98

After the above frequencies were determined, students' incorrect answers for each item were analyzed to identify their errors regarding the concepts of mean, median and mode given in bar graph representations. In order to identify the students' errors regarding the mean concept, items 1-a, 1-b, 2-ab, 2-c, 3-a, 4-b and 6 were analyzed. According to the results, operational error, finding total number, incorrect usage of averaging algorithm, accepted close values, looking pattern and turned to find smaller or larger numbers, not meeting all requirements of a problem, forming incomplete data set, incorrect reading of values from the graph were identified as errors regarding mean concept. Secondly, according to analysis of students' incorrect answers, errors regarding the median concept were observed in the items 3-b and 5. Results of the analysis showed that students' identified errors regarding median were operational error, wrong decision on unordered data set, wrong decision on ordered data set, wrong interpretation of graph, forming incomplete data set, incorrect reading of values from graph, and incorrect largest and/or smallest data values. Lastly, errors regarding mode concepts were observed in items 4-a and 5. According to analysis of the results, students’ three error types were identified, namely inappropriate usage of averaging algorithm, forming incomplete data set and incorrect largest and/or smallest data values. Besides, misinterpretations, which were also coded as 1, were analyzed from items 2-d, 3-c and 4-c in order to determine participants' misinterpretations related to the concepts of mean, median and mode given in bar graph representations.

3.8 Assumptions and Limitations of the Study

In this section, assumptions and limitations of the study are discussed. First, the study was based on several assumptions. It was assumed that the participants of the study answered questions in the SAQ carefully, sincerely and willingly. It was also assumed that the participants' ages, levels of intelligence and socioeconomic backgrounds were similar. The last assumption was that the data collectors did not have any bias.

On the other hand, there were limitations due to the sampling method of the study, which was convenience sampling. Since the convenience sampling is not one of the random sampling methods, the selected sample could not be representative of a large population. According to Fraenken and Wallen (2006) "convenience samples cannot be considered representative of any population and should be avoided if at all possible" (p.100). However, Fraenkel and Wallen (2006) also explained that demographic information and other characteristics of subjects should be included in such cases. Thus, the results of the study were reported accompanied with demographic information of the participants. However, the results of the study could not be generalized to a large population. Additionally, the findings regarding the students' interpretation of the mean, the median and the mode concepts was limited with the questions of SAQ since when different questions were asked related to these concepts, different findings could be found. Furthermore, the findings of the present study were limited with the participants' ability of self-expression since items of the SAQ required answers of their solving process or their reasoning for given answer.

3.9 Internal Validity and External Validity

3.9.1 Internal Validity

Internal validity of a study refers to the differences observed in the dependent variable, which occur due to the independent variable (Fraenkel & Wallen, 2006). In survey research, location, mortality (loss of subjects) and instrumentation are defined as the main internal validity threats (Fraenkel & Wallen, 2006). Thus, the possibility of each threat was examined.

Location may be a threat for research studies when the location where the data are collected may affect the results of the study by producing different explanations (Fraenkel & Wallen, 2006). In the study, location was not a possible threat since the achievement test was implemented in the students' own classrooms, which are public schools and have similar conditions.

Mortality may be a threat to cause alternative explanations on results of the study when some participants of a study are lost in the study process (Fraenkel & Wallen, 2006). Since the achievement test was conducted at one point of time, mortality was not a possible internal validity threat for the study.

The instrumentation threat includes instrument decay, data collector characteristics and data collector bias (Fraenkel & Wallen, 2006). Instrument decay may be a validity threat for a study when the instrument or its scoring rubric is altered (Fraenkel & Wallen, 2006). In the present study, since data were collected through the instrument which had open-ended questions, a rubric was prepared and the papers were evaluated by two scorers using the rubric. Thus, the instrument decay threat was eliminated. Data collector characteristics may be a validity threat if results of a study produce different explanations due to characteristics of data collector(s) (Fraenkel & Wallen, 2006). The data of this study were collected by teachers from different branches. In order to control the threat, the data collection procedure was standardized and then the data collectors were informed about the implementation procedure of the instrument. Lastly, the data collector bias threat is a potential of data collectors' unintended distortion of the results of the study (Fraenkel & Wallen, 2006). In the present study, interaction between data collectors and the participants were not allowed during the application of the test. Thus, data collector bias threat was taken under control.

3.9.2 External Validity

Fraenkel and Wallen (2006) defined external validity as the degree to which the findings of a study can be generalized to other situations. Population generalizability and ecological generalizability are parts of external validity of a study. Population generalizability is the degree to which the sample of a study represents the target

population of the study (Fraenkel & Wallen, 2006). In this study, the results were not easily generalizable to the target population because the sampling method of the study was convenience sampling.

On the other hand, ecological generalizability of a study is defined as "the degree [of] generalizability of results of the study to other settings and conditions—the setting—under which a study takes place" (Fraenkel & Wallen, 2007, p.106). Thus, the results of the present study could be accepted as generalizable to the seventh grade elementary students who have similar conditions with the sample of the study.

CHAPTER 4

FINDINGS

In this chapter, the findings of the present study are presented in three sections based on the purposes of the study. The first purpose of the study was to investigate seventh grade students' solution strategies while solving questions regarding measures of central tendencies, that is the mean, the median and the mode concepts given in bar graph representations. Therefore, the first section of this chapter presents the findings obtained by analyzing students' correct answers in the Statistics Achievement Questionnaire (SAQ), which they had completed, to identify the solution strategies they resorted to while solving each question.

In addition, the second purpose of this study was to investigate seventh grade students' possible errors regarding the mean, the median and the mode concepts given in bar graph representations. Therefore, the second section of this chapter presents findings regarding the students' wrong answers, which were analyzed from the SAQ to identify their errors while providing answers to the questions regarding the concepts through bar graphs.

Lastly, the third purpose of this study was to investigate seventh grade students' possible misinterpretations regarding the concepts of mean, median and mode given in bar graph representations. Therefore, the third section of this chapter presents findings regarding the students' misinterpretations, which were analyzed from the SAQ to identify their misinterpretations while providing answers to the questions regarding the concepts given in bar graph representations.

4.1 Solution strategies

In order to identify the students' solution strategies, correct answers accompanied with an acceptable solution strategy were analyzed from the SAQ, which the students had completed. The students' solution strategies were categorized for each concept. More specifically, for the mean concept, *using balance model*, *using average formula*, and *guess and check* strategies were identified. Moreover, for the median concept, *using numerical procedures*, *depending on graph representations*, and *guess and check* strategies were identified. Lastly, for the mode concept, *using numerical procedures*, *depending on graph representations* and *guess and check* strategies were detected based on the analysis of solution strategies employed in the correct answers with an acceptable solution strategy. Table 4.1 presents frequencies of the solution strategies employed by the students, for each item regarding the mean, the median and the mode concepts.

Table 4.1 Frequencies (and percentages) of solution strategies employed by students providing correct responses

	Mean			Median			Mode		
STR. ITEMS	**Using Balance Model (%)**	**Using Average Formula (%)**	**Guess and Check (%)**	**Using Numerical Procedures (%)**	**Depend-ing on Graph Represen-tations (%)**	**Guess and Check (%)**	**Using Numerical Procedures (%)**	**Depending on Graph Representations (%)**	**Guess and Check (%)**
Item 1-a	23 (31.1%)	43 (58.1%)	8 (10.8%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Item 1-b	19 (30.2%)	38 (60.3%)	6 (9.5%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Item 2-ab	8 (9.1%)	60 (68.2%)	20 (22.7%)	-	-	-	-	-	-
Item 2-c	2 (2.7%)	73 (97.3%)	0 (0%)	-	-	-	-	-	-
Item 2-d	-	-	-	-	-	-	-	-	-
Item 3-a	0	94 (100%)	0	-	-	-	-	-	-
Item 3-b	-	-	-	50 (94.3%)	3 (5.7%)	0 (0%)	-	-	-
Item 3-c	-	-	-	-	-	-	-	-	-
Item 4-a	-	-	-	-	-	-	6 (4.7%)	122 (95.3%)	0 (0%)
Item 4-b	0 (0%)	107 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Item 4-c	-	-	-	-	-	-	-	-	-
Item 5	-	-	-	22 (78.6%)	0 (0%)	6 (21.4%)	22 (78.6%)	0 (0%)	6 (21.4%)
Item 6	4 (6.2%)	58 (89.2%)	3 (4.6%)	-	-	-	-	-	-

As can be observed in Table 4.1, three solution strategies were identified for the questions regarding each concept. Findings indicate that the majority of the students applying the appropriate solution strategy for questions regarding the mean used *using average formula* as a solution method. For example, 100% of the participants used *using average formula* solution strategy to solve item 3-c and 97.7% of them also used the solution strategy to solve item 2-c. On the other hand, *using numerical procedures*

solution strategy is the most frequently used solution strategy for questions regarding the median. To illustrate, 94.3% of the participants solved item 3-b by means of *using numerical procedures* solution strategy. Additionally, in item 4-a, which is related to the mode concept, 95.3% of the participants solved it with *depending on graph representations* solution strategy since data were given on a graph. Examples of students' appropriate solution strategies are presented below under the categories based on each type of measurement of central tendency question.

4.1.1 Solution Strategies for the Questions Regarding the Mean

Solution strategies to solve questions regarding the mean were addressed in items 1-a, 1-b, 2-ab, 2-c, 3-a, 4-b and 6 since the questions required calculation of the mean of a data set or construction of a possible data set for a given mean. When solution strategies of the students who answered correctly with an acceptable solution strategy were analyzed, three types of solution strategies were identified which were *using balance model*, *using average formula*, and *guess and check*.

Using Balance Model: The students who used using balance model as a solution strategy, accepted the mean as a point of balance. For example, in one question, the mean of a data set is 5, and three of four data of the data set are 2, 4, and 6. While solving the question, when a student says that the mean of 4 and 6 is 5, then other data is 8 because 2 is balanced with 8 to obtain 5 as mean of the data set. Therefore, in this solution strategy, a value from a data set is balanced with the other value in order to obtain the mean of the data set.

According to Table 4.1, using balance model was most frequently used solution strategy for item 1-a. It was seen that 23 students (31.1%) among 74 students used using balance model solution strategy to solve the item. Moreover, 19 students (30.2%) among 63 students also used using balance model as a solution strategy for item 1-b. In addition, 8 students (9.1%) among 88 students for item 2-ab, 2 students (2.7%) among 75 students for item 2-c, and 4 students (6.2%) among 65 students for item 6 used the strategy to solve the questions correctly. To illustrate, the solution strategy employed by Participant 19 for item 1-a is presented below:

Participant 19:

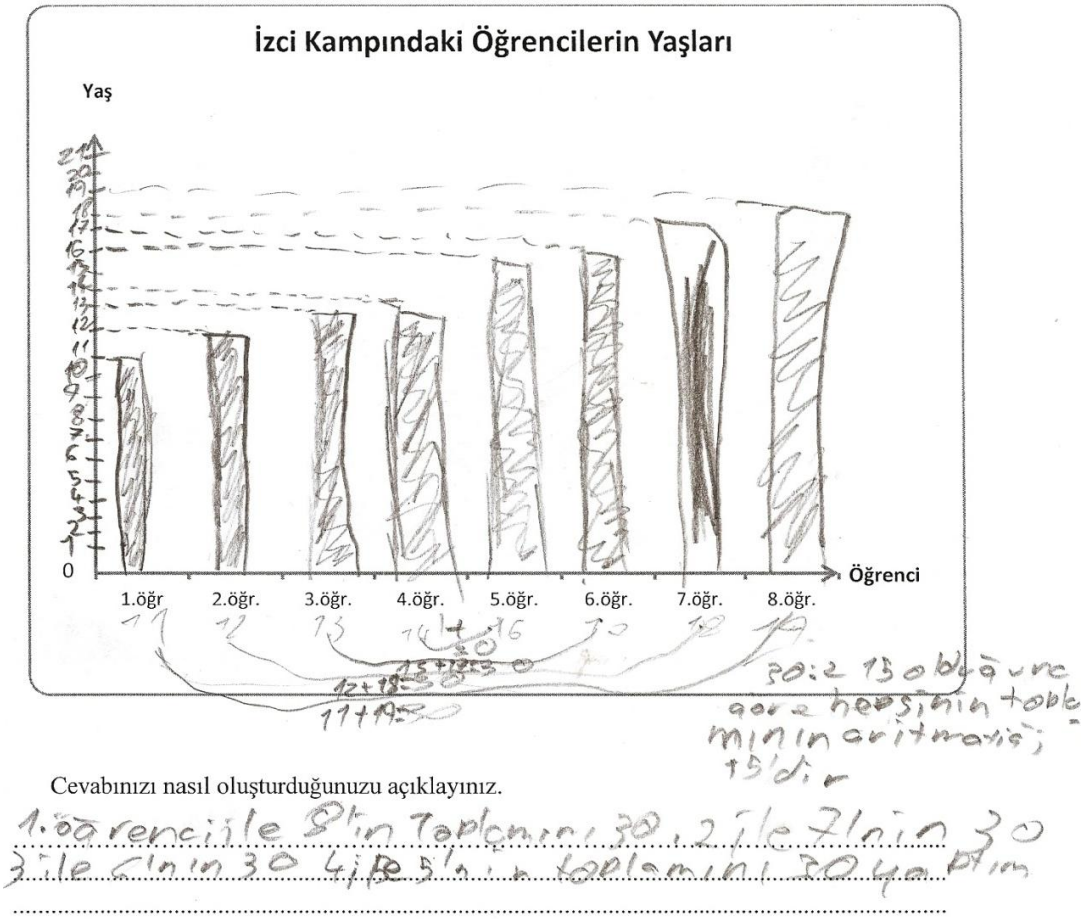


Figure 4.1 Answer of Participant 19 to item 1-a

Participant 19 stated that each value in the data set was balanced with the other values in the data set while constructing a possible data set for the given mean. Age of student 1 was decided as 11 and the value was balanced with age of student 8 as 19 since he believed that the mean of 11 and 19 was 15. In addition, the age of student 2 was decided as 12 and it was balanced with the 7th student's age as 18 since the mean of 12 and 18 was 15. Participant 19 identified the ages of the other students using the same procedure.

Like the previous example, Participant 129 solved item 6 by using balance model solution strategy and it is presented below:

Participant 129:

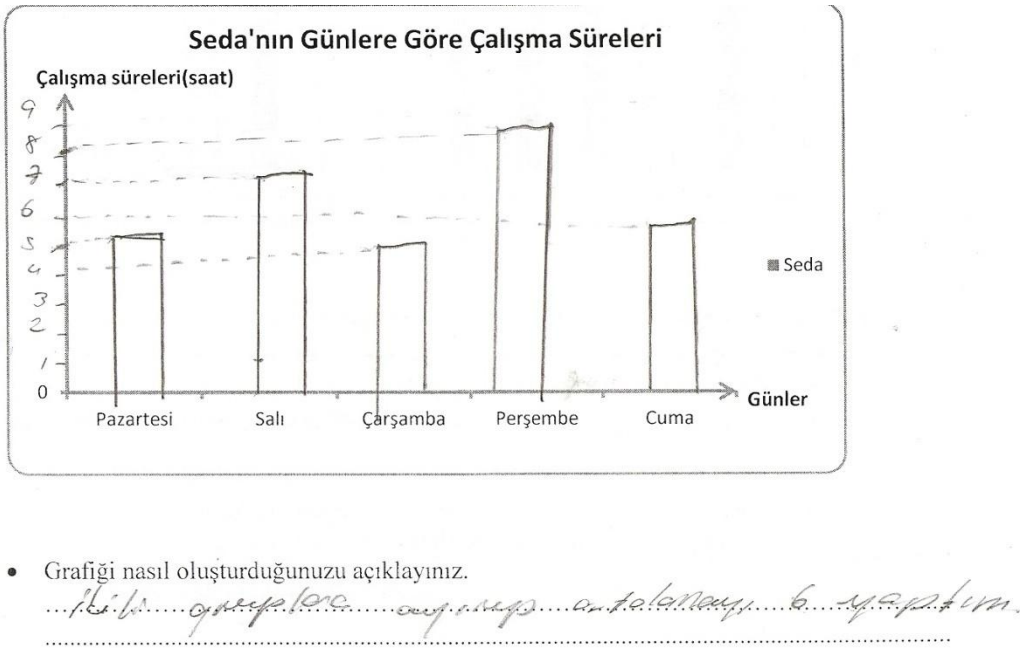


Figure 4.2 Answer of Participant 129 to item 6

Participant 129 said that the working duration of Seda was chosen by dividing the groups of two working days. Her working durations on Monday and Tuesday were chosen as 5 and 7, respectively since the mean of the two values is 6. Moreover, she decided that her working duration on Wednesday was 4 and that on Thursday was 8 to obtain a mean value of 6 for the two values. The last value was chosen as 6 since it was the last value and the mean of the other values was 6.

In general, using balance model solution strategy was most frequently used in the questions requiring either the construction of a possible data set according to the given mean or finding a missing value of a data set for the given mean and other values of the data set.

Using Average Formula: Students who used using average formula solution strategy while solving the questions regarding the mean concept used the "add and then divide" average formula. Thus, when the average of a data set was asked, the students added all the values in a data set and then divided it by the number of values. On the other hand, when a question was asked with a missing value from a data set, first the total number of

values was found and then the total number of the given values in the data set was subtracted.

As can be seen in Table 4.1, the solution strategy was most frequently used in items 3-a and 4-b. More specifically, it was observed that 94 students (100%) in item 3-a and 107 students (100%) in item 4-b had used the solution strategy. Furthermore, to solve the questions using average formula was also found to be used by 43 students (58.1%) among 74 students in item 1-a, 38 students (60.3%) among 63 students in item 1-b, 60 students (68.2%) among 88 students in item 2-ab, 73 students (97.7%) among 75 students in item 2-c, and 58 students (89.2%) among 65 students in item 6. For example, Participant 154 used the solution strategy to solve item 3-a as presented below:

Participant 154:

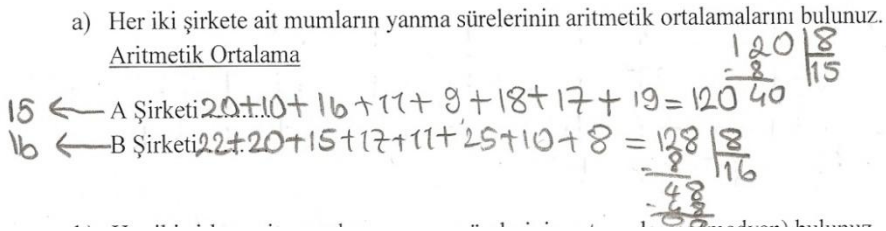
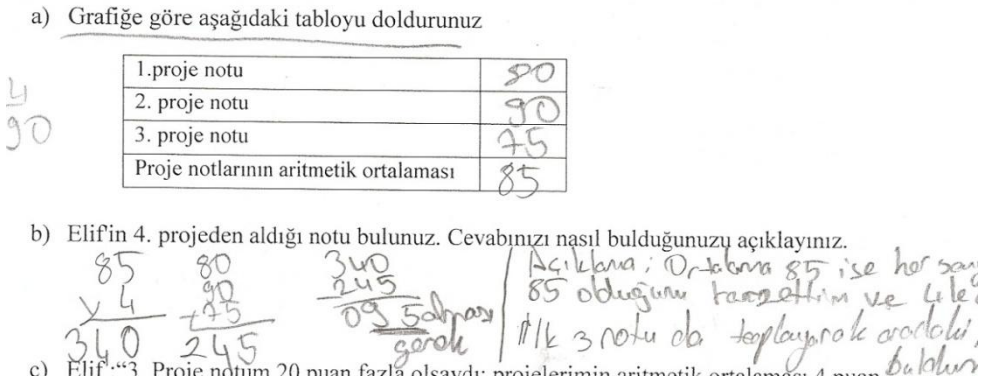


Figure 4.3 Answer of Participant 154 to item 3-a

As can be observed in the solution of Participant 154, to find arithmetic average of the burning times she added all the burning times of the candles for each company and then divided the sum by the number of candles in the companies. Thus, the participant used using average formula solution strategy. Additionally, it was seen that Participant 18 had solved item 2-ab correctly by using average formula strategy as presented below.

Participant 18:


a) Grafiğe göre aşağıdaki tabloyu doldurunuz

1.proje notu	80
2. proje notu	90
3. proje notu	75
Proje notlarının aritmetik ortalaması	85

b) Elif'in 4. projeden aldığı notu bulunuz. Cevabınızı nasıl bulduğunuzu açıklayınız.

85 × 4 = 340; 80 + 90 + 75 = 245; 340 − 245 = 95 almaya gerek

Açıklama; Ortalama 85 ise her sayı 85 olduğunu tahmettim ve 4 le ilk 3 notu da toplayarak aradaki bulduk

c) Elif'in 3. Proje notum 20 puan fazla olsaydı; projelerimin aritmetik ortalaması 4 puan

Figure 4.4 Answer of Participant 18 to item 2-ab

As Participant 18 explained, first the given mean was multiplied by the number of values and then the sum of the given values was subtracted from the total number of all values. Thus, 85 was multiplied by 4 and 340 was reached. Then, the sum of given values, which was 245, was subtracted from 340 and the result was found as 95.

In general, using average formula solution strategy was most frequently used in the questions requiring the calculation of the mean or finding a missing value of a data set for the given mean and other values of the data set.

Guess and Check: Another strategy that the students used in addition to the solution strategy to solve the questions related to the mean was guess and check strategy. First they chose a value(s) in a data set and then checked the correctness or appropriateness of the value(s) for the mean of the data set. If the value(s) was not correct, another value(s) was chosen and checked again, until the value(s) was correct to obtain the given mean.

As can be observed in Table 4.1, guess and check solution strategy was most frequently used in item 2-ab by 20 students (22.7%) among 88 students. Moreover, the solution strategy were used to solve item 1-a, 1-b, and 6 by 8 students (10.8%) among 74 students, 6 students (9.5%) among 63 students, and 3 students (4.6%) among 65 students in item 6, respectively.

To illustrate, the solution strategy used by Participant 28 is presented below:

Participant 28:

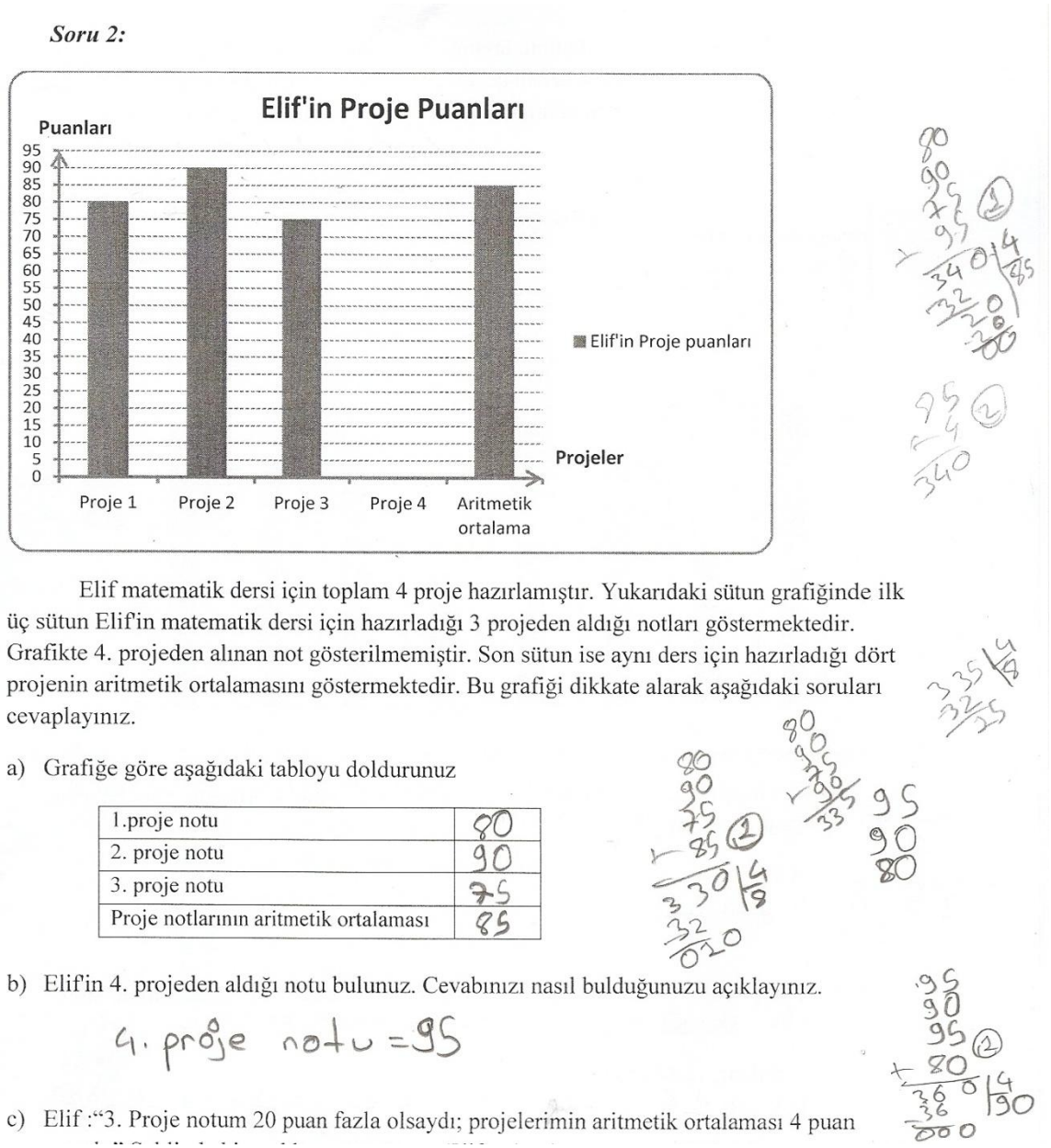


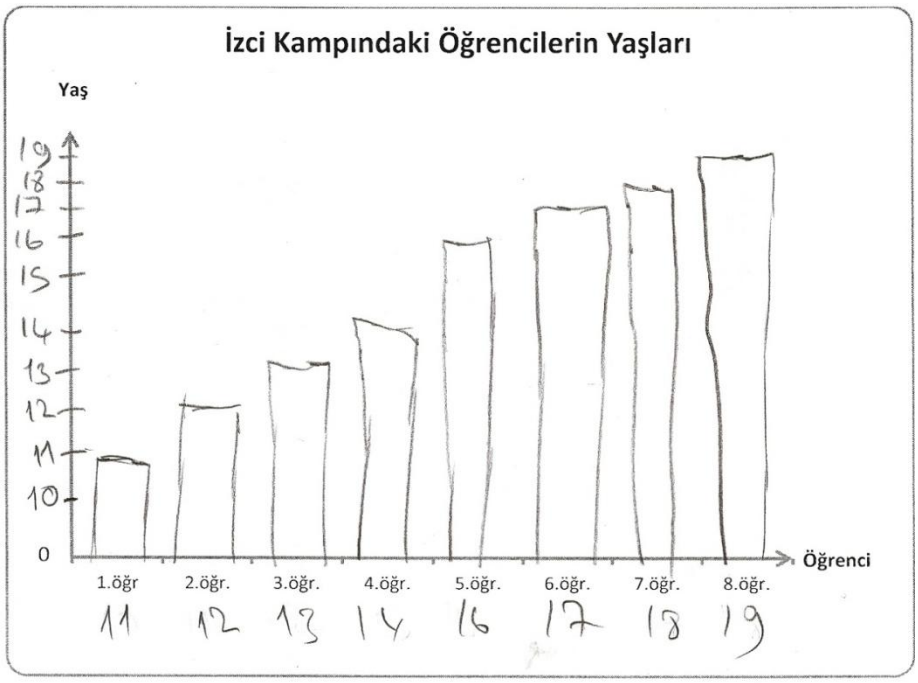
Figure 4.5 Answer of Participant 28 to item 2-ab

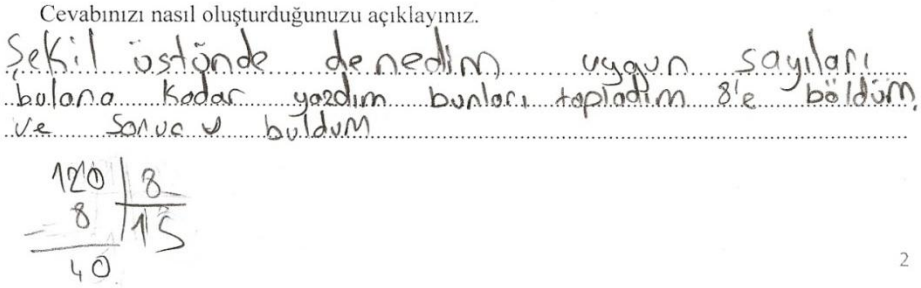
In order to solve item 2-ab, Participant 28 chose 80 for the missing project score and then recalculated the average of Elif's project scores to check whether the average was 85 for the score. However, the score average was not 85. Then, 85 and 90 were checked to find the missing project score of Elif, but these were also not appropriate. Lastly, 95

was checked and it was seen that the value was appropriate to obtain 85 as the arithmetic average of her scores.

In addition, the solution strategy used by Participant 208 in item 1-b is another example that is presented below.

Participant 132:




Cevabınızı nasıl oluşturduğunuzu açıklayınız.

Şekil üstünde denedim uygun sayıları bulana kadar yazdım bunları topladım 8'e böldüm ve sonuca ulaştım buldum

120 | 8
8 | 15
40

2

Figure 4.6 Answer of Participant 208 to item 1-b

As participant 132 stated, he initially guessed the ages of each student and then found the average of the constructed data set to see whether the average of the values was 15. Then, he concluded that the constructed data set was appropriate for the given average.

In general, the results of the study showed that the questions which required the construction of a possible data set according to the given mean, or finding a missing value of a data set for the given mean could be solved through guess and check solution strategy.

4.1.2 Solution Strategies of the Questions Regarding the Median

Solution strategies of questions related to the median were addressed in items 1-a, 1-b, 3-b, 4-b and 5 because item 1-a and 1-b required the construction of a possible data set for the given averages and the other questions required finding the median of a data set or constructing a possible data set for a given median. However, none of the students used the median as an average for items 1-a, 1-b and 4-b because all the students tried to find a possible data set as if the given average was the mean of the data set. The solution strategies of the students who answered correctly using an acceptable solution strategy were analyzed and it was found that there were three types of solution strategies used, namely *using numerical procedures, depending on graph representations*, and *guess and check.*

Using Numerical Procedures: The using numerical procedures solution strategy is used in the following way: Numerical values of a data set are used in the strategy to find median of a data set. If a data set is given on a graph, students obtain values of the data set from given graph. Then, when the values are listed in order, the median of a distribution that includes an odd number of values is the middle score. To illustrate, the median of the distribution 2, 4, 6, 8 and 10 is 6. On the other hand, when the values are listed in order, the median of a distribution that includes an even number of values is the midpoint of the two middle scores. For example, in the distribution 1, 3, 4, 6, 7, 8, the median is 5.

As can be observed in Table 4.1, using numerical procedures solution strategy was used by most of the students who correctly answered using an acceptable solution strategy or made an operational error using an acceptable solution strategy in the questions including the calculation of the median. Moreover, the solution strategy was most frequently used to solve item 3-b by 50 students (94.3%) among 53 students. In addition,

22 students (78.6%) among 28 students used using numerical procedures solution strategy for item 5. To illustrate, Participant 75 ordered numerical values of the given data sets and then found the middle score of the data sets to solve item 3-b. The solution strategy used by Participant 75 is presented below:

Participant 75:

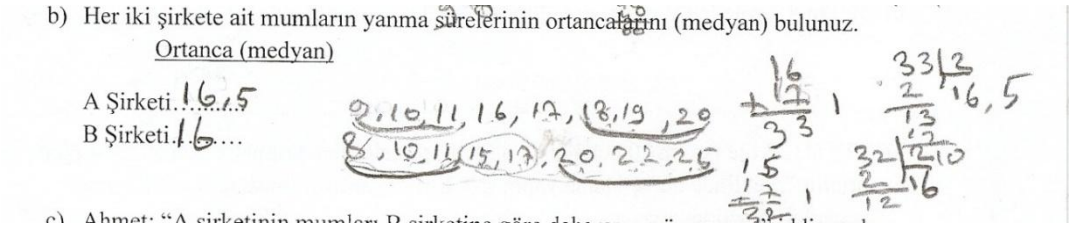


Figure 4.7 Answer of Participant 75 to item 3-b

As can be seen from the solution strategy used by Participant 75, first she ordered the burning times of the candles for each company and then found the half mid-point of the middle values for these data sets. Similarly, Participant 163 used the solution strategy and it is presented below:

Participant 163:

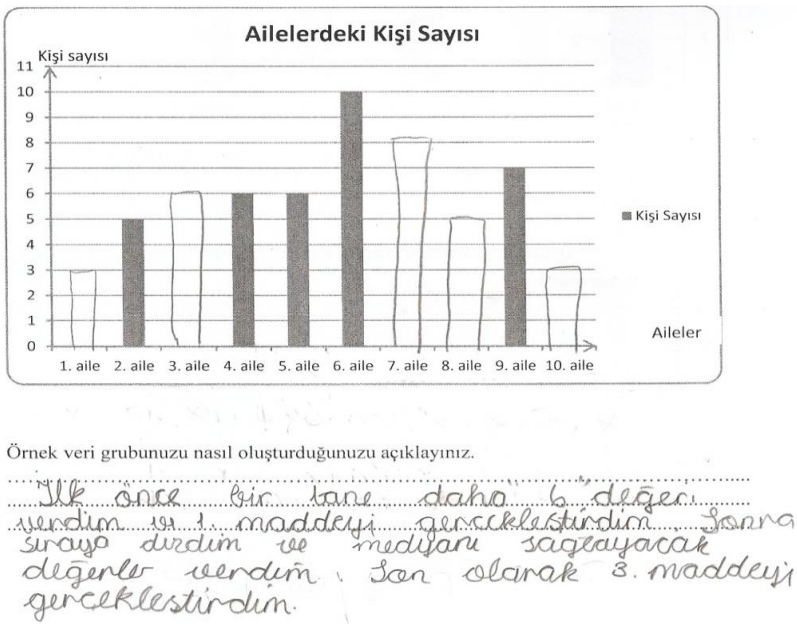


Figure 4.8 Answer of Participant 163 to item 5

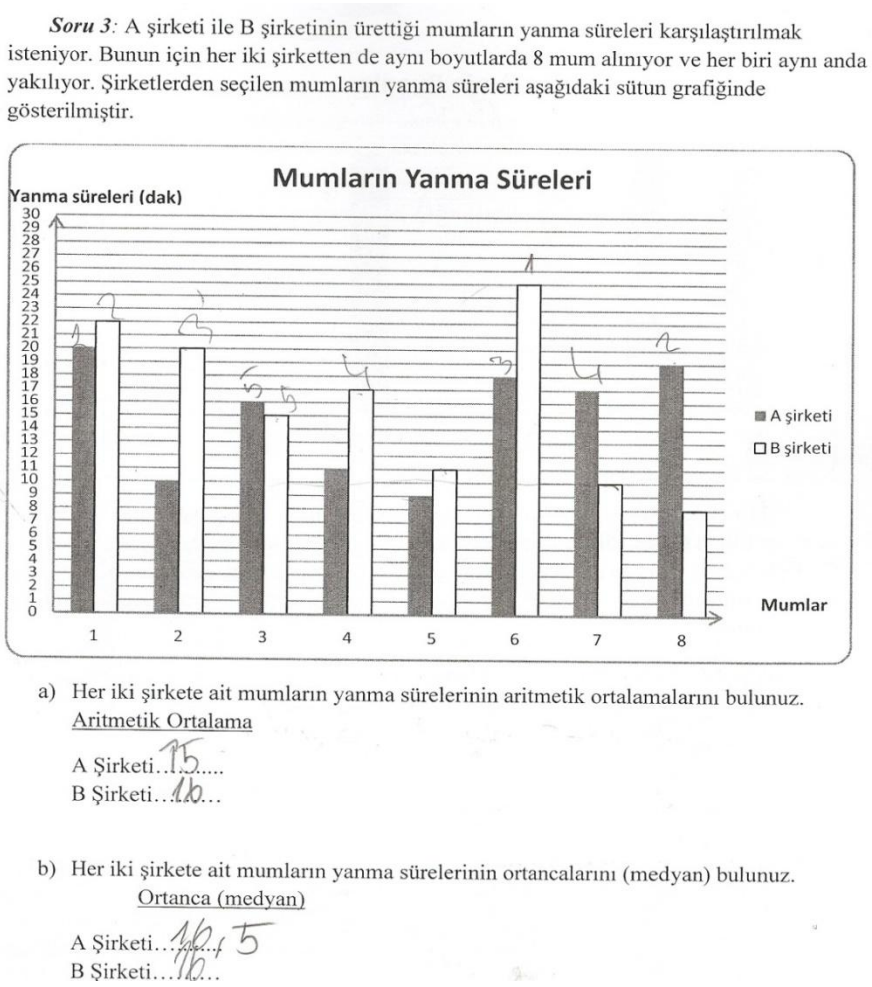
As Participant 163 stated, while constructing a possible data set in item 5, values were ordered to check the median of the constructed data set.

According to results of the study, using numerical procedures solution strategy was most frequently used in the questions which required directly finding the median of a data set. The other solution strategy was depending on graph representations and the details of this strategy is given below.

Depending on Graph Representations: The students who used depending on graph representations solution strategy to correctly solve the questions related to the median found the median of the data set directly based on the given graph. The solution strategy can be used when a data set is given on a graph.

As can be observed in Table 4.1, only 3 students (5.7%) among 53 students used depending on graph representations solution strategy to solve item 3-b. For example, Participant 18 ordered the given data set on the graph and then found the median of the data set to solve item 3-b. To illustrate, the solution strategy of Participant 18 is presented below:

Participant 18:



Soru 3: A şirketi ile B şirketinin ürettiği mumların yanma süreleri karşılaştırılmak isteniyor. Bunun için her iki şirketten de aynı boyutlarda 8 mum alınıyor ve her biri aynı anda yakılıyor. Şirketlerden seçilen mumların yanma süreleri aşağıdaki sütun grafiğinde gösterilmiştir.

a) Her iki şirkete ait mumların yanma sürelerinin aritmetik ortalamalarını bulunuz.
Aritmetik Ortalama

A Şirketi…15…
B Şirketi…16…

b) Her iki şirkete ait mumların yanma sürelerinin ortancalarını (medyan) bulunuz.
Ortanca (medyan)

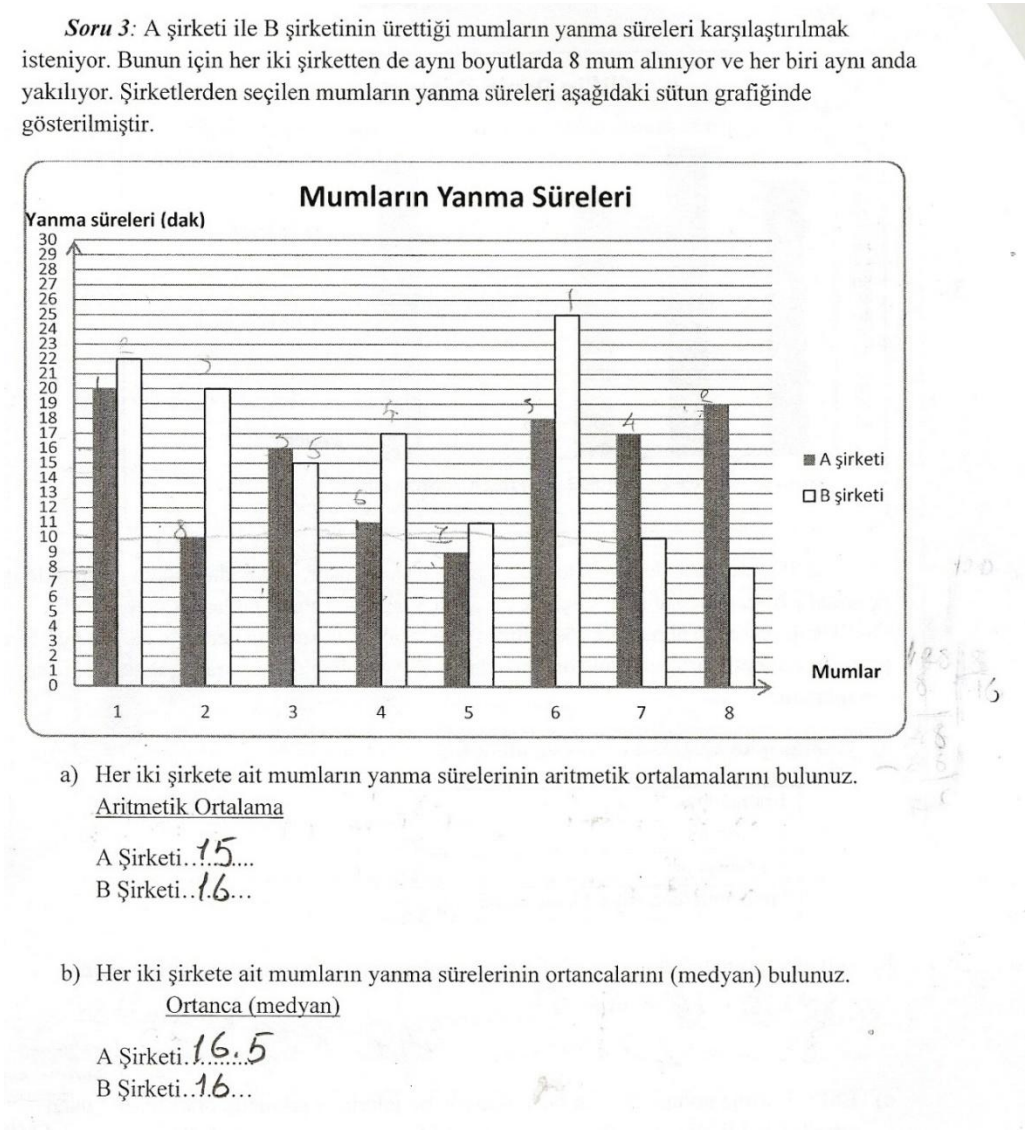
A Şirketi…16…5
B Şirketi…16…

Figure 4.9 Answer of Participant 18 to item 3-b

As can be observed in the solution of Participant 18, he ordered the values for each data set on the graph by assigning numbers to the bars. Then, the average of the fourth and fifth values was found for each data set to find the median of the data sets.

Similarly, Participant 54 used the same solution strategy as Participant 18 and it is presented below:

Participant 54:



Soru 3: A şirketi ile B şirketinin ürettiği mumların yanma süreleri karşılaştırılmak isteniyor. Bunun için her iki şirketten de aynı boyutlarda 8 mum alınıyor ve her biri aynı anda yakılıyor. Şirketlerden seçilen mumların yanma süreleri aşağıdaki sütun grafiğinde gösterilmiştir.

a) Her iki şirkete ait mumların yanma sürelerinin aritmetik ortalamalarını bulunuz.
Aritmetik Ortalama

A Şirketi..15....
B Şirketi..16...

b) Her iki şirkete ait mumların yanma sürelerinin ortancalarını (medyan) bulunuz.
Ortanca (medyan)

A Şirketi..16.5
B Şirketi..16...

Figure 4.10 Answer of Participant 54 to item 3-b

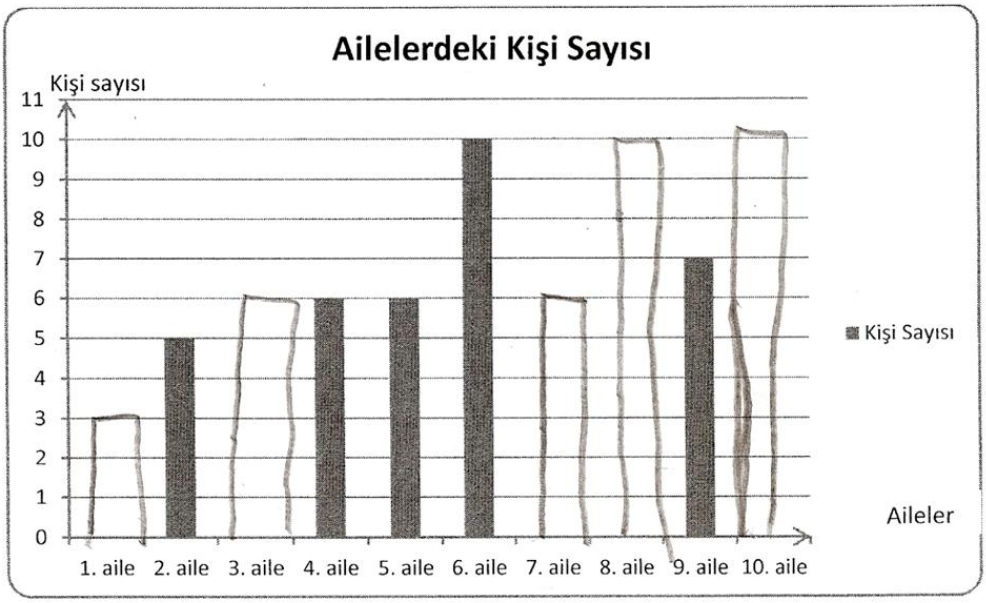
In general, the results of this study showed that when a data set was given on a graph, the median of the data set could be found directly based on this graph. Furthermore, another solution strategy was guess and check to solve the questions regarding the median. The details regarding this strategy is presented below.

Guess and Check: Similar to guess and check solution strategy that was encountered in the solutions provided to the mean questions, the students who used guess and check solution strategy first guessed the answer of the question and then checked the

correctness or appropriateness of their answer based on the given conditions in the question. To illustrate, a question required the construction of a possible data set for a given median. In order to solve the question via this strategy, first a possible data set was guessed and then the correctness of the data set was checked for the given median.

According to Table 4.1, it was seen that only 6 students (21.4%) among 28 students used guess and check solution strategy to solve item 5. To illustrate, Participant 139 stated that she initially guessed a possible data set and then checked the correctness of the constructed data set for the given median. Thus, the participant used guess and check solution strategy to solve item 5.

Participant 139:



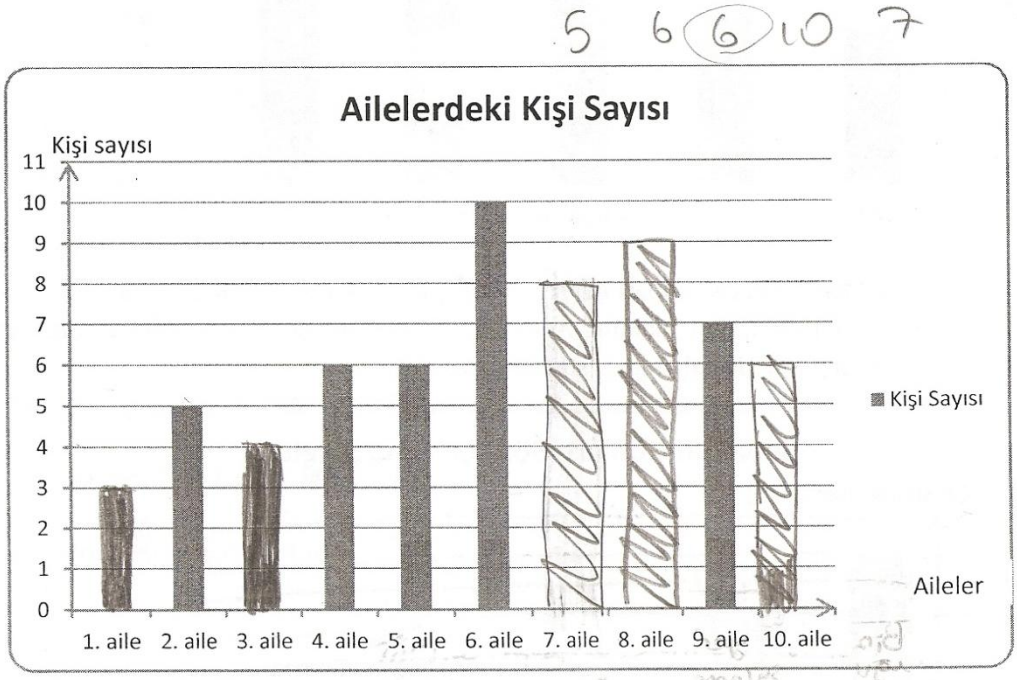
- Örnek veri grubunuzu nasıl oluşturduğunuzu açıklayınız.

Onas rastyele dersler verdim sonra şeri kalanar göre tamamladım

Figure 4.11 Answer of Participant 139 to item 5

According to the explanation of Participant 21, it was observed that he had guessed the remaining values and then checked the correctness of the data set for the given median. The explanation and answer of Participant 21 is given below:

Participant 21:



- Örnek veri grubunuzu nasıl oluşturduğunuzu açıklayınız.
  Ailedeki kişi sayılarına bakarak kalanlara göre tahmin yaptım.

Figure 4.12 Answer of Participant 21 to item 5

In general, the results of the study showed that guess and check solution strategy was used to solve item 5 requiring the construction of a possible data set for the given median of the data set.

4.1.3. Solution Strategies of the Questions Regarding the Mode

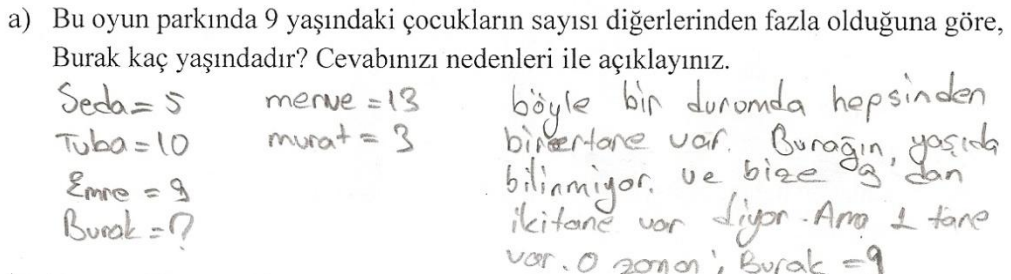
Solution strategies regarding the mode questions were identified in items 1-a, 1-b, 4-a, 4-b and 5 because items 1-a and 1-b required the construction of a possible data set for the given averages and 4-b required finding the average of a data set; the remaining

questions required the construction of a possible data set for a given mode. However, none of the students accepted the mode as an average for items 1-a, 1-b and 4-b since all of the students used the mean as the average to solve these items. According to the results of the study, the questions related to the mode could be solved by means of three types of strategies, namely *using numerical procedures*, *depending on graph representations*, and *guess and check*.

Using Numerical Procedures: The students who used using numerical procedures solution strategy to correctly solve the questions related to the mode found the most frequent value in a given data set based on numerical values of a data set.

As presented in Table 4.1, this solution strategy by used by 6 students (4.7%) among 128 students to solve item 4-a and by 22 students (78.6%) among 28 students to solve item 5 correctly. For example, Participant 4 used the strategy in order to solve item 4-a as presented below:

Participant 4:


a) Bu oyun parkında 9 yaşındaki çocukların sayısı diğerlerinden fazla olduğuna göre, Burak kaç yaşındadır? Cevabınızı nedenleri ile açıklayınız.

Seda = 5
Tuba = 10
Emre = 9
Burak = ?

merve = 13
murat = 3

böyle bir durumda hepsinden birertane var. Burağın yaşında bilinmiyor. ve bize 9 dan ikitane var diyor. Ama 1 tane var. O zaman, Burak = 9

Figure 4.13 Answer of Participant 4 to item 4-a

As can be observed in the solution provided by Participant 4, first she wrote each datum from graph as 5, 10, 9, 13 and 3; and then she concluded that in order to obtain the age of 9 as most frequent datum, Burak must be 9 years old.

As another example, the solution of Participant 104 is presented below:

Participant 104:

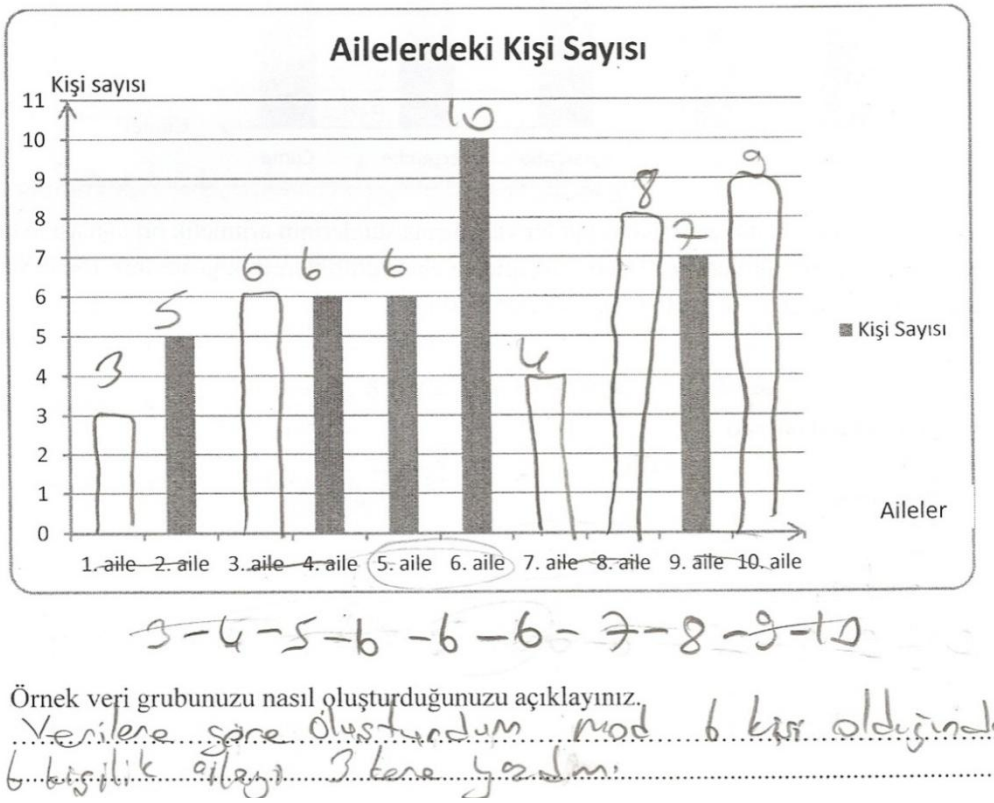


Figure 4.14 Answer of Participant 104 to item 5

As Participant 104 stated, she selected 6 individuals from the third family to have 6 as the most frequent datum in the constructed data set. Another strategy used to solve the questions regarding the mode was depending on graph representations. The details of this strategy are presented below.

Depending on Graph Representations: The students who used depending on graph representations solution strategy to correctly solve the questions related to the mode calculated the mode of a data set directly based on a given graph. The solution strategy was used when a data set was given on a graph.

As can be observed in Table 4.1, 122 students (95.3%) among 128 students used this strategy to correctly solve item 4-a. For example, Participant 52 directly used the values on the graph to find the missing value in item 4-a. Thus, according to the graph, the age of Burak was predicted to be 9 to obtain the most frequent age in the group. The solution strategy utilized by Participant 52 for item 4-a is presented below:

Participant 52:

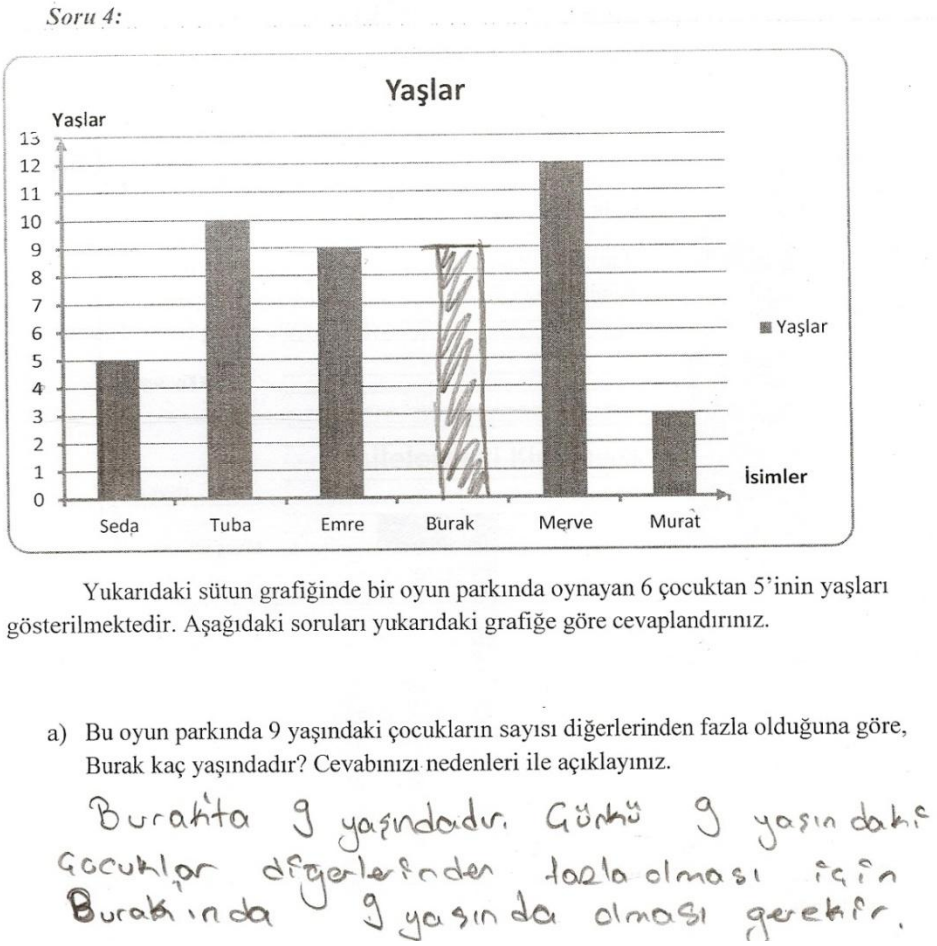


Figure 4.15 Answer of Participant 52 to item 4-a

Similarly, Participant 161 identified the age of Burak as 9 directly based on the given bar graph since he said that based on given bar graph, there were one student who was 9 years old but number of children's who was 9 years old was more than other. The explanation of the solution strategy used by Participant 161 is given below:

Participant 161:

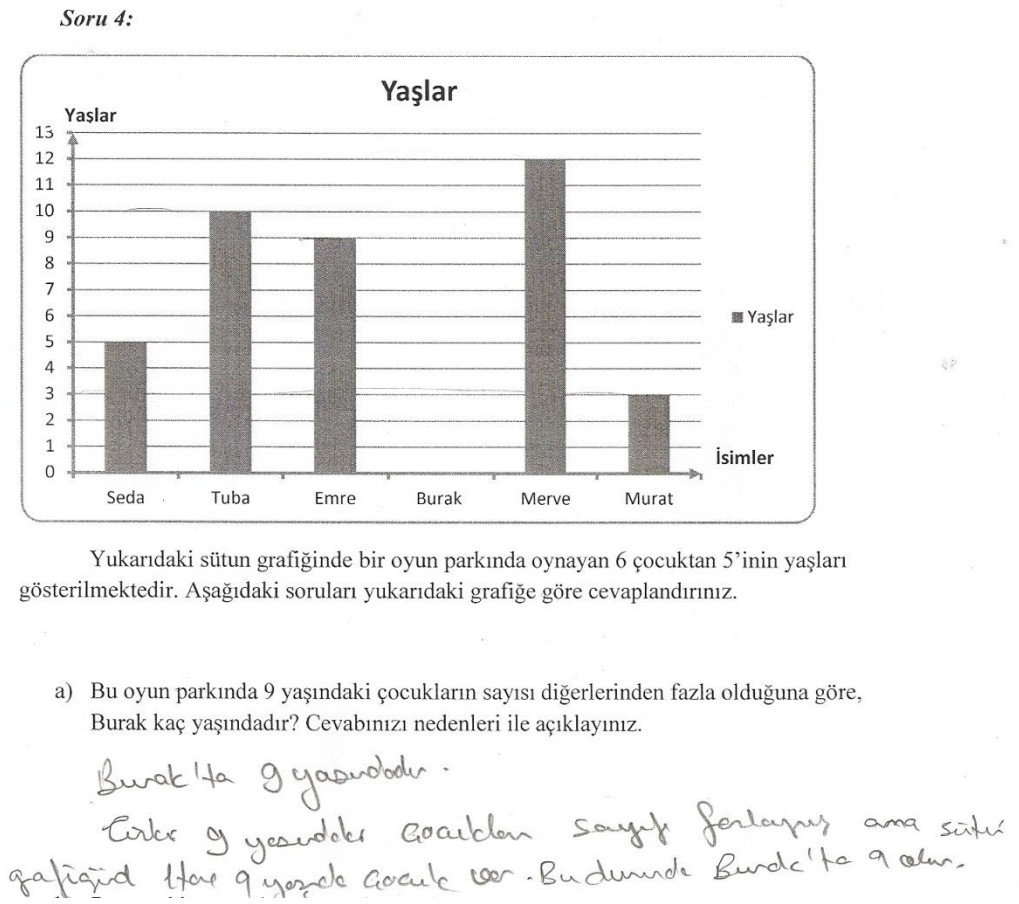


Figure 4.16 Answer of Participant 161 to item 4-a

In general, according to the results of the study, the depending on graph representations solution strategy was most frequently used to solve the questions provided that each value was given on a graph.

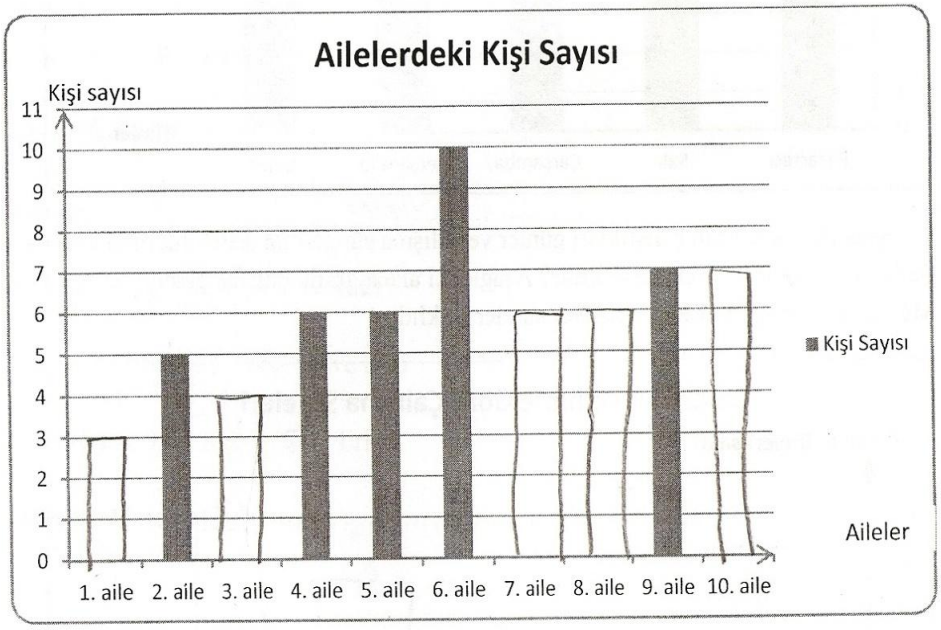
Another strategy to solve the questions regarding the mode was guess and check strategy, which is explained under the following heading.

Guess and Check: Similar to guess and check strategy used to solve questions regarding the mean and the median, when guess and check solution strategy was used to solve mode questions, first the answer of a question was guessed and then the correctness or appropriateness of the answer was checked. For example, when a question required the construction of a possible data set, students who used this solution strategy

first guessed a possible data set, then checked the correctness of the data set for the given conditions.

According to Table 4.1, it was seen that only 6 students (21.4%) among 28 students used the guess and check solution strategy to solve item 5. To illustrate, according to the explanation of Participant 9, it can be understood that he used this solution strategy to solve item 5.

Participant 9:



- Örnek veri grubunuzu nasıl oluşturduğunuzu açıklayınız.
örnek verdim baktım doğru çıktı ben de tamam dedim.

Figure 4.17 Answer of Participant 9 to item 5

Based on the explanation of Participant 9, first a data set was constructed, then the appropriateness for the given mode and median of the data set was checked.

In general, the results of the study showed that item 5 requiring the construction of a possible data set for the given mode of the data set was solved by means of guess and

check strategy. In addition, some of the median and mean questions were solved via this solution strategy when the construction of a possible data set was required. On the other hand, it was also used to find a missing value in a question asking for the mean.

4.2 Errors Regarding the Concepts of Mean, Median and Mode

The second purpose of this study was to investigate seventh grade students possible errors regarding the concepts of mean, median and mode given in bar graph representations. In accordance with this purpose, this section included errors in the problems regarding the concepts of mean, median and mode.

In order to identify the students' errors regarding the concepts of mean, median and mode, wrong answers and constructed inappropriate data sets were analyzed from the SAQ that the students had completed. In this part, the students' errors were categorized under three main groups for mean, median and mode, namely *algorithmically based errors, errors based on formal knowledge* and *other errors.* Firstly, algorithmically based error is an error type which is generally made in the arithmetic operations (Ashlock, 1990). In this study, many students made these errors while doing addition, subtraction, multiplication or division operations. Secondly, errors based on formal knowledge are errors caused by limited conceptions or inadequate knowledge regarding the properties of the concepts of mean, median and mode (Tirosh, 2000). Lastly, in this study, the group of other errors includes the students' errors other than those in the first and the second group of errors. Each error group was further categorized under three headings for each concept. More specifically, for the mean concept, the sub-group of algorithmically based errors was *operational error*. In addition, the sub-groups of errors based on formal knowledge were *finding total number, incorrect usage of averaging algorithm, accepted close values, looking pattern,* and *turned to find smaller or larger number*. Lastly, the sub-groups of other errors were *not meeting all requirements of a problem, forming incomplete data set* and *incorrect reading of values from graph*. Moreover, for the median concept, *operational error* was also defined as a sub-group of algorithmically based errors; *wrong decision on unordered data set, wrong decision on ordered data set* and *wrong interpretation of graph* were defined as sub-groups of errors based on formal knowledge; and sub-groups of other errors were *forming incomplete*

data set, incorrect reading of values from graph and *incorrect largest and/or smallest data values.* In addition to these, the students' errors in the questions regarding the mode concept, *inappropriate usage of averaging algorithm* was observed as a sub-group of errors based on formal knowledge, and *forming incomplete data set* and *incorrect largest and/or smallest data values* were sub-groups of other errors. Errors regarding each concept are presented in the following parts.

4.2.1 Errors Regarding the Mean Errors related to the mean were identified in items 1-a, 1-b, 2-ab, 2-c, 3-a, 4-b and 6 since the calculation of the mean of a data set or the construction of a possible data set for a given mean were required in the items. In order to identify the students' errors, wrong answers were identified in items 2-ab, 2-c, and 3-a, which required the calculation of the mean, and in item 4-b, which required the calculation of the average of a data set. Moreover, the construction of inappropriate data sets were identified in items 1-a, 1-b, and 6 since the items required the construction of a possible data for the given average in the SAQ, which the students completed. According to the findings of the analyses, *operational error* was observed as an algorithmically based errors. Moreover, *finding total number*, *incorrect usage of averaging algorithm*, *accepted close values*, *looking pattern* and *turned to find smaller or larger numbers* were observed as errors based on formal knowledge. Lastly, *not meeting all requirements of a problem*, *forming incomplete data set* and *incorrect reading of values from graph* were accepted as other errors of the questions regarding the mean. Frequencies and percentages of these errors committed by the participants are presented in Table 4.2 below.

Table 4.2 Frequencies (and percentages) of the students' errors regarding the mean

ERRORS / ITEMS	Algorithmically Based Errors	Errors Based on Formal Knowledge					Other Errors			
	Operational Error	Finding Total Number	Incorrect Usage of averaging algorithm	Accepted Close Values	Looking Pattern	Turned to Find Smaller or Larger Numbers	Not Meeting All Requirements of a Problem	Form-ing Incomplete Data Set	In-correct Reading of Values from Graph	Total
Item 1-a	0 (0%)	17 (24.0%)	5 (7.0%)	3 (4.2%)	0 (0%)	41 (57.8%)	4 (5.6%)	1 (1.4%)	0 (0%)	71
Item 1-b	0 (0%)	11 (22.4%)	1 (2.0%)	9 (18.4%)	0 (0%)	26 (53.1%)	2 (4.1%)	0 (0%)	0 (0%)	49
Item 2-ab	11 (14.1%)	0 (0%)	59 (75.6 %)	1 (1.3%)	3 (3.9%)	0 (0%)	0 (0%)	0 (0%)	4 (5.1%)	78
Item 2-c	5 (10.2%)	0 (0%)	19 (38.8%)	0 (0%)	0 (0%)	0 (0%)	25 (51.0%)	0 (0%)	0 (0%)	49
Item 3-a	46 (74.2%)	9 (14.5%)	2 (3.2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	5 (8.1%)	62
Item 4-b	23 (39.0%)	8 (13.5%)	26 (44.1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (3.4%)	59
Item 6	3 (3.1%)	0 (0%)	0 (0%)	0 (0%)	7 (7.1%)	0 (0%)	83 (84.7%)	2 (2.0%)	3 (3.1%)	98
Total	68	45	112	13	10	67	114	3	14	

As it can be seen in Table 4.2, there was one type of errors under the heading of algorithmically based errors, five types of errors in the category of errors based on formal knowledge and three types of errors in the other errors category related to the mean. According to the table, the most common error type was the 'not meeting all requirements of a problem'. In fact, 83 students among 98 students made the error while they were solving item 6. Moreover, the 'incorrect usage of averaging algorithm' was the

second frequently made error that the students had committed. It was seen that 59 out of 78 students in item 2-ab and 26 out of 59 students in item 4-b used the averaging algorithm incorrectly. In addition to these, the third common error type was the 'operational error'. For example, 46 students (74.2%) among 62 students made operational error in item 3-a. Examples of students' wrong answers and their constructed inappropriate data sets are presented below under the category headings given in Table 4.2.

4.2.1.1 Algorithmically Based Errors

In this study, algorithmically based errors were made due to basic operational errors.

Operational Error: Some of the students made calculation errors while doing addition, subtraction, multiplication or division. Thus, the students made operational errors since if they had done the operations correctly in an item, they would have solved the item correctly. As given in Table 4.2, operational error was most frequently observed in item 3-a, and it was made by 46 students (74.2%) among 62 students. In addition, this error was seen in items 2-ab, 2-c, 4-b and 6. In item 2-ab, 11 students (14.1%) among 78 students; in item 2-c, 5 students (10.2%) among 49 students; in item 4-b, 23 students (39.0%) among 59 students made an operational error. Lastly, an operational error was made by 3 students (3.1%) among 98 students in item 6. To illustrate, the response of Participant 93 to item 3-a is given below:

Participant 93:

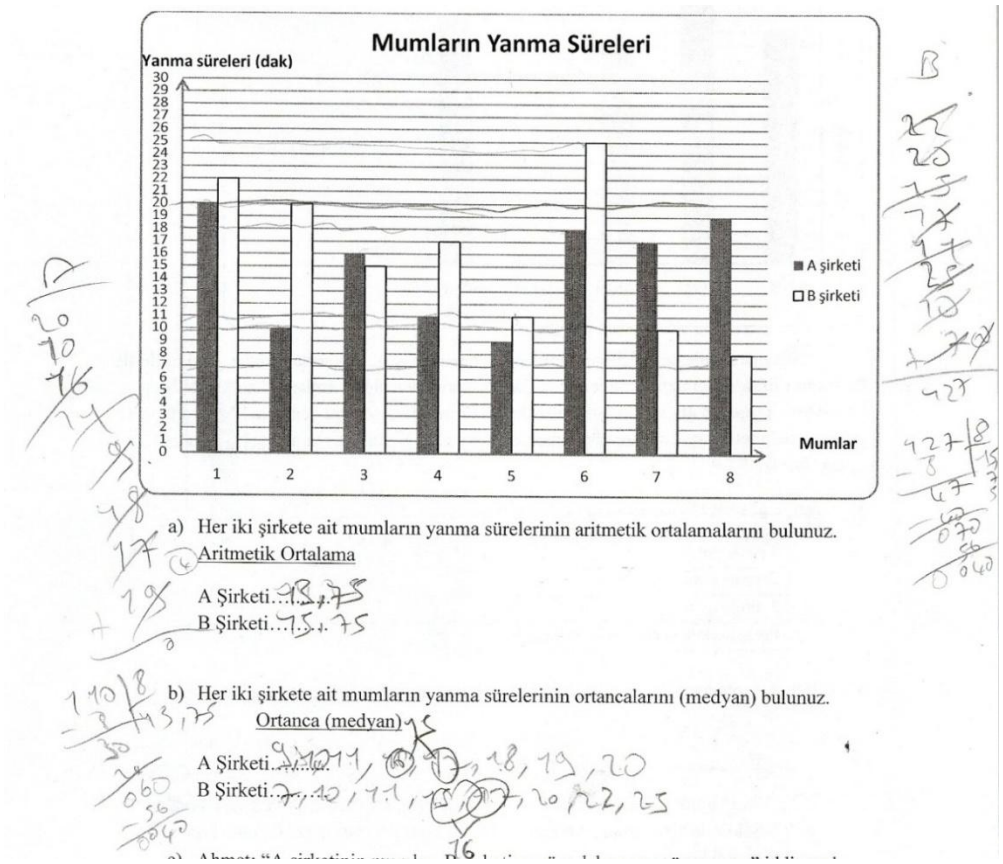


Figure 4.18 Answer of Participant 93 to item 3-a

As can be seen in the solution provided by Participant 93, the Participant wrote burning times of each company correctly but she did not find the sum of the times correctly. More specifically, the Participant find the sum of burning times of Company A as 110 instead of 125 and find the sum of burning times of Company B as 427 instead of 425. Because of that error, the mean of each company was not found correctly. Thus, an operational error was made by Participant 93. Furthermore, as it is seen below, Participant 27 was one of the students who made this error in item 4-b.

Participant 27:

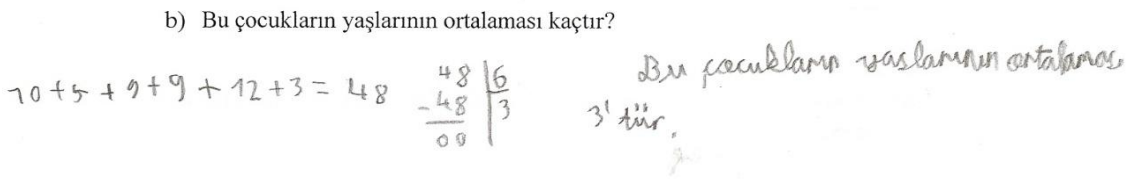


Figure 4.19 Answer of Participant 27 to item 4-b

According to the response of Participant 27, it can be observed that he added each value in the data set but he made an operational error while dividing 48 by 6.

4.2.1.2 Errors Based on Formal Knowledge

In this study, *finding total number*, *incorrect usage of averaging algorithm*, *accepted close values*, *looking pattern*, and *turned to find smaller or larger numbers* were observed as errors based on formal knowledge category.

Finding Total Number: The students who made this error found the total number of all the values in a data set correctly based on the mean of the data set. However, when an item required the construction of a possible data set, they could not construct a possible data set according to the total value. In addition, when the arithmetic average of a data set was asked for, the students who made this error found the total number of values as the answer. For example, the mean of a data set is given as 15 and the sample size of the data set is 4. The student who made this error found the answer as 60 since 15x4=60. On the other hand, when the values of a data set were 1, 2, 3, and 5; the student found 1+2+3+5=11 as the mean of the data set.

According to Table 4.2, this error type was made in items 1-a, 1-b, 3-a, and 4-b, and this error was most frequently seen in items 1-a and 1-b. In item 1-a, it was observed that 17 students (24.0%) among 71 students had made this error. Additionally, it was observed in the responses that 11 students (22.4%) among 49 students in item 1-b, 9 students (14.5%) among 62 students in item 3-a , 8 students (13.5%) among 59 students in item 4-b had made this error. For example, the error of *finding total number* was seen in the response provided by Participant 57 in item 3-a, which is presented below:

Participant 57:

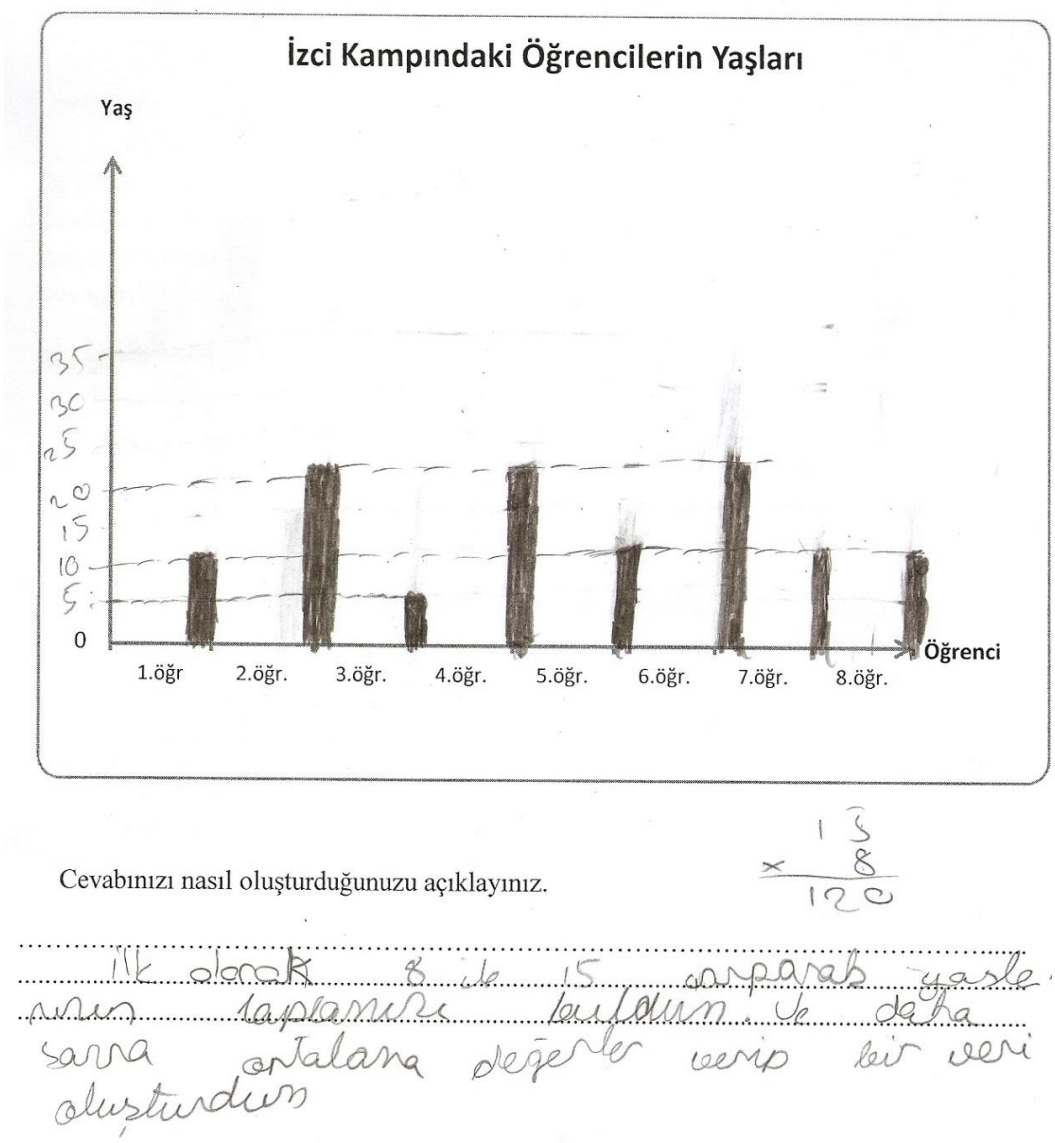
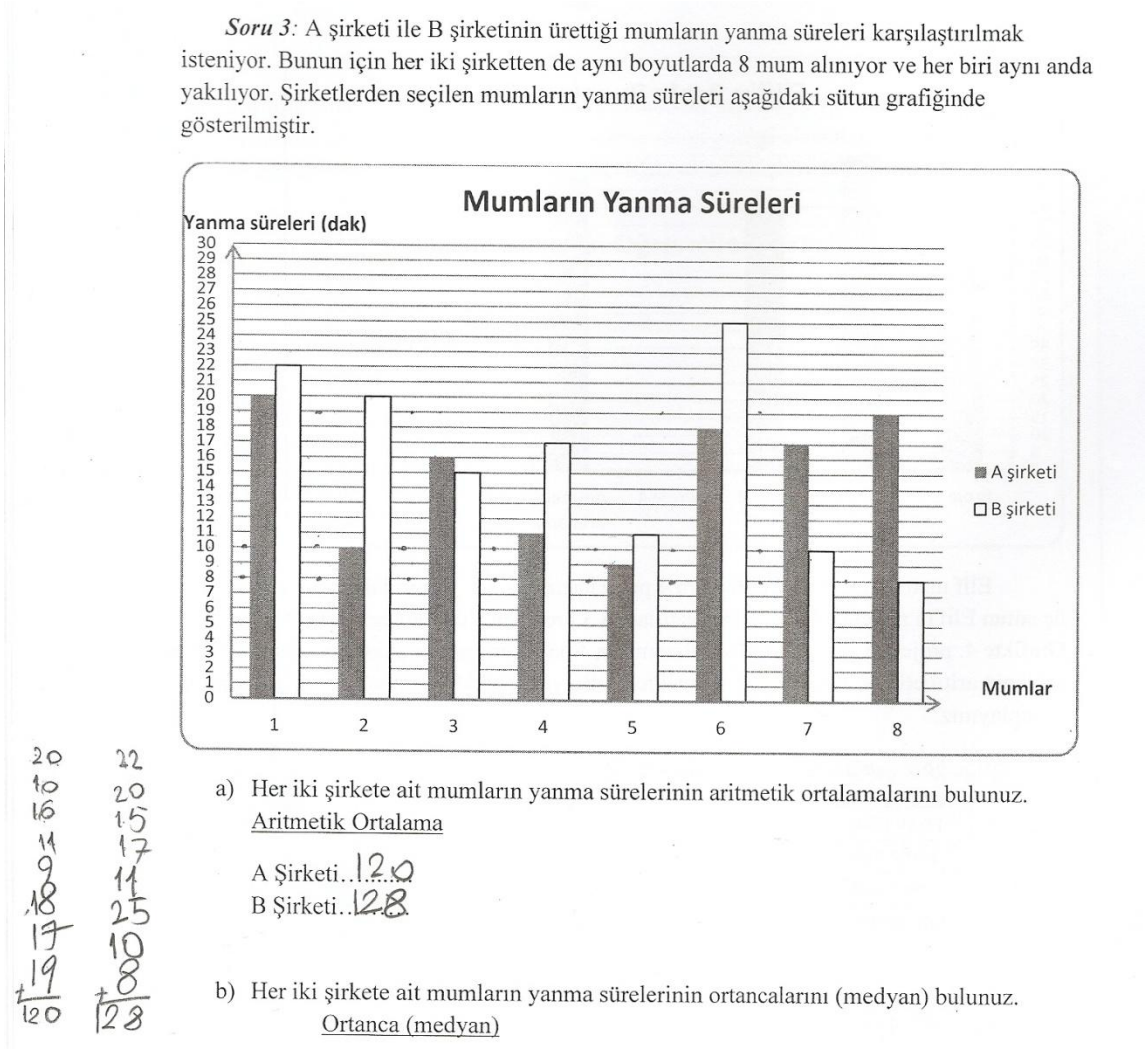


Figure 4.20 Answer of Participant 57 to item 3-a

As Participant 57 stated, she found the total number of values as 120 first and then tried to find a possible data set. However, when the constructed data set was checked, it was seen that the data set was not appropriate for the given mean.

Another example is that Participant 17 made the *finding total number* error type in item 1-a as presented below:

Participant 17:



Soru 3: A şirketi ile B şirketinin ürettiği mumların yanma süreleri karşılaştırılmak isteniyor. Bunun için her iki şirketten de aynı boyutlarda 8 mum alınıyor ve her biri aynı anda yakılıyor. Şirketlerden seçilen mumların yanma süreleri aşağıdaki sütun grafiğinde gösterilmiştir.

20 22
10 20
16 15
11 17
9 11
18 25
17 10
19 8
+ +
120 128

a) Her iki şirkete ait mumların yanma sürelerinin aritmetik ortalamalarını bulunuz.
Aritmetik Ortalama

A Şirketi..120
B Şirketi..128

b) Her iki şirkete ait mumların yanma sürelerinin ortancalarını (medyan) bulunuz.
Ortanca (medyan)

Figure 4.21 Answer of Participant 17 to item 1-a

Based on the response of Participant 17, the total number of candle's burning times for each company was found as the answer. Because of that error, mean of Company A and Company B was found 120 and 128, respectively.

Another error based on formal knowledge was the *incorrect usage of averaging algorithm*. Definition and examples of this error is presented under the following heading.

Incorrect Usage of Averaging Algorithm: The students who made this error tried to use computational algorithm of the mean to solve the items regarding the mean but their application was incorrect. The error was made in items 1-a, 1-b, 2-ab, 2-c, 3-a, and 4-b. In addition, as seen in Table 4.2 the incorrect usage of averaging algorithm error was most frequently made in item 2-ab by 59 students (75.6%) among 78 students. Moreover, the error was done by 5 students (7.0%) among 71 students in item 1-a, 1 students (2.0%) among 49 students in item 1-b, 19 students (38.8%) among 49 students in item 2-c, 2 students (3.2%) among 62 students in item 3-a, 26 students (44.1%) among 59 students in 4-b. As an example, the answer of Participant 29 for item 4-b can be given.

Participant 29:

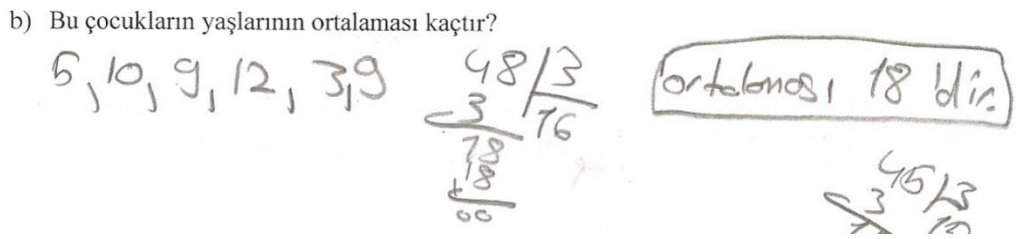


Figure 4.22 Answer of Participant 29 to item 4-b

As seen from the above solution, Participant 29 tried to find the arithmetic average of 5, 10, 9, 12, 3 and 9. First, the total number of values was found, but the result was divided by 3 in order to find the arithmetic average of the data set. Thus, the Participant tried to use the averaging algorithm of the mean but the application was incorrect. Similarly, in the below example, Participant 171 made the *incorrect usage of averaging algorithm* error in 3-a since first the total number was found and then the total number was divided by 10, which was not the number of values for the data set.

Participant 171:

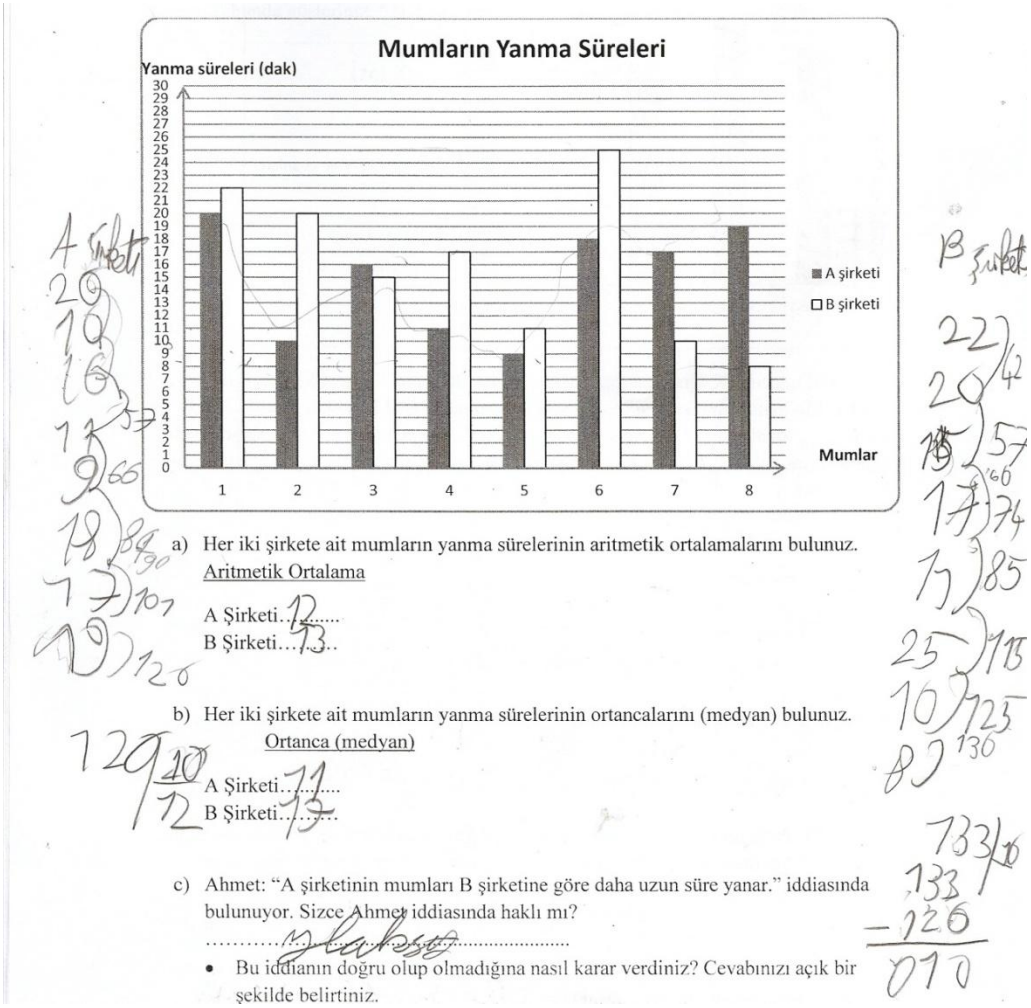
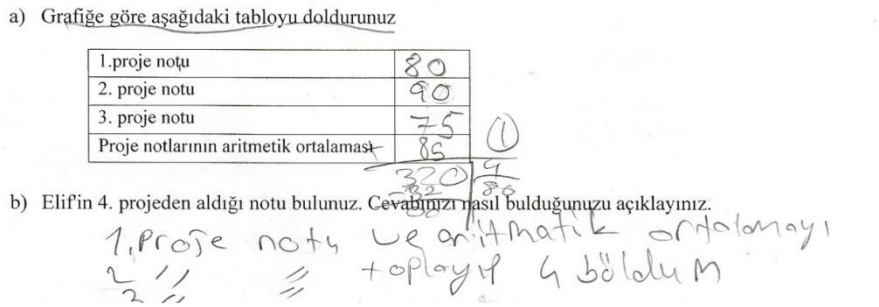


Figure 4.23 Answer of Participant 171 to item 3-a

In addition to these, there were different versions of the *incorrect usage of the averaging algorithm* error type in item 2-ab. Each version is explained with examples. To illustrate, the response of Participant 98 is given below.

Participant 98:


a) Grafiğe göre aşağıdaki tabloyu doldurunuz

1.proje notu	80
2. proje notu	90
3. proje notu	75
Proje notlarının aritmetik ortalaması	85

b) Elif'in 4. projeden aldığı notu bulunuz. Cevabınızı nasıl bulduğunuzu açıklayınız.

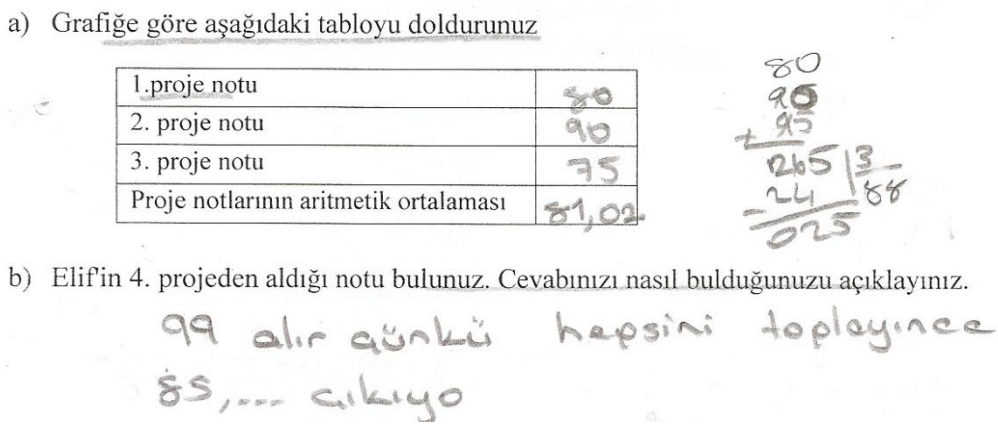
1.Proje notu ve aritmatik ortalamayı toplayıp 4 böldüm

Figure 4.24 Answer of Participant 98 to item 2-ab

Participant 98 added the first, second, and third project scores and the mean of these scores; then it was divided by 4, and 80 was obtained. Thus, the student gave 80 as the answer for the fourth project score.

The second version of this error was also made in item 2-ab. As an example, the response of Participant 48 is presented below:

Participant 48:


a) Grafiğe göre aşağıdaki tabloyu doldurunuz

1.proje notu	80
2. proje notu	90
3. proje notu	75
Proje notlarının aritmetik ortalaması	81,02

b) Elif'in 4. projeden aldığı notu bulunuz. Cevabınızı nasıl bulduğunuzu açıklayınız.

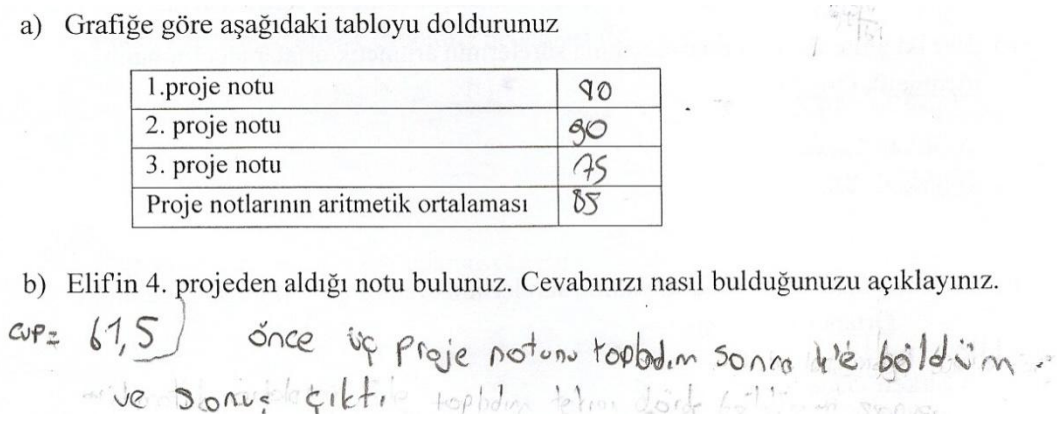
99 alır çünkü hepsini toplayınca 85,... çıkıyo

Figure 4.25 Answer of Participant 48 to item 2-ab

Participant 48 added the first, second and third project scores of Elif and then divided the sum by 3. Then, the student found the answer 88 as the fourth project score.

In addition, the third version was also made in item 2-ab. For example, Participant 7 added the first, second and third project scores; then the sum was divided by 4 to find the fourth project score of Elif. The solution of Participant 7 is given below:

Participant 7:



a) Grafiğe göre aşağıdaki tabloyu doldurunuz

1.proje notu	90
2. proje notu	90
3. proje notu	75
Proje notlarının aritmetik ortalaması	85

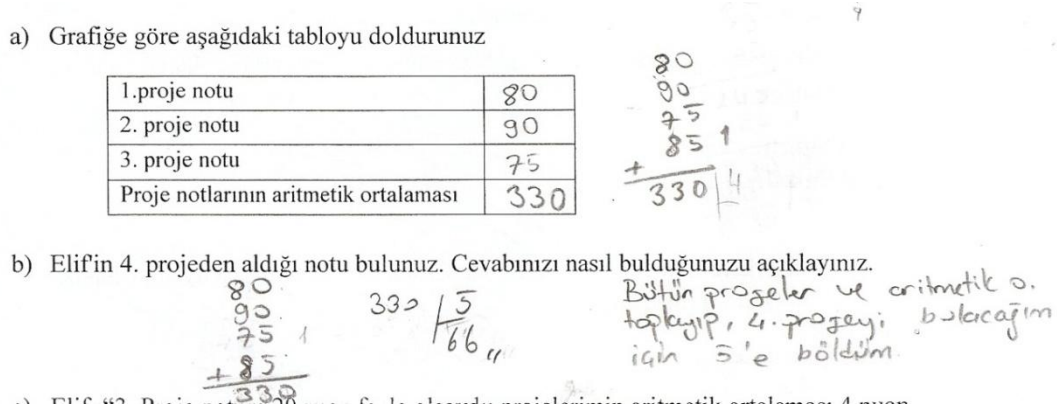
b) Elif'in 4. projeden aldığı notu bulunuz. Cevabınızı nasıl bulduğunuzu açıklayınız.

cvp= 61,5 önce üç proje notunu topladım sonra 4'e böldüm ve sonuç çıktı

Figure 4.26 Answer of Participant 7 to item 2-ab

On the other hand, Participant 1 made another version of this error in item 2-ab. As Participant 1 stated, when he tried to solve the item, he added the first, second, third project scores and the given arithmetic average; then the sum was divided by 5 since there were five values together with the fourth project score. The solution of Participant 1 is presented below.

Participant 1:



a) Grafiğe göre aşağıdaki tabloyu doldurunuz

1.proje notu	80
2. proje notu	90
3. proje notu	75
Proje notlarının aritmetik ortalaması	330

80
90
75
85
+
330 | 4

b) Elif'in 4. projeden aldığı notu bulunuz. Cevabınızı nasıl bulduğunuzu açıklayınız.

80
90
75
+ 85
330

330 | 5
66

Bütün projeler ve aritmetik o. toplayıp, 4. projeyi bulacağım için 5'e böldüm

Figure 4.27 Answer of Participant 1 to item 2-ab

Another error, which was a sub-group of errors based on formal knowledge regarding the concepts of mean, was *accepted close values* and this error type is defined under the following heading.

Accepted Close Values: This error type was observed when the students tried to use the guess and check solution strategy. While a student was trying to construct a possible data set for the given mean and sample size, the student stopped trying to construct a data set although the remainder was not zero. Thus, the students accepted a close value as the given mean. For example, the mean of a data set was given as 5 and the sample size of the data set was 4. A student said that values of the data set might be 3, 4, 6, and 8. However, when the mean of the data set was calculated, the quotient was 5 but the remainder was not zero. Thus, there was an error since the data set was not appropriate for the given mean.

According to Table 4.2, the *accepted close values* error was rarely made when compared with the other error types. This error was made by 3 students (4.2%) among 71 students, 9 students (18.4%) among 49 students and 1 student (1.3%) among 78 students in item 1-a, 1-b and 2-ab respectively.

The following response of Participant 74 for item 1-a is an example for this error.

Participant 74:

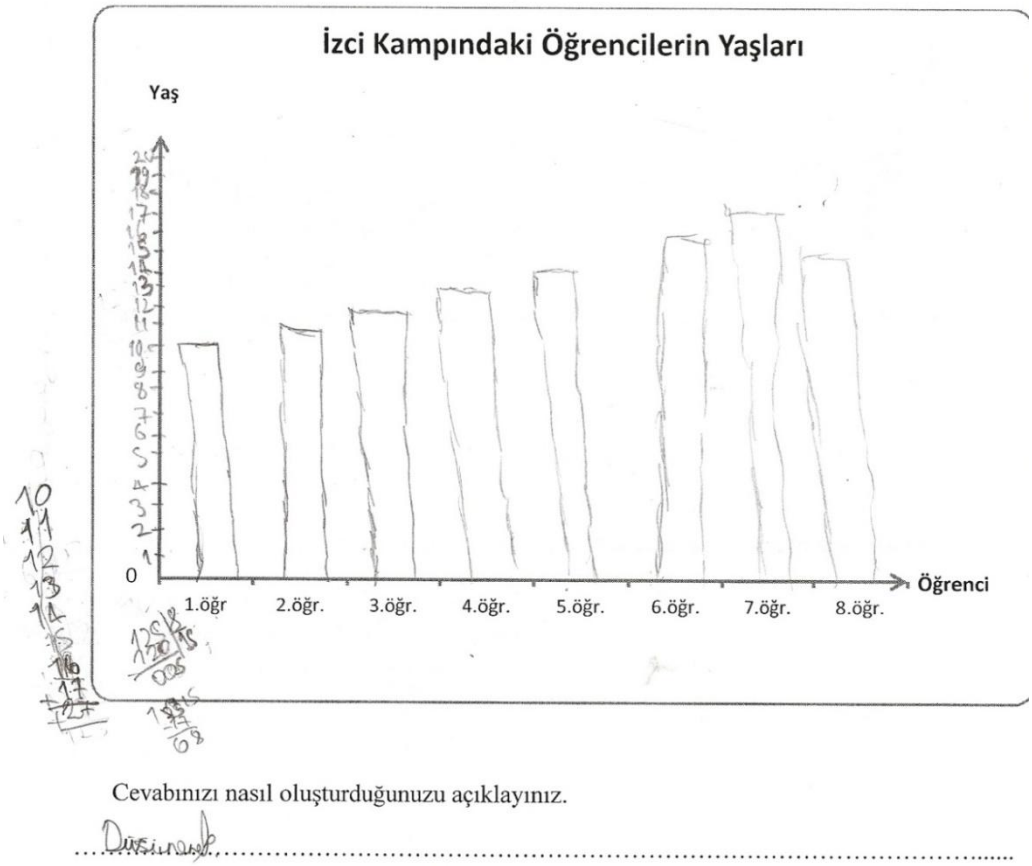


Figure 4.28 Answer of Participant 74 to item 1-a

As seen in the solution of Participant 74, first a possible data set was constructed and then the data set was checked for its appropriateness for the given mean. However, the total number of values in the data set was found to be 125. Then, 125 was divided by 8 and the quotient was found as 15 but the remainder was not 0. Thus, this error was categorized as *accepted close values* error type since the mean of the constructed data set was not 15, but the total number values were close to the actual total number of values. Moreover, the solution of Participant 98 for item 1-b, which is presented below, was similar that of Participant 74.

Participant 98:

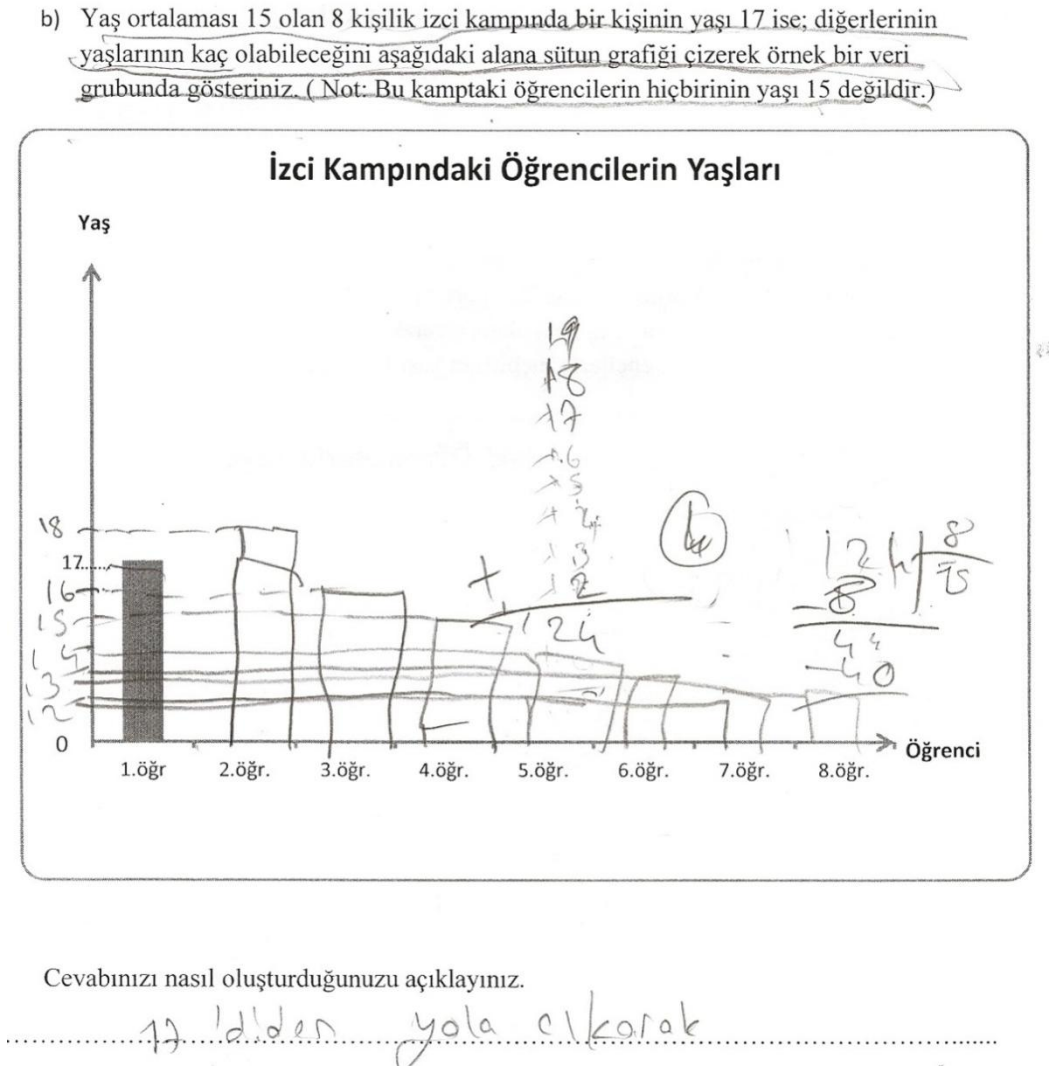


Figure 4.29 Answer of Participant 98 to item 1-b

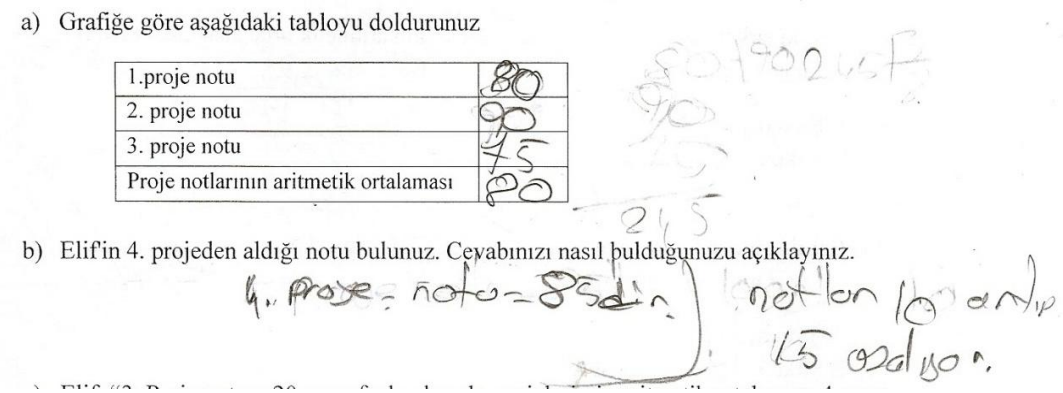
As can be seen in the solution provided by Participant 98, a data set was constructed and the sum of the data set was found to be 124. Subsequently, the sum was divided by 8, and 15 was found as the mean of the data set. However, the remainder of the division operation was 4. Thus, Participant 98 made the accepted close values error in item 1-b since the quotient was found to be 15 but remainder was not zero.

Another sub-error of errors based on the formal knowledge category was *looking pattern*. The details regarding this error is given below.

Looking Pattern: The students who made this error tried to find a pattern for the values in a data set although there was no pattern. According to the results of the study, this error was seen when some of the values in a data set were given and a missing value was asked for.

As can be observed in Table 4.2, the *looking pattern* error was committed in item 2-ab and 6 by 3 students (3.9%) among 78 students and 7 students (7.1%) among 98 students, respectively. To illustrate, Participant 231 made the *looking pattern* error while trying to solve item 2-ab. His response is presented below:

Participant 231:


a) Grafiğe göre aşağıdaki tabloyu doldurunuz

1.proje notu	80
2. proje notu	90
3. proje notu	75
Proje notlarının aritmetik ortalaması	80

b) Elif'in 4. projeden aldığı notu bulunuz. Cevabınızı nasıl bulduğunuzu açıklayınız.

4. proje notu = 85 dir. notlar 10 artıp 15 azalıyor.

Figure 4.30 Answer of Participant 231 to item 2-ab

It was claimed in the explanation of Participant 231 that there were a pattern among Elif's project scores. The pattern was claimed to be an increase of 10 point from the first to the second score and then a decrease of 15 points from the second to the third project score. Thus, the claimed pattern was an increase of 10, a decrease of 15, an increase of 10 etc.. According to the pattern, Elif's project score was an increase of 10 points, thus yielding 85. However, when the fourth project score was 85, the mean of these four project scores was not equal to 85. Thus, the *looking pattern* error type was made since there was not any pattern among the project scores of Elif.

On the other hand, some students made this error in item 5. As an example, the answer of Participant 168 is given below.

Participant 168:

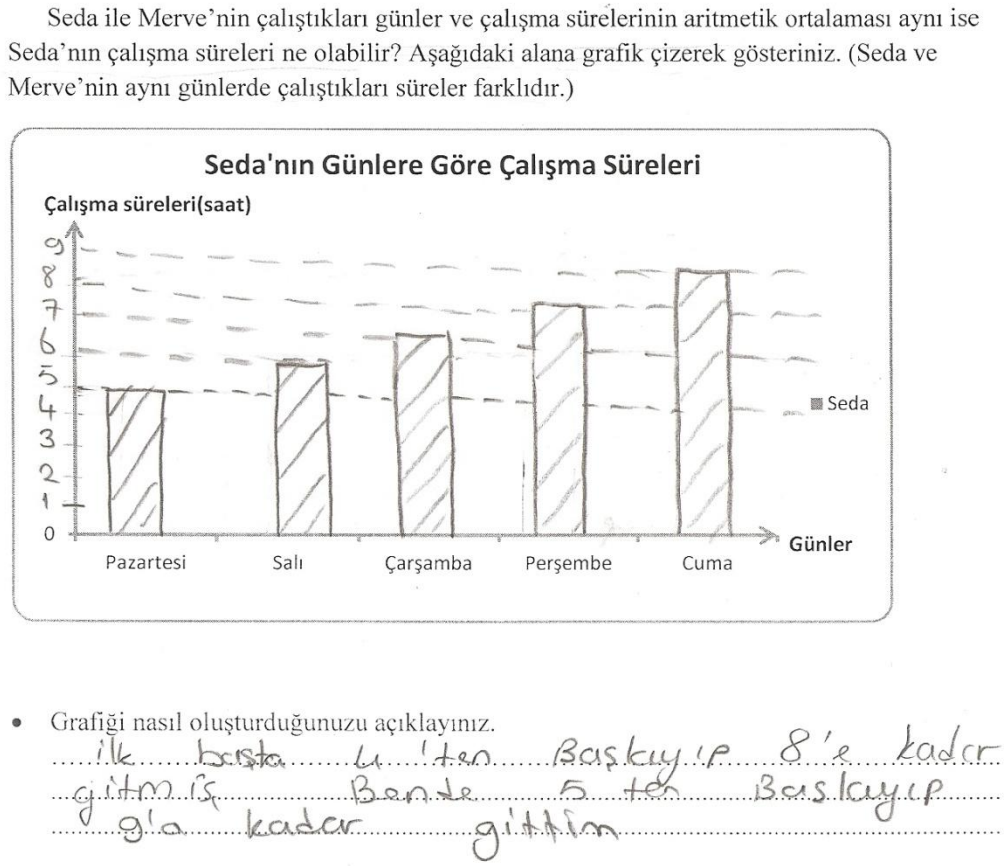


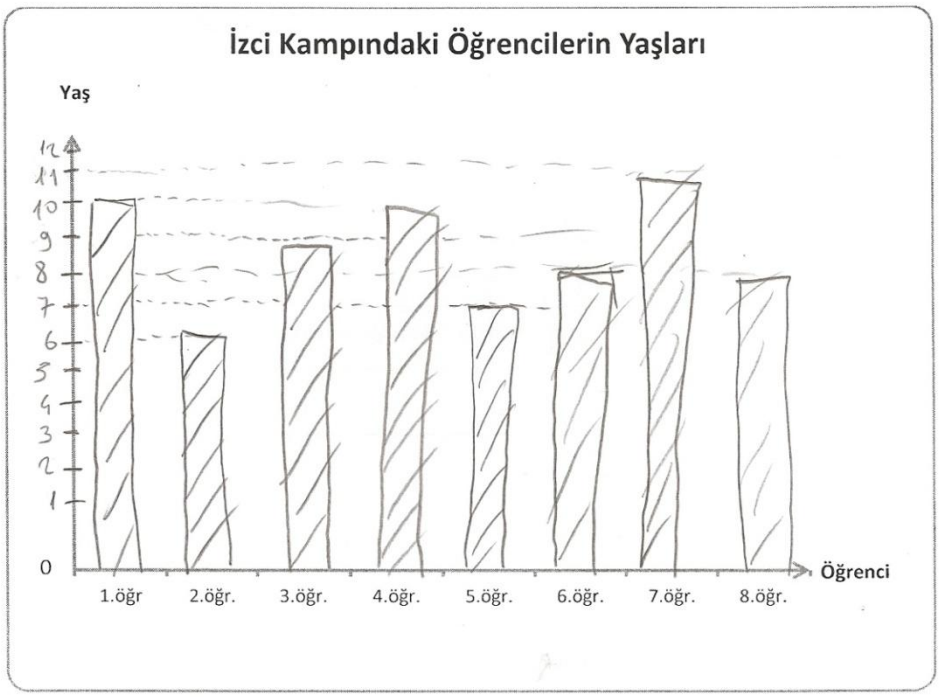
Figure 4.31 Answer of Participant 168 to item 6

As Participant 168 stated, 5 hours was selected as the working duration of Seda on Monday and then continued as 6, 7, 8 and 9 for the other days since Merve's working duration on Monday was 4 and increased one hour on Tuesday to be 5 and then continued increasing one by one from day to day. Thus, the participant accepted the working durations of Merve as a pattern and applied this pattern in her method of solution. However, the accepted pattern was not appropriate to reach the acceptable data set regarding the working durations of Seda since the means of working durations of had to be the same in order to be an acceptable data set.

Turned to Find Smaller or Larger Numbers: When the construction of a possible data set was required, the students who made this error selected all the values as smaller than the mean of the data set or larger than the mean of the data set.

As can be observed in Table 4.2., this error was made by 41 students (57.8%) among 71 students in item 1-a and 26 students (53.1%) among 49 students n item 1-b. For example, Participant 207 was one of the students who made this error in item 1-a.

Participant 207:



Cevabınızı nasıl oluşturduğunuzu açıklayınız.

15 yaşın altında ortalama verdim

Figure 4.32 Answer of Participant 207 to item 1-a

As Participant 207 said, each value was selected according to the given mean. In more detail, all values of the constructed data set were smaller than 15 which was the mean of the data set. Thus, the error of *turned to find smaller or larger numbers* was made. The following response of Participant 226 exemplifies this error in item 1-b.

Participant 226:

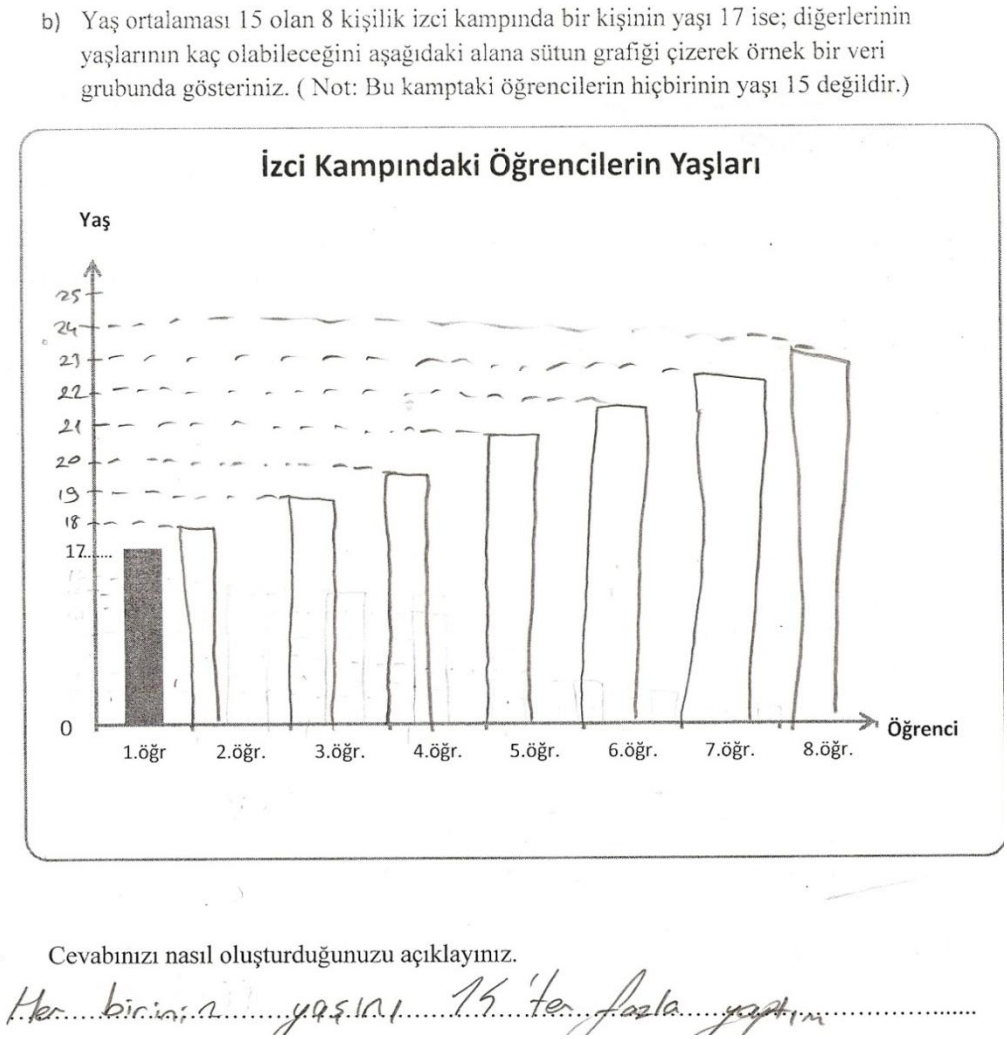


Figure 4.33 Answer of Participant 226 to item 1-b

As Participant 226 stated, each value of the constructed data set was larger than the given mean in the problem context. Thus, the *turned to find smaller or larger numbers* error was made in the response of Participant 226.

In the following part, each error in the other errors category related to the mean is defined and examples regarding these errors are given.

4.2.1.3 Other Errors

In this study, other errors regarding the mean concept were *not meeting all requirements of a problem*, *forming incomplete data set* and *incorrect reading of values from graph*.

Not Meeting All Requirements of a Problem: The students who made this error did not meet all the requirements of a given item. For example, in item 6, answers of many students were not appropriate to the written note or expression in the problem context. In more detail, in item 6, although there was a note which stated 'Seda and Merve have different working durations on the same day', many students decided at least one of their working durations was the same on the same days. Thus, this error was made in item 6. As can be observed in Table 4.2, 83 students (84.7%) among 98 students made the *not meeting all requirements of a problem* error in item 6. For instance, the response of Participant 193, which is presented below, exemplifies this error:

Participant 193:

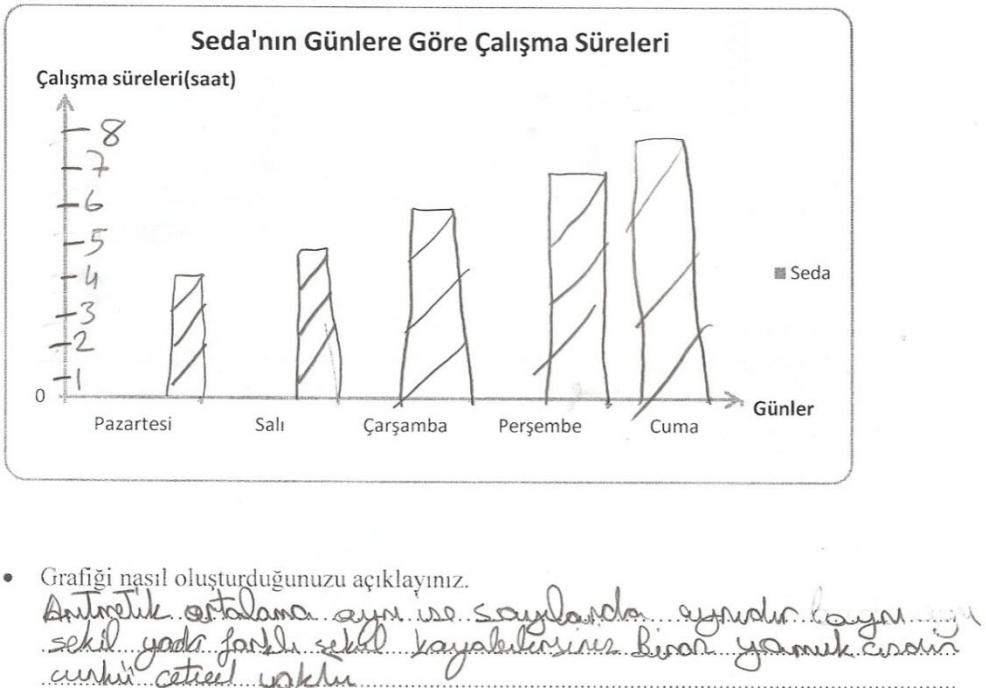


Figure 4.34 Answer of Participant 193 to item 6

As can be seen from the graph constructed by Participant 193, each value was selected the same with the given data set in the question part of item 6. More specifically, the

working durations of Seda was stated to be the same with that of Merve on the same days. In fact, the mean of the data set was equal to the given data set, but the item required that Merve and Seda have different working durations on the same days. Thus, the error of *not meeting all requirements of a problem* was made.

Participant 233 also made this error since the working duration of Seda was the same with that of Merve on Wednesday. The constructed data set for item 6 is given below.

Participant 233:

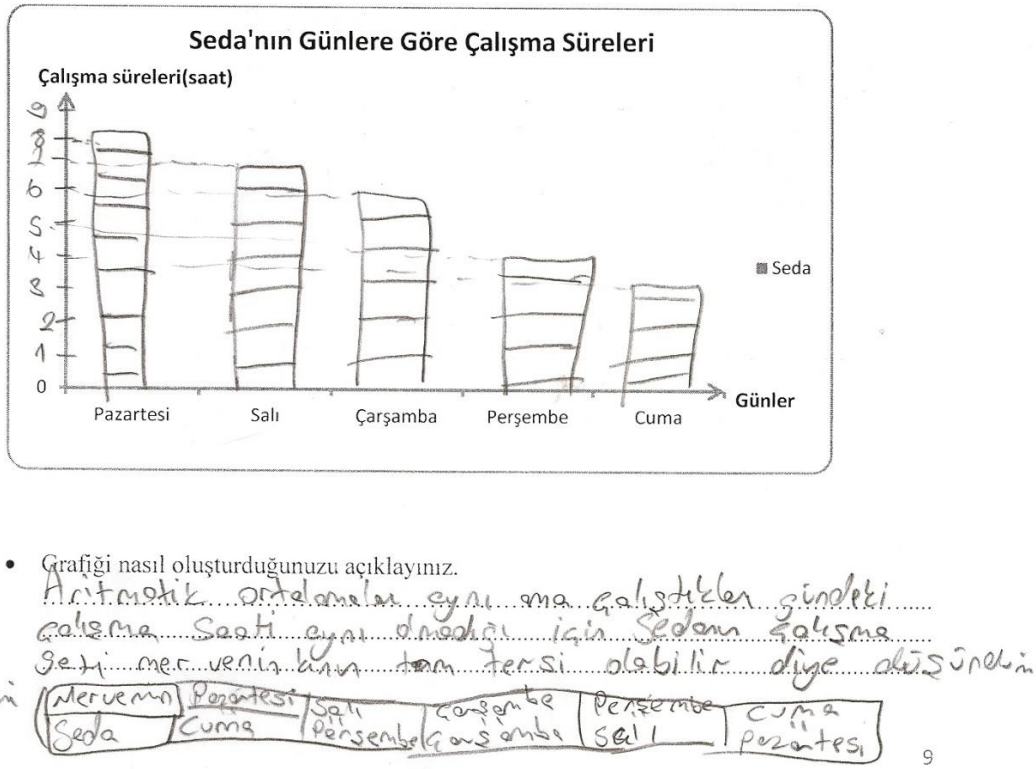
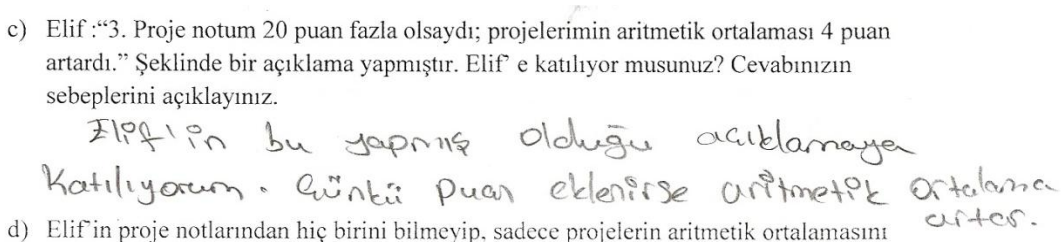


Figure 4.35 Answer of Participant 233 to item 6

Furthermore, this error was made by 4 students (5.6%) among 71 students and 2 students (4.1 %) among 49 students in item 1-a and 1-b respectively. On the other hand, this error was made in item 2-c by 25 students (51.0%) among 49 students since the students evaluated whether or not the given mean was increased but the item required the calculation of this increase. Thus, this error occurred. To illustrate, Participant 181 made the error of *not meeting all requirements of a problem* in item 2-c as presented below.

Participant 181:



c) Elif :"3. Proje notum 20 puan fazla olsaydı; projelerimin aritmetik ortalaması 4 puan artardı." Şeklinde bir açıklama yapmıştır. Elif' e katılıyor musunuz? Cevabınızın sebeplerini açıklayınız.

Elif'in bu yapmış olduğu açıklamaya katılıyorum. Çünkü puan eklenirse aritmetik ortalama artar.

d) Elif'in proje notlarından hiç birini bilmeyip, sadece projelerin aritmetik ortalamasını

Figure 4.36 Answer of Participant 181 to item 2-c

As Participant 181 stated, the mean of Elif's project scores were evaluated in terms of only increase or decrease. However, the statement of item 2-c required the identification of the amount of increase in Elif's project scores. Thus, this error was made since in her response she only stated that the arithmetic average of her project scores would increase; however, the response statement expected was the quantity of the increase.

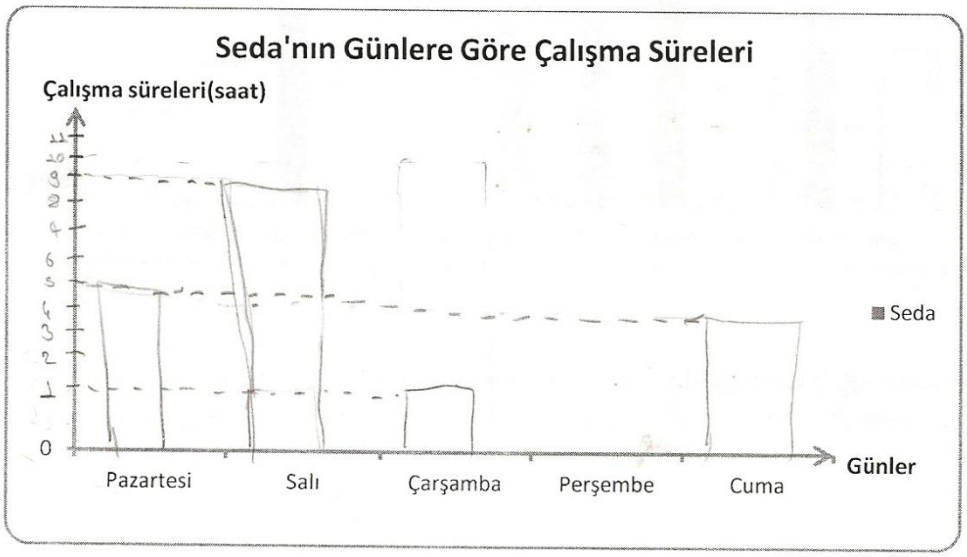
On the other hand, *forming incomplete data set* was another error type, which is defined and exemplified under the following heading.

Forming Incomplete Data Set: The students who made this error did not construct an appropriate data set for a given sample size or did not use all the values of a given data set in order to find the average of the data set. To illustrate, in a problem, sample size of a data set was given as 5 and mean of the data set was given and the construction of a possible data set was required. The student who made this error constructed a data set but the data set did not include five values. Thus, the constructed data set was incomplete.

As seen from Table 4.2, 1 student (1.4%) among 71 students in item 1-a and 2 students (2.0%) among 98 students in item 6 made this error. For example, the answer of Participant 80 for item 6 is presented below.

Participant 80:

Seda ile Merve'nin çalıştıkları günler ve çalışma sürelerinin aritmetik ortalaması aynı ise Seda'nın çalışma süreleri ne olabilir? Aşağıdaki alana grafik çizerek gösteriniz. (Seda ve Merve'nin aynı günlerde çalıştıkları süreler farklıdır.)



- Grafiği nasıl oluşturduğunuzu açıklayınız.
  Birsaa tahmin etmeye çalıştım birsaa yaborda bırakarak yapmaya çalıştım.

Figure 4.37 Answer of Participant 80 to item 6

As can be observed in the constructed data set of Participant 80, Seda's working duration on Thursday could not be decided. Thus, the data set was incomplete. Moreover, Participant 226 also made this error while trying to construct a possible data set for the working durations of Seda.

Participant 226:

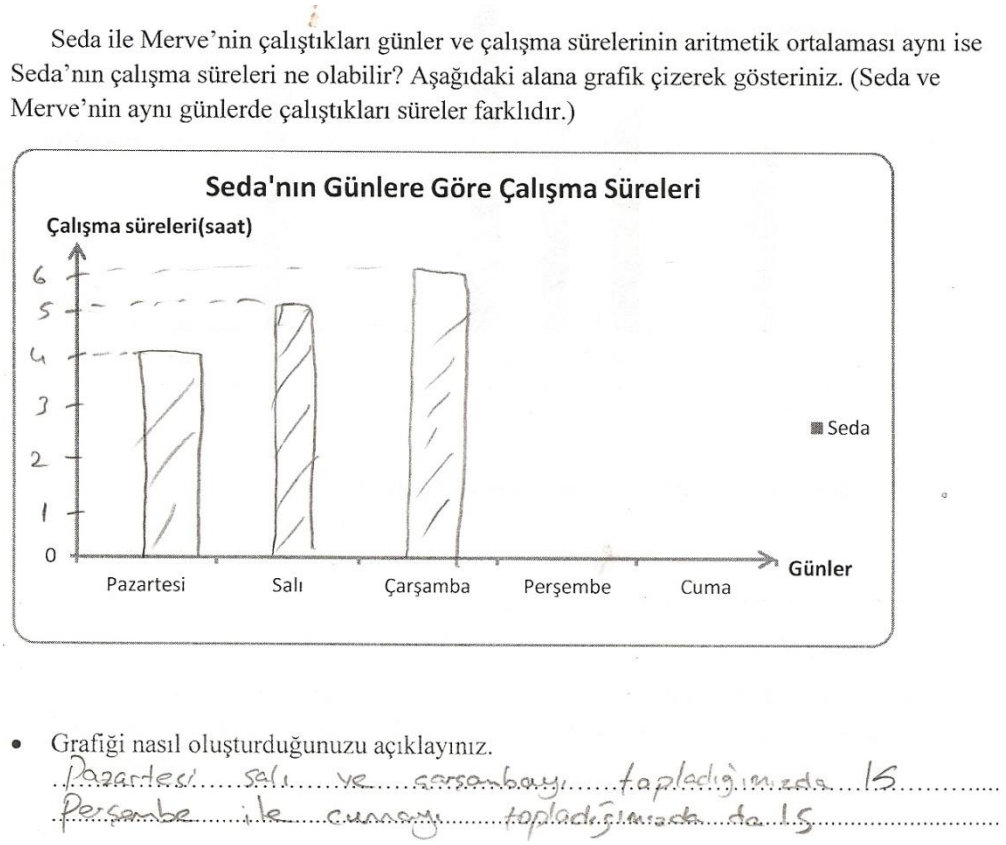


Figure 4.37 Answer of Participant 226 to item 6

According to the constructed data set by Participant 226, Seda's working durations on Thursday and Friday could not decided. Although she explained that the sum of the values of Thursday and Friday was 15, the data set was incomplete since these two pieces of data could not decided.

Incorrect Reading of Values from Graph: When values were given on a bar graph, the students who made this error read one or more values incorrectly from the graph. Thus, when one or more values of a data set was incorrect, the mean of the data set would have been different from the correct value.

As can be seen from Table 4.2, the error of *incorrect reading of values from graph error* was made in items 2-ab, 3-a, 4-b and 6. It was seen that 4 students (5.1%) among 78 students in item 2-ab, 5 students (8.1%) among 62 students in item 3-a, 2 students

(3.4%) among 59 students in item 4-b, and 3 students (3.1%) among 98 students in item 6 made this error while trying to solve these items. For instance, the error of *incorrect reading of values from graph* was seen in the response of Participant 31 in item 3-a, which is presented below:

Participant 31:

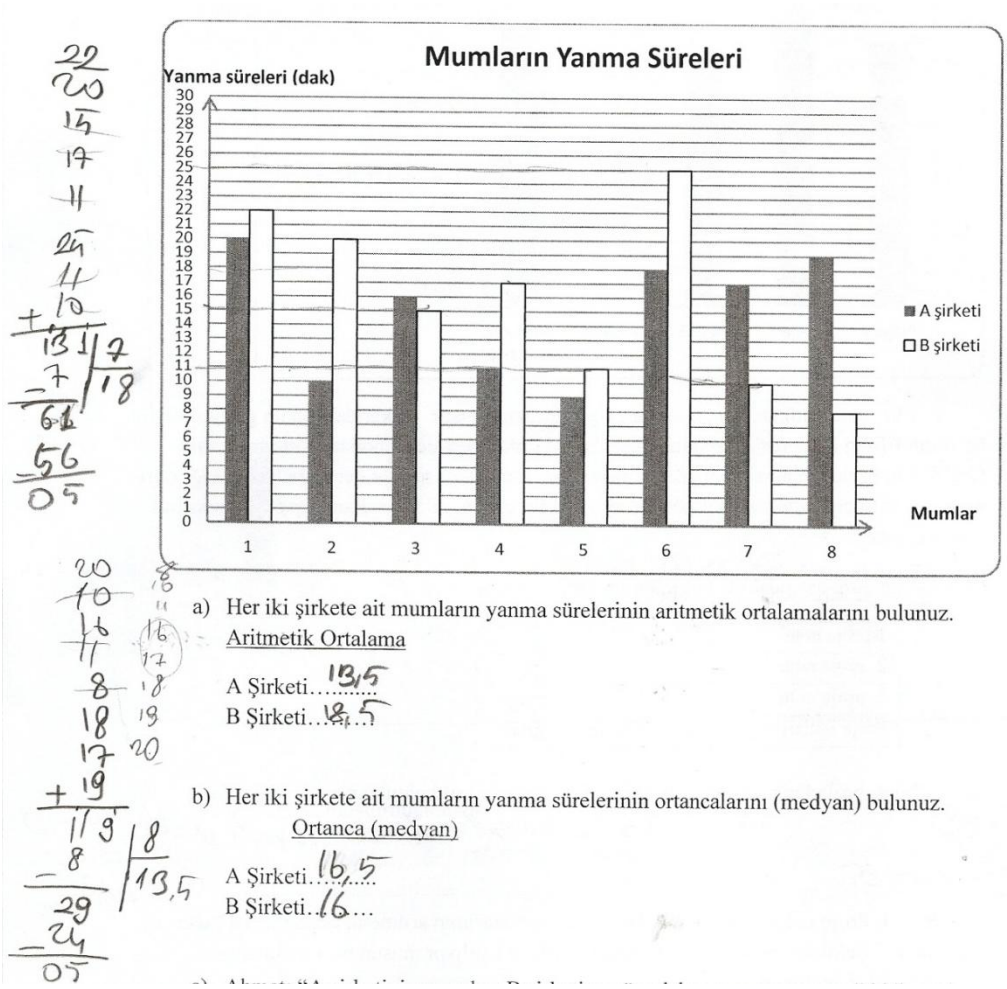


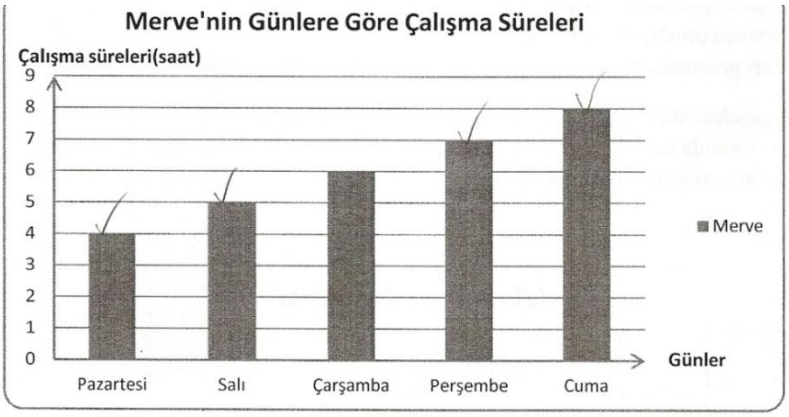
Figure 4.38 Answer of Participant 31 to item 3-a

As can be observed in the solution of Participant 31 for item 3-a, the burning duration of the fifth candle of Company A were read incorrectly since the burning duration of the fifth candle was 9 but it was read as 8. For this reason, the sum of the burning durations of candles was found to be 119 instead of 120; then the arithmetic average of the burning durations of the candles of Company A was found incorrectly. Moreover, the

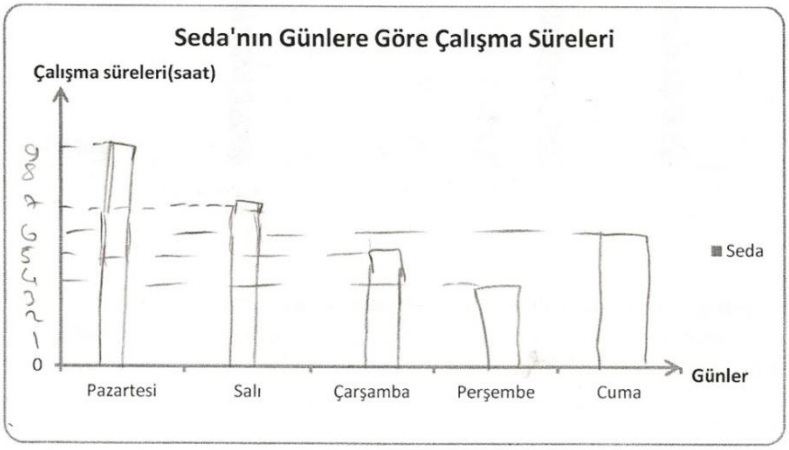
same error was also made in the calculation of the arithmetic average of the burning durations of the candles for Company B.

The response of Participant 3 can be given as another example of this error. Her response is given below.

Participant 3:



Seda ile Merve'nin çalıştıkları günler ve çalışma sürelerinin aritmetik ortalaması aynı ise Seda'nın çalışma süreleri ne olabilir? Aşağıdaki alana grafik çizerek gösteriniz. (Seda ve Merve'nin aynı günlerde çalıştıkları süreler farklıdır.)



- Grafiği nasıl oluşturduğunuzu açıklayınız.

sadece günleri değiştirdim

Figure 4.39 Answer of Participant 3 to item 6

As Participant 3 stated, she used the same working durations of Merve in order to find the possible working durations of Seda. However, she read Merve's working duration on

Friday as 9 although there was no day which had 9 working hours. Thus, Participant 3 had read a value from the given graph incorrectly.

In the following part, the students' errors regarding the median concept are discussed. The frequencies and percentages of the errors made by the students related to the median are presented.

4.2.2 Errors Regarding the Median

Errors regarding the median were encountered in the responses given to items 3-b and 5. The calculation of the medians of each data set was required in item 3-b and the construction of a possible data set for the given median and mode was required in item 5. However, errors were observed in the students' wrong answers for item 3-b and inappropriate data sets were constructed for item 5. Analysis results of this study showed that errors related to the median could be categorized into three main groups, namely *algorithmically based errors, errors based on formal knowledge* and *other errors*. Moreover, a sub-group of algorithmically based errors is *operational error* and sub-groups of errors based on formal knowledge are *wrong decision on unordered data set, wrong decision on ordered data set* and *wrong interpretation of graph*. On the other hand, *forming incomplete data set, incorrect reading of values from graph* and *incorrect largest and/or smallest data values* were accepted as sub-groups of other errors. The frequencies and percentages participants committing these errors in item 3-b and 5 are presented in Table 4.3 below.

Table 4.3 Frequencies (and percentages) of the students' errors regarding the median

	Algorithmically Based Errors	**Errors Based on Formal Knowledge**			**Other Errors**			
ERRORS ITEMS	Operational Error	Wrong Decision on Unordered Data Set	Wrong Decision on Ordered Data Set	Wrong Interpretation of Graph	Forming Incomplete Data Set	Incorrect Reading of Values from Graph	Incorrect Largest and/or Smallest Data Values	Total
Item 3-b	9 (9.6%)	23 (24.5%)	26 (27.7%)	29 (30.8%)	5 (5.3%)	2 (2.1%)	0 (0%)	94
Item 5	0 (0%)	0 (0%)	0 (0%)	0 (0%)	62 (81.6%)	0 (0%)	14 (18.4%)	76
Total	9	23	26	29	67	2	14	

As it can be observed in Table 4.3, there is one type of error within algorithmically based errors, three types of errors within errors based on formal knowledge and three types of errors within other errors related to the median. According to the table, the majority of the students who had given a wrong answer while trying to find the median of a given data set in item 3-b made an error of *wrong interpretation of graph* error. Moreover, this error was made by 29 students (30.8%) among 94 students in item 3-b. In addition, the most common error type in item 5 was *forming incomplete data set* since the students had not constructed an appropriate data set for the given sample size. It was seen that 62 students (81.6%) among 76 students had made this error in item 5. Examples of the students' wrong answers for each addressed error type is presented below under the category headings given in Table 4.3.

4.2.2.1 Algorithmically Based Errors

Operational Errors: As defined in errors regarding the mean, the students made calculation errors while doing addition, subtraction, multiplication or division operations. In fact, the solution strategy used by students was correct but they made an operational error.

As can be observed in Table 4.3, this error was made by 9 students (9.6%) among 94 students in item 3-b. As an example, the response of Participant 5 is presented below.

Participant 5:

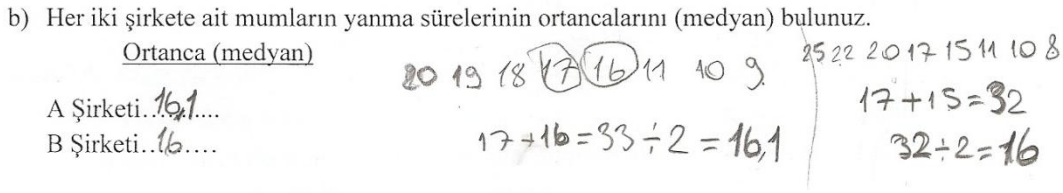


Figure 4.40 Answer of Participant 5 to item 3-b

As can be seen in the response of Participant 5, each data set was ordered from the largest to smallest. Then, the middle values of the two ordered data sets were added and the sums were divided by 2. However, an operational error was made while trying to find the median of the data set regarding Company A since the sum of the middle values of the data set was not divided correctly by 2. Furthermore, this error was also made by Participant 142 in item 3-b.

Participant 142:

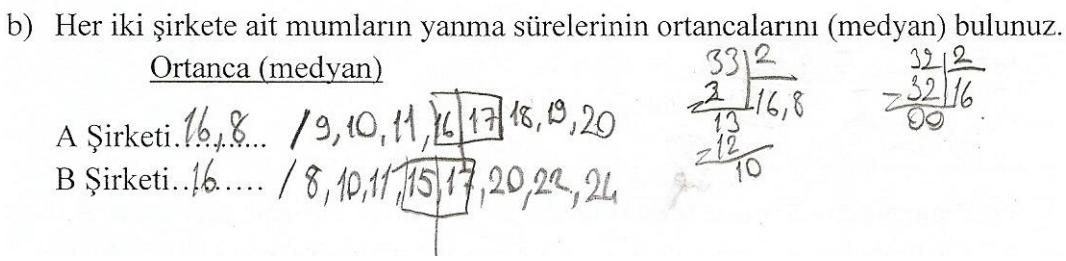


Figure 4.41 Answer of Participant 42 to item 3-b

Similar to Participant 5, Participant 142 initially ordered each data set. Secondly, the middle values of the data sets were added since the data sets had an even number of values. Lastly, the sums were divided by 2 in order to find the median of each data set. However, while dividing the sum of the middle values of the burning durations of the candles of Company A by 2, an operational error was made.

4.2.2.2 Errors Based on Formal Knowledge

There were three sub-groups of errors based on formal knowledge, namely *wrong decision on unordered data set*, *wrong decision on ordered data set*, and *wrong interpretation of graph*.

Wrong Decision on Unordered Data Set: This error type was seen when the students tried to find the median of a data set in item 3-b. The students did not order the obtained data set from the given bar graph in item 3-b to find the median of the data set. Thus, the median of the data set was not found correctly. According to Table 4.3, 23 students (24.5%) among 94 students made this error in item 3-b. When the response of the students was investigated, some of the students who had made this error type found the mid-point of the two middle values of the unordered data set since there was an even number of values.

Different versions of the *wrong decision on unordered data set* error were made by the students, which are described below with example responses for each version.

Firstly, some of the students gave two middle values from an unordered data set as the answer. To illustrate, Participant 117 wrote the burning durations of the candles of Company A and Company B without ordering them. Then, he found the median of the data set of Company A as '11, 9' and median of the data set of Company B as '17, 11'. Thus, he found two middle scores in the unordered data sets as the median of these two data sets. His response is presented below.

Participant 117:

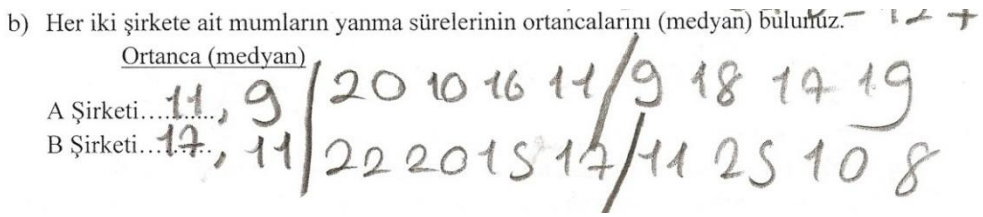


Figure 4.42 Answer of Participant 117 to item 3-b

Secondly, a student subtracted the small middle value from the large middle value of an obtained unordered even data set and then the results were given as the answers. In item 3-b, Participant 59 wrote the burning durations of the candles for each company based on the given bar graph. 9 and 11 were the middle scores of unordered data set for Company A and then he subtracted 9 from 11 and found the median of burning durations of candles of Company A as 2. In the same way, the participant subtracted 11 from 17 and found the median of the burning durations of candles of Company B as 6 and the solution is presented below:

Participant 59:

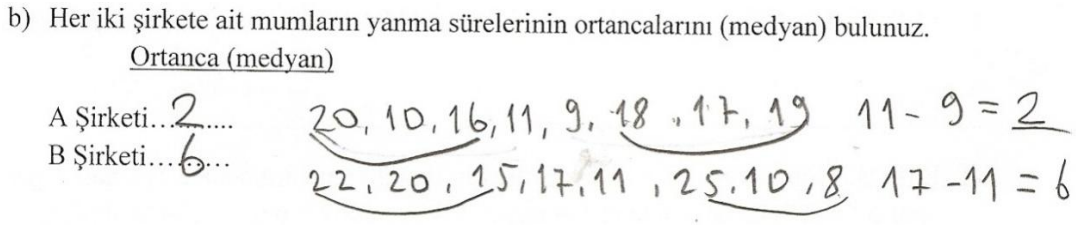

b) Her iki şirkete ait mumların yanma sürelerinin ortancalarını (medyan) bulunuz.
Ortanca (medyan)
A Şirketi…2….
B Şirketi…6…
20, 10, 16, 11, 9, 18, 17, 19 11 - 9 = 2
22, 20, 15, 17, 11, 25, 10, 8 17 - 11 = 6

Figure 4.43 Answer of Participant 59 to item 3-b

Thirdly, some of the students selected one of the middle values in the unordered even data set as the answer. For example, Participant 95 wrote each data set based on the given bar graph and found the median of the first data set as 11 and the second data set as 17 as given below:

Participant 95:

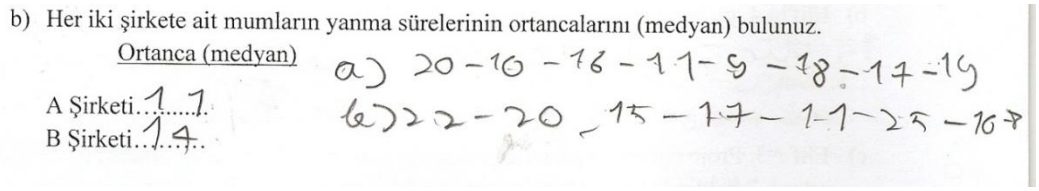
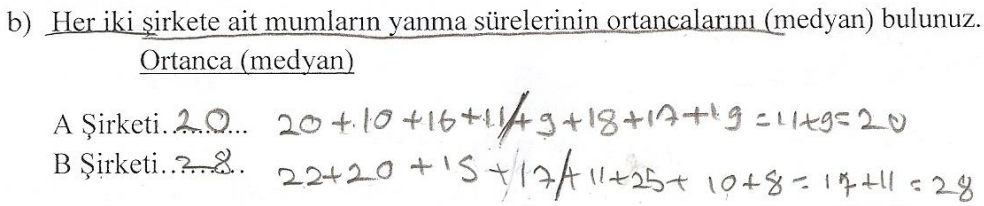

b) Her iki şirkete ait mumların yanma sürelerinin ortancalarını (medyan) bulunuz.
Ortanca (medyan)
A Şirketi…11…
B Şirketi…17…
a) 20 - 10 - 16 - 11 - 9 - 18 - 17 - 19
b) 22 - 20 - 15 - 17 - 11 - 25 - 10 - 8

Figure 4.44 Answer of Participant 95 to item 3-b

Lastly, some of the students added the middle values of the unordered even data set and then the result was given as the answer. For instance, the response provided by Participant 150 showed that for the first data set, the middle scores in the unordered data

set, 11 and 9, were added and 20 was given as the median of the data set. Furthermore, the same procedure was applied in the second data set. The solution of Participant 150 is stated below.

Participant 150:


b) Her iki şirkete ait mumların yanma sürelerinin ortancalarını (medyan) bulunuz.
Ortanca (medyan)

A Şirketi..20... 20+10+16+11+9+18+17+19=11+9=20
B Şirketi..28.. 22+20+15+17+11+25+10+8=17+11=28

Figure 4.45 Answer of Participant 150 to item 3-b

On the other hand, another formally based error was *wrong decision on ordered data set*. Details regarding this error are mentioned under the following heading.

Wrong Decision on Ordered Data Set: When the median of a data set was asked, the students who made this error first ordered the data set but they could not find its median. This error was seen in item 3-b since the item asked for the median of two data sets given on a bar graph. As seen in Table 4.3, this error was made by 26 students (27.7%) among 94 students in item 3-b. The analysis of item 3-b indicated that there were different versions of this error of *wrong decision on ordered data set* since the data set included an even number of values.

Different versions of the *wrong decision on ordered data set* error were made by students. Each version is described with examples below.

In the first version of the error, the student had given two middle values from the ordered data set as the answer for each data set in item 3-b. For instance, Participant 21 was one of the students who had made first version of this error and response is given below.

Participant 21:

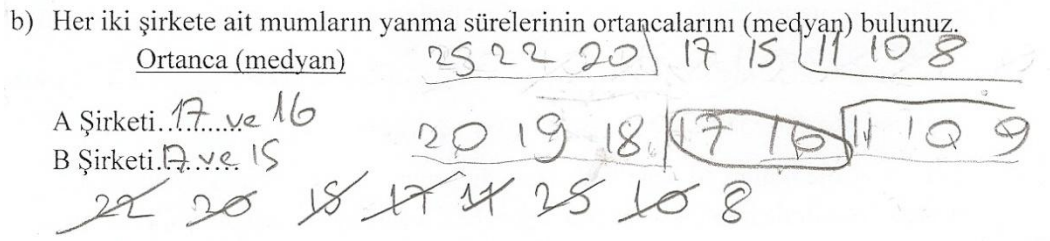


Figure 4.46 Answer of Participant 21 to item 3-b

As can be seen in the response of Participant 21, each data set was ordered from large to small values but the two middle values for each data set was given as the answers.

In the second version of the error, the student subtracted the small middle value from the large middle value within the ordered values of the data sets; then the results were divided by two in order to find the median of the data sets in item 3-b. The following response of Participant 2 exemplifies the second version of the *wrong decision on ordered data set error.*

Participant 2:

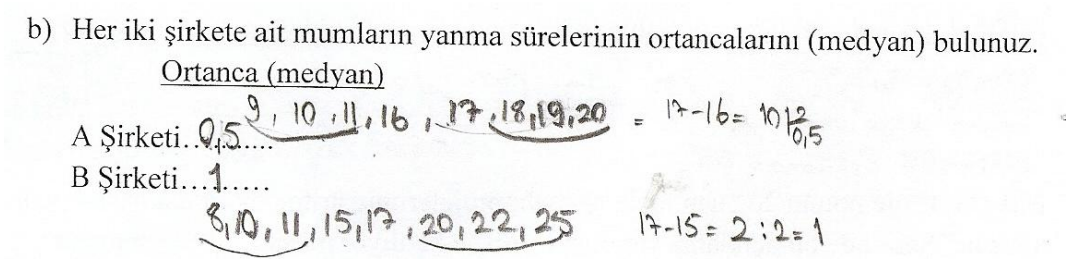


Figure 4.47 Answer of Participant 2 to item 3-b

As can be seen above in the response provided by Participant 21, at the beginning, the values of each data were ordered from small to large. Subsequently, 16 was subtracted from 17, yielding 1. Then, 1 was divided by 2 and 0.5 was found as the median of the burning durations of the candles for Company A. Additionally, the same procedure was applied for Company B. Thus, this version of the *wrong decision on ordered data set* error was made by Participant 21.

In the third version of this error, the student selected the smaller one of the middle values from each ordered data set as the answer(s). As an example, the following response of Participant 158 can be given.

Participant 158:

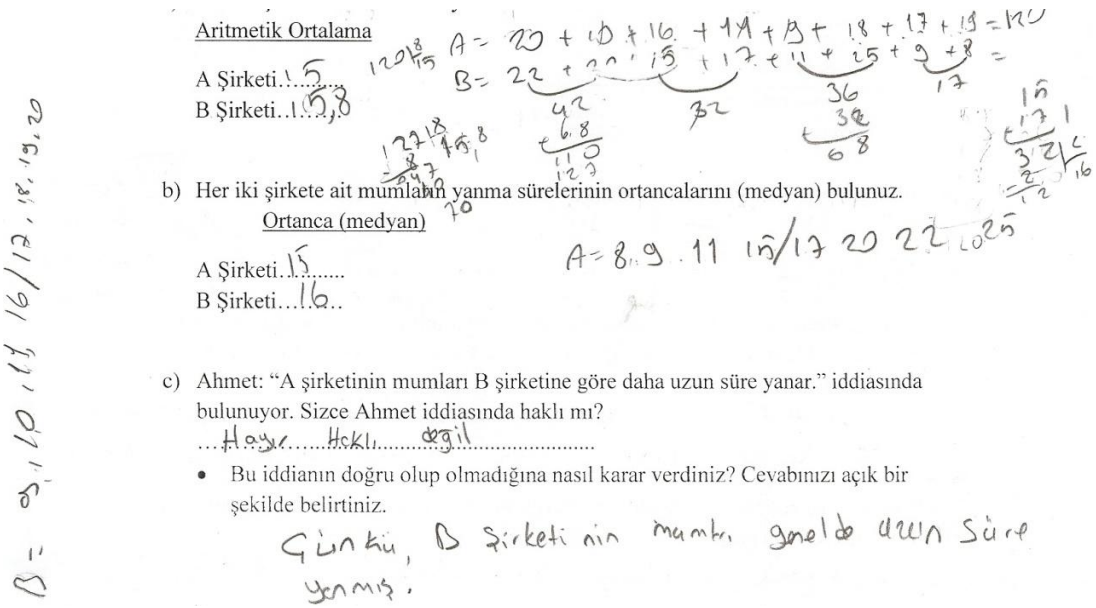
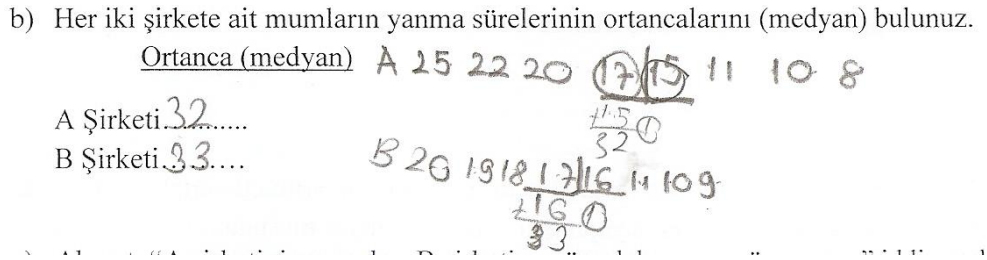


Figure 4.48 Answer of Participant 158 to item 3-b

Based on the response of Participant 158 given above, it can be said that first the data set of Company A was ordered, then 15, which was smaller one of the middle values of the data set was found as the median of the data set. The median of the data set regarding Company B was found in the same way.

In the four version, the two middle values of the ordered data sets were added and the results were given as the answer. For instance, Participant 184 made the fourth version of this error as presented below.

Participant 184:


b) Her iki şirkete ait mumların yanma sürelerinin ortancalarını (medyan) bulunuz.

Ortanca (medyan)

A Şirketi 32

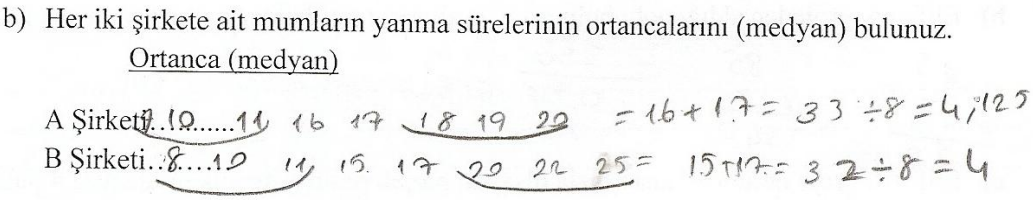
B Şirketi 33

Figure 4.49 Answer of Participant 184 to item 3-b

According to the response of Participant 184, it can be said that the fourth version of *wrong decision on ordered data set* error was made since 17 and 15 which were middle values of the ordered values of Company A was added and median of the data set was found as 32. Moreover, the median of the burning durations of candles of Company B was found in the same way.

In the last version of this error, the student first added two middle scores of an ordered even data set but then divided it by the number of values in the data set. For example, in order to find the median of the burning durations of the candles of Company B in item 3-b, Participant 61 added 15 and 17; then the sum divided by 8, which was the number of values, thus 4 was given as the answer. The response of Participant 61 is presented below.

Participant 61:


b) Her iki şirkete ait mumların yanma sürelerinin ortancalarını (medyan) bulunuz.

Ortanca (medyan)

A Şirketi 7 10 11 16 17 18 19 20 = 16 + 17 = 33 ÷ 8 = 4,125

B Şirketi 8 10 11 15 17 20 22 25 = 15 + 17 = 32 ÷ 8 = 4

Figure 4.50 Answer of Participant 61 to item 3-b

Another sub-group of errors based on formal knowledge regarding the median was *wrong interpretation of graph*. Details and examples of this error are presented under the following heading.

Wrong Interpretation of Graph: The student who made this error tried to find the median of a data set directly from the given bar graph, but the answer found was not correct. This error was made in 3-b since the item asked for the median of two data sets given on a bar graph. As presented in Table 4.3, the error of *wrong interpretation of graph* was the most frequently made error in item 3-b; 29 students (30.8%) among 94 students had made this error in the item. For example, Participant 187 made this error in item 3-b as can be seen in the response of the participant given below:

Participant 187:

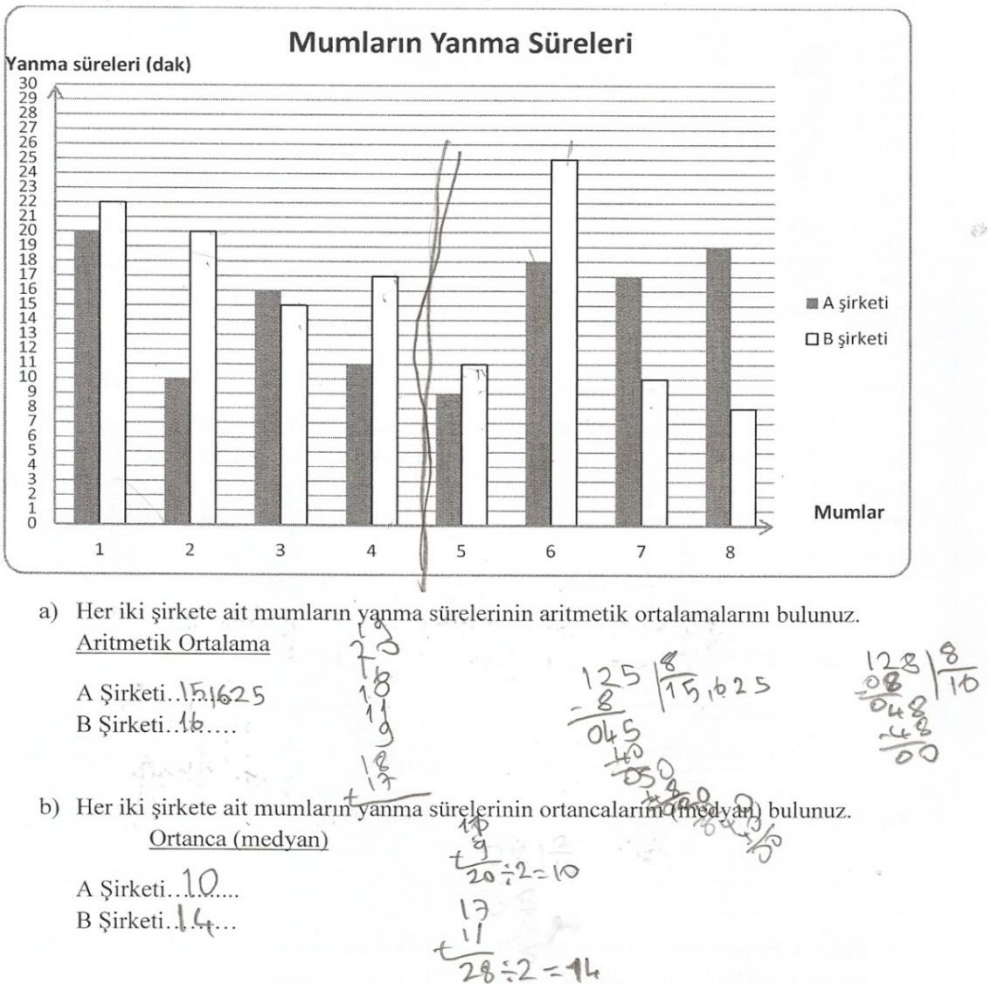


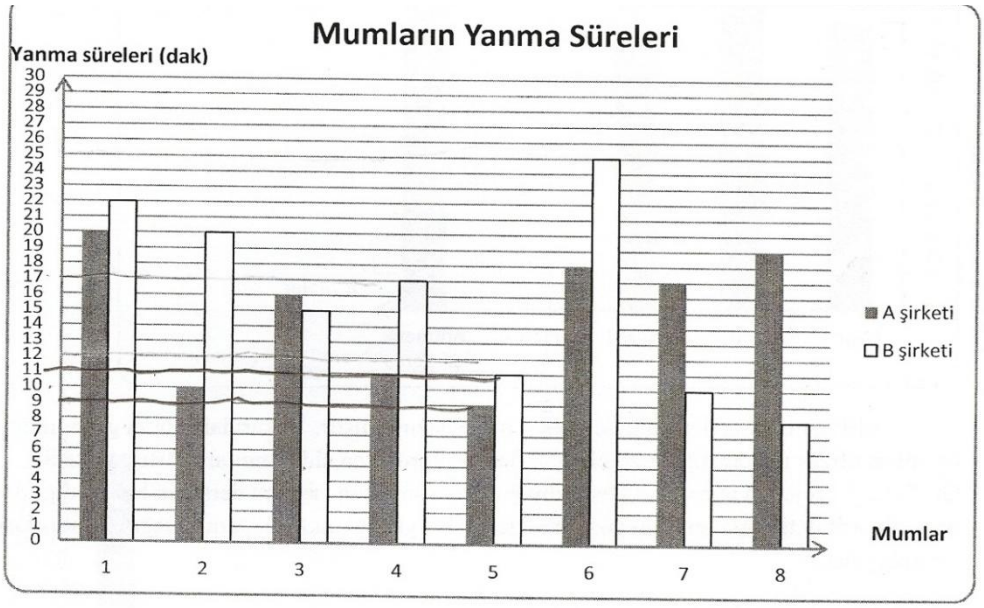
Figure 4.51 Answer of Participant 187 to item 3-b

As can be seen in the response of Participant 187 for item 3-b, a line on the given bar graph was drawn. Thus, the median of the data sets regarding Company A and Company

B was found from the unordered data sets since the data sets were not ordered on the given bar graph. For example, in order to find the median of the data set regarding the burning durations of the candles of Company A; 11, which was the burning duration of the fourth candle and 9, which was the burning duration of the fifth candle was added, and the sum was divided by 2. Subsequently, 10 was given as the median of the data set. Furthermore, the median of the burning durations of the candles of Company B was found using the same method of solution. Thus, the participant made this error.

On the other hand, there were other versions of the *wrong interpretation of graph* error. In the first version, the students directly selected one of the middle values of the given bar graph as the median of the data sets regarding Company A and Company B. To illustrate, the response of Participant 42 to item 3-b is presented below as an example of this version.

Participant 42:



a) Her iki şirkete ait mumların yanma sürelerinin aritmetik ortalamalarını bulunuz.
Aritmetik Ortalama

A Şirketi. 20
B Şirketi. 22

b) Her iki şirkete ait mumların yanma sürelerinin ortancalarını (medyan) bulunuz.
Ortanca (medyan)

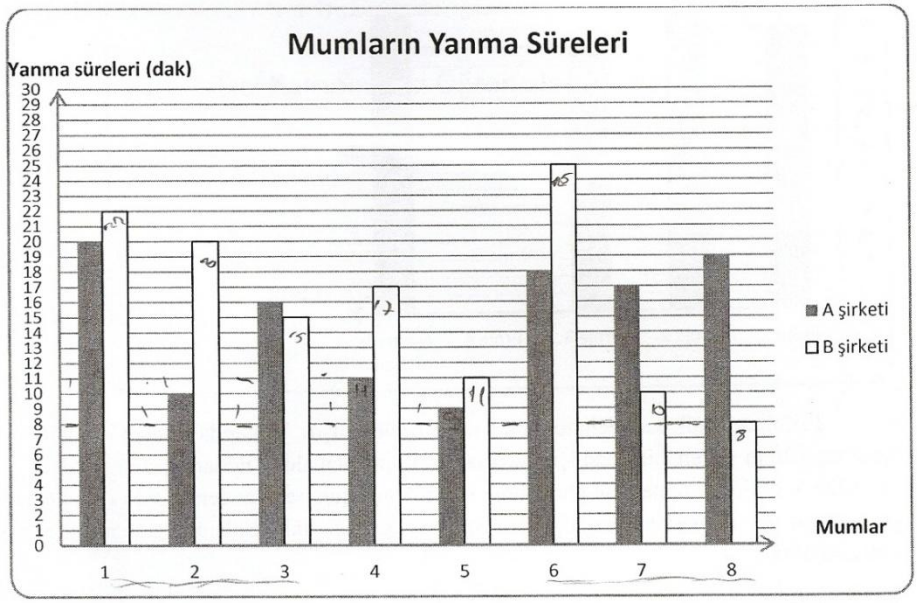
A Şirketi. 11
B Şirketi. 9

Figure 4.52 Answer of Participant 42 to item 3-b

According to the drawings of Participant 42 on the given bar graph, the fifth values of each data set was directly selected from the given bar graph and then 11 and 9 were found as the median of the burning durations of the candles of Company A and B, respectively.

In another version of this error, when a given data set had an even number of values, some of the students directly added the middle values of the data set to find its median. For instance, Participant 46 made this version of the *wrong interpretation of graph* error and his response is presented below.

Participant 46:



a) Her iki şirkete ait mumların yanma sürelerinin aritmetik ortalamalarını bulunuz.
Aritmetik Ortalama

A Şirketi..15.....
B Şirketi..16.....

b) Her iki şirkete ait mumların yanma sürelerinin ortancalarını (medyan) bulunuz.
Ortanca (medyan)

A Şirketi..20...
B Şirketi..28..

Figure 4.53 Answer of Participant 46 to item 3-b

According to the response of Participant 46 in item 3-b, 11, which was fourth value of the data set regarding Company A, and 9, which was fifth value of the same data set, was added and 20 was found as the median of the burning durations of the candles of Company A. Moreover, the median of the burning durations of the candles of Company B was found in the same way and 28 was given as the answer.

On the other hand, there was another version of the *wrong interpretation of graph* error. In this version, two middle values of each data set was directly selected according to the order of the given graph. Thus, two middle values were selected from the unordered even data sets since the values were given as unordered data sets on a bar graph.

Participant 136:

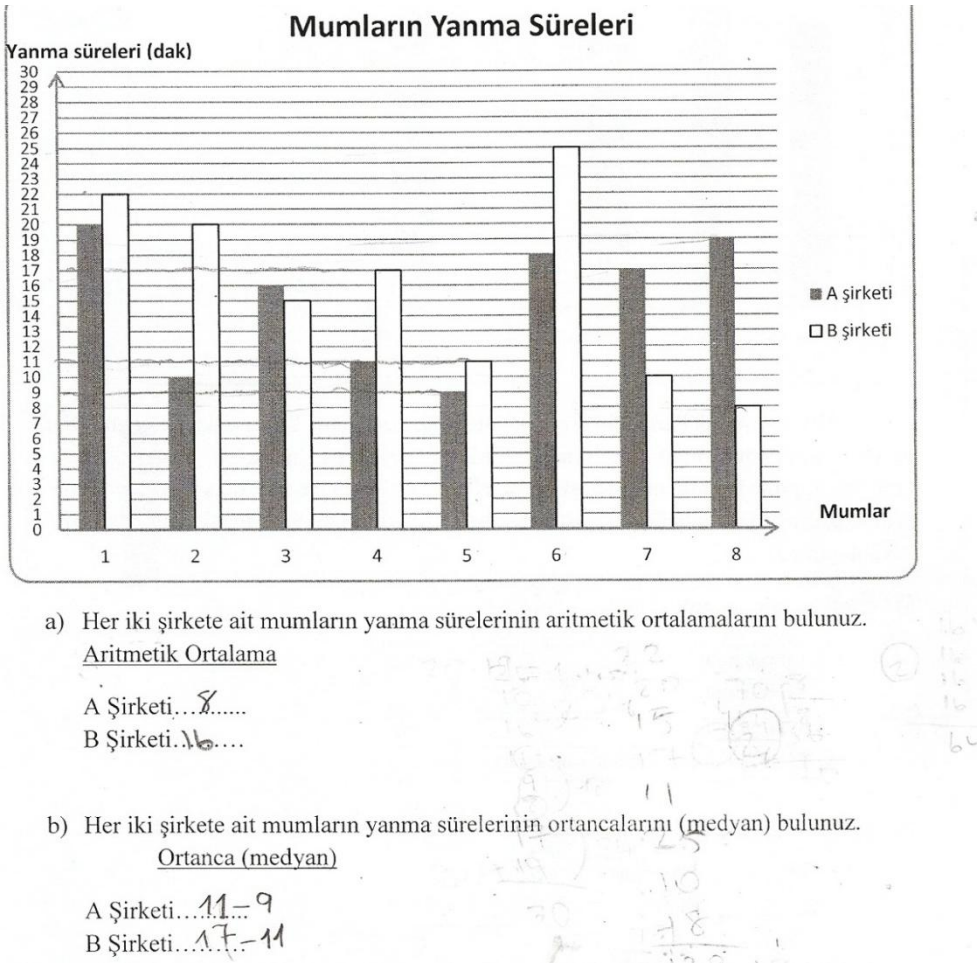


Figure 4.54 Answer of Participant 136 to item 3-b

As can be seen from the drawings on the given graph and in the response of Participant 136 to item 3-b, the fourth and fifth values from the data set, which represented eight candles' burning durations of Company A, was directly found based on the given graph and 11-9 was given as the median of this data set. Furthermore, 17-11 was given as the median of data set regarding the candles' burning durations of Company B by applying the same procedure.

In the following part of this chapter, the sub-groups of the other errors regarding the median is given with definitions and examples.

4.2.2.3 Other Errors

Students' responses in the SAQ showed that their errors related to the median under the sub-group of other errors can be categorized as: *Forming incomplete data set, incorrect reading of values from graph*, and *incorrect largest and/or smallest data values.*

Forming Incomplete Data Set: As defined in the errors regarding the mean concept of this study, the students who made this error did not use all the values of the given data set in order to find the median of the data set or did not construct an appropriate data set for the given sample size. This error was made in items 3-b and 5. As can be observed in Table 4.3, 5 students (5.3%) among 94 students and 62 students (81.6%) among 76 students made the *forming incomplete data set* error in item 3-b and 5, respectively. More specifically, in item 3-b, the students did not use one or more values from the data set given on a bar graph while trying to find the median of the data set. Thus, finding the correct answer was impossible.

The response of Participant 18 to item 3-b could be given as an example for the *forming incomplete data set* error:

Participant 18:

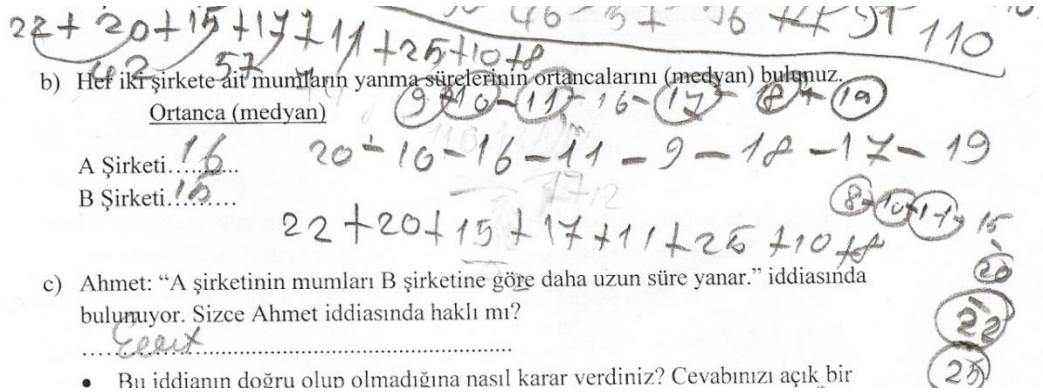
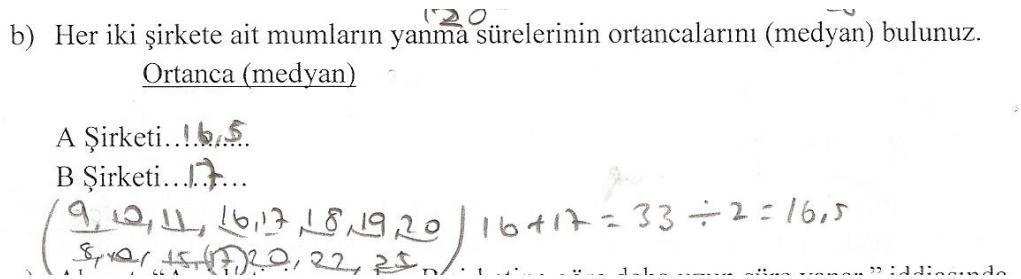


Figure 4.55 Answer of Participant 18 to item 3-b

As can be seen in the response provided by Participant 18, when the median of the data set which included the burning durations of eight candles of Company A was sought for, one value from the data set was missing. Then, median was found as 16 for the remaining seven values of the data set. Thus, the median of the data set was not found

correctly due to an incomplete data set. Furthermore, the same error was made when the participant tried to find the median of the data set of Company B. The response of Participant 134 to item 3-b is another example of the *forming incomplete data set* error:

Participant 134:


b) Her iki şirkete ait mumların yanma sürelerinin ortancalarını (medyan) bulunuz.

Ortanca (medyan)

A Şirketi…16,5…

B Şirketi…17…

(9, 10, 11, 16, 17, 18, 19, 20) 16+17 = 33 ÷ 2 = 16,5

Figure 4.56 Answer of Participant 134 to item 3-b

As can be seen in the response of Participant 134, seven values were used in order to find the median of the burning durations of the candles of Company B. Thus, the *forming incomplete data set* error was made since although there were eight values in the data set, only seven of them were used.

On the other hand, in item 5, some of the students who made this error constructed a possible data set which had one or more missing values according to the required sample size. Thus, these constructed data sets were also incomplete. When a student constructed a possible data set with one or more missing value, evaluating the correctness of the median of the data set was not appropriate since the constructed data set was incomplete.

For instance, the data set constructed by Participant 6 in item 5 had one missing value as presented below.

Participant 6:

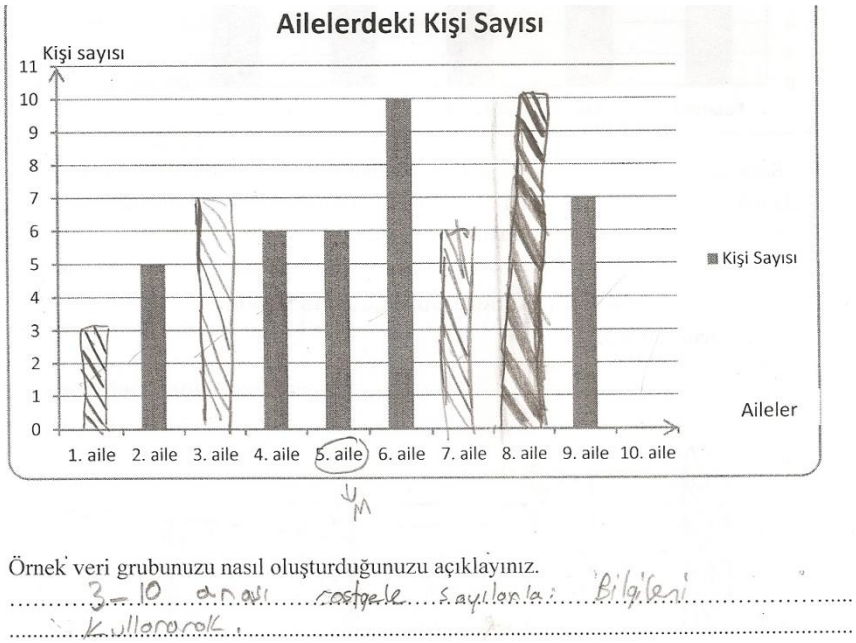
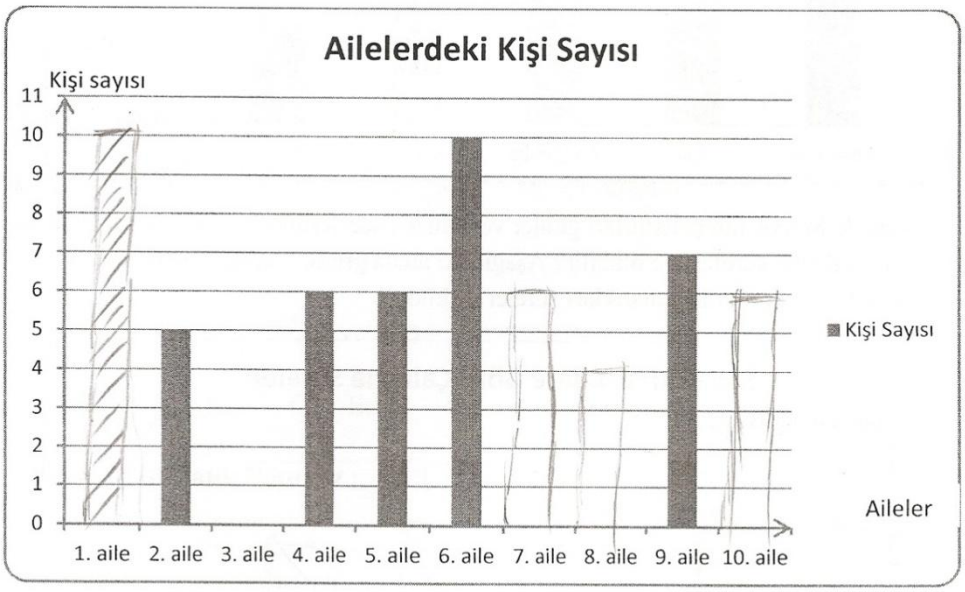


Figure 4.57 Answer of Participant 6 to item 5

As can be seen from the bar graph drawn, there were 9 values instead of 10. Thus, the median of the constructed data set could not be evaluated since the data set was incomplete according to the given sample size in the problem. Furthermore, the constructed data set of Participant 23 for item 5 was incomplete as presented below.

Participant 23:



- Örnek veri grubunuzu nasıl oluşturduğunuzu açıklayınız.
  Tepe değer ve moda göre belirledim.

Figure 4.58 Answer of Participant 23 to item 5

According to drawn bar graph by Participant 23, the constructed data set was also incomplete since a possible value was not selected for the number of individuals in the third family. Thus, the appropriateness of the given median of the constructed data set was not evaluated since the *forming incomplete data set* error was made.

Another sub-group within other errors regarding the median was *incorrect reading of values from graph*. This error is defined in detail under the following heading.

Incorrect Reading of Values from Graph: The *incorrect reading of values from graph* error was also made in the items regarding the mean and it is defined as follows: When values were given on a bar graph, the students who made this error type read the values on the graph incorrectly. Thus, the median of the data set was found incorrectly due to one or more incorrect readings of values. This error was made in item 3-b since the item asked for the median of two data sets that were given on a bar graph. As seen in Table

4.3, this error was made by 2 students (2.1%) among 94 students. As an example, the response of Participant 193 to item 3-b is presented below:

Participant 193:

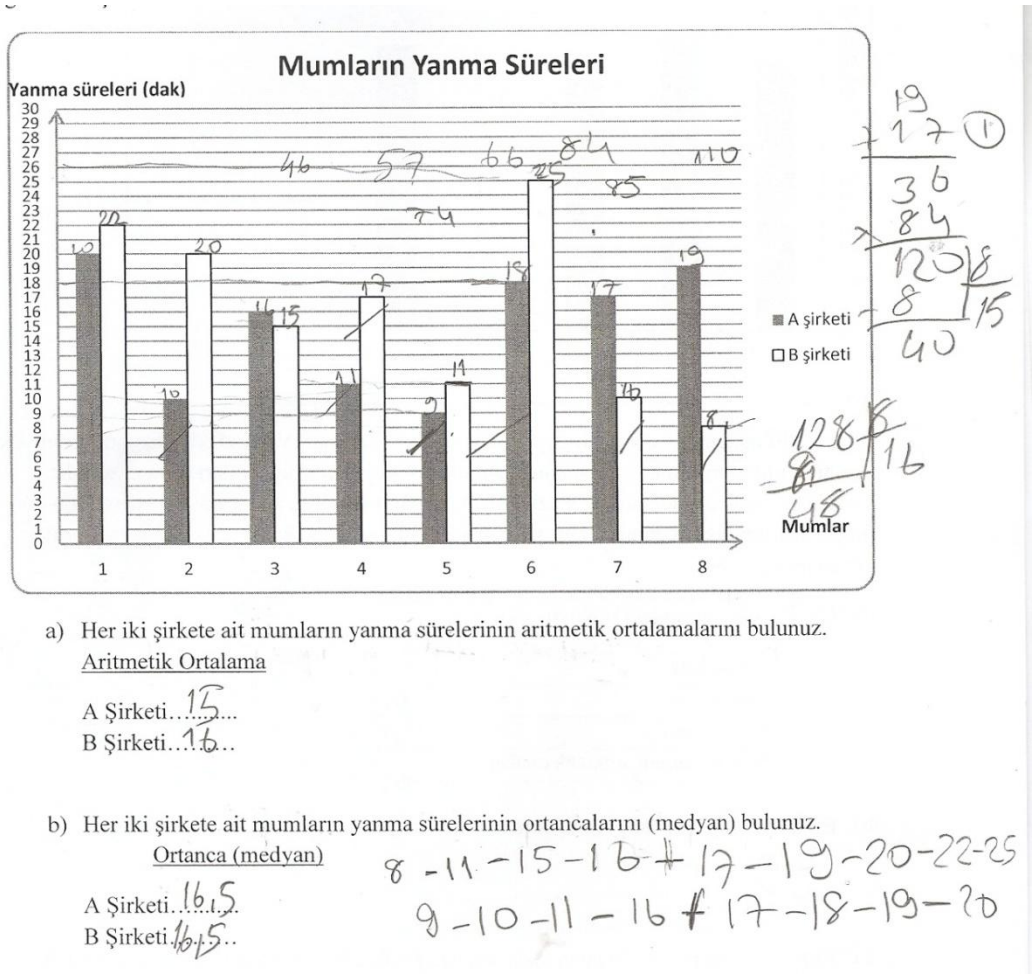
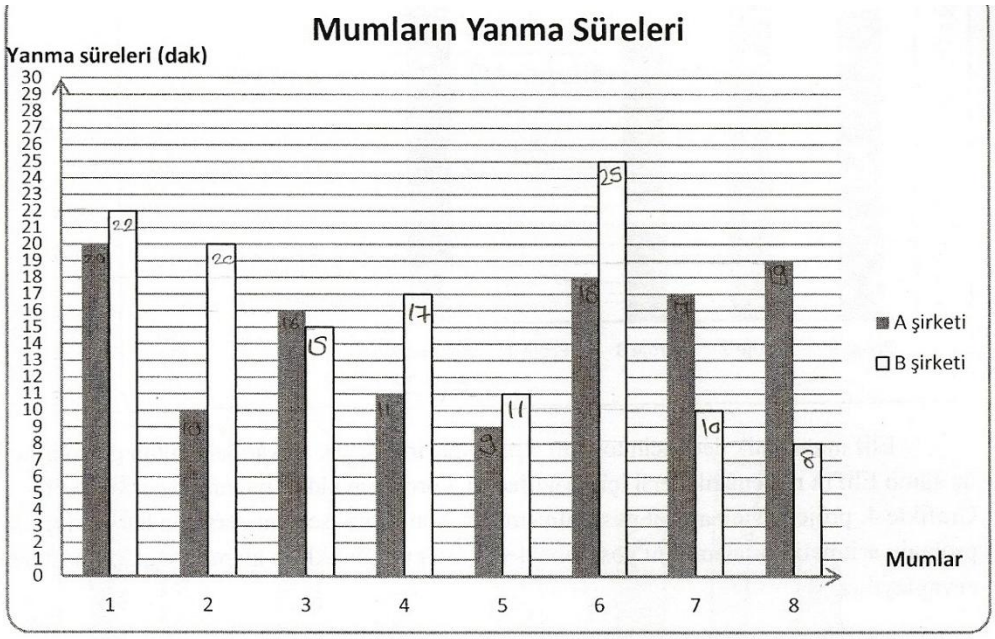


Figure 4.59 Answer of Participant 193 to item 3-b

According to the response of Participant 193 to item 3-b, the median of the data set regarding Company B was found incorrectly since the seventh value was read 16 instead of 10. Thus, the data set was changed and the median of the real data set could not be found.

Furthermore, Participant 52 also made this error in his response, which is presented below:

Participant 52:



a) Her iki şirkete ait mumların yanma sürelerinin aritmetik ortalamalarını bulunuz.
Aritmetik Ortalama

A Şirketi..15.....
B Şirketi..16...

b) Her iki şirkete ait mumların yanma sürelerinin ortancalarını (medyan) bulunuz.
Ortanca (medyan)

A Şirketi..16.....
B Şirketi..16...

9 - 10 - 11 - 16 - 16 - 17 - 19 - 20
8 - 10 - 11 - 15 - 17 - 20 - 22

Figure 4.60 Answer of Participant 52 to item 3-b

When the written values provided by Participant 52 to answer item 3-b was investigated, it was seen that the sixth value of the data set of the burning durations of the candles of Company A was read 16 instead of 18. Thus, the median of this data set was found incorrectly due to the error.

Incorrect Largest and/or Smallest Data Values: The students who made this error constructed possible data sets which were not appropriate for the given largest and smallest data values. In more detail, the construction of a possible data set was expected in the problem, but the smallest and largest values of the data set were included in the

problem context. When a student did not construct an appropriate data set for the given smallest and largest values, the error of *incorrect largest and/or smallest data values* was made.

As seen in Table 4.3, 14 students (18.4%) among 76 students made this error in item 5. To illustrate, Participant 30, whose response is presented below, made this error in item 5.

Participant 30:

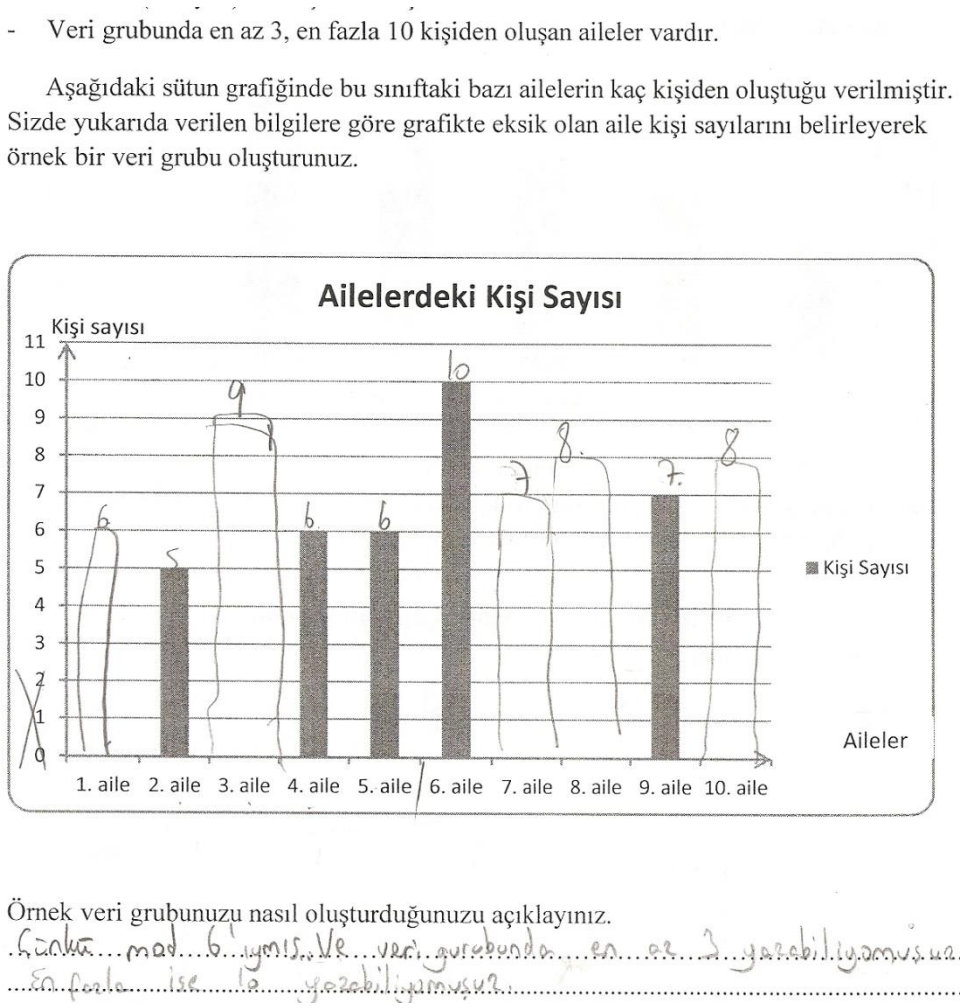


Figure 4.61 Answer of Participant 30 to item 5

As can be observed in the response of Participant 30, the smallest data value was 5 and the largest data value was 10 in the constructed data set. However, the item required the

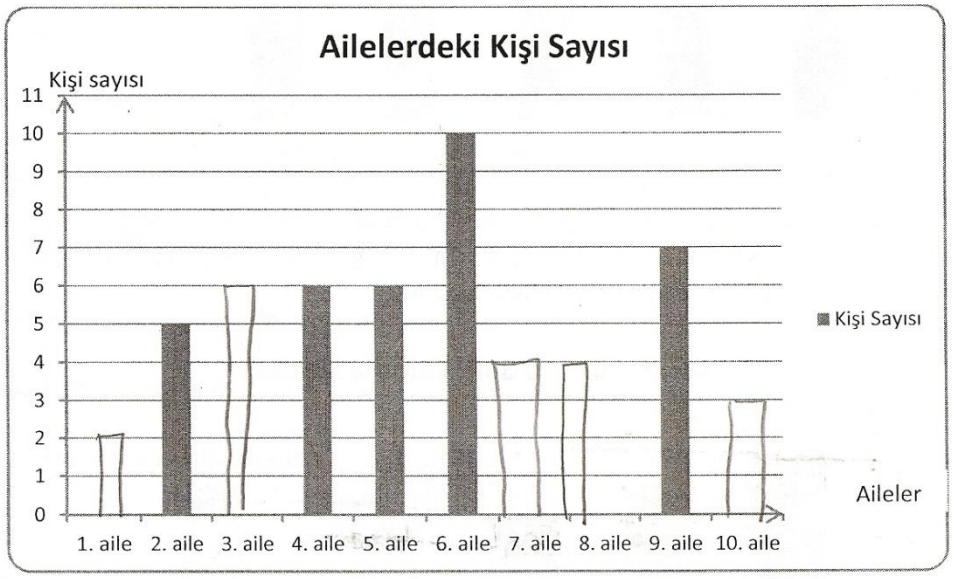
smallest data value to be 3 and the largest data value to be 10. Thus, the *incorrect largest and/or smallest data values* error was made. Moreover, the median of this constructed data set was 7 instead of 6.

The response provided by Participant 62 to item 5, which is presented below, was another example regarding this error.

Participant 62:

- Tepe değer (mod) = 6 kişiden oluşan aile
- Ortanca (medyan) = 6 kişiden oluşan aile
- Veri grubunda en az 3, en fazla 10 kişiden oluşan aileler vardır.

Aşağıdaki sütun grafiğinde bu sınıftaki bazı ailelerin kaç kişiden oluştuğu verilmiştir. Sizde yukarıda verilen bilgilere göre grafikte eksik olan aile kişi sayılarını belirleyerek örnek bir veri grubu oluşturunuz.



- Örnek veri grubunuzu nasıl oluşturduğunuzu açıklayınız.

Oluşturamadım bunu zor oldum.

Figure 4.62 Answer of Participant 62 to item 5

Based on the constructed data set it can be observed that Participant 62 made this error since the smallest data value was 2 and the largest value was 10. However, the item

required the smallest value to be 3 and the largest value to be 10 in the constructed data set. Moreover, the median of this constructed data set was 5.5, not 6.

In the following part, the students' errors regarding the mode concept are categorized and defined with examples.

4.2.3 Errors Regarding the Mode

In order to identify errors related to the mode, students' wrong answers with unacceptable solution strategy in item 4-a and the responses of students who constructed inappropriate data sets in item 5 were examined. Based on the analysis results, there was one sub-group in the category of errors based on formal knowledge, namely *inappropriate usage of averaging algorithm*. In addition, two types of sub-groups within the other errors category was defined, namely *forming incomplete data set* and *incorrect largest and/or smallest data values.* Frequencies and percentages of students who made errors regarding the mode in their responses to items 4-a and 5 are given in Table 4.4 below.

Table 4.4 Frequencies (and percentages) of the students' errors regarding the mode

ERRORS ITEMS	**Errors Based on Formal Knowledge**	**Other Errors**		**TOTAL**
	Inappropriate Usage of Averaging Algorithm	Forming Incomplete Data Set	Incorrect Largest and/or Smallest Data Values	
Item 4-a	21 (100%)	0 (0%)	0 (0%)	21
Item 5	0 (0%)	62 (81.6%)	14 (18.4%)	76
TOTAL	21	62	14	

As can be seen in Table 4.4, the types of errors made in the responses to items 4-a and item 5 regarding the mode concept do not show similarities in terms of the sub-groups. The error of *inappropriate usage of averaging algorithm* was the most common error in the responses given to item 4-a. To be more specific, 21 students (100%) among 21 students made this error. On the other hand, the error *forming incomplete data set* was the most frequently made error while the students tried to solve item 5. Particularly, this

error was made by 62 students (81.6%) among 76 students in item 5. Definitions of each error type with sample students responses are given under the category headings given in Table 4.4.

4.2.3.1 Errors Based on Formal Knowledge

Inappropriate Usage of Averaging Algorithm: Although there was no question asking for the mean of a data set or was not given the mean of a data set, the students used the computational algorithm of the mean. This error was made in item 4-a by 21 students (100%). The students who made this error found the mean of the given values of a data set to find the missing value in the data set in item 4-a. Participant 1, whose response is presented below, made this error in the response he gave to item 4-a.

Participant 1:

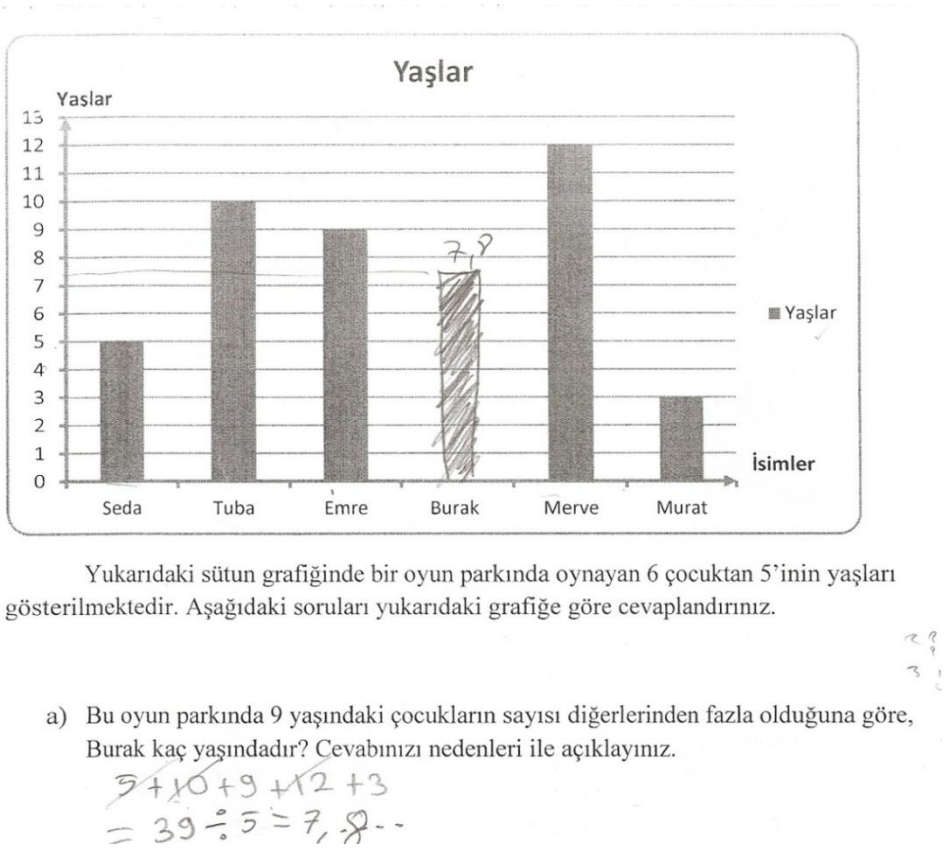


Figure 4.63 Answer of Participant 1 to item 4-a

As can be seen in the response provided by Participant 1, even though a missing value was asked for in item 4-a, the mean of the given values on a bar graph was found in order to find the missing value. Thus, the *inappropriate usage of averaging algorithm* error was made by Participant 1. The response of Participant 8 is another example of this error and is as follows:

Participant 8:

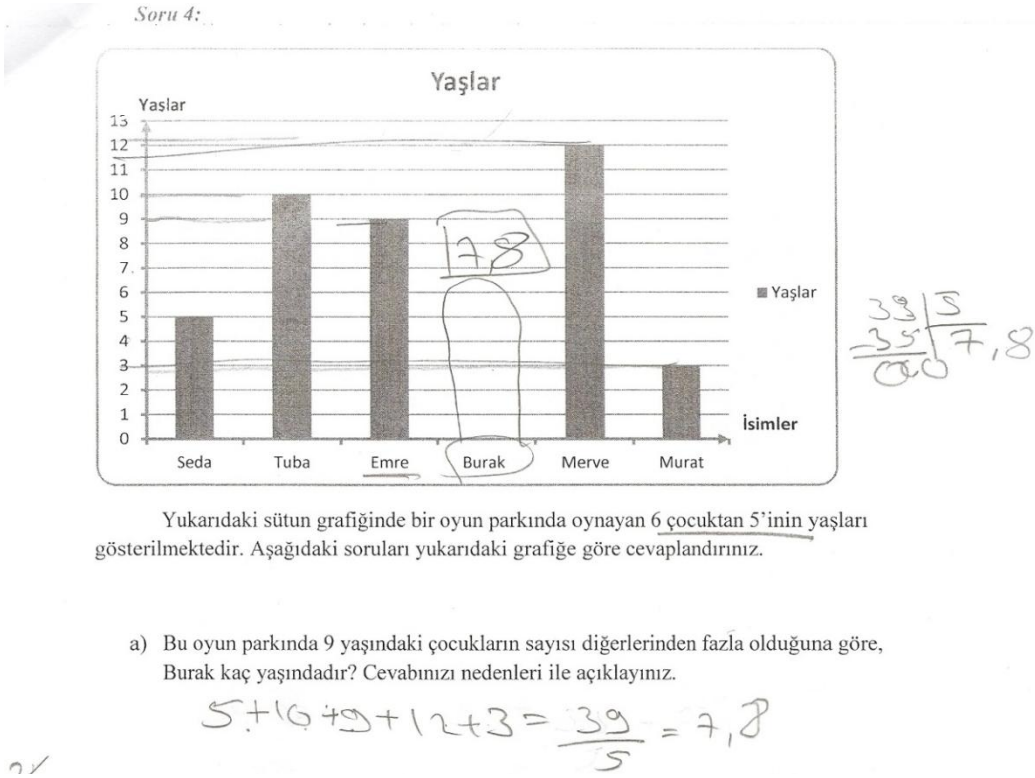


Figure 4.64 Answer of Participant 8 to item 4-a

Similar to the response of Participant 1, Participant 8 found the mean of the given values to find a missing value of the data set. The usage of the averaging algorithm was not appropriate in finding the missing value in the data set, thus this error was made.

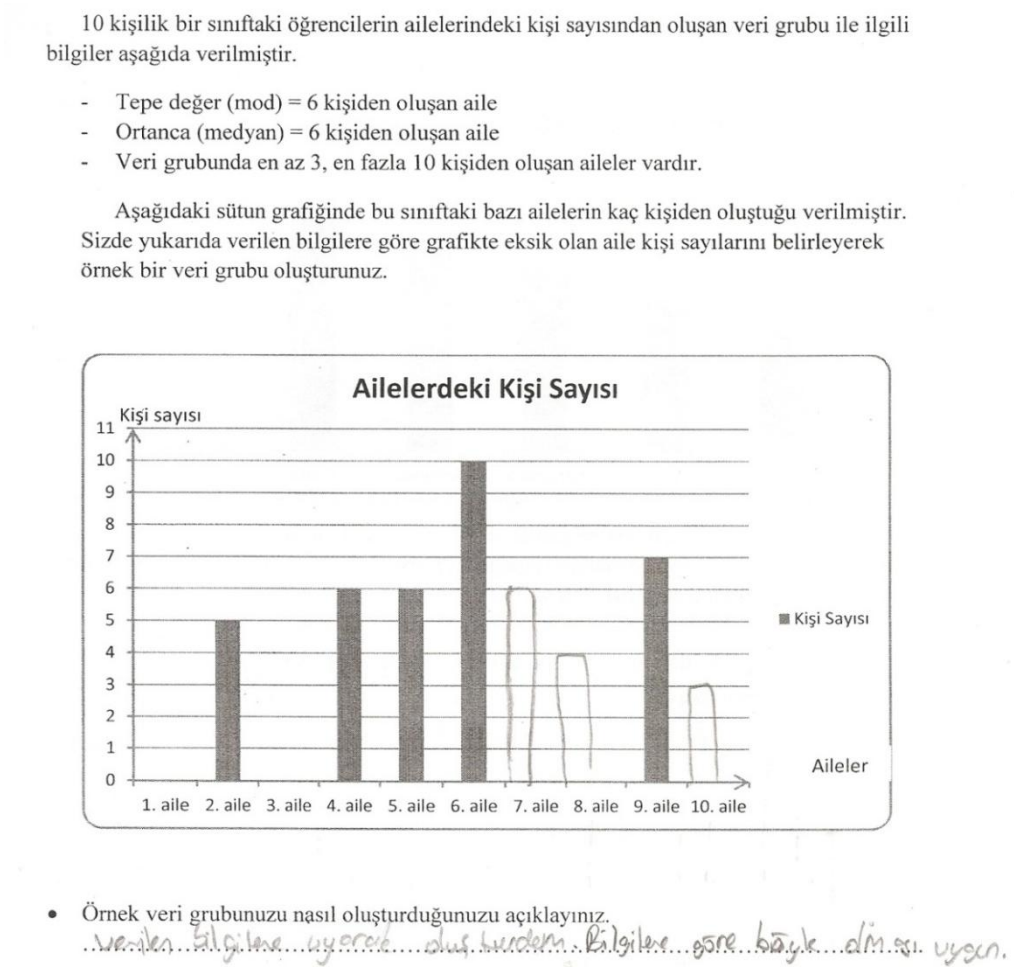
In addition to errors based on formal knowledge, other errors regarding the mode are explained with definitions and examples under the following heading.

4.2.3.2 Other Errors

Errors of *forming incomplete data set* and *incorrect largest and/or smallest data values* were also identified within other errors regarding the mode concept, which were made by the participants of the study.

Forming Incomplete Data Set: This error was also mentioned within errors regarding the concepts of mean and median. The error of *forming incomplete data set* was made when the students did not use all the values of the given data set in order to find its average or they did not construct an appropriate data set for the given sample size. This error was made in item 5 by 62 students (81.6%) among 76 students. The constructed data sets of the students were not appropriate in terms of the given sample size. Furthermore, evaluating the correctness of the mode of the data set was not appropriate since the data sets were incomplete. The response of Participant 13 to item 5 was an example of this error as presented below.

Participant 13:



10 kişilik bir sınıftaki öğrencilerin ailelerindeki kişi sayısından oluşan veri grubu ile ilgili bilgiler aşağıda verilmiştir.

- Tepe değer (mod) = 6 kişiden oluşan aile
- Ortanca (medyan) = 6 kişiden oluşan aile
- Veri grubunda en az 3, en fazla 10 kişiden oluşan aileler vardır.

Aşağıdaki sütun grafiğinde bu sınıftaki bazı ailelerin kaç kişiden oluştuğu verilmiştir. Sizde yukarıda verilen bilgilere göre grafikte eksik olan aile kişi sayılarını belirleyerek örnek bir veri grubu oluşturunuz.

- Örnek veri grubunuzu nasıl oluşturduğunuzu açıklayınız.

verilen bilgilere uyarak oluşturdum. Bilgilere göre böyle olması uygun.

Figure 4.65 Answer of Participant 13 to item 5

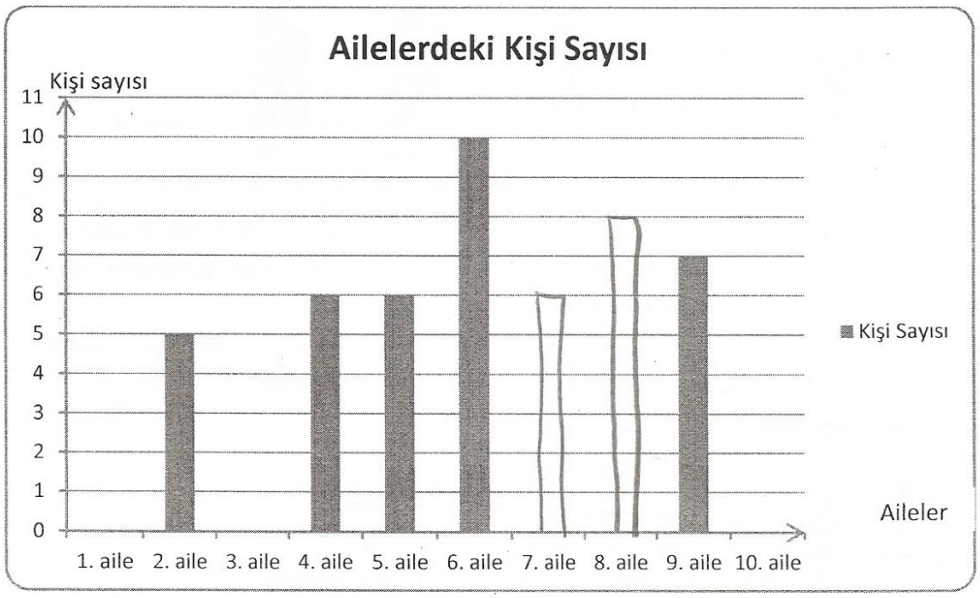
According to the data set constructed on the bar graph by Participant 13, the number of individuals in the first and third families were missing. Thus, the data set was incomplete. This error was also made by Participant 67 in item 5 and it is given below.

Participant 67:

10 kişilik bir sınıftaki öğrencilerin ailelerindeki kişi sayısından oluşan veri grubu ile ilgili bilgiler aşağıda verilmiştir.

- Tepe değer (mod) = 6 kişiden oluşan aile
- Ortanca (medyan) = 6 kişiden oluşan aile
- Veri grubunda en az 3, en fazla 10 kişiden oluşan aileler vardır.

Aşağıdaki sütun grafiğinde bu sınıftaki bazı ailelerin kaç kişiden oluştuğu verilmiştir. Sizde yukarıda verilen bilgilere göre grafikte eksik olan aile kişi sayılarını belirleyerek örnek bir veri grubu oluşturunuz.



- Örnek veri grubunuzu nasıl oluşturduğunuzu açıklayınız.

7. Aileyi 6 kişi buldum Çünkü tepe değer olması için gerekliydi. 8. Aile ise 6'dan küçük olamazdı.

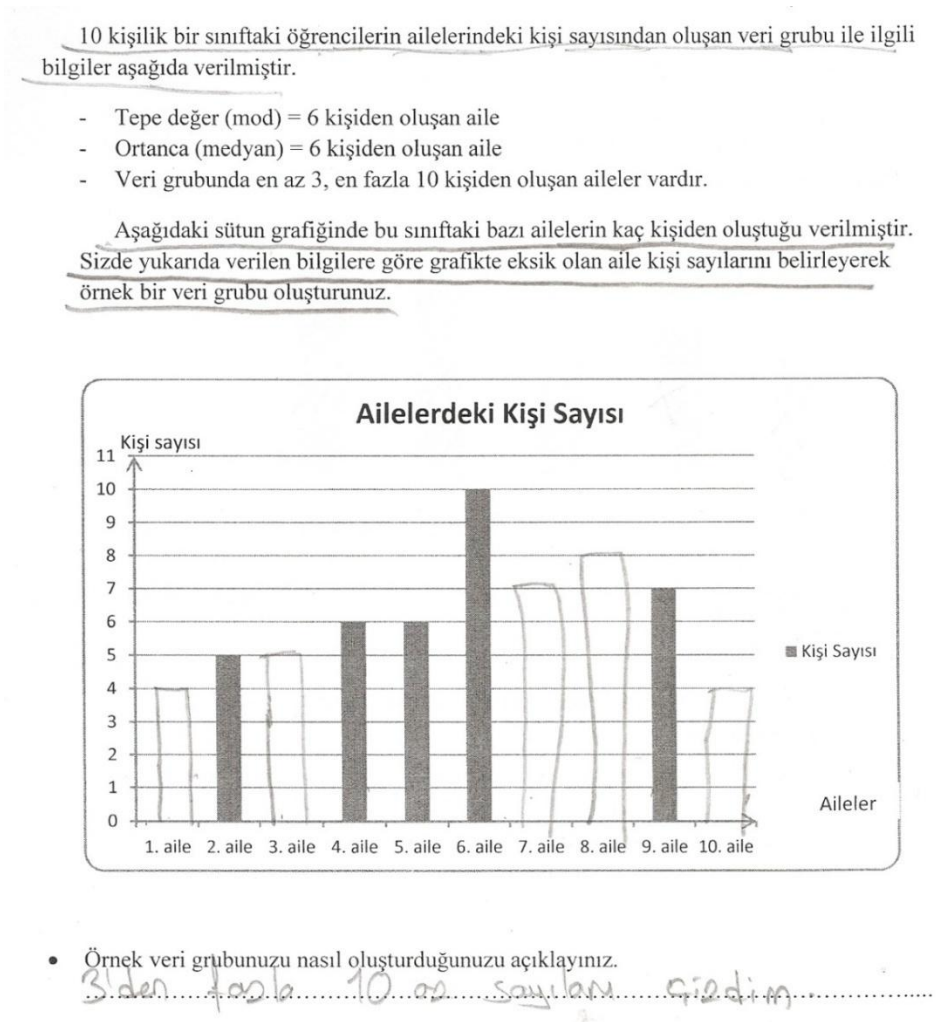
Figure 4.66 Answer of Participant 67 to item 5

As can be seen in the answer of Participant 67, the number of individuals in the first, third and tenth families were not decided. Thus, this error was made since the constructed data set was incomplete.

Another sub-error of this category was the *incorrect largest and/or smallest data values*. Details regarding this error are presented under the following heading.

Incorrect Largest and/or Smallest Data Values: This error was the same with errors regarding the median concept of the study. In fact, since this error was independent of the concepts of median and mode, this error was made in item 5 in the same way in terms of the mode and the median. When the students made this error, their constructed possible data sets were not appropriate in terms of the largest and/or the smallest data values which were given in the problem context. The response of Participant 214 was an example of this error as given below.

Participant 214:


10 kişilik bir sınıftaki öğrencilerin ailelerindeki kişi sayısından oluşan veri grubu ile ilgili bilgiler aşağıda verilmiştir.

- Tepe değer (mod) = 6 kişiden oluşan aile
- Ortanca (medyan) = 6 kişiden oluşan aile
- Veri grubunda en az 3, en fazla 10 kişiden oluşan aileler vardır.

Aşağıdaki sütun grafiğinde bu sınıftaki bazı ailelerin kaç kişiden oluştuğu verilmiştir. Sizde yukarıda verilen bilgilere göre grafikte eksik olan aile kişi sayılarını belirleyerek örnek bir veri grubu oluşturunuz.

- Örnek veri grubunuzu nasıl oluşturduğunuzu açıklayınız.

3'den fazla 10 ar sayıları çiedim

Figure 4.67 Answer of Participant 214 to item 5

As can be observed in the constructed data set of Participant 214, this error was made since the smallest data value was found to be 4 and the largest value was found to be 10. However, the item required 3 as the smallest and 10 as the largest data value of the data set. Moreover, the mode of the constructed data set was found to be 6, which is not correct.

Participant 79:

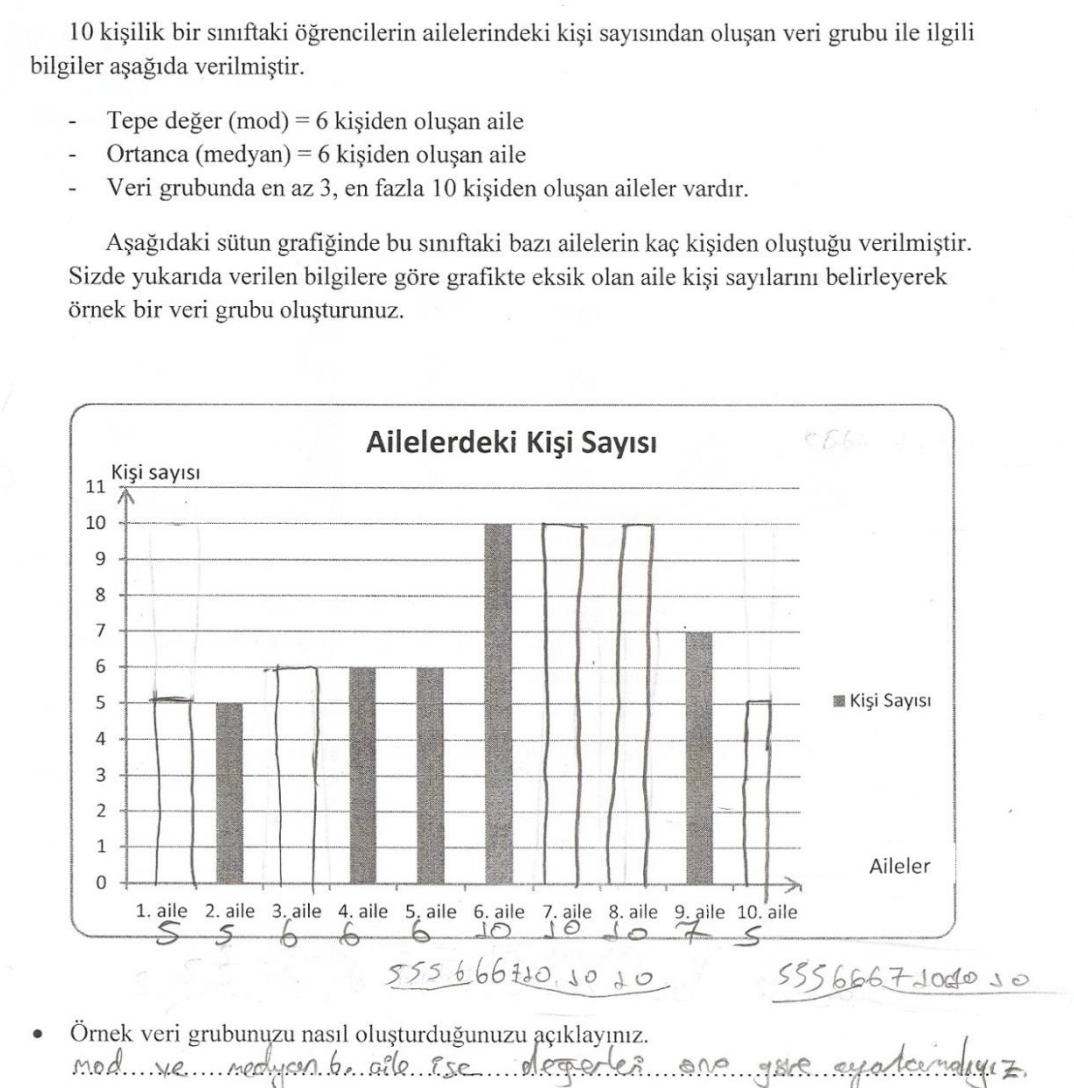


Figure 4.68 Answer of Participant 79 to item 5

As can be seen in the response of Participant 79, the smallest data value was found to be 5 and the largest data value was found to be 10 in the constructed data set. As the

smallest value in the data set 5 is not appropriate since the item required the smallest data value to be 3. Thus, the error of *incorrect largest and/or smallest data values* was made. In addition, the mode of the constructed data set was found to be 6, which is incorrect.

4.3 Misinterpretations Regarding the Concepts of Mean, Median and Mode

In order to identify their misinterpretations while providing answers to the questions regarding the concepts of mean, median and mode given in bar graph representations, the third section of this chapter presents findings regarding the students' misinterpretations, which were analyzed from the Statistics Achievement Questionnaire. The SAQ included three interpretation questions, which were item 2-d, item 3-c and item 4-c. In this section, the results of these interpretation items are presented. In the items, students were expected to interpret a situation or decide whether a statement was true or not and state the reasons underlying their decisions. Since one of the purposes was to investigate to what extent seventh grade students' interpret the concepts of mean, median and mode given in bar graph representations, the participants' answers were categorized into three groups: correct, incorrect and no interpretation. Subsequently, the incorrect interpretations were examined in detail. The frequencies of these errors encountered in the responses of the participants to items 2-d, 3-c and 4-c are presented in Table 4.5 as follows:

Table 4.5 Frequencies (and percentages) of the students' correct, incorrect and no interpretations in items 2-d, 3-c and 4-c

	Correct Interpretation	**Incorrect Interpretation**	**No Interpretation**	**TOTAL**
Item 2-d	98 (42.1%)	48 (20.6%)	87(37.3%)	233
Item 3-c	143 (61.4%)	30 (12.9%)	60 (25.7%)	233
Item 4-c	103 (44.2%)	80 (34.3%)	50 (21.5%)	233

According to Table 4.5, while 143 students (61.4%) made correct interpretation in item 3-c, 103 students (44.2%) made correct interpretations in item 4-c. In addition, 98 students made correct interpretations and 48 students made incorrect interpretations in

item 2-d. Details regarding misinterpretations of the students for items 2-d, 3-c and 4-c are given under the following headings.

4.3.1 Misinterpretations in Responses to Item 2-d

In item 2-d, the students' interpretation of the mean was investigated. When the mean of a data set was given, it was asked what could be said about values of the data set. More specifically, the item was worded as such: when we know the arithmetic average (mean) of Elif's project scores as 85, what could be said about Elif's project scores?

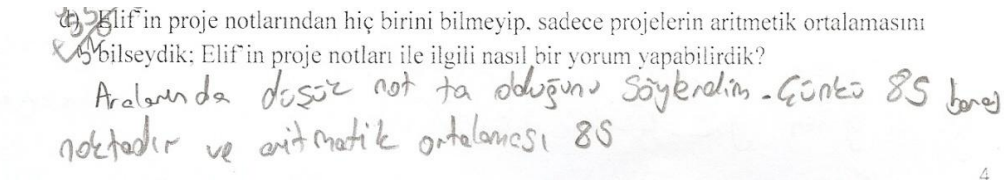
Misinterpretations were made by 48 students in item 2-d. These misinterpretations were categorized as given in Table 4.6 below:

Table 4.6 Frequencies of each misinterpretations in item 2-d

Misinterpretations	**Frequency**
All project scores are smaller than 85.	3
Based on incorrect use of averaging algorithm	15
Elif's project sores are at least 4	2
Elif's project scores are close to each other	2
Elif's project sores are at least 85	2
Elif's project scores are medium	3
Others	20
TOTAL	48

As can be seen in Table 4.6, 3 students among 33 students said that all the project scores of Elif were smaller than the mean of these project scores. The interpretation was not correct since the mean was the average of these project scores. Thus, the arithmetic average had to be in balance with the scores. All scores could not be smaller than the arithmetic average. To give an example, the response of Participant 233 is given below.

Participant 233:


d) Elif'in proje notlarından hiç birini bilmeyip, sadece projelerin aritmetik ortalamasını bilseydik; Elif'in proje notları ile ilgili nasıl bir yorum yapabilirdik?

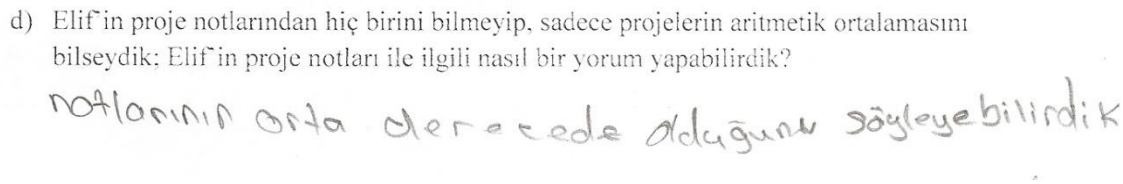
Aralarında düşük not ta olduğunu söyledim. Çünkü 85 barajı noktadır ve aritmetik ortalaması 85

4

Figure 4.69 Answer of Participant 233 to item 2-d

Moreover, the misinterpretation of another 3 students was that Elif's project scores were medium. For example, Participant 84 made this misinterpretation is that it could be said that Elif's scores are medium and her comment was as follows:

Participant 84:

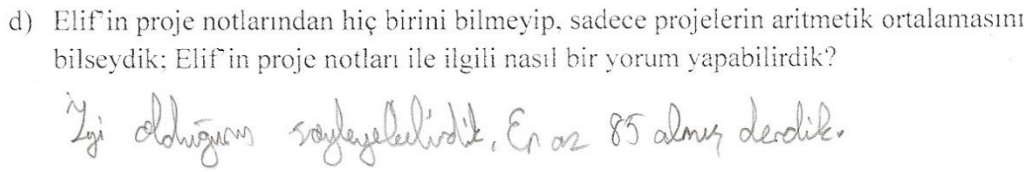

d) Elif'in proje notlarından hiç birini bilmeyip, sadece projelerin aritmetik ortalamasını bilseydik; Elif'in proje notları ile ilgili nasıl bir yorum yapabilirdik?

notlarının orta derecede olduğunu söyleyebilirdik

Figure 4.70 Answer of Participant 84 to item 2-d

Furthermore, 2 students said that when Elif's project scores were at least 4, the mean of these scores could be 85. However, the interpretation was incorrect since the mean could be 85 when the values smaller than 4 was balanced with values higher than the mean.

Participant 142:


d) Elif'in proje notlarından hiç birini bilmeyip, sadece projelerin aritmetik ortalamasını bilseydik; Elif'in proje notları ile ilgili nasıl bir yorum yapabilirdik?

İyi olduğunu söyleyebilirdik, En az 85 almış dedik.

Figure 4.71 Answer of Participant 142 to item 2-d

On the other hand, 15 students made misinterpretations based on the incorrect use of the averaging algorithm. In addition, 2 students said that Elif's project sores were higher

than the mean and the other 2 students stated that Elif's project scores were close to each other.

4.3.2 Misinterpretations in the Responses to Item 3-c

In item 3-c, the students' conceptual knowledge regarding the mean and the median was investigated. In the question, the burning durations of eight candles of two companies were given on a bar graph accompanied with a statement, which was followed with a question asking whether the statement was true or not. The statement was as follows: "Ahmet claims that the candles of Company A have longer burning durations than those of Company B. Do you agree with Ahmet? How did you decide whether Ahmet is right or not? Explain your reasoning." In order to decide the correctness of the statement, the burning durations of the candles had to be compared.

Some of the students made misinterpretations while evaluating the correctness of the statement. The frequencies of five different misinterpretations are presented in Table 4.7 below.

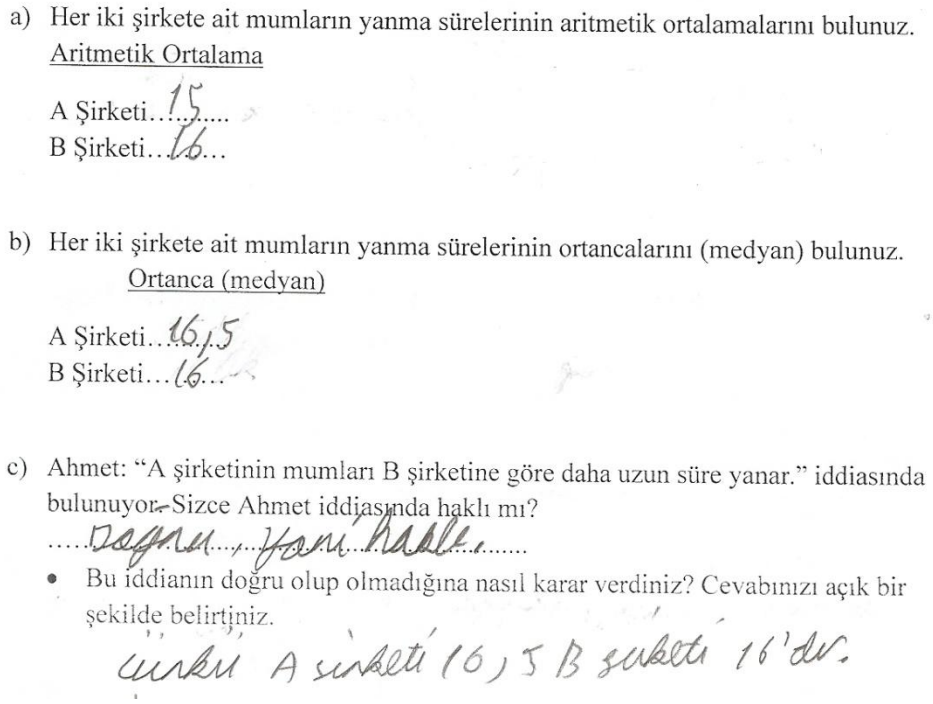
Table 4.7 Frequencies (and percentages) of each misinterpretations in item 3-c

Misinterpretations	Frequency
Since the median of Company A was higher than Company B	7
Since the candle which had the longest burning duration in Company B	9
Since there was an equal number of candles	2
Since the burning durations of the candles in Company B were higher than those of Company A	6
Since the sum of the mean and the median of Company B was higher than that of Company A	2
Others	4
TOTAL	30

According to Table 4.7, 30 students among 233 students made misinterpretations in item 3-c. Among these 30 students, 7 students used the median of the given burning durations of eight candles of each company to evaluate the correctness of the given statement. This

interpretation was incorrect since the data sets did not include any outliers, and the median was not the appropriate average type to compare the two data sets. To illustrate, Participant 32 made the misinterpretation presented below.

Participant 32:



a) Her iki şirkete ait mumların yanma sürelerinin aritmetik ortalamalarını bulunuz.
Aritmetik Ortalama

A Şirketi…15…
B Şirketi…16…

b) Her iki şirkete ait mumların yanma sürelerinin ortancalarını (medyan) bulunuz.
Ortanca (medyan)

A Şirketi…16,5
B Şirketi…16…

c) Ahmet: "A şirketinin mumları B şirketine göre daha uzun süre yanar." iddiasında bulunuyor. Sizce Ahmet iddiasında haklı mı?
…Doğru…, Yani haklı…

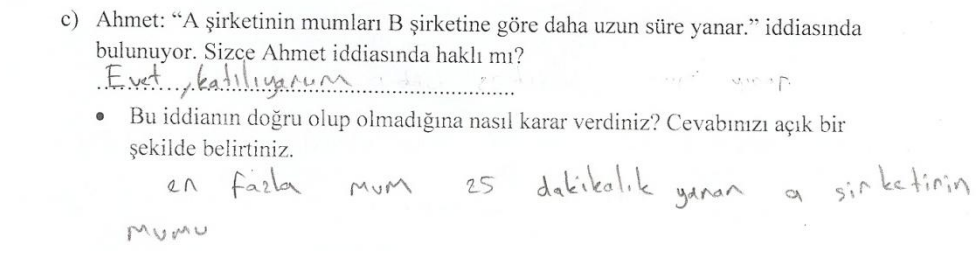
- Bu iddianın doğru olup olmadığına nasıl karar verdiniz? Cevabınızı açık bir şekilde belirtiniz.

Çünkü A şirketi 16,5 B şirketi 16'dır.

Figure 4.72 Answer of Participant 32 to item 3-c

Moreover, 9 students made misinterpretations while trying to decide which company had more burning durations since they based their decision only on a candle which had the longest burning durations among the candles of these two data sets. Thus, they said that " the burning durations of the candles of Company B were longer than those of Company A since the candle which had the longest burning duration was a candle of Company B". To give an example for this interpretation, the response of Participant 9 is given below.

Participant 9:


c) Ahmet: "A şirketinin mumları B şirketine göre daha uzun süre yanar." iddiasında bulunuyor. Sizce Ahmet iddiasında haklı mı?

Evet, katılıyorum

- Bu iddianın doğru olup olmadığına nasıl karar verdiniz? Cevabınızı açık bir şekilde belirtiniz.

en fazla mum 25 dakikalık yanan a şirketinin mumu

Figure 4.73 Answer of Participant 9 to item 3-c

On the other hand, 6 students decided their general overview regarding each data set in order to compare them. Furthermore, 2 students added the mean and the median of each data set to arrive at a conclusion regarding which Company's candles had longer burning durations.

4.3.3 Misinterpretations in the Responses to Item 4-c

In item 4-c, when a value was subtracted from a data set, how the students interpreted the average of the remaining values of the data set was examined.

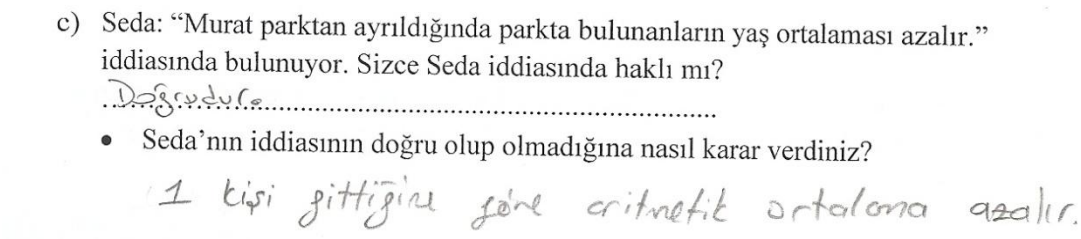
Table 4.8 Frequencies (and percentages) of each misinterpretations in item 4-c

Misinterpretations	Frequency
The average was decreased when a person was left from the group	57
The average was decreased since the sum of the age of the group was decreased.	11
The average was decreased since number of value in the data set was decreased.	2
The average did not change since the age of the person was small	2
Whether or not the average decreased or not changed based on the found incorrect average.	8
TOTAL	80

As presented in Table 4.8, there were five different misinterpretations of the students. The most common misinterpretation was that "the average was decreased when a person

left the group". This misinterpretation was made by 57 students among 80 students. The answer of Participant 1 is presented as an example below.

Participant 1:


c) Seda: "Murat parktan ayrıldığında parkta bulunanların yaş ortalaması azalır." iddiasında bulunuyor. Sizce Seda iddiasında haklı mı?

Doğrudur.

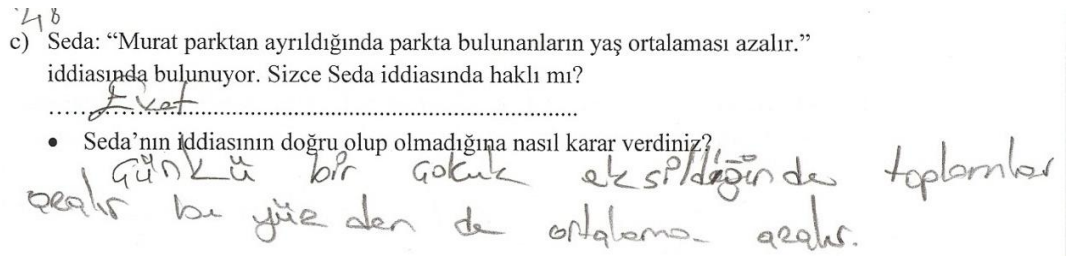
- Seda'nın iddiasının doğru olup olmadığına nasıl karar verdiniz?

1 kişi gittiğine göre aritmetik ortalama azalır.

Figure 4.74 Answer of Participant 1 to item 4-c

In addition, the second common misinterpretation in item 4-c was that "the average decreased since the sum of the age of the group decreased". 11 students among 80 students made this interpretation. The students were not taking into consideration the effect of the number of individuals on average. To illustrate, Participant 82 made this interpretation and it is given below.

Participant 82:


c) Seda: "Murat parktan ayrıldığında parkta bulunanların yaş ortalaması azalır." iddiasında bulunuyor. Sizce Seda iddiasında haklı mı?

Evet

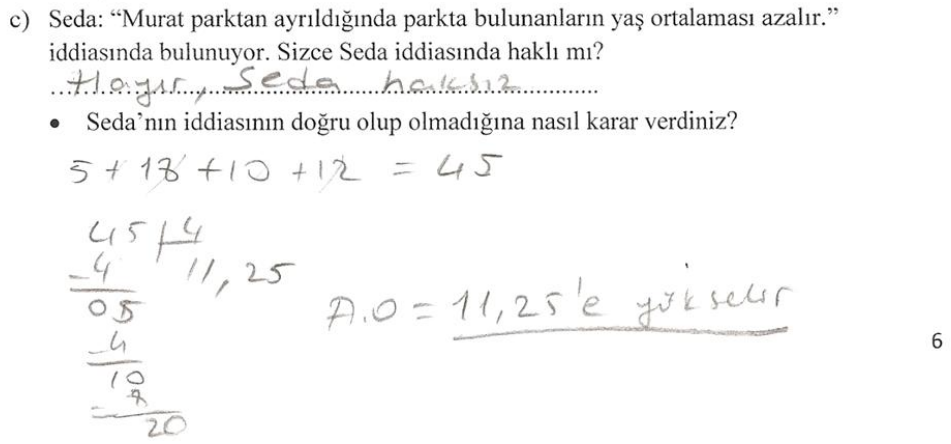
- Seda'nın iddiasının doğru olup olmadığına nasıl karar verdiniz?

Çünkü bir çocuk eksildiğinde toplamlar azalır bu yüzden de ortalama azalır.

Figure 4.75 Answer of Participant 82 to item 4-c

On the other hand, 2 students said that the average of the children's age decreased since the number of children in the park decreased. Moreover, 2 students said that the average did not change since the age of the child was small. In addition, 8 students made wrong interpretations since they found the average of a given data set due to the incorrect usage of averaging algorithm error. The response of Participant 11 is presented below.

Participant 11:


c) Seda: "Murat parktan ayrıldığında parkta bulunanların yaş ortalaması azalır." iddiasında bulunuyor. Sizce Seda iddiasında haklı mı?

Hayır, Seda haksız

- Seda'nın iddiasının doğru olup olmadığına nasıl karar verdiniz?

5 + 18 + 10 + 12 = 45

45 | 4
-4 11,25
05
-4
10
-8
20

A.O = 11,25'e yükselir

6

Figure 4.76 Answer of Participant 11 to item 4-c

As can be observed in the response of Participant 11, the average was found incorrectly due to the incorrect usage of averaging algorithm error. More specifically, the sum was divided by 4 although there were 5 children in the park. Thus, based on the incorrect average, a misinterpretation was made.

4.4 Summary of Findings

The first purpose of this study was to investigate middle school seventh grade students' solution strategies while solving questions regarding the mean, the median and the mode given in bar graph representations. According to the findings of the study, the seventh grade students used three different solution strategies to solve the questions regarding each concept. More specifically, in order to solve the questions regarding the mean, while most of the students who answered correctly used the using average formula solution strategy, the least used solution strategy was the guess and check. Additionally, more common solution strategy used to solve the question regarding the median was the using numerical procedures solution strategy and the least common solution strategy was the depending on graph representations. Lastly, the students used the depending on graph representations solution strategy most frequently to solve the questions regarding the mode.

The second purpose of the present study was to investigate the errors made by the students related to the concepts of mean, median and mode concepts given in bar graph representations. According to the results, it was found that the students made errors while solving the questions related to the mean, the median and the mode concepts. Their errors regarding each concept were categorized into three groups: *Algorithmically based errors, errors based on formal knowledge* and *other errors*. *Operational error* mentioned as sub-group of algorithmically based error were encountered in the responses to questions regarding each concept. Moreover, there were five error types made as sub-groups of errors based on formal knowledge regarding the mean concept. More specifically, the error of *incorrect usage of averaging algorithm* was made by many students. Besides, the error of *looking pattern* had the lowest percentage among the errors based on formal knowledge. Furthermore, the students also made some errors from the category of other errors regarding the mean. In this category, the most common error was the *not meeting all requirements of a problem*. On the other hand, while the students tried to solve the questions regarding the median, they made some errors from each category. As sub-groups of errors based on formal knowledge regarding the median, *wrong decision on unordered data set, wrong decision on ordered data set* and *wrong interpretation of graph* were identified. Nearly an equal number students made these errors. Furthermore, the *forming incomplete data set* was a common error within the three sub-groups of the other errors category. Lastly, errors were also encountered in the responses of the students for the questions regarding the mode. The error of *inappropriate usage of averaging algorithm* was made errors based on formal knowledge related to the mode. Besides, the errors of *forming incomplete data set* and *incorrect largest and/or smallest data values* were made regarding the mode.

The third purpose was to investigate misinterpretations made by the students related to the concepts of mean, median and mode given in bar graph representations. According to the results, most of the students made judgments based on the magnitude of a data set based on a single value instead of an average of this data set in item 3-c. Furthermore, "the average decreased when a person left from the group" was most frequently expressed as a misinterpretation by the students in their responses to item 4-c.

CHAPTER 5

DISCUSSION, IMPLICATIONS AND RECOMMENDATIONS

The aim of the present study was to investigate the solution strategies used by seventh grade students and errors made by the students while solving questions regarding the measures of central tendencies, that is the mean, the median and the mode concepts given in bar graph representations. Also, the possible misinterpretations made by seventh grade students regarding the concepts of mean, median and mode given in bar graph representations were examined.

This chapter includes the summary of the findings in accordance with the purposes of the study and a discussion of the findings with regard to previous studies. Furthermore, implications and recommendations for further studies are presented.

5.1 Solution Strategies for the Questions Regarding the Concepts of Mean, Median and Mode

The Statistics Achievement Questionnaire (SAQ) regarding the mean concept includes two types of questions, namely finding the mean of a given data set and constructing a possible data set for the given mean. The findings obtained by analyzing the students' correct answers in the Statistics Achievement Questionnaire (SAQ) showed that the questions regarding the mean were solved by means of three different solution strategies, namely using balance model, using average formula and guess and check.

More specifically, in the present study, while finding the mean of a data set or a missing value of a data set for the given mean, using average formula solution strategy was used.

This finding was in agreement with the study of Mokros and Russell (1995), from which one of the SAQ items was adapted. In the study of Mokros and Russell (1995), the average of a data set was given and the construction of a possible data set was required. One of the five different constructions of representativeness was algorithm (arithmetic average). More precisely, most of the students used the using average formula solution strategy to solve the items regarding the mean. This finding was also consistent with the results of studies by Cai (1998; 2000), which stated that most of the participants also used using average formula strategy to solve averaging problems. The students had the tendency to use using average formula solution strategy to solve the items; this may be attributed to the fact that only this strategy might have been taught by the mathematics teacher of the students to find the mean of a data set during their lessons. Moreover, the middle school mathematics textbooks in Turkey include only using average formula solution strategy to solve questions regarding the mean. Thus, frequent use of algorithms to find the mean of the given data is not a surprising finding for the present study.

On the other hand, some of the students used using balance model and guess and check solution strategy to solve these kinds of questions. Using balance model and guess and check solution strategies were generally used while constructing a possible data set for the given mean when compared with the solution of the questions which required finding the mean of a data set. In general, students who used guess and check strategy first guessed a possible data set, and then checked the correctness of the data set by using the averaging algorithm of the mean. Guess and check solution strategy was observed in the study of Cai (1998; 2000), but it was found that participants used guess and check solution strategies while trying to find a missing value of a data set for the given mean and the other values of the data set. The findings of this study may have revealed the use of this strategy since students get used to using this strategy while solving questions related to the other topics of mathematics. For example, in order to find a missing value of an equation, firstly a value might be guessed, and then the correctness of the value could be checked. Thus, students' familiarity with this strategy from the previous classes might be the reason underlying their use of this strategy in solving average problems.

On the other hand, as previously mentioned, using balance model strategy was used to solve the items regarding the mean in the present study. This finding was also in agreement with the study of Mokros and Russell (1995) since the results of the study indicated that one of the constructions of representativeness of average was accepted as the balance point. Furthermore, similar to findings of this study, the results of the studies of Cai (1998; 2000) showed that participants used the levelling solution strategy to solve questions regarding the mean, which was similar to using balance model solution strategy in the present study. Using balance model solution strategy could be considered as an invented strategy since the mathematics' textbooks in Turkey do not include using balance model as a solution strategy to solve the mean questions. Inventing an alternative strategy to solve a question related to the mean might be an indicator of students' conceptual understanding of the concept (Cai, 2000). Therefore, the students who used this strategy as a solution method could be categorized as a group of students that have conceptual understanding regarding the mean concept. In other words, the students might understand that when the mean is accepted as a point on a number line, the data of each side of the point on the number line are balanced (Van de Walle, 2013).

In terms of the median, the questions of the SAQ assessed the students' solution strategies while finding the median of a data set and constructing a possible data for the given median. Furthermore, each data set in the SAQ questions was shown on a bar graph. The findings of the study showed that three types of solution strategies were used by the students, which were using numerical procedures, depending on graph representations, and guess and check. While trying to find the median of a data set, the majority of the students solved by using numerical procedures as a solution strategy. When this strategy was used, first the data set was ordered, and then the mid-point of the two middle values of the data sets was found since the data sets of the SAQ items regarding the median included an even number of values. The participants of the study have generally preferred this strategy since their teacher might generally have used the ordered given data set strategy to solve questions regarding the median during their lessons or their teacher might not have asked questions regarding the median concept that did not lend themselves to be solved with different solution strategies. Moreover, they might generally be using this strategy since their textbooks include only this

strategy as a solution strategy for questions regarding the median. In addition to using numerical solution strategy, some of the participants used depending on graph representations to solve the questions regarding the median. This solution strategy might have been preferred since the data sets were presented on a bar graph.

The SAQ also included items regarding the mode concept. One of them asked for the missing value of a data set when the mode of the data set was given. Furthermore, another item required the construction of a possible data set based on the given mode. The data sets of each question in the SAQ were given on a bar graph. Findings showed that there were three types of solution strategies to solve the questions regarding the mode, which were using numerical procedures, depending on graph representations, and guess and check. In order to find a missing value of a data set according to the given mode, most of the students used the solution strategy of depending on graph representations. The solution strategy of depending on graph representations might have been preferred by majority of the students since the data sets of each item were given on a bar graph. However, if the data set had not been given on a graph, the generally preferred solution strategy might have been the solution strategy of using numerical procedures because students generally use this strategy when a data set is not given on a graph during their regular mathematics lessons. Moreover, some of the students also used guess and check solution strategy while constructing a possible data set for the given mode. This strategy was used to find a missing value was asked when the mode and other values of the data set were given. As mentioned before, they may have used this strategy since they were familiar with guess and check solution strategy to solve problems regarding other mathematics topics. In addition, one of the reasons why the findings of the present study revealed the use of different solution strategies might be that the students were not familiar with items given in the SAQ. More specifically, the items regarding the mode did not directly ask for the mode of a data set; one of them asked the missing value of a data set when the mode of the data set was indirectly given and the other one required the construction of a data set. Thus, the students might have generated alternative solution strategies to solve the questions regarding the mode since they were not used to these kinds of problems and because data sets of the questions were given on bar graphs.

5.2 Students' Errors Regarding the Concepts of Mean, Median and Mode

Data analyses revealed that the students experienced problems in constructing data sets for the given averages, finding a missing data based on the given values and averages, calculating the average of a data set and interpreting the averages. Therefore, they made errors and wrong interpretations while trying to find answers to the given items in the SAQ. Findings showed that the students made different errors for each average type. These errors were categorized under three main groups, namely algorithmically based errors, errors based on formal knowledge and other errors.

5.2.1 Students' Errors Regarding the Mean Concept

The students made errors in the items that assessed their knowledge regarding the mean. The sub-group of algorithmically based errors for the mean concept was operational error. In addition, the sub-groups of errors based on formal knowledge were finding total number, the incorrect usage of averaging algorithm, the accepted close values, the looking pattern, and the turned to find smaller or larger numbers. Lastly, the sub-groups of the other errors were not meeting all requirements of a problem, forming incomplete data set and the incorrect reading of values from graph. More specifically, operational errors were made while calculating the mean of a data set. In fact, even though they were aware of the procedures involved in computing the mean, they made errors while adding, subtracting, multiplying or dividing numbers (e.g. item 2-ab, 2-c, 3-a and 4-b).

Errors based on formal knowledge revealed information regarding the students' conceptual knowledge. The most common error based on formal knowledge was the incorrect usage of averaging algorithm. For instance, a student found the sum of values of the given data set but the sum was not divided by the number of values to find the mean of the data set as in item 4-b. Also, when the mean and all of the values of a data set, except for one missing value, were given, first the student added the given values and the mean, then the sum was divided by a value which included the given number of values and the mean as in item 2-ab. This finding was also consistent with the findings of previous research studies in which students' errors regarding computational algorithm of the mean were reported (Cai, 1998; 2000; Maverach, 1983; Pollatsek et al., 1981;

Watson & Moritz, 1999). The reason of this error might be students' lack of conceptual knowledge regarding the mean since if they had conceptualized the mean as a fair share, they might not have made this error. In other words, the students who made this error may merely have had the knowledge that all values needed to be added and then divided by a number; however, they might not have known which values had to be added and by which values they were to be divided by due to lack of conceptual knowledge. Moreover, they might also have had problems while using the inverse of the averaging algorithm. In item 2-ab, a missing value of a data set was asked when the other values and the mean of the data set were given on a bar graph. Finding the missing value of a data set when the other values and the mean of the data set were given on a bar graph might not have been a routine problem for them since they might not have solved these kinds of questions during their lessons. From this perspective, the reason of the incorrect usage of averaging algorithm might have derived from the students' inadequate conceptual knowledge regarding the mean (Cai, 2000; Hiebert & Carpenter, 1992). More specifically, if students had understood the mean as "... a number that represents what all of the data items would be if they were leveled off" (Van de Walle, 2013, p.446), they might not have made the incorrect usage of averaging algorithm error.

On the other hand, the second common error based on formal knowledge was the turned to find smaller or larger numbers error type, which was made while constructing a possible data set for the given mean. All values were selected either smaller or larger than the given mean by the students who made this error. This error might have occurred due to lack of conceptual knowledge since most probably they regarded the mean as the maximum or minimum value of a data set rather than a balance point. To state it differently, when they understand the mean as a balance point, they might think that the data set includes some values which are higher than the mean and some values which are lower than the mean, but each side of the point are balanced. Furthermore, the construction of a possible data set for the given mean might be a non-routine problem for participants of the study since they might not have solved these kinds of questions during their mathematics lessons. However, if students had complete conceptual knowledge regarding the mean concept, they would have used their knowledge to solve

non-routine problems regarding the concept (Hiebert & Carpenter, 1992; Watson & Moritz, 2009).

Lastly, the third common error based on formal knowledge regarding the mean was the finding total number. The students who made this error found the sum of the given values and gave the sum as the mean of the given data set. In the present study, this error was made while either finding the mean of a data set or constructing a possible data set for the given mean. This error was also made by the participant of the study of Cai (2000) while trying to find a missing value of a data set for the given mean and other values of the data set. Since some of the students added all the values of a data set and gave the sum as the mean of the data set, the reason underlying this error can be attributed to students' lack of procedural knowledge regarding the mean concept, which necessitates the knowledge of the "add-and-divide" algorithm to find the mean. The accepted close values error was another error based on formal knowledge. The accepted close values error was also observed in the studies of Cai (1998; 2000). The students who made the accepted close values error stopped trying to construct a data set although the remainder was not zero while trying to construct a possible data set for the given mean and sample size. In fact, this error type was observed when the students tried to use the guess and check solution strategy. Thus, it is highly likely that the students guessed a possible data set, found the sum of the values, and then divided the sum by the number of values and checked whether or not the quotient was equal to the given mean without considering the remainder of the division. The looking pattern error was another type of observed error based on formal knowledge. The looking pattern error was made when the students selected the missing values of a data set or constructed a possible data set as if there were a pattern among the values of the data sets although there was no pattern. The finding was consistent with observed errors in the study of Cai (2000). The reason of this error might be the overgeneralization of the exploring pattern solution strategy to find a value which was missing in a set of value (Cai, 2000). In fact, although one of the ways of doing mathematics is exploring patterns, which helps students build mathematical relationships (NCTM, 1989), students examine whether or not there are any patterns among the given values in order to effectively use exploring patterns as a solution strategy. In general, the reason underlying the errors based on formal

knowledge, which the present study yielded, might be the format of the items given in the SAQ since the students might have been unfamiliar with these types of questions.

On the other hand, there were errors other than the algorithmically based errors and errors based on formal knowledge categories, which was mentioned under the category of other errors. The most common type of error within this category was not meeting all requirements of a problem. The reason underlying this error might be careless reading of the items since the given answers were not appropriate for a written note or expression. Furthermore, there was another error type, which was the forming incomplete data set. In the present study, the students who made this error did not construct appropriate data sets for the given sample size. This finding might be because of the students' careless reading of the sample size of the required data set. Also, the students might not have constructed appropriate data sets for the given sample size due to their lack of knowledge. More specifically, the participants might not have known the meaning of sample size or they might have had difficulty while trying to construct a possible data set not only for the given mean but also for the sample size. Additionally, the last sub-group of other errors category was the incorrect reading of values from given graph. In this error, some of the values which were given on a bar graph were not read correctly. This may have resulted from making a minor mistake while reading a value since most of the students who made this error read only one or two values incorrectly.

In addition to these, in the present study, when the average of a data set was asked for or when the average of a data set was given and a possible data set for the given average was asked for, all of the students accepted only the average as the mean of the data sets. This finding was also supported by Brown and Silver (1989), Uçar and Akdoğan (2009) and Randall (2006). The main reason behind this error might be the commonalty between the word ‘mean’ and other average words. More specifically, the mean is presented as arithmetic average in Turkish mathematics textbooks. Thus, the term ‘arithmetic average’ includes the word ‘average’. Furthermore, poorly-constructed questions in some textbooks may be contributing to this error since while asking for the mean of a data set, the term ‘average’ might be used instead of the term ‘mean’.

The analysis of the students' answers indicated that the students experienced more problems while trying to find a missing value of a data set when other values and the mean of the data set were given than while finding the mean of a data set. The results of previous studies showed that whereas most of the students could easily calculate the mean of a given data set, they experienced problems while working backwards to find the missing values in a data set by using the given mean (Cai, 1995; Mokros & Russell, 1995).

5.2.2 Students' Errors Regarding the Median Concept

SAQ included two items to assess the students' knowledge regarding the median. One of them asked for the median of a data set and the other one required the construction of a possible data set for the given median. The findings of the study indicated that the students made errors while trying to solve the SAQ items regarding the median concept. Based on the observations of a study by Russell, Schifter and Bastable (2002), students have difficulties while learning the median concept.

As previously mentioned, operational errors were made by the students while adding, subtracting, multiplying or dividing numbers. The students were aware of the procedures to find the median of a data set but they made errors while doing operations. In addition, the students made errors based on formal knowledge regarding the median. Three types of errors based on formal knowledge were observed, namely the wrong decision on unordered data set, the wrong decision on ordered data set and the wrong interpretation of graph. These errors were made in item 3-b, which asked for the median of a data set. In this item, the data sets were given on a bar graph and the data sets included an even number of values. Among the three error types, the most common one was the wrong interpretation of graph. The reason underlying this situation might be that the students who made this error directly tried to find the median of the data set from the given graph. In other words, their answers were not correct since the data set was not ordered on the graph and there was an even number of values. The other reason of this error might be that the students who made this error did not need to order the given data set, so they thought that making a decision based on the given graph was appropriate. If they

had not had the need to order the given data set on a bar graph, they might not have understood that half of all values must be above or at the median and half of them must be below or at the median (Van de Walle, 2013). Thus, the students might not have made the error of wrong interpretation of graph when the data set was not given on a bar graph in item 3-b. Hence, a graphical representation might have been the cause of this error. The second error based on formal knowledge was the wrong decision on ordered data set. The students who made this error ordered the given data sets but they could not find their median. The main reason behind this error might be that the data sets in item 3-c included an even number of values. Thus, the students knew that the given values needed to be ordered first and then the middle point of the ordered values was to be selected in order to find the median of a data set. However, they confronted a problem in that there was no one middle value. Because of this reason, they might have made wrong decisions on their ordered data sets. Lastly, the error of the wrong decision on unordered data set was made when the students tried to find the median of a data set in item 3-b. The students did not order the obtained data set to find the median of the data set. The reason of this error might be similar with that of the error of wrong interpretation of graph.

On the other hand, there were errors within the category of other errors regarding the median concepts since these errors were not directly related to the median but they affected their answers while providing answers to the items that assessed the students' knowledge of the median. The most common error in the category of other errors regarding the median concept was the forming incomplete data set. Item 5 required the construction of a possible data set for the given median. The sample size of the data set was also included in the wording of the item. The students who made this error did not construct the appropriate data set for the given sample size. The reason of this error might be similar with that of the error regarding the mean. The students might not have constructed appropriate data sets for the given sample size due to their lack of knowledge. More specifically, students may not have known the meaning of sample size or they might have had difficulty while trying to construct a possible data set not only for the given median and mode but also for the sample size. Furthermore, the error of incorrect reading of values from graph was made in item 3-b because the data sets were

given on a bar graph. In fact, only some of the values of the data sets were read incorrectly, not all of them. Therefore, the reason underlying this error might have derived from the carelessness of the students while reading the given bar graph in item 3-b. Lastly, the incorrect largest and/or smallest data values error was made since the students who made this error did not construct the appropriate data sets for the given range in the problem context. The first reason of this error might be their careless reading of item 5 before constructing a possible data set for the given median. The second reason of this error might be the students' inadequate reading abilities. They might have misunderstood the given expression of the problem context. Thirdly, item 5, which required the construction of a possible data set, might have been a non-routine problem for them. Thus, they might have experienced problems while solving the item.

5.2.3 Students' Errors Regarding the Mode Concept

SAQ included two items to assess the students' knowledge regarding the mode. One of them gave the mode of a data set indirectly and asked for the missing value of the data set. Furthermore, the other required the construction of a possible data set for the given mode. The findings of the study indicated that the students made errors while trying to solve the SAQ items regarding the mode concept. Based on the analysis results, there was one sub-group in the category of errors based on formal knowledge, which was the inappropriate usage of averaging algorithm and two types of sub-groups within the other errors category were defined, namely the forming incomplete data set and the incorrect largest and/or smallest data values.

According to the findings, there was one sub-group of errors based on the formal knowledge category regarding the mode, which was the inappropriate usage of averaging algorithm. This error was made in item 4-c. In this item, all the values of a data set except one value and the most frequent value were given; the missing value of the data set was asked for. The students who made this error found the mean of the given values of the data set. This situation may have resulted from the fact that they focused on using the computational algorithm of the mean when a missing value of a data set was asked for based on the given values and average. To state it differently, finding the missing value of a data set when the mode of the data set was given might have been a

non-routine type of problem for them. If they had a more complete conceptual understanding of the mode, that the mode is the most frequent data within a data set, they might not have experienced difficulties while solving non-routine problems regarding the mode (Cai, 2000).

On the other hand, the forming incomplete data set and the incorrect largest and/or smallest data values errors were the sub-groups of the other errors category regarding the mode. The forming incomplete data set error was mentioned within the other errors regarding the concepts of the mean and the median. Moreover, the latter was also defined in the other errors category regarding the median concept. In fact, since these errors were made in item 5 and they were independent of the concepts of the median and mode, and as the item required the construction of a data set for these concepts, these errors were considered as the sub-groups of other errors category for the two concepts.

5.3 Misinterpretations Regarding the Concepts of Mean, Median and Mode

SAQ included three interpretation items, namely item 2-d, item 3-c, and item 4-c. In these items, students were expected to interpret a situation or decide whether a statement was true or not and state the reasons underlying their decisions since one of the purposes of this study was to investigate seventh grade students' misinterpretations regarding the concepts of mean, median and mode given in bar graph representations. Their misinterpretations were examined in detail. More specifically, the percentages of misinterpretations for items 2-d, 3-c and 4-c were 20.6%, 12.9% and 34.3%, respectively.

In item 2-d, it was asked what could be said about the values of the data set when the mean of a data set was given. More specifically, the item was worded as follows: “When we know the arithmetic average (mean) of Elif's project scores as 85, what could be said about Elif's project scores?” Based on the analysis results of the students' misinterpretations for item 2-d, it was seen that most of the students had used the computational algorithm of the mean but incorrectly. For example, the given average was divided by the number of values and an interpretation was made for the result. The reason of this misinterpretation might be the students' lack of understanding regarding

the mean, that the mean is a summary statistics of a data set. If the students had had this understanding, they might not have used the computational algorithm of the mean incorrectly. Besides, the finding might merely be the result of the incorrect usage of averaging algorithm. On the other hand, the remaining misinterpretations were made by almost equal number of students. To illustrate, 3 of 48 students said that 'all project scores were smaller than 85' and 2 of them said that 'Elif's project scores were at least 85'. These findings might have derived from a misconception held by students, which is that all values of a data set must be either smaller or larger than the mean of the data set. When the students know the mean as a balance point, they might not make this wrong interpretation (Konold & Pollatsek, 2002; Mokros & Russell, 1995) because if they could understand the mean as a balance point, they might think that a data set can include not only smaller values but also larger values than the mean of the data set. These misinterpretations were consistent with the turned to find smaller or larger numbers error, which was defined in the errors regarding the mean concept of the present study. Besides, 2 students among 48 students said that 'all project scores must be at least 4' and 2 of them said that 'Elif's project scores are close to each other'. The students might have thought that when the students had a score which was smaller than four or when there was a gap among the scores of Elif, the mean could not be 85. Thus, the reason behind this error might be students' lack of conceptual understanding related to the mean because if the mean had been understood as a fair share or a balance point, they would not have made these interpretations (Konold & Pollatsek, 2002; Mokros & Russell, 1995). In other words, since fair share is accepting the mean as equal distribution of a data set and the balance point is accepting the mean as a point of balance in a data set, the students who possessed these understandings might not have made the misinterpretations.

In item 3-c, the students' conceptual knowledge regarding the mean and the median was investigated. In this item, a statement was given and then it was asked whether the statement was true or not. The statement was as follows: "Ahmet claims that the candles of Company A have longer burning durations than those of Company B. Do you agree with Ahmet? How did you decide whether Ahmet is right or not? Explain your reasoning." The correctness of the statement could be decided by comparing the average

of the burning durations of the candles. Results of research studies showed that students do not know how to compare two groups (Shaughnessy, 1992; Watson & Moritz, 1999). Findings of the present study indicated that while making a decision, the students made incorrect interpretations. For example, 9 among 30 students said that 'the statement was incorrect since the candle which had the longest burning duration was the one in Company B'. The students who made this interpretation made their decision by using the burning duration of a candle which had the longest burning duration within the data sets. In fact, their decision was correct, but the interpretation was incorrect because having the largest value in a data set did not provide a summary of the data set. Most probably, the students did not think that the averages are summary statistics of a data set. Furthermore, this misinterpretation might have derived from giving data sets on a bar graph since when deciding which company's candles have longer burning durations, the students might have based their response on the largest value in data sets by using the peaks of the given bar graph. On the other hand, although some students used averages to describe a group of data, most of them did not use them to compare two data sets (Konold & Pollatsek, 2002; Watson & Moritz, 1999). Thus, since the averages are used to summarize a data set and they can also use to compare two data sets (Shaughnessy, 1992; Watson & Moritz, 1999), students who used an appropriate average type might make correct interpretations while comparing two data sets. In addition, 7 students among 30 students said that 'the statement was correct since the median of Company A was higher than that of Company B'. Although the students used an average in order to decide about the correctness of the statement, their interpretation was incorrect since the median was not an appropriate summary statistics for the data set. Since the data set did not include any outlier, the mean was the appropriate summary statistics for the data set. In other words, when a data set has extreme values, the median is the most appropriate average type to summarize the data set (Fraenkel & Wallen, 2006). Therefore, if the students had known that the median is appropriate summary statistics when a data set has extreme values, they might not have made the misinterpretation.

In item 4-c, how the students interpreted the average of the remaining values of the data set when a value was subtracted from a data set was examined. Five different misinterpretations were observed. The most frequent misinterpretation for the item was

that 'the average decreased when a person left the group', which was made by 57 students among 80 students. This misinterpretation might have derived from a misconception that when a value was subtracted from a data set, the average always decreased. The students who made this misinterpretation might not have regarded the mean as a point of balance. In other words, they might not have understood that some of values of a data set was smaller than the mean of the data set and some of them was larger than the mean of the data set (Hardiman, Well & Pollatsek, 1984). Therefore, they might not have evaluated the effect of the subtracted value of the data set on the mean of the data set since a student who understands the mean as a point of balance might know that when a value which is smaller than the mean is subtracted, the mean is increased. Furthermore, 11 students among 80 students said that 'the average decreased since the sum of the age of the group decreased'. The students who made this interpretation were not taking into consideration the change of the sample size. On the other hand, 2 students among 80 students said that 'the average of the children's age decreased since the number of children in the park decreased'. The reason of the last two misinterpretations might be lack of procedural and conceptual understanding regarding computational algorithm of the mean. To state it differently, they might not have regarded the mean as a fair-share since the fair-share accepts the mean as equal distribution of a data set (Mokros & Russell, 1995). If a student does not understand the mean as a fair-share, s/he might not understand the computational algorithm of the mean and they might only focus on the sum of values or the number of values.

5.4 Implications

This study offers valuable information to teachers, teacher educators, textbook writers and curriculum developers about the solution strategies middle school students resorted to while solving questions regarding the concepts of mean, median and mode given in bar graph representations; and their possible errors and misinterpretations regarding these concepts. Findings of this study revealed that the seventh grade students used three different solution strategies to solve the questions regarding each concept and the students made errors and misinterpretations while solving the questions related to the mean, the median and the mode concepts. Teachers, teacher educators, textbook writers

and curriculum developers should take the solution strategies, errors and misinterpretations into consideration in order to prepare effective teaching environment and learning materials while teaching the measures of central tendencies.

More specifically, the results of the study can help teachers to gain insight into middle school students' possible solution strategies, errors and misinterpretations regarding the concepts of mean, median and mode. Firstly, teachers could benefit from middle school seventh grade students' possible solution strategies while solving questions regarding the measures of central tendencies by means of bar graphs in order to provide information about the students' procedural and conceptual understanding regarding these concepts. When they are informed about the possible solution strategies used by middle school students to solve questions regarding the concepts of mean, median and mode, they might prepare appropriate questions to reveal the different solution strategies middle school students used in relation to the concepts they dealt with duirng their mathematics lessons. In this way, students' conceptual and procedural knowledge regarding these concepts might be developed. Moreover, middle school students' understanding of averages might be understood through their solution strategies because the findings of the study indicated that some of the students understood average as balance point, some of them as fair share and some of them have both of the understandings.

Secondly, teachers could benefit from findings of the study regarding middle school students' possible errors and misinterpretations of the concepts of mean, median and mode through seminars or in-service programs to create awareness by which teachers can prepare appropriate teaching plans which could eliminate the errors and prevent the misinterpretations. To illustrate, teachers can solve additional questions regarding the concepts in order to develop middle school students' procedural and conceptual understanding about these concepts. Thus, their errors and misinterpretations can be eliminated. Furthermore, some of their errors and misinterpretations might be due to their misconceptions. The teachers can become aware of their students' errors and misinterpretations and in this way they can investigate the reasons underlying their students' errors and misinterpretations regarding the average concepts. Thus, they can find appropriate remedies and overcome those difficulties and misconceptions.

In addition, teacher educators can also benefit from the findings of the study. More specifically, pre-service middle school mathematics teachers can be informed about middle school students' possible errors and misinterpretations regarding the concepts of mean, median and mode. In this way, since pre-service teachers will be aware of the defined errors and misinterpretations, they can prepare appropriate teaching plans to eliminate the errors and to prevent the misinterpretations when they are in-service teachers. Moreover, pre-service teachers can be informed about students' possible solution strategies regarding the concepts of mean, median and mode by teacher educators. In this way, prospective teachers can be aware of different solution strategies to solve questions regarding these concepts and they can welcome these strategies through appropriate problems.

In addition to teachers and teacher educators, textbook writers and curriculum developers can also benefit from the findings of the study. The teacher guides of mathematics textbooks may benefit from the findings of the present study. More specifically, different solution strategies for a question regarding the mean, the median and the mode could be added to textbooks to raise awareness on alternative solution methods regarding these concepts. Textbooks could include problems enabling middle school students' to generate different solution strategies through multiple thinking. In this way, students' conceptual and procedural knowledge regarding these concepts could be developed (Hiebert & Carpenter, 1992; Wang, Dogan & Lin, 2006) Furthermore, questions similar to the ones used in the present study that support the conceptual understanding of the concepts of mean, median and mode could be added to mathematics textbooks. To illustrate, a missing value problem could be asked when values of the data set was given on a bar graph or a possible data set could be generated for the given average. With these questions students' conceptual understanding might improve since the questions may be non-routine for middle school students.

5.5 Recommendations for Further Research Studies

The present study has limitations due to the sampling method of the study, which was convenience sampling. Since the sampling method is not one of the random sampling methods, the findings of the study could not be generalized to a large population. In

order to generalize the findings of the study to a population, further research might be conducted to a sample from middle school seventh grade students in Turkey, selected with a random sampling method. Moreover, findings of the present study was limited with the questions asked in SAQ since when different questions were asked related to the concepts of mean, median and mode, different findings could be reached. Furthermore, a similar study might be conducted in private schools to investigate private middle school students' understandings regarding the average concepts. Besides, further study might be conducted to investigate the reasons and misconceptions behind the errors and the misinterpretations of middle school seventh grade students regarding the concepts of mean, median and mode. In this way, since teachers could be aware of the students' misconceptions behind their errors and wrong interpretations, they could have a high level of readiness regarding the average concepts. Moreover, a longitudinal study could be conducted to investigate the development of middle school students' procedural and conceptual knowledge regarding the mean, the median and the mode concepts based on their grade level.

REFERENCES

Ashlock, R. B. (1994). *Error patterns in computation*. New York, Macmillan.

Askew, M. & Wiliam, D. (1995). *Ofsted Reviews of Research: Recent research in mathematics education,* 5-16. London: HMSO.

Baroody, A. J., Feil, Y., & Johnson, A. R. (2007). An alternative reconceptualization of procedural and conceptual knowledge. *Journal for Research in Mathematics Education, 38*(2), 115 – 131.

Becker, J. P. (Ed.), 1992. *Report of U.S.–Japan Cross-National Research on Students' Problem Solving Behaviors*. Columbus, OH: ERIC/SMEAC Clearing House, ED 351/204.

Behr, M. J., Harel, G., Post, T., & Lesh, R. (1992). Rational number, ratio, and proportion. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning*, 296-333. New York: Macmillan.

Bisanz, J., & LeFevre, J. (1992). Understanding elementary mathematics. In J. I. D. Campbell (Ed.), *The Nature and Origins of Mathematical Skills,* (pp. 113-136). Amsterdam: North-Holland Elsevier Science Publishers.

Bright, G. W. & Friel, S. N. (1998). Graphical representations: Helping students interpret data. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K–12* (pp. 63–88). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bulut, S., Ekici, C. & Iseri, A. I. (1999). Bazı olasılık kavramlarının gelişimi için çalışma yapraklarının geliştirilmesi. *Hacettepe University Journal of Education, 15,* 129-136.

Cai, J. (1995). Beyond the computational algorithm: Students' understanding of the arithmetic average concept. In L. Meria & D. Carraher (Eds.), *Proceeding of the 19th Psychology of Mathematics Education Conference* (Vol. 3. pp. 144-151).

Cai, J. (1998). Exploring students' conceptual understanding of the averaging algorithm. *School Science and Mathematics, 98*(2), 93-98.

Cai, J. (2000). Understanding and representing the arithmetic averaging algorithm: an analysis and comparison of US and Chinese students' responses. *International Journal of Mathematical Education in Science and Technology, 31*(6), 839-855.

Capraro, R. M., Kulm, G., Hammer, M., & Capraro, M. M. (2002). The origin and persistence of misconceptions in statistical thinking. In D. S. Mewborn, P. Sztajn, D. Y. White, H. G. Wiegel, R. L. Bryant, & K. Nooney (Eds.), *Proceedings of the 24th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 3. pp. 1339-1340). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Capraro, M. M., Kulm, G. & Capraro, R. M. (2005). Middle grades: Misconceptions in statistics thinking. *Journal of Schools Science and Mathematics, 105*(4), 165-174.

Carr, J., & Begg, A. (1994). Introducing box and whisker plots. In J. Garfield (Ed.), *Research Papers from the Forth International Conference on Teaching Statistics*. Minneapolis, MN.

Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics, 50*(1), 66-71.

Conners, F. A., McCown, S. M., & Roskos-Ewoldsen, B. (1998). Unique challenges in teaching undergraduate statistics. *Teaching of Psychology, 25*(1), 40-42.

Cox, L.S.(1975). Systematic Errors in the Four Vertical Algorithms in Normal and Handicapped Population. *Journal for Research in Mathematics Education*, 6(4), 202-220.

Demirci, N. & Uyanık, F. (2009). The correlation between tenth grade students' understanding and interpreting graphs and their kinematics achievement. *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education, 3*(2), 22-51.

Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education*. New York, the USA: McGraw-Hill.

Franklin, C. A., Mewborn, D., Moreno, J., Peck, R. Perry, M., & Scheaffer, R. (2005). *Guidelines for assessment and instruction in statistics education (GAISE) report.* Alexandria, VA: American Statistical Association.

Friel, S. N., & Bright, G. W. (1998). Teach-Stat: A model for professional development in data analysis and statistics for teachers K-6. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching assessment in grades K-12* (pp.89-117). Mahwah, NJ: Erlbaum.

Friel, S. N., Curcio, F. R. & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal of Research in Mathematics Education, 32,* 12-158.

Fry, E. (1984). *A theory of graphs for reading comprehension and writing communication*. New Brunswick, NJ: Rutgers University.

Garcia Cruz, J. A., & Garrett, A. J. (2008). Understanding the arithmetic mean: A study with secondary and university students. *Journal of the Korea Society of Mathematics Education Series D: Research in Mathematical Education, 12*(1), 49-66.

Green, M., Piel, J. A., & Flowers, C. (2008).Reversing education majors' arithmetic misconceptions with short-term instruction using manipulatives. *Journal of Educational Research, 101*(4), 234-242.

Groth, R. E. & Bergner, J. A. (2006). Preservice elementary teachers' conceptual and procedural knowledge of mean, median and mode. *Mathematical Thinking and Learning, 8*(1), 37-63.

Hawkins, A. S., & Kapadia, R. (1984). Children's conceptions of probability - A psychological and pedagogical review. *Educational Studies in Mathematics, 15*, 349-377.

Hardiman, P., Well, A. D., & Pollatsek, A. (1984) The usefulness of a balance model in understanding the mean. *Journal of Educational Psychology, 76*, 792-801.

Hewson, P. W. & Hewson, M. A. G. (1984). The role of conceptual conflict in conceptual change and the design of science instruction. *Instructional Science, 13*, 1-13.

Hiebert, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (ed.), *Handbook of Research on Mathematics Teaching and Learning* (New York: Macmillan), pp. 65–97.

Hiebert, J. & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1–28). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Jacobbe, T., & Carvalho, C. (2011). Teachers' understanding of averages. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics – Challenges for teaching and teacher education* (pp. 199–209). Dordrecht: Springer.

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

Kathleen, M.S. (1994). *The development and validation of a categorization of misconceptions in the learning of chemistry*. Unpublished Doctoral Dissertation, University of Massachusetts.

Kennedy, L.M., Tipps, S., & Johnson, A., *Guiding Children's Learning of Mathematics.* Thomson Wadsworth, USA, 2008.

Konold, C. & Pollatsek, A. (2002). Data analysis as a research for signals in noisy processes. *Journal for Research in Mathematics Education, 33*, 259-289.

Konold, C & Higgins, T. L. (2003), "Reasoning about data", In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics*, Reston, VA: National Council of Teachers of Mathematics, pp.193-215.

Koshy, V. (2000) Children's mistakes and misconceptions. In Koshy, V., Ernest, P.andCasey, R. *Mathematics for Primary Teachers* London: Routledge.

Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology, 3*, 185–226.

Lappan, G., Fey, J. T., Fitzgerald, W. M., Friel, S. N. & Philips, E. D. (2006). *Connected mathematics 2: Streching and shrinking: Understanding similarity*. Boston: Pearson Prentice Hall (School Division).

Leavy, A. M. (2004). Indexing distributions of data: Preservice teachers' notions of representativeness. *School Science and Mathematics, 104*(3), 119-133.

Leon, M. and Zawojewski, J. (1990). Use of Arithmetic Mean: An Investigation of Four Properties Issues and Preliminary Results. In D. Vere-Jones (Ed.), *Proceedings of the III International Conference on Teaching Statistics* (pp. 302-306). New Zealand.

Luneta, K & Makonye, P.J. (2010). Learners errors and misconceptions in elementary analysis: A case study of a Grade 12 class in South Africa, *Acta Didactica Napocenia 3*(3), 36-45.

Mathews, D. and Clark, J.M. (2007, July, 25), *Successful Students' Conceptions of Mean, Standard Deviation, and The Central Limit Theorem*. Unpublished Paper. Retrieved from http://www1.hollins.edu/faculty/clarkjm/stats1.pdf.

Mevarech, Z. (1983). A deep structure model of students' statistical misconceptions. *Educational Studies in Mathematics, 14*, 415-429.

Ministry of National Education [MoNE]. (2005*). İlkögretim matematik dersi ögretim programı 6-8. sınıflar: Ögretim programı ve kılavuzu* Ankara,Turkey

Ministry of National Education [MoNE]. (2013). *İlköğretim Okulu Ders Programları: Matematik Programı 5-6-7-8 (Elementary Curricula Programs:Mathematics Curricula Program for Middle Grades)*.Retrieved from http://ttkb.meb.gov.tr/www/guncellenen-ogretim-programlari/icerik/151.

Mokros, J. & Russell, S. J. (1995). Children's concepts of averages and representativeness. *Journal for Research in Mathematics Education, 26*, 20-39.

National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: The Council

National Council of Teacher of Mathematics (2000). *Principles and standard for school mathematics*. Reston, VA: Author

Nesher, P. (1987). Towards an instructional theory: The role of learners' misconception for the learning of mathematics. *For the Learning of Mathematics*, 7(3), 33-39.

Nisbet, S. (2002a). Year seven students' representation of numerical data: The influence of sample size. In B. Barton, K. Irwin, M. Pfankuch, and M. Thomas(Ed.), *Mathematics Education in the South Pacific, Proceedings of the 25th Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 520-527). Auckland: MERGA.

Nisbet, S. (2002b). Representing numerical data: The influence of sample size. In A. Cockburn and E. Nardi (Eds.), *Proceedings of the 26th Annual Conference of the International Group for the Psychology of Mathematics Education* (pp. 417- 424). Norwich, UK: University of East Anglia.

Nisbet, S., Jones, G., Thornton, C., Langrall, C., & Mooney, E. (2003). Children's Representation and Organisation of Data. *Mathematics Education Research Journal, 15*(1), 42-58.

Olivier, A. (1989). *Handling pupils misconceptions.* Presidential address delivered at the Thirteenth National Convention on Mathematics, Physical Science and Biology Education, Pretoria, 3-7 July 1989. Retrieved from http://academic.sun.ac.za/mathed/174/Misconceptions.pdf

Oxford Dictionaries. (2014, November 13). The second dictionary search result for the entered words 'bar graph' [HTML document]. Retrieved from http://www.oxforddictionaries.com/definition/american_english/bar-graph

Pereira-Mendoza, L., & Mellor, J. (1990). Students' concepts of bar graphs: Some preliminary findings. In D. Vere-Jones (Eds.), *Proceedings of the Third International Conference on Teaching Statistics* (pp. 150–157). Voorburg, The Netherlands: International Statistical Institute.

Pollatsek, A., Konold, C., Well, A. & Lima, S. (1984). Beliefs underlying random sampling. *Cognition and Instruction, 12*, 395-401.

Pollatsek, A., Lima, S., & Well, A. D. (1981). Concept or computation: Students' understanding of the mean. *Educational Studies in Mathematics, 12*, 191-204.

Pratt, D. (2005). Probability in teacher education and development. In G. A. Jones (Ed.), *Exploring probability in school* (pp. 171-190). New York: Springer Science+Business Media, Inc.

Radatz, H. (1979). Error analysis in mathematics education. *Journal of Research in Mathematics Education*, *10*(3),163–172.

Radatz, H. (1980). Students' Errors in the Mathematical Learning Process: A Survey *For the Learning of Mathematics*, *1*(1), 16-20.

Randall, G. (2006). An Exploration of Studentsaticsrors in the Mathema*Teaching Statistics*, *28*(1), 17-21.

Riccomini, P. J. (2005). Identification and remediation of systematic error patterns in subtraction. *Learning Disability Quarterly*, *28*(3), 233-242.

Russell, S. J. & Mokros, J. (1996). Research into practice: What do children understand about average? *Teaching Children Mathematics, 2*(6), 360-364.

Russell, S. J., Schifter, D. & Bastable, V. (2002). *Developing mathematical ideas: Working with data*. Parsippany, NJ: Dale Seymour Publications.

Russell, S. J. (2006). What does it mean that "5 has a lot"? From the world to data and back. In G. F. Burrill & P. C. Elliott(Eds.*), Thinking and reasoning about data and chance: 68th NCTM yearbook* (pp. 17-30). Reston VA: NCTM.

Schneider, M. & Stern, E. (2010). The developmental relations between conceptual and procedural knowledge: A multimethod approach. *Developmental Psychology, 46*(1), 178-192.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In Grouws, D.A.(Ed.). *Handbook of research on mathematics teaching and learning* (pp. 465-494). New York: Macmillan.

Smith, J. P., diSessa A. A., & Roschelle, J. (1993). Misconceptions Reconceived: A Constructivist analysis of Knowledge in Transition. *The Journal of Learning Sciences, 3(2)*, 115-163.

Strauss, S., & Bichler, E. (1988). The development of children's concept of arithmetic average. *Journal for Research in Mathematics Education, 19*, 64-80.

Suco, E., Samere, M., & Hong, S. L. (2012). Enabling students to understand measures of central tendency & variation. *Mathematics Teaching, 231*, 20-22.

Tarr, J. E. & Shaughnessy, J. M. (2007). *Data and Chance. Results from the 2003 National Asssessment of Educational Process*. Reston, VA: NCTM.

Taşar, M. F., İngeç, Ş. K., & Güneş, P. Ü. (2009). Grafik çizme ve anlama becerisinin saptanması. *Ahmet Keleşoğlu Eğitim Fakültesi Dergisi, 27,* 71-83.

Temiz, B. K. & Tan, M. (2009a). Assessing the graph drawing abilities with a check list. *Journal of Ahmet Keleşoğlu Education Faculty, 27*, 71-83.

Temiz, B. K. & Tan, M. (2009b). The abilities of first grade students to interpret the graaphs at high school. *Journal of Ahmet Keleşoğlu Education Faculty, 28*, 31-43.

Tirosh, D. (2000). Enhancing prospective teachers' knowledge of children's conceptions: The case of division fractions. *Journal of Research in Mathematics Education, 31*(1), 5-25.

Tortop, T. (2011). *7th-grade students' typical errors and possible misconceptions in graphs concept before and after the regular mathematics instruction (*Master's thesis). Retrieved from the National Thesis Center of the Higher Education Council database. (No. 270552)

Uçar, Z. T. & Akdoğan, E. N. (2009). Middle school students' understanding of average. *Elementary Education Online, 8(*2), 391-400.

Van de Walle, J.A. (2013). *Elementary and middle school mathematics: Teaching developmentally.* New York: Pearson Education, Inc.

Watson, J. M., & Moritz, J. B. (1999). Interpreting and predicting from bar graphs. *Australian Journal of Early Childhood*, *24*(2), 22-27.

Watson, J. M. & Moritz, J. B. (2001). Development of reasoning associated with pictographs: Representing, interpreting, and predicting. *Educational Studies in Mathematics, 48*, 47-81.

Watson, J. M. & Moritz, J. B. (2009). The longitudinal development of understanding of average. *Mathematical Thinking and Learning, 2*, 11-50.

Wood, R. (1968). Objectives in the teaching of mathematics. *Educational Research, 10*, 83-98.

APPENDICES

APPENDIX A. PERMISSION OBTAINED FROM METU APPLIED ETHICS RESEARCH CENTER

ÖĞRENCİ İŞLERİ DAİRE BAŞKANLIĞI
REGISTRAR'S OFFICE

ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

DUMLUPINAR BULVARI 06800
ÇANKAYA ANKARA/TURKEY
T: +90 312 210 34 17
F: +90 312 210 79 60
oidb@metu.edu.tr
www.oidb.metu.edu.tr

545690
07/02/2014

SAYI: 54850036-300 – 541

001276

31.01.2014

ÇANKAKALE VALİLİĞİNE
(İl Milli Eğitim Müdürlüğü)

Üniversitemiz İlköğretim Fen ve Matematik Eğitimi Ana Bilim Dalı Yüksek Lisans Programı öğrencisi Didem Enisoğlu'nun 2013-2014 Eğitim Öğretim yılı II. Döneminde yüksek lisans tezi kapsamında *"Yedinci Sınıf Öğrencilerinin Verilen Sütun Grafiğinde Aritmetik Ortalama, Ortanca ve Tepe Değer Kavramlarını Yorumlarken Yaptıkları Hatalar ve Kavram Yanılgıları"* başlıklı araştırma çalışmasına ilişkin hazırlanan anketi Çanakkale İl Milli Eğitim Müdürlüğü'ne bağlı ekli listede belirtilen İlköğretim okullarında uygulama yapmak için, öğrencinin isteği doğrultusunda görevlendirilmesi Etik Komite onayı ile uygun görülmüştür.

Uygulamanın yapılabilmesi için gereğini arz ederim.

Saygılarımla.

Prof. Dr. M. Volkan Atalay
Rektör Yardımcısı

6665

İl Milli Eğit. Md.
06.02.2014
Vali a

Bu evrâkın 3070 sayılı Kanun gereğince elektronik imzamızla imzalandığı tasdik olunur.
06.02.2014
Sukuri SAVAŞ
Şef

Ekler:
1- Öğrencinin Dilekçesi
2- Danışmanın Dilekçesi
3- İAEK Proje Bilgi Formu
4- Okul Listesi
5- Veri Toplama Araçları

APPENDIX B. PERMISSION OBTAINED FROM MINISTRY OF EDUCATION

FORM: 2

T.C.
MİLLÎ EĞİTİM BAKANLIĞI

ARAŞTIRMA DEĞERLENDİRME FORMU

ARAŞTIRMA SAHİBİNİN	
Adı Soyadı	Didem ENİSOĞLU
Kurumu / Üniversitesi	Orta Doğu Teknik Üniversitesi İlköğretim Fen ve Matematik Eğitimi Ana Bilim Dalı
Araştırma yapılacak iller/ilçeler	Çanakkale Gelibolu
Araştırma yapılacak eğitim kurumu ve kademesi	Ortaokul
Araştırmanın konusu	Yedinci Sınıf Öğrencilerinin Verilen Sütun Grafiğinde Aritmetik Ortalama, Ortanca, Teğe Değer Kavramlarını Yorumlarken Yaptıkları Hatalar Ve Kavram Yanılgıları
Üniversite / Kurum onayı	Var
Araştırma/proje/ödev/tez önerisi	Tez Önerisi
Veri toplama araçları	Anket
Görüş istenilecek Birim/Birimler	Öğrenciler
KOMİSYON GÖRÜŞÜ	
UYGUNDUR	
Komisyon kararı	Oybirliği ile alınmıştır.
Muhalif üyenin Adı ve Soyadı:	

KOMİSYON

10/02/2014

Komisyon Başkanı	Üye	Üye
Mehmet Atik EKİN	Zekiye KILIÇ	Raşit TURAN

T.C.
ÇANAKKALE VALİLİĞİ
İl Millî Eğitim Müdürlüğü

Sayı : 60305806/44/649674 13/02/2014
Konu: Anket Çalışması

MİLLİ EĞİTİM MÜDÜRLÜĞÜNE

İlgi : Orta Doğu Teknik Üniversitesinin 31/01/2014 tarihli ve 541 sayılı yazısı.

Orta Doğu Teknik Üniversitesi İlköğretim Fen ve Matematik Eğitimi Anabilim Dalı Yüksek Lisans Öğrencisi Didem ENİSOĞLU tarafından "Yedinci Sınıf Öğrencilerinin Verilen Sütun Grafiğinde Aritmetik Ortalama, Ortanca, Teğe Değer Kavramlarını Yorumlarken Yaptıkları Hatalar Ve Kavram Yanılgıları" başlıklı araştırma çalışması kapsamında, ekli listede adı bulunan Ortaokullarda okuyan öğrencilere yönelik anket uygulaması yapılması isteği ilgi yazısı ile teklif edilmektedir.

Söz konusu anket çalışması Müdürlüğümüz Anket-Araştırma İnceleme Komisyonunca incelenerek uygun görülmüştür.

Makamlarınızca da uygun görülmesi halinde olurlarınıza arz ederim.

Mehmet Atik EKİN
Şube Müdürü

OLUR
13/02/2014

Dr. Şaban KARATAŞ
Milli Eğitim Müdürü

Güvenli Elektronik İmzalı
Aslı İle Aynıdır
14/02/2014
Memur

Bu belge, 5070 sayılı Elektronik İmza Kanununun 5 inci maddesi gereğince güvenli elektronik imza ile imzalanmıştır.
Evrak teyidi http://evraksorgu.meb.gov.tr adresinden 432a-3d76-36bc-b9bc-d768 kodu ile yapılabilir.

Çanakkale İl Milli Eğitim Müdürlüğü Ek Binası
Strateji Geliştirme Bölümü Merkez/ÇANAKKALE
e-posta: istatistik17 *a* meb.gov.tr

Ayrıntılı bilgi için: Özlem Emine AYDIN V.H.K.İ.
Tel: (0 286) 217 46 93-130

APPENDIX C. STATISTICS ACHIEVEMENT QUESTIONNAIRE

Sevgili Öğrenciler,
Bu çalışmanın amacı sizlerin verilen sütun grafiğinde aritmetik ortalama, ortanca ve tepe değer kavramlarını nasıl yorumladığınızla ilgilidir. Sorulara vereceğiniz yanıtlar, bilimsel bir araştırmada kullanılacak ve gizli tutulacaktır. Lütfen soruları dikkatlice okuyarak eksiksiz yanıtlayınız.

Teşekkür ederim.

Didem ENİSOĞLU
Orta Doğu Teknik Üniversitesi

KİŞİSEL BİLGİLER

1.Adınız:…….

2. Okulunuz: ……

3. Sınıfınız : ☐ 7. Sınıf ☐ 8. Sınıf

4. Yaşınız: ☐10 ☐ 11 ☐12 ☐13 ve üstü

5. Cinsiyetiniz: ☐Kız ☐ Erkek

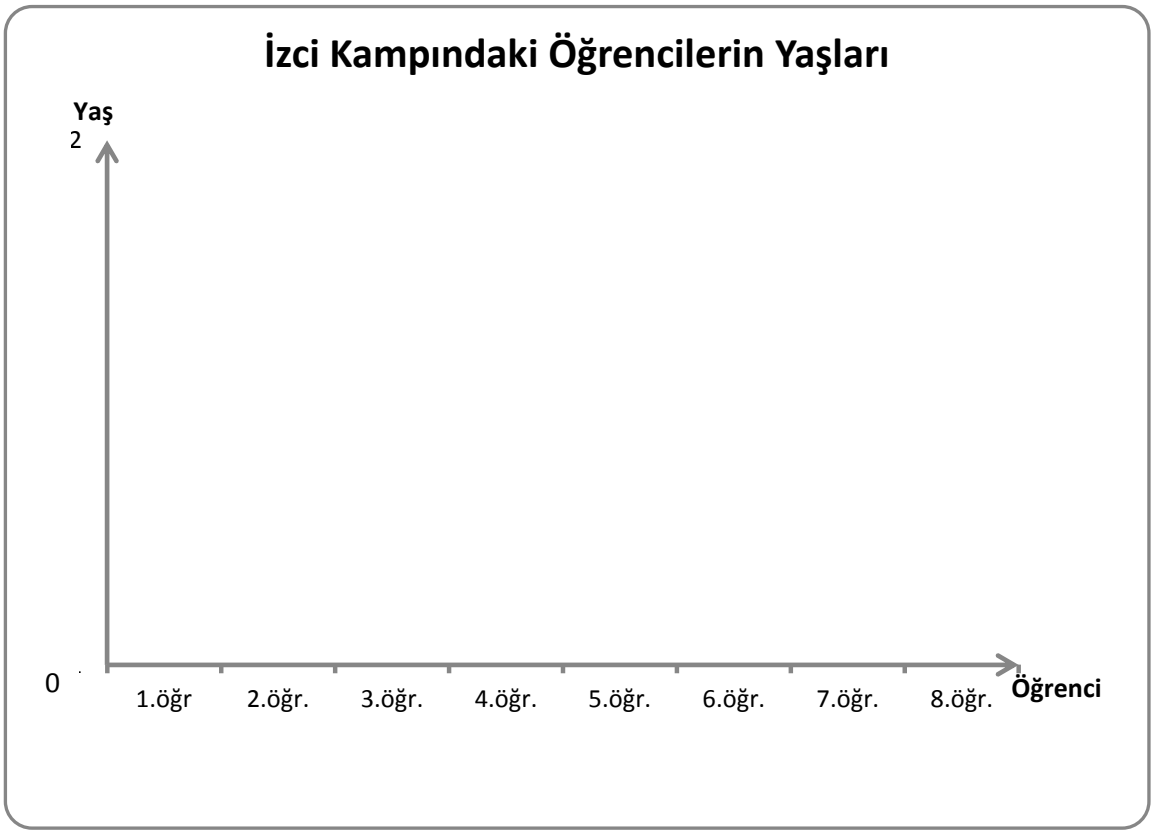
6. 1. Dönem Matematik Notunuz : ……..

İSTATİSTİK BAŞARI TESTİ

Bu test istatistikle ilgili 6 soru içermektedir. Bazı sorular bir ya da birkaç alt soru içermektedir. Bazılarında ise açıklama yapmanız istenmektedir. **Lütfen tüm soruları cevaplamaya çalışınız.**

Soru 1:

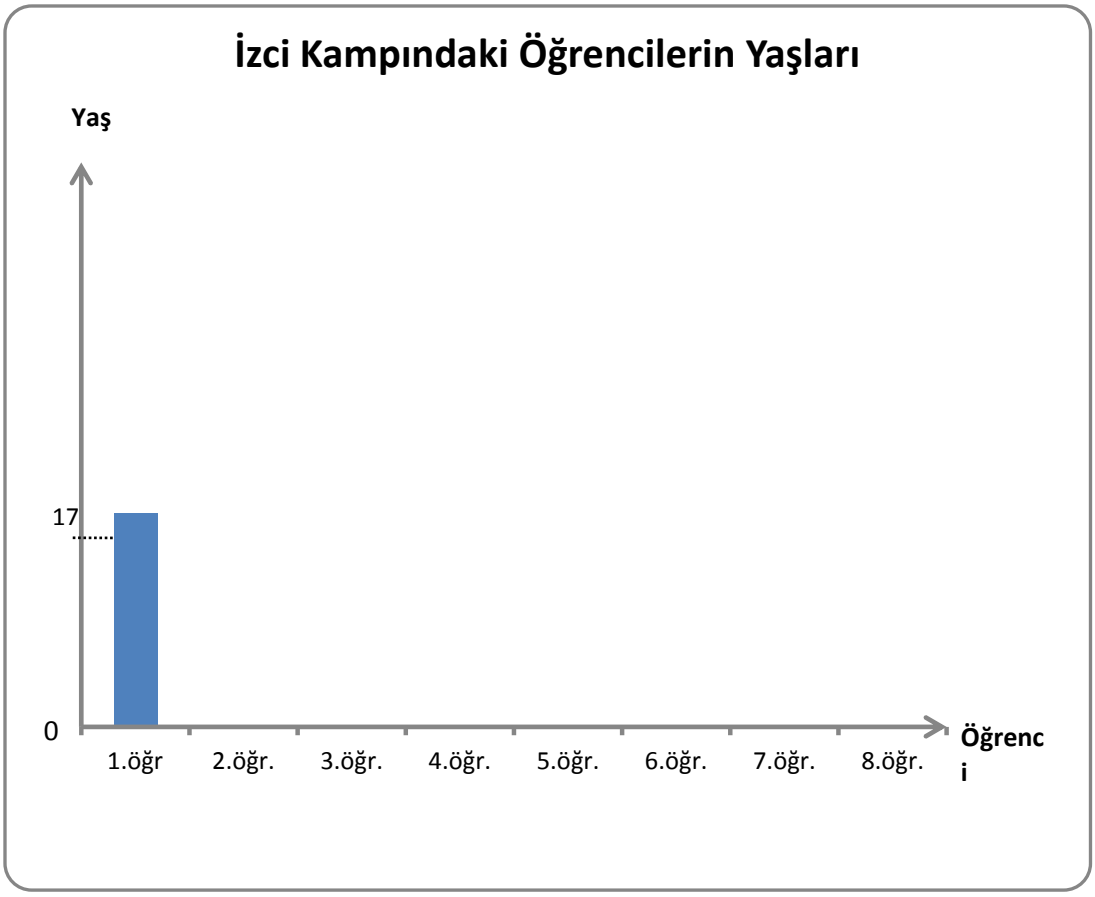
a) Bir izci kampında farklı yaşlarda 8 öğrenci vardır. Bu gruptaki öğrencilerin yaşlarının ortalaması 15'tir. Verilen bu bilgilere göre bu öğrencilerin yaşlarının kaç olabileceğini gösteren sütun grafiğini aşağıdaki alana çizerek örnek bir veri grubu oluşturunuz. (Not: Bu kamptaki öğrencilerin hiçbirinin yaşı 15 değildir.)



Cevabınızı nasıl oluşturduğunuzu açıklayınız.

……
…….....
.....
.....
.....

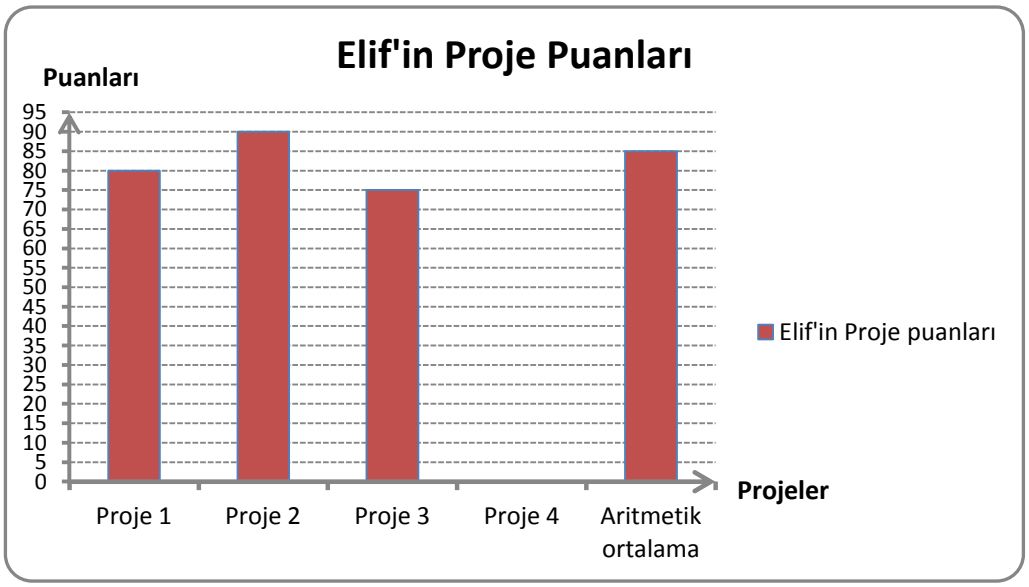
b) Yaş ortalaması 15 olan 8 kişilik izci kampında bir kişinin yaşı 17 ise; diğerlerinin yaşlarının kaç olabileceğini aşağıdaki alana sütun grafiği çizerek örnek bir veri grubunda gösteriniz. (Not: Bu kamptaki öğrencilerin hiçbirinin yaşı 15 değildir.)



Cevabınızı nasıl oluşturduğunuzu açıklayınız.

……
…….....
.....
.....
.....

Soru 2:



Elif matematik dersi için toplam 4 proje hazırlamıştır. Yukarıdaki sütun grafiğinde ilk üç sütun Elif'in matematik dersi için hazırladığı 3 projeden aldığı notları göstermektedir. Grafikte 4. projeden alınan not gösterilmemiştir. Son sütun ise aynı ders için hazırladığı dört projenin aritmetik ortalamasını göstermektedir. Bu grafiği dikkate alarak aşağıdaki soruları cevaplayınız.

a) Grafiğe göre aşağıdaki tabloyu doldurunuz

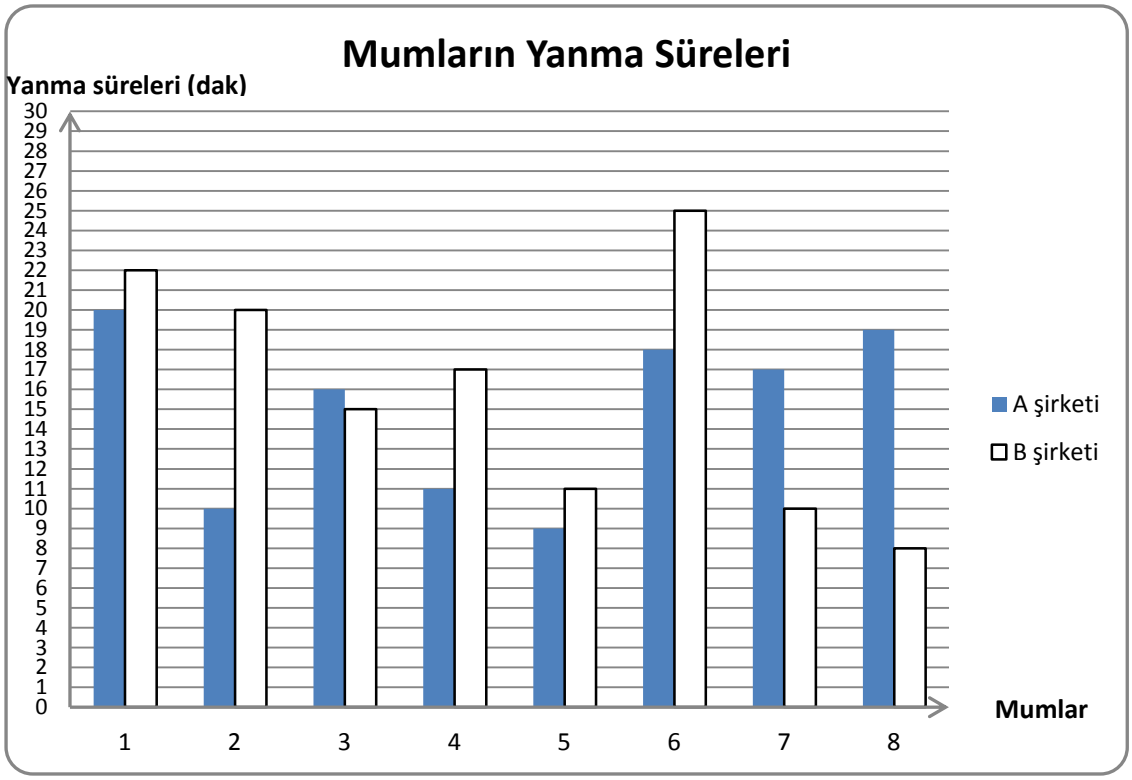
1.proje notu	
2. proje notu	
3. proje notu	
Proje notlarının aritmetik ortalaması	

b) Elif'in 4. projeden aldığı notu bulunuz. Cevabınızı nasıl bulduğunuzu açıklayınız.

c) Elif :"3. Proje notum 20 puan fazla olsaydı; projelerimin aritmetik ortalaması 4 puan artardı." Şeklinde bir açıklama yapmıştır. Elif' e katılıyor musunuz? Cevabınızın sebeplerini açıklayınız.

d) Elif'in proje notlarından hiç birini bilmeyip, sadece projelerin aritmetik ortalamasını bilseydik; Elif'in proje notları ile ilgili nasıl bir yorum yapabilirdik?

Soru 3: A şirketi ile B şirketinin ürettiği mumların yanma süreleri karşılaştırılmak isteniyor. Bunun için her iki şirketten de aynı boyutlarda 8 mum alınıyor ve her biri aynı anda yakılıyor. Şirketlerden seçilen mumların yanma süreleri aşağıdaki sütun grafiğinde gösterilmiştir.



a) Her iki şirkete ait mumların yanma sürelerinin aritmetik ortalamalarını bulunuz.

<u>Aritmetik Ortalama</u>

A Şirketi….....

B Şirketi……

b) Her iki şirkete ait mumların yanma sürelerinin ortancalarını (medyan) bulunuz.

<u>Ortanca (medyan)</u>

A Şirketi….....

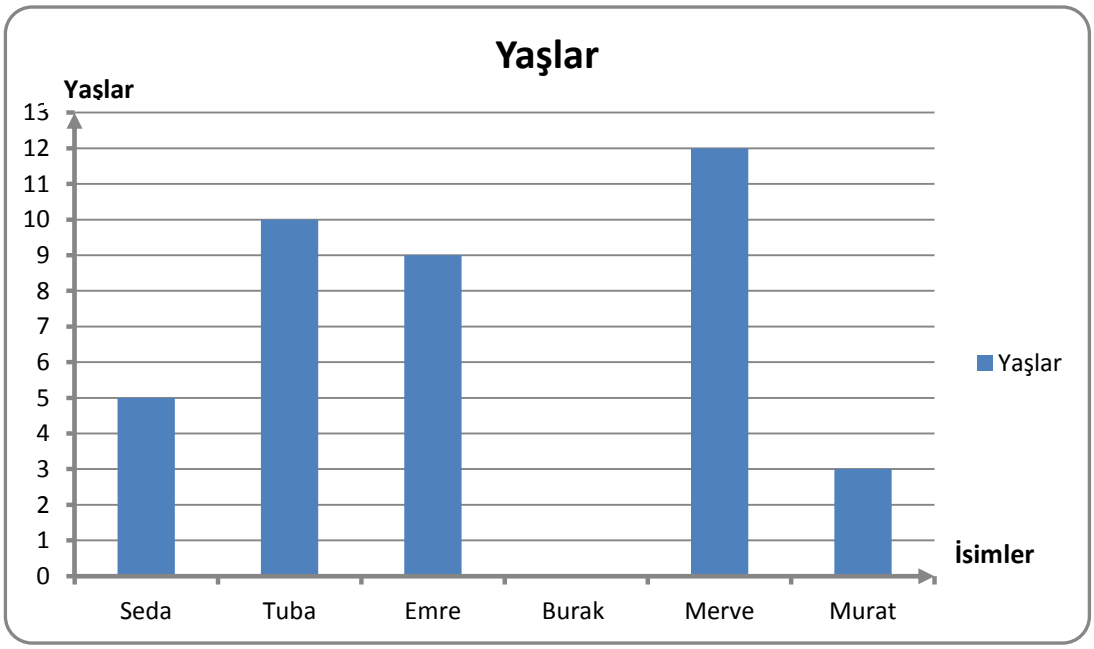
B Şirketi……

c) Ahmet: "A şirketinin mumları B şirketine göre daha uzun süre yanar." iddiasında bulunuyor. Sizce Ahmet iddiasında haklı mı?

…….....

- Bu iddianın doğru olup olmadığına nasıl karar verdiniz? Cevabınızı açık bir şekilde belirtiniz.

Soru 4:



Yukarıdaki sütun grafiğinde bir oyun parkında oynayan 6 çocuktan 5'inin yaşları gösterilmektedir. Aşağıdaki soruları yukarıdaki grafiğe göre cevaplandırınız.

a) Bu oyun parkında 9 yaşındaki çocukların sayısı diğerlerinden fazla olduğuna göre, Burak kaç yaşındadır? Cevabınızı nedenleri ile açıklayınız.

b) Bu çocukların yaşlarının ortalaması kaçtır?

c) Seda: "Murat parktan ayrıldığında parkta bulunanların yaş ortalaması azalır." iddiasında bulunuyor. Sizce Seda iddiasında haklı mı?

…….....

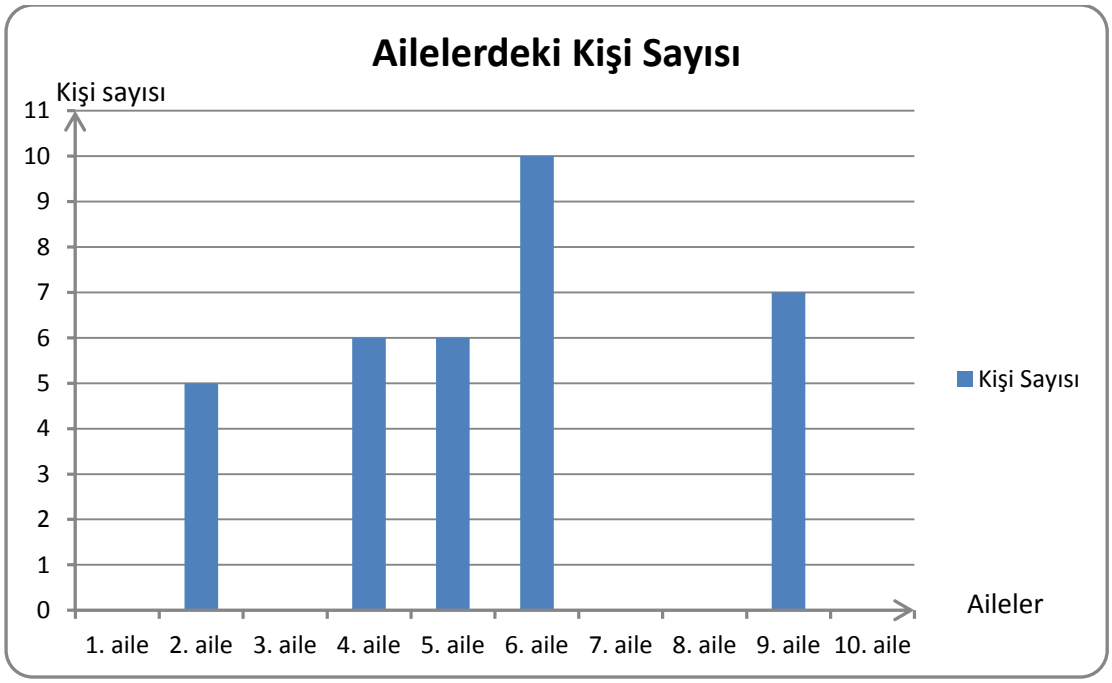
- Seda'nın iddiasının doğru olup olmadığına nasıl karar verdiniz?

Soru 5:

10 kişilik bir sınıftaki öğrencilerin ailelerindeki kişi sayısından oluşan veri grubu ile ilgili bilgiler aşağıda verilmiştir.

- Tepe değer (mod) = 6 kişiden oluşan aile
- Ortanca (medyan) = 6 kişiden oluşan aile
- Veri grubunda en az 3, en fazla 10 kişiden oluşan aileler vardır.

Aşağıdaki sütun grafiğinde bu sınıftaki bazı ailelerin kaç kişiden oluştuğu verilmiştir. Sizde yukarıda verilen bilgilere göre grafikte eksik olan aile kişi sayılarını belirleyerek örnek bir veri grubu oluşturunuz.

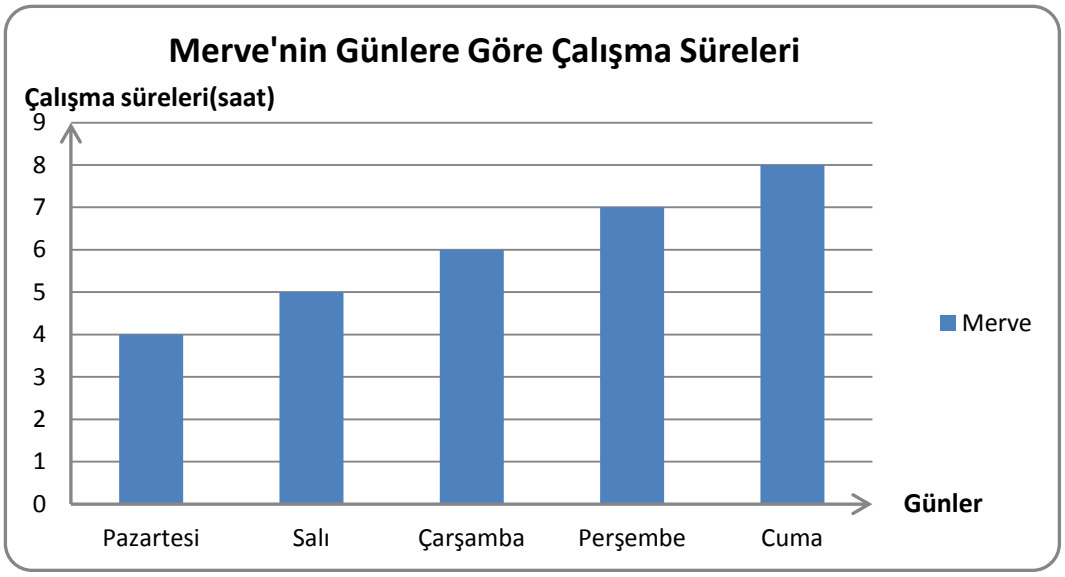


- Örnek veri grubunuzu nasıl oluşturduğunuzu açıklayınız.

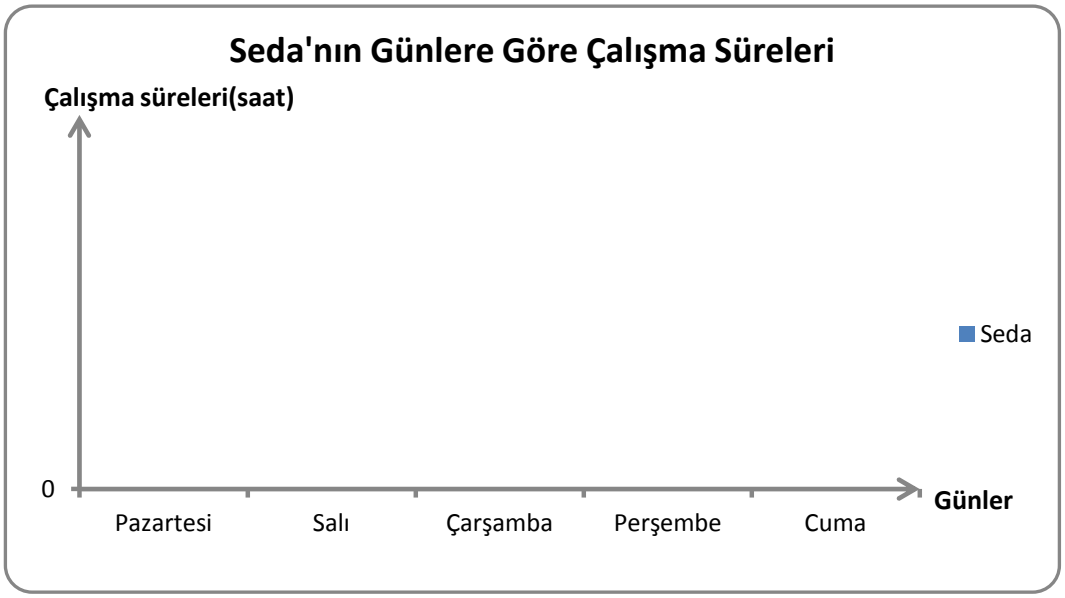
……
…….....
.....
.....
.....

Soru 6:

Merve ve Seda aynı işyerinin iki çalışanıdır. Aşağıdaki grafikte Merve'nin günlere göre çalışma süreleri gösterilmektedir.



Seda ile Merve'nin çalıştıkları günler ve çalışma sürelerinin aritmetik ortalaması aynı ise Seda'nın çalışma süreleri ne olabilir? Aşağıdaki alana grafik çizerek gösteriniz. (Seda ve Merve'nin aynı günlerde çalıştıkları süreler farklıdır.)



- Grafiği nasıl oluşturduğunuzu açıklayınız.

……....

.....

APPENDIX D. TURKISH SUMMARY

Giriş

Son yıllarda, uluslararası akımlardan dolayı istatistiğin matematik eğitimindeki önemi artmaktadır (Jacobbe & Carvalho, 2011). İstatistiğin günlük hayatın önemli bir parçası olması, artan önemin sebeplerinden biri olabilir. Diğer sebep ise istatistiğin tarih, fen ve coğrafya gibi alanlar arasında bağlantı kurmasıdır (Konold & Higgins, 2003).

İstatistiğin birçok alanla bağlantılı olmasından dolayı artan önemiyle birlikte, NCTM (National Council of Teachers of Mathematics, 2000) veri analizi ve olasılık konularını Amerika Birleşik Devletleri'nin matematik müfredatının 5 alanından biri olarak kabul etmiştir (Pratt, 2005). Benzer olarak da, Türkiye'de 1990 ve 1992 yıllarında matematik müfredatında yapılan değişikliklerle, istatistik konusuyla ilgili birçok kavram matematik müfredatına eklenmiştir. Ayrıca, bu yıllarda yapılan değişikliklerle, yeni ders kitaplarına merkezi eğilim ölçüleri ve yayılma ölçüleri gibi temel kavramlar eklenmiştir (Bulut, Ekici & İşeri, 1999). Buna ek olarak 2005 yılında yapılan değişikliklerle; öğrencilerin istatistik bilgilerini kullanabilecekleri günlük yaşam problemleri Türkiye matematik müfredatına eklenmiş, böylece günlük yaşam problemleri aracılığıyla öğrencilere istatistik ile ilgili yorum yapma ve karar verme becerilerinin kazandırılması amaçlanmıştır (MEB, 2005). Son olarak, 2013 yılında Türkiye müfredatında yapılan değişiklikle birlikte ortaokul matematik müfredatının her kademesine istatistik kavramlarının öğretimi eklenmiştir (MEB, 2013).

İstatistikteki veri analiz kısmının öneminden dolayı, eğitimde yapılan reform hareketleriyle öğrencilerin istatistik kavramlarını anlayışları ilgi konusu haline gelmiştir. Ancak, öğrencilerin aritmetik ortalama, ortanca ve tepe değer gibi istatistik kavramları ilgili hatalar yaptıkları, birçok zorlukla karşılaştıkları ve kavram yanılgılarına düştükleri

görülmüştür (Bright & Friel, 1998; Cai, 2000; Strauss & Bichler, 1988; Smith, diSessa, & Roschelle, 1993). Ayrıca, öğrencilerin ağırlıklı aritmetik ortalama hesaplarken zorluklarla karşılaştığı ifade edilmiştir (Pollatsek, Lima & Well, 1981). Cai (2000) öğrencilerin aritmetik ortalama ile ilgili karşılaştığı zorlukların asıl sebebinin öğrencilerin aritmetik ortalama ile ilgili kavramsal bilgilerinin eksikliği olduğunu ileri sürmüştür. Bu nedenle; öğrencilerin aritmetik ortalama, ortanca ve tepe değer kavramları ile ilgili işlemsel ve kavramsal bilgileri hakkında bilgi edinmek için bu konularla ilgili soruları çözerken yaptıkları hatalar incelenmelidir. Diğer taraftan, öğrenciler verileri ve grafikleri yorumlarken de hatalar yapmaktadır (Bright & Friel, 1998; Capraro, Kulm & Capraro, 2005; Tarr & Shaughnessy, 2007). Bundan dolayı, araştırmacılar öğrencilerin aritmetik ortalama, ortanca ve tepe değer kavramlarını hesaplarken hata yapabilecekleri ve öğrencilerin bu kavramları yanlış yorumlayabilecekleri konusunda hemfikirdir.

Öğrencilerin problemleri çözme stratejileri ile ilgili erişilebilir yazında bir çok çalışma varken (Becker, 1992; Cai, 1995; 1998), öğrencilerin istatistik kavramları ile ilgili soruları çözerken kullandıkları stratejiler ile ilgili sınırlı sayıda çalışma vardır (Cai, 2000). Öğrencilerin olası çözüm stratejilerini incelemek önemlidir çünkü eğer bir öğrenci aritmetik ortalama ile ilgili kavramsal bilgiye sahipse, bu öğrenci aritmetik ortalama ile ilgili soruları çözerken ortalama algoritmasını doğru bir şekilde uygulayabilir (Cai, 1998; Watson & Moritz, 2000).

Öğrencilerin aritmetik ortalama, ortanca ve tepe değer kavramlarıyla ilgili anlayışları önemli olmasına rağmen; Türkiye'deki öğrencilerin bu kavramlarla ile ilgili soruları çözerken kullandıkları stratejilerle ilgili çalışmalar, yaptıkları hatalar ve yanlış yorumlarla ilgili yapılan çalışmalar gibi yetersizdir (Uçar & Akdoğan, 2009).

Çalışmanın Amaçları

Bu çalışmanın amacı üç kısımdan oluşmaktadır: Çalışmanın birinci amacı, yedinci sınıf öğrencilerinin sütun grafiği gösteriminde verilen aritmetik ortalama, ortanca ve tepe değer kavramları ile ilgili soruları çözerken kullandıkları olası çözüm stratejilerini incelemektir. Çalışmanın ikinci amacı, yedinci sınıf öğrencilerinin sütun grafiği gösteriminde verilen aritmetik ortalama, ortanca ve tepe değer kavramları ile ilgili

soruları çözerken yaptıkları hataları incelemektir. Çalışmanın üçüncü amacı, öğrencilerin sütun grafiği gösteriminde verilen aritmetik ortalama, ortanca ve tepe değer kavramları ile ilgili sorularda yaptıkları olası yanlış yorumlarını incelemektir.

Çalışmanın Önemi

Matematik eğitiminde, aritmetik ortalama, ortanca ve tepe değer kavramlarının önemi, uygun durumlarda kullanımı vurgulanarak ABD, Avustralya ve İngiltere gibi bir çok ülke tarafından kabul edilmiştir (Watson & Moritz, 2009). Buna ek olarak, Türkiye matematik programındaki reformlarda, öğrencilerin istatistiksel bilgilerine verdikleri önemi göstermektedir (MEB, 2005; MEB, 2013).

Öğrenciler istatistiksel kavramlarla ilgili soruları çözerken farklı çözüm stratejileri kullanabilirler (Cai, 2000). Öğrencilerin kullandıkları bu çözüm stratejileri onların aritmetik ortalama, ortanca ve tepe değer kavramları ile ilgili izledikleri yol ve kavramsal bilgileri ile ilgili bilgi verebilir. Aslında, öğrencilerin çözüm stratejilerini incelemek onların ortalama kavramlarını anlayışları ile ilgili de bilgi verecektir (Mokros & Russell, 1995; Watson & Moritz, 2009). Diğer yandan, öğrencilerin çözüm stratejilerini analiz etmek onların hataları hakkında da bilgi verebilir. Daha önce de bahsedildiği gibi, yapılan çalışmalara göre merkezi eğilim ölçüleri öğrenciler için kolay anlaşılan kavramlar değildir (Smith, diSessa, & Roschelle, 1993; Strauss & Bichler, 1988) ve öğrenciler bu kavramlarla ile ilgili zorluklarla karşılaşmakta ve hatalar yapmaktadır (Garcia & Garrett, 2008). Bu sebeple, ortaokul öğrencilerinin çözüm stratejilerini incelemek önemlidir. Bu konuda yapılan çalışmaların Türk alan yazınında yetersiz olması ise bu çalışmanın önemini artırmaktadır.

Aritmetik ortalama, ortanca ve tepe değer kavramlarını yorumlama, araştırmaların sonuçlarını yorumlayabilmek için önemlidir. Ancak, öğrenciler bu kavramları yorumlarken kavram yanılgısına sahip olmaya eğilimlidir (Tarr & Shaughnessy, 2007). Bu sebepten dolayı, bu çalışma aritmetik ortalama, ortanca ve tepe değerle ilgili yorumları içerdiği için önemli görülmektedir. Ayrıca, ortalama kavramları ile ilgili yorumları öne çıkaran çalışmaların yazında yetersiz sayıda olduğu ise bu çalışmanın önemini artırmaktadır.

İstatistiksel kavramları anlamak öğrencilerin hem günlük hayatını hem de eğitim ve öğretim hayatını etkilediği için önemlidir (Konold & Higgins, 2003). Ayrıca, bu çalışmada kullanılan testteki sorularda, veri setleri sütun grafikleri üzerinde verilmiştir çünkü sütun grafikleri günlük hayatta sıklıkla karşılaşılan grafik çeşididir. Örneğin gazetelerde, analiz sonuçlarında ve Türkiye'deki yedinci sınıf matematik ders kitaplarındaki ortalama kavramlarıyla ilgili sorularda sütun grafikleri kullanılmaktadır. Ancak, öğrencilerin aritmetik ortalama, ortanca ve tepe değer kavramlarıyla ilgili sorularda kullandıkları çözüm stratejileri ve yaptıkları hatalar ile ilgili araştırmalar yapılmasına rağmen (Cai, 2000; Pollatsek, Lima, & Well, 1981; Smith, diSessa, & Roschelle, 1993), bu konuda sınırlı sayıda çalışma vardır (Cai, 2000; Uçar & Akdoğan, 2009). Ayrıca, Türk alan yazınında aritmetik ortalama, ortanca ve tepe değer kavramları ile ilgili yeterli sayıda çalışma yoktur. Bu durumlar çalışmanın önemini artırmaktadır.

Önemli Terimlerin Tanımları

Aritmetik Ortalama: Aritmetik ortalama, bir dağılımdaki bütün skorların ortalaması olarak tanımlanmıştır. Aritmetik ortalama, bir veri grubundaki bütün skorların toplamının veri sayısına bölünmesi ile bulunur. Aritmetik ortalama, toplanmış veri grubunun tanımlayıcılarından biridir (Fraenkel & Wallen, 2006).

Ortanca (Medyan): Ortanca (medyan) bir dağılımdaki skorların orta noktasıdır. Vverilerin yüzde 50'si bu noktanın yukarısında, yüzde 50'si ise aşağısındadır. Ortanca (medyan), toplanmış veri grubunun tanımlayıcılarından biridir ve ortanca (medyan) dağılımdaki terim sayısından etkilenir (Fraenkel & Wallen, 2006).

Tepe Değer (Mod) : Tepe değer (mod) istatistikteki ortalamayı bulmanın diğer yoludur ve bir dağılımdaki en sık tekrar eden skor olarak tanımlanmıştır. Bazı dağılımlar birden fazla tepe değere sahip olabilir, bunlar çift tepeli (bimodal) dağılımlar olarak adlandırılır.

Grafik: İki boyutlu bir yüzey üzerinde bir nokta, çizgi ya da alanın pozisyonu tarafından iletilen bilgi demektir (Fry, 1984).

Hata: Bir hata, kayma, falso, yanlışlık ya da kesinlikten sapmadır (Luneta & Makonye, 2010).

Yöntem

Evren ve Örneklem

Bu çalışmanın örneklemini Çanakkale'nin Gelibolu ilçesindeki devlet okullarına devam eden 233 yedinci sınıf öğrencisi oluşturmaktadır. Bu öğrenciler Çanakkale'nin Gelibolu ilçesindeki 2 devlet okulundan uygun örnekleme yöntemiyle seçilmiştir. Bu öğrencilerin temel karakteristikleri aşağıda Tablo 1'de verilmiştir.

Tablo 1 Çalışmanın Katılımcıları ve Temel Karakteristikleri

Sınıflar	Sayı	Yaş	Cinsiyet	
		Ortalama	Erkek	Kız
7-A	32	11.88	17 (53.1%)	15(46.9%)
7-B	33	11.94	15 (45.5%)	18 (54.5%)
7-C	32	11.84	18 (56.3%)	14 (43.7%)
7-A	29	12.03	17 (58.6%)	12 (41.4%)
7-B	27	11.96	16 (59.3%)	11 (40.7%)
7-C	27	12.11	14 (51.9%)	13 (48.1%)
7-D	25	12.08	15 (60.0%)	10 (40.0%)
7-E	28	11.89	15 (53.6%)	13 (46.4%)
Toplam	233	11.97	127 (54.5%)	106 (45.5%)

Araştırma Soruları

Bu çalışmanın üç tane araştırma sorusu vardır.

1. Yedinci sınıf öğrencilerinin sütun grafiği gösteriminde verilen aritmetik ortalama, ortanca ve tepe değer kavramları ile ilgili soruları çözerken kullandıkları olası çözüm stratejileri nelerdir?

2. Yedinci sınıf öğrencilerinin sütun grafiği gösteriminde verilen aritmetik ortalama, ortanca ve tepe değer kavramları ile ilgili sorularda yaptıkları olası hatalar nelerdir?

3. Yedinci sınıf öğrencilerinin sütun grafikleri gösteriminde verilen aritmetik ortalama, ortanca ve tepe değer kavramları ile ilgili sorularda yaptıkları olası yanlış yorumlar nelerdir?

Araştırma Yöntemi

Araştırmada tarama tipi araştırma yöntemi kullanılmıştır.

Veri Toplama Aracı

Çalışmanın verileri öğrencilerin başarı testine verdikleri cevaplar aracılığıyla toplanmıştır.

İstatistik Başarı Testi

Katılımcıların aritmetik ortalama, ortanca ve tepe değer kavramları ile ilgili soruları çözerken kullandıkları çözüm stratejilerini, yaptıkları hataları ve yanlış yorumları incelemek için veri gruplarına sütun grafikleri gösteriminde verilen bir başarı testi hazırlanmıştır. Bu başarı testi 6 adet açık uçlu sorudan oluşmaktadır. Bu sorulardan birinin tamamı, ikincisinin de bir kısmı alan yazınından adapte edilmiştir. Geriye kalan sorular ise araştırmacı tarafından geliştirilmiştir. Bu istatistik başarı testinin hazırlanma sürecinde yedinci sınıf ortaokul matematik programında yer alan kazanımlar göz önüne alınmış ve belirtke tablosu hazırlanmıştır. Hazırlanan belirtke tablosu aşağıda Tablo 2'de verilmiştir.

Tablo 2: İstatistik Başarı Testindeki Sorular ile İlgili İçerik Tablosu

	Kazanımlar
Aritmetik Ortalama	Öğrenciler verilerin aritmetik ortalamasını hesaplar. 1a, 1b, 2b, 2c, 3a, 4b, 6
	Öğrenciler verilerin aritmetik ortalamsını yorumlar. 1a, 1b, 2d, 3c, 4c, 6
Ortanca (Medyan)	Öğrenciler verilerin ortancasını hesaplar. 3b, 4b
	Öğrenciler verilerin ortancasını yorumlar. 3c, 5
Tepe değer (mod)	Öğrenciler verilerin tepe değerini hesaplar. 4a, 4b
	Öğrenciler verilerin tepe değerini yorumlar. 4a, 5

Başarı testi belirtke tablosuna göre hazırlandıktan sonra üç uzman görüşü alınmıştır ve 72 sekizinci sınıf öğrencisi ile pilot çalışma yapılmıştır. Bu pilot çalışma sonucuna göre, verilen 45 dakika sürenin yetersiz olduğu sonucuna varılmış ve testin tamamlanması için uygun olan sürenin 60 dakika olmasına karar verilmiştir. Bunlara ek olarak, 35 yedinci sınıf öğrencisinin cevapları matematik eğitimi yüksek lisans öğrencisi olan ikinci kişi tarafından değerlendirilmiştir. Araştırmacı ve bu kişinin verdiği skorlar arasındaki korelasyon %98'dir.

Veri Toplama Süreci

Çalışmanın verileri 2013-2014 eğitim öğretim yılının bahar döneminde toplanmışlar. Veriler toplanmadan önce gerekli etik izinler alınmıştır. Aralık ayında pilot çalışma

yapılmış, Mart ayında ise çalışmanın asıl verileri toplanmıştır. Asıl verilerin toplanma aşamasında, İstatistik Başarı Testi iki okulda ayrı zamanlarda yapılmıştır. Okulların birinde 5 sınıfta da eş zamanlı uygulama yapılmıştır. Diğerinde de 3 sınıfta eş zamanlı uygulama yapılmıştır. Her iki okulda da araştırmacı birer sınıfa uygulama yapmıştır. Diğer sınıflara da farklı branşlardan öğretmenler uygulama yapmıştır. Diğer sınıflarda uygulama yapacak öğretmenler, uygulama öncesi araştırma ile ilgili bilgilendirilmişlerdir.

Veri Analizi

Öğrencilerin sütun grafikleri gösteriminde verilen aritmetik ortalama, ortanca ve tepe değer kavramları ile ilgili sorularda kullandıkları çözüm stratejilerinin, yaptıkları hataların ve yanlış yorumların belirlenmesi için öğrencilerin başarı testine verdiği cevaplar derinlemesine incelenmiştir.

Öncelikli olarak öğrencilerin başarı testindeki sorulara verdiği cevaplara göre belirlenen dereceli puanlama anahtarında, her problem için verilen cevaplar 0 ile 4 arası kodlanmıştır. Genel olarak, kabul edilebilir çözüm stratejisi içeren doğru cevaplar 4; kabul edilebilir doğru çözüm stratejisi içeren yarım cevaplar 3; açıklama içermeyen doğru cevaplar 2; kabul edilemeyen çözüm stratejileri içeren yanlış çözümler 1; alakasız cevaplar veya boş bırakılan sorular 0 olarak kodlanmıştır. Aritmetik ortalama, ortanca ve tepe değer kavramlarıyla ilgili soruları çözerken kullanılan stratejileri belirlemek için 4 kodu verilen cevaplar derinlemesine incelenmiştir. Öncelikle her soruyu için kaç kişinin 4 kodu aldığı belirlenmiştir. Daha sonra, her kavramla ilgili çözüm stratejilerini belirlemek için kavramlarla ilgili sütun grafikleri gösteriminde verilen sorular kendi içinde değerlendirilmiştir. Aritmetik ortalama, ortanca ve tepe değer kavramlarıyla ilgili her soru için yaptıkları hataları ve yanlış yorumları belirlemek için ise her soru için 1 kodu verilen katılımcı sayısı belirlenmiş ve bu katılımcıların cevapları derinlemesine incelenerek her kavramla ilgili yapılan hatalar ve yanlış yorumlar belirlenmiştir.

Araştırmanın Varsayımları ve Sınırlılıkları

Araştırmanın ilk varsayımı öğrencilerin İstatistik Başarı Testi'ni cevaplarken dikkatli, içten ve açık yürekli olduğudur. Ayrıca, öğrencilerin yaşlarının, zeka düzeylerinin ve sosyoekonomik geçmişinin benzer olduğu varsayılmıştır.

Çalışmanın katılımcılarının bulabildiğini örnekleme yoluyla seçilmesi sonuçların daha geniş bir popülasyona genellenmesini sınırlandırmaktadır. Ayrıca, elde edilen sonuçlar İstatistik Başarı Testi'ndeki sorularla sınırlıdır çünkü farklı sorular için farklı bulgular elde edilebilir. Bunlara ek olarak, bulguların elde edilmesinde öğrencilerin soruların çözümleriyle ilgili yaptığı açıklamalar önemli bir yere sahip olduğu için, bulgular öğrencilerin kendini ifade edebilme yeteneğiyle sınırlıdır.

Bulgular ve Tartışma

Bu çalışmanın üç amacı bulunmaktadır. Katılımcıların aritmetik ortalama, ortanca ve tepe değer kavramları ile ilgili sütun grafiği gösteriminde verilen soruları çözerken kullandıkları çözüm stratejilerinin belirlenmesi birinci amaçtır. İkinci amaç ise; katılımcıların aritmetik ortalama, ortanca tepe değer kavramları ile ilgili sütun grafiği gösteriminde verilen soruları çözerken yaptıkları hataların incelenmesidir. Üçüncü ve son amaç ise; öğrencilerin bu kavramlarla ilgili yaptıkları yanlış yorumların belirlenmesidir.

Çalışmanın bulguları öğrencilerin aritmetik ortalama, ortanca ve tepe değer kavramlarından her biriyle ilgili soruları çözerken üç farklı strateji kullandıklarını göstermiştir. Aritmetik ortalama ile ilgili denge modelini kullanma, ortalama formülünü kullanma ve tahmin ve tahmin ve kontrol stratejileri belirlenmiştir. Ortalama formülünü kullanma stratejisi öğrencilerin aritmetik ortalamayı bulurken veya verilen aritmetik ortalamaya göre olası veri grubu oluştururken bütün verilenlerin toplanıp veri sayısına bölme algoritmasını kullanmasıdır. Bu stratejiyi kullanan katılımcı sayısının diğerlerini kullananlardan fazla olduğu görülmüştür. Bu bulgu geçmiş çalışmaların sonuçları ile de tutarlılık göstermektedir (Cai, 1998; 2000). Bu stratejinin diğer stratejilerden daha fazla kullanılmasının sebebi katılımcıların bu stratejiyi matematik derslerinde kullanmış olmaları olabilir. Ayrıca bu durumun diğer sebebi de katılımcıların kendi matematik ders

kitaplarında bu stratejinin aritmetik ortalama ile ilgili soruların çözümünde kullanılması olabilir. Diğer yandan, bazı öğrenciler özellikle veri grubu oluştururken denge modelini kullanma ve tahmin ve kontrol stratejilerini kullanmıştır. Tahmin ve kontrol stratejisi Cai (1998; 2000)'in çalışmalarında da gözlemlenmiştir. Öğrencilerin tahmin ve kontrol stratejisini kullanmasının sebebi daha önceden matematikle ilgili başka konularda kullanmış olmaları olabilir. Denge modeli ise öğrencilerin aritmetik ortalamayı bir veri grubunun denge noktası kabul ederek veri grubu oluşturması veya eksik veriyi bu denge noktasına göre bulmasıdır. Yani, verilerin aritmetik ortalaması bu denge noktası olacak şekilde ikili gruplara ayırarak eksik veriyi bulmasıdır veya olası veri grubu oluşturmasıdır. Denge modelini kullanımı ise Mokros and Russell'in (1995) çalışmasında da gözlemlenmiştir. Ayrıca, diğer çalışmalarda da bu stratejiye benzer bir strateji gözlemlenmiştir (Cai,1998; 2000). Bu stratejinin kullanılmasının sebebi öğrencilerin aritmetik ortalama ile ilgili kavramsal bilgiye yeterli düzeyde sahip olması olabilir (Cai, 2000). Buna ek olarak, öğrenciler ortanca ile ilgili soruları çözerken de 3 farklı strateji kullanmışlardır. Bunlar sayısal yöntemleri kullanma, grafik temsillerine bağlı olma ve tahmin ve kontrol stratejileridir. Katılımcılar çoğunlukla sayısal yöntemleri kullanma stratejisini kullanmışlardır. Bu yöntem sütun grafiği gösteriminde verilen veri grubunu sıralayıp ortancanın bulunmasıdır. Öğrencilerin derste bu strateji sıklıkla kullanmaları araştırma sorularını da çözerken tercih etmelerinin sebebi olabilir. Son olarak, tepe değer kavramıyla ilgili de üç strateji kullanılmıştır. Bu stratejiler de sayısal yöntemleri kullanma, grafik temsillerine bağlı olma ve tahmin ve kontrol stratejileridir. Tepe değer soruları çözülürken, öğrencilerden büyük bir kısmı grafik temsillerine bağlı olma stratejisini kullanmıştır. Bunun sebebi de verilerin grafik üzerinde verilmiş olması olabilir.

Bu çalışmada, öğrencilerin aritmetik ortalama, ortanca ve tepe değer ile ilgili soruları çözerken yaptıkları hatalar da incelenmiştir. Katılımcıların bu üç kavramla ilgili yaptığı hatalar üç ana grup altında toplanmıştır. Bu ana gruplar: Algoritmaya dayanan hatalar, kavramsal bilgiye dayanan hatalar, ve diğer hatalardır. Bulgular göstermiştir ki aritmetik ortalama kavramı için belirlenen hatalar; işlem hatası, toplam sayıyı bulma, ortalama algoritmasını yanlış kullanma, daha küçük veya daha büyük sayılara yönelme, problemin bütün gerekliliklerini yerine getirmeme, eksik veri grubu oluşturma ve

grafikten değerleri yanlış okumadır. Katılımcıların tarafından en çok yapılan hata ortalama algoritmasını yanlış kullanmadır. Yani, öğrenciler aritmetik ortalamayı algoritma kullanarak bulurken sorun yaşamaktadırlar. Bu bulgu literatürdeki diğer çalışmalarla da tutarlıdır (Cai, 1998; 2000; Maverach, 1983; Pollatsek vd., 1981; Watson & Moritz, 1999). Bu hatanın sebebi öğrencilerin aritmetik ortalama ile ilgili kavramsal bilgisindeki eksik olabilir. Çünkü eğer öğrenciler aritmetik ortalamanın seviyeleri eşitleme yorumunu anlamış olsaydı, bu hatayı yapmayabilirlerdi. Ayrıca İstatistik Başarı Testi'ndeki sorular katılımcıların alışık olmadığı sorular olduğu için bu hata yapılmış olabilir.

Katılımcılar tarafından ikinci sıklıkla yapılan hata daha küçük veya daha büyük sayılara yönelmedir. Bu hatayı yapan öğrenciler, verilen aritmetik ortalama için uygun veri grubu oluştururken ya tüm verileri aritmetik ortalamadan küçük ya da tüm verileri aritmetik ortalamadan büyük seçmektedir. Bu hatanın da temel sebebi öğrencilerin kavramsal bilgilerindeki eksik olabilir. Çünkü eğer öğrenci aritmetik ortalamayı denge noktası olarak yorumlamak yerine bir veri grubunun en büyük veya en küçük değeri olarak yorumlarsa bu hatayı yapabilir. Verilen aritmetik ortalama için olası veri grubu oluşturmak alışık olmadıkları bir soru tipi olabilir. Ancak, kavramsal bilgisi tam olan öğrenciler, alışık olmadıkları soru tipini de çözebilir (Hiebert & Carpenter, 1992; Watson & Moritz, 2009).

Diğer yandan, katılımcıların ortanca kavramıyla ilgili soruları çözerken de hata yaptığı gözlemlenmiştir. Araştırmanın bulgularına göre ortanca kavramı için belirlenen hatalar; işlem hatası, sıralanmamış veri grubu üzerinden yanlış karar, sıralanmış veri grubu üzerinden yanlış karar, grafiği yanlış yorumlama, eksik veri grubu oluşturma, grafikten değerleri yanlış okuma ve en büyük ve/veya en küçük değeri yanlış tespit etmedir. Bu hatalar içinde en çok gözlemlenen hata grafiği yanlış yorumlama olmuştur. Bu hatanın sebebi öğrencilerin verilen sütun grafiği üzerinden direkt veri grubunun ortancasını bulmaya çalışması olabilir. Yani, veriler grafikte sıralı değildir ve öğrenciler direkt verilen grafik üzerindeki sekiz verinin ortancasını bulmak için dördüncü ve beşinci verinin ortalamasını bulmuştur. Bu durumda grafiği yanlış yorumlama hatası yapılmıştır. Belki de veri grupları sütun grafiği üzerinde verilmeseydi, öğrenciler bu hatayı

yapmayacaktı. Ancak, bu hatayı yapan öğrenciler verilerin sıralı olması gerektiğini düşünmüyor da olabilir.

Bunlara ek olarak, katılımcıların tepe değer ile ilgili sorularda da hata yaptığı gözlemlenmiştir. Araştırmanın bulgularına göre tepe değer kavramı ile ilgili sorularda yapılan hatalar: ortalama algoritmasını yanlış yerde kullanma, eksik veri grubu oluşturma ve en büyük ve/veya en küçük değeri yanlış tespit etme. Katılımcılardan bir kısmı tepe değeri verilen bir veri grubunda eksik değeri bulurken aritmetik ortalama algoritmasını yanlış yerde kullanmıştır ve yapılan bu hata ortalama algoritmasını yanlış yerde kullanma olarak adlandırılmıştır. Öğrencilerin bu hatayı yapmasının sebebi, daha önce bir veri grubundaki eksik değerin sorulduğu sorularda aritmetik ortalama algoritmasını kullanmış olmaları olabilir. Ancak, eğer öğrenciler tepe değer ve aritmetik ortalama kavramları ile ilgili yeterli kavramsal bilgiye sahip olsaydı, bu hatayı yapmamış olabilirdi (Cai, 2000). Diğer yandan, bazı katılımcılarda eksik veri grubu oluşturma ve en büyük ve/veya en küçük değeri yanlış tespit etme hatalarını yapmışlardır. Ancak, bu hatalar olarak tepe değer kavramından bağımsızdır. Ayrıca, bu hatalar aynı şekilde aritmetik ortalama ve ortanca ile ilgili sorularda da yapılmıştır.

Katılımcıların tamamı ortalama sorulduğunda aritmetik ortalamaya yönelmiştir. Bu bulgu da diğer çalışmalarla tutarlıdır (Randall, 2006; Uçar & Akdoğan, 2009). Bu hatanın sebebi aritmetik ortalama ifadesinin ortalama kelimesini içermesi olabilir veya bazı iyi hazırlanmamış kitapların ortalamayı sorarak aritmetik ortalamayı buldurması olabilir. Ayrıca, katılımcılar bir veri grubuna ait ortalama ve bir değer hariç bütün veri grubunu verdiğinde, eksik değeri bulurken zorlanmaktadırlar.

Bu bulgulara ek olarak, 2-d, 3-c, ve 4-c soruları ortalama kavramları ile ilgili yorumlama gerektirdiğinden, öğrencilerin yaptığı yanlış yorumlar bu sorulara verdikleri yanlış cevaplar incelenerek belirlenmiştir. Öğrencilere soru 2-d'de bir veri grubundaki verilerin hiç birini bilmeyip sadece bu veri grubunun aritmetik ortalamasını bilseydik, verilerle ilgili nasıl yorum yapılabilirdi diye soruldu. Öğrencilerin bu soruya verdiği cevaplar incelendiğinde öğrencilerin bir kısmı ortalama algoritmasını yanlış kullanma hatasını yaparak yanlış bir değer bulup, bu değer üzerinden yanlış yorum yapmıştır. Bunun yanı sıra, bu yanlış yorumun sebebi öğrencilerin aritmetik ortalama ile ilgili bilgi eksiklikleri

olabilir çünkü eğer öğrenciler aritmetik ortalamayı bir veri grubunun temsilcisi olarak görseydi, bu yanlış yorumu yapmamış olabilirdi. Ayrıca, öğrencilerden bir kısmı değerlerin tamamının aritmetik ortalamadan küçük olması gerektiğini söylerken, bir kısmı da tüm değerlerin aritmetik ortalamadan büyük olması gerektiğini söylemiştir. Bu yanlış yorumu yapan öğrenciler aritmetik ortalamayı denge noktası olarak görseydi bu hatayı yapmayabilirdi (Konold & Pollatsek, 2002; Mokros & Russell, 1995). Çünkü eğer öğrenciler aritmetik ortalamayı denge noktası olarak algılarsa, bazı verilerin aritmetik ortalamadan küçük bazı verilerin ise aritmetik ortalamadan büyük olması gerektiğini anlayabilir. Buna ek olarak soru 3-c'de öğrencilerden sütun grafiği gösteriminde verilen iki veri grubunu karşılaştırması istenmiştir. Önceki çalışmalar göstermiştir ki öğrenciler iki veri grubunu karşılaştırırken sorun yaşamaktadır (Shaughnessy, 1992; Watson & Moritz, 1999). Ancak, katılımcılardan bazıları veri gruplarındaki bir veriye göre karşılaştırma yapmışlardır. Bu öğrenciler ortalama kavramlarını bir veri grubunu temsil eden değerler olarak görmüyor olabilir. Ayrıca, bazı öğrenciler verilen veri gruplarını, ortancayı kullanarak karşılaştırmış ve yanlış karar vermişlerdir. Verdikleri yanlış kararın sebebi ise yanlış ortalamayı tercih etmeleridir çünkü ortanca veri grubunda uç değerler olduğu zaman veri grubunu özetleyen değerdir (Fraenkel & Wallen, 2006) Son olarak soru 4-d de bir veri grubundan bir değer çıkarıldığında ortalamanın bundan nasıl etkilendiği sorulmuştur. Sorudaki çıkarılan bu değer, veri grubunun aritmetik ortalamasından küçüktür. Katılımcılardan büyük bir kısmı, bir veri grubundan bir değer çıkarıldığında aritmetik ortalamanın her zaman azalacağını söylemiştir. Katılımcıların yaptıkları bu yanlış yorumun sebebi aritmetik ortalamayı denge noktası olarak görmemeleri olabilir. Çünkü eğer bir öğrenci aritmetik ortalamayı denge noktası olarak algılarsa, aritmetik ortalamadan küçük bir değer veri grubundan çıkarıldığında aritmetik ortalamanın artacağını düşünebilir.

Doğurgular

Bu çalışmanın sonuçları matematik öğretmenleri, öğretmen eğiticileri, program geliştiriciler ve ders kitabı yazarları için önemli bilgiler sunmaktadır.

Bu çalışmanın bulguları göstermiştir ki; ortaokul yedinci sınıf öğrencileri sütun grafiği gösteriminde verilen aritmetik ortalama, ortanca ve tepe değer kavramlarından her

biriyle ilgili soruları çözerken üç farklı strateji kullanmışlardır. Ayrıca, bu öğrenciler sütun grafiği gösteriminde verilen aritmetik ortalama, ortanca ve tepe değerle ilgili soruları çözerken hatalar ve yanlış yorumlar yapmışlardır. Öğretmenler, öğretmen eğiticileri ve ders kitabı yazarları bu öğrencilerin çözüm stratejilerini, yaptıkları hataları ve yanlış yorumları dikkate alarak daha etkili öğrenme ortamı hazırlamaları önerilmektedir.

Öğrencilerin bu kavramlarla ilgili yaptıkları hatalar ve yanlış yorumlarla ilgili bilgilendirme yapmak için okullardaki öğretmenlere ve öğretmen adaylarına yönelik seminerler organize edilebilir. Böylece öğretmenler öğrencilerin hatalarını ve yanlış yorumlarını fark ederek uygun ders planı hazırlayabilir ve öğrencilerin hatalarını ortadan kaldırarak, yanlış yorumları engelleyebilir. Aslında, öğrencilerin yaptıkları hataların en büyük sebebi kavramsal bilgilerindeki eksikliktir. Öğrencilere belli kalıplarda sorular sormak da öğrencilerin kavramsal gelişiminin eksik kalmasına sebep olabilir. Bu nedenle, öğrencilerin bu kavramlarla ilgili çok yönlü düşünmelerini sağlayacak sorular sorulması önerilmektedir.

Öneriler

Aynı çalışma olasılıklı örnekleme yöntemlerinden biriyle seçilerek tekrarlanabilir. Ayrıca, aynı çalışma özel okulda okuyan yedinci sınıf öğrencilerine de yapılabilir.

Çalışmanın yönteminde değişiklik yapılarak da farklı çalışmalar yapılabilir. İlk olarak, boylamsal bir çalışma yürütülebilir. İkinci olarak, merkezi eğilim ölçüleriyle ilgili testler geliştirilerek, bu çalışmayla aynı amacı taşıyan başka bir çalışma yürütülebilir. Son olarak, öğrencilerin yaptıkları hataların ve yanlış yorumların sebepleri ve öğrencilerin bu kavramlarla ilgili kavram yanılgıları incelenebilir.

Kaynakça

Bright, G. W. & Friel, S. N. (1998). Graphical representations: Helping students interpret data. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K–12* (pp. 63–88). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bulut, S., Ekici, C. & Iseri, A. I. (1999). Bazı olasılık kavramlarının gelişimi için çalışma yapraklarının geliştirilmesi. *Hacettepe University Journal of Education, 15,* 129-136.

Cai, J. (1995). Beyond the computational algorithm: Students' understanding of the arithmetic average concept. In L. Meria & D. Carraher (Eds.), *Proceeding of the 19th Psychology of Mathematics Education Conference* (Vol. 3. pp. 144-151).

Cai, J. (1998). Exploring students' conceptual understanding of the averaging algorithm. *School Science and Mathematics, 98*(2), 93-98.

Cai, J. (2000). Understanding and representing the arithmetic averaging algorithm: an analysis and comparison of US and Chinese students' responses. *International Journal of Mathematical Education in Science and Technology, 31*(6), 839-855.

Capraro, M. M., Kulm, G. & Capraro, R. M. (2005). Middle grades: Misconceptions in statistics thinking. *Journal of Schools Science and Mathematics, 105*(4), 165-174.

Jacobbe, T., & Carvalho, C. (2011). Teachers' understanding of averages. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics – Challenges for teaching and teacher education* (pp. 199–209). Dordrecht: Springer.

Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education*. New York, the USA: McGraw-Hill.

Hiebert, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (ed.), *Handbook of Research on Mathematics Teaching and Learning* (New York: Macmillan), pp. 65–97.

Konold, C & Higgins, T. L. (2003), "Reasoning about data", In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics*, Reston, VA: National Council of Teachers of Mathematics, pp.193-215.

Konold, C. & Pollatsek, A. (2002). Data analysis as a research for signals in noisy processes. *Journal for Research in Mathematics Education, 33*, 259-289.

Mevarech, Z. (1983). A deep structure model of students' statistical misconceptions. *Educational Studies in Mathematics, 14*, 415-429.

Milli Eğitim Bakanlığı [MEB], 2005. *İlkögretim matematik dersi ögretim programı 6-8. sınıflar: Ögretim programı ve kılavuzu* Ankara,Turkey.

Milli Eğitim Bakanlığı [MEB], 2013. *Ortaokul Matematik Dersi (5, 6, 7 ve 8. Sınıflar) Öğretim Programı* Ankara: MEB Yayıncılık.

Mokros, J. & Russell, S. J. (1995). Children's concepts of averages and representativeness. *Journal for Research in Mathematics Education, 26*, 20-39

National Council of Teacher of Mathematics (2000). *Principles and standard for school mathematics*. Reston, VA: Author.

Pollatsek, A., Lima, S., & Well, A. D. (1981). Concept or computation: Students' understanding of the mean. *Educational Studies in Mathematics, 12*, 191-204.

Pratt, D. (2005). Probability in teacher education and development. In G. A. Jones (Ed.), *Exploring probability in school* (pp. 171-190). New York: Springer Science+Business Media, Inc.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In Grouws, D.A.(Ed.). *Handbook of research on mathematics teaching and learning* (pp. 465-494). New York: Macmillan.

Smith, J. P., diSessa A. A., & Roschelle, J. (1993). Misconceptions Reconceived: A Constructivist analysis of Knowledge in Transition. *The Journal of Learning Sciences, 3(2)*, 115-163.Strauss, S., & Bichler, E. (1988). The development of children's concept of arithmetic average. *Journal for Research in Mathematics Education, 19*, 64-80.

Tarr, J. E. & Shaughnessy, J. M. (2007). *Data and Chance. Results from the 2003 National Asssessment of Educational Process*. Reston, VA: NCTM.

Uçar, Z. T. & Akdoğan, E. N. (2009). Middle school students' understanding of average. *Elementary Education Online, 8(*2), 391-400.

Watson, J. M., & Moritz, J. B. (1999). Interpreting and predicting from bar graphs. *Australian Journal of Early Childhood, 24*(2), 22-27.

Watson, J. M. & Moritz, J. B. (2001). Development of reasoning associated with pictographs: Representing, interpreting, and predicting. *Educational Studies in Mathematics, 48*, 47-81.

Watson, J. M. & Moritz, J. B. (2009). The longitudinal development of understanding of average. *Mathematical Thinking and Learning, 2*, 11-50.

APPENDIX E. TEZ FOTOKOPİSİ İZİN FORMU

ENSTİTÜ

Fen Bilimleri Enstitüsü	
Sosyal Bilimler Enstitüsü	X
Uygulamalı Matematik Enstitüsü	
Enformatik Enstitüsü	
Deniz Bilimleri Enstitüsü	

YAZARIN

Soyadı : Enisoğlu
Adı : Didem
Bölümü : İlköğretim Fen ve Matematik Eğitimi

TEZİN ADI (İngilizce) : Seventh Grade Students' Possible Solution Strategies, Errors and Misinterpretations Regarding the Concepts of Mean, Median and Mode Given in Bar Graph Representations

TEZİN TÜRÜ : Yüksek Lisans [X] Doktora [ ]

1. Tezimin tamamından kaynak gösterilmek şartıyla fotokopi alınabilir. [X]
2. Tezimin içindekiler sayfası, özet, indeks sayfalarından ve/veya bir bölümünden kaynak gösterilmek şartıyla fotokopi alınabilir. [ ]
3. Tezimden bir bir (1) yıl süreyle fotokopi alınamaz. [ ]

TEZİN KÜTÜPHANEYE TESLİM TARİHİ: