AN INVESTIGATION OF 7TH GRADE STUDENTS' STATISTICAL LITERACY ABOUT THE CONCEPTS OF AVERAGE AND VARIATION ON BAR AND LINE GRAPHS

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

BY

EMİNE ÇATMAN AKSOY

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

SEPTEMBER 2018

Approval of the Graduate School of Social Sciences

Prof. Dr. Tülin GENÇÖZ
Director

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

Assoc. Prof. Dr. Elvan ŞAHİN
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.

Prof. Dr. Mine IŞIKSAL BOSTAN
Supervisor

Examining Committee Members

Assoc. Prof. Dr. Elif YETKİN ÖZDEMİR	(Hacettepe Uni., MFBE)	
Prof. Dr. Mine IŞIKSAL BOSTAN	(METU, MSE)	
Assist. Prof. Dr. Şerife SEVİNÇ	(METU, MSE)	

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: Emine ÇATMAN AKSOY

Signature :

ABSTRACT

AN INVESTIGATION OF 7TH GRADE STUDENTS' STATISTICAL LITERACY ABOUT THE CONCEPTS OF AVERAGE AND VARIATION ON BAR AND LINE GRAPHS

Çatman Aksoy, Emine

M.S., Department of Elementary Science and Mathematics Education

Supervisor: Prof. Dr. Mine IŞIKSAL BOSTAN

September 2018, 288 pages

The aim of the present study was to analyze statistical literacy of seventh grade students on the concepts of "average" and "variation" on bar and line graphs related to the data obtained from social or scientific contexts. More specifically, seventh grade students' statistical literacy levels in terms of the concepts of "average" and "variation" on bar and line graphs related to the data obtained from social or scientific contexts were determined and how students at different statistical literacy levels define, interpret and evaluate the concepts of average and variation was investigated.

Participants of the study were 164 seventh grade students from two public middle schools in Akyurt and Çankaya district of Ankara. Data of the study were collected via Statistical Literacy Test (SLT) during the spring semester of 2017-2018 academic year. Obtained data were analyzed using the statistical literacy framework of Watson and Callingham (2003) which consists of six hierarchical levels, beginning from Level 1 to Level 6.

The findings of the study indicated that statistical literacy levels of the students were generally higher in the average concept when compared to determined statistical literacy levels related to the concept of variation. While most of the students generally performed at Level 3-4 in the framework of Watson and Callingham (2003) in the questions related to average concept, most of them performed at Level 1-2 in the questions related to the concept of variation. To state it differently, while most of the students could interpret the concept of average on bar and line graphs, most of them had difficulty in interpreting the variation concept on bar and line graphs related to the data obtained from social or scientific contexts. Moreover, almost all students had difficulty in evaluation of the average and variation concepts on bar and line graphs related to the data obtained from social or scientific contexts. Lastly, while students' interpretations and evaluations of the average concept when data were presented on bar and line graphs did not differ much from the current studies in the literature, bar and line graphs seems to be helpful for students in evaluation of the concept of variation.

Keywords: Statistical Literacy, Average, Variation, Bar and Line Graphs, Middle School Students

ÖZ

YEDİNCİ SINIF ÖĞRENCİLERİNİN
ORTALAMA VE DEĞİŞİM KAVRAMLARI İLE İLGİLİ İSTATİSTİKSEL
OKURYAZARLIKLARININ SÜTUN VE ÇİZGİ GRAFİĞİNDE İNCELENMESİ

Çatman Aksoy, Emine

Yüksek Lisans, İlköğretim Fen ve Matematik Alanları Eğitimi Bölümü
Tez Yöneticisi: Prof. Dr. Mine IŞIKSAL BOSTAN

Eylül 2018, 288 sayfa

Bu çalışmanın amacı ortaokul yedinci sınıf öğrencilerinin ortalama ve değişim kavramları ile ilgili istatistiksel okuryazarlıklarını sütun ve çizgi grafiğinde sosyal veya bilimsel bağlamlarda sunulan veriler kullanarak incelemektir. Bu bağlamda, bu çalışma yedinci sınıf öğrencilerinin ortalama ve değişim kavramları ile ilgili istatistiksel okuryazarlık seviyelerini sütun ve çizgi grafiğinde sosyal veya bilimsel bağlamlarda sunulan veriler kullanarak belirlemeyi ve farklı istatistiksel okuryazarlık seviyelerine sahip öğrencilerin ortalama ve değişim kavramları ile ilgili tanımlarını, yorumlarını ve değerlendirmelerini araştırmaktadır.

Çalışmaya Ankara'nın Akyurt ve Çankaya ilçesinden 164 devlet okulu öğrencisi katılmıştır. Veriler İstatistiksel Okuryazarlık Testi (İOT) aracılığıyla 2017-2018 öğretim yılı bahar döneminde toplanmıştır. Öğrencilerin cevapları Watson ve Callingham (2003)'ın Seviye 1'den başlayan ve hiyerarşik olarak Seviye 6'ya kadar devam eden istatistiksel okuryazarlık çerçevesi kullanılarak incelenmiştir.

Çalışmanın bulguları öğrencilerin ortalama kavramı ile ilgili istatistiksel okuryazarlık seviyelerinin değişim kavramına nazaran daha yüksek olduğunu göstermiştir. Ortalama kavramını içeren sorularda çoğu öğrencinin istatistiksel okuryazarlık seviyesi Seviye 3-4 olarak belirlenmiştir. Fakat değişim kavramını içeren sorularda çoğu öğrencinin istatistiksel okuryazarlık seviyesinin çerçevede ilk iki seviye olarak belirtilen Seviye 1-2 olduğu gözlemlenmiştir. Diğer bir deyişle, çoğu öğrenci ortalama kavramını sütun ve çizgi grafiğinde sosyal veya bilimsel bağlamlarda sunulan verilerde yorumlayabilirken, öğrencilerin değişim kavramının yorumlanmasında zorluk yaşadıkları görülmüştür. Ayrıca, hemen hemen bütün öğrencilerin ortalama ve değişim kavramlarını sütun ve çizgi grafiğinde sosyal veya bilimsel bağlamlarda sunulan verileri değerlendirmede kullanmakta zorluk yaşadıkları sonucuna ulaşılmıştır. Son olarak, verilerin sütun veya çizgi grafiği üzerinde sunulmasının öğrencilerin ortalama kavramının yorumlanması ve değerlendirilmesinde çok bir etkisi gözükmezken, sütun ve çizgi grafiklerinin öğrencilerin değişim kavramı ile ilgili değerlendirme yapmasında yardımcı olduğu görülmüştür.

Anahtar kelimeler: İstatistiksel Okuryazarlık, Ortalama, Değişim, Sütun ve Çizgi Grafiği, Orta Okul Öğrencileri

To My Family and My Lovely Husband

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to my advisor Prof. Dr. Mine IŞIKSAL-BOSTAN. I always felt that you are like my mother who helped me in both academic and personal issues. Thanks for answering my endless questions even in your holidays, your valuable feedbacks and continuous support throughout this process. This study would not have been come to this point without you. Thanks a million for everything.

I would also thank to my committee members, Assoc. Prof. Dr. Elif YETKİN-ÖZDEMİR and Assist. Prof. Dr. Şerife SEVİNÇ for their feedback, suggestions and encouragement for my study.

Special thanks go to my mother, Selime ÇATMAN, my father, Zeki ÇATMAN and my little brother, Sefa ÇATMAN. I always felt your love, encouragement and believe in me throughout my life. I wish to express my sincere thanks to my mother, who did not leave me alone even in a minute and listened and encouraged me every time. This study would not have been finished without you.

I also want to present my deepest appreciation to my brother, İsmail ÇATMAN, who became both a father and mother to me during my undergraduate education and encouraged me to continue graduate education. Thank you for your continuous love and support to me throughout my life. I also want to thank to the lovely wife and son of my brother, Reyhan ÇATMAN and Mehmet Alp ÇATMAN.

I wish to express my sincere thanks to my sweet cousin, Ceren Kezban GÜL who always believed and encouraged me during this process.

I would like to extend my appreciation to my dear friends, Mukaddes İNAN, Mehtap ÖZEN, Ayça ALAN and Semanur KANDİL who supported me in every step of this study. Semanur KANDİL, thank you for your help in this study and answering my endless questions.

Special thanks go to my cute friends Elçin ERBASAN and Nurbanu DURAN. You have always been there for your help. Thank you for not leaving me alone in this process. I am very lucky to have friends like you.

I would also like to teachers, Zelal YALVAÇ, Didem YORGANCI, Kürşat YORGANCI, Mestan Anıl DEMİR and Hüseyin ŞENEL, for separating their valuable time for my study. I also thank to lovely students who contributed to this study.

I want to present my deepest appreciation to Mustafa KAPUKAYA, my dear teacher. Thank you for your continuous support even in your holiday. I am so lucky to have a teacher like you.

I wish to express my thanks to TUBITAK for the financial support during my master degree study.

Last, but by no means least, extra special thanks go to my lovely husband Fethi Barış AKSOY. Thank you for your everlasting love, support and patient. Thank you for sharing my tears every time. This study would not have been finished without you. I also thank to the family of AKSOY, İbrahim AKSOY, Ümmü AKSOY and Nazlı AKSOY for their support and patient during this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AEC	Australian Educational Council
GAISE	Guidelines and Assessment for Statistics Education Report
IOT	Istatistiksel Okuryazarlık Testi
MoNE	Ministry of National Education
NAEP	National Assessment of Educational Progress
NCTM	National Council of Teachers of Mathematics
OECD	Organization for Economic Co-Operation and Development
SLT	Statistical Literacy Test

CHAPTER 1

INTRODUCTION

"The ultimate goal: Statistical Literacy" is stated at the beginning of the Guidelines and Assessment for Statistics Education Report (GAISE, 2005, p.1.). According to this report, the aim was to ensure that students graduate "from the high school with sufficient statistical reasoning to cope with requirements of citizenship, employment and family and to have a healthy, happy, and productive life" (2005, p.1). In other words, one of the goals of the schools stated by GAISE (2005) is to raise individuals who are statistically literate.

In their daily life, individuals need to make decisions about many areas such as education, economy, politics and health (Halpern, 1997). In all of these areas, people encounter percentages, some type of graphs, charts, averages, rates or probabilities (Wallman, 1993). To put it differently, individuals encounter statistical information everywhere in their daily life and an informed citizen should be able to understand this information and then be able to make decisions based on that understanding (Towsend, 2006; Wallman, 1993). However, statistical information emerging in daily life or the media could involve bias or subjectivity (Gal, 2004). Some statistics are attached to advertisements, arguments, or suggestions just to make them seem more reliable (Ben-Zvi & Garfield, 2004) and those statistics are generally produced by the owners of those advertisements or arguments (Shield, 1999). Therefore, individuals need to not only understand the statistical information presented, but also critically evaluate those biased, misleading statistical claims (Gal, 2004; Watson, 2006). In other words, researchers maintain that individuals need to be statistically literate.

Statistical literacy is essential for people to critically question the vast amount of information presented to them (Rumsey, 2002) and a key ability that people living in information societies should acquire (Gal, 2002). Furthermore, it is essential for children to make decisions in the twenty-first century since it is very easy for them to be misinformed with different kinds of data displays (Watson & English, 2015). As previously stated, statistical literacy is an expected outcome of schooling (GAISE, 2005; Gal, 2002; Ben-Zvi & Garfield, 2004). Thus, many curriculum documents from several countries, which have a huge role in the development of statistical literacy of students (Gal, 2004; Watson, 2006), emphasize the importance of statistical literacy and thus have integrated concepts of statistics and probability into their curricula. For instance, the National Council of Teachers of Mathematics (NCTM, 2000) implied the importance of statistical literacy stating that if the aim is to raise consumers who could make critical and informed decisions, the knowledge of statistics is essential. Moreover, the Australian Educational Council (1991) added into their curricula the strands of understanding and explaining social uses of chance, and understanding the impacts of statistics on daily life in order to raise statistically literate individuals. Similarly, the mathematics curriculum in Turkey includes a content domain called “data analysis”, which is one of the five content domains in the curriculum. In the Turkish curriculum, the aim is to ensure that by the end of the eight grade students should be able to form research questions, collect appropriate data, represent and analyze the collected data using measures of central tendency and spread and lastly interpret the results obtained. Interpreting statistics in real life contexts and making decisions according to those interpretations were also emphasized in the Turkish curriculum (MoNE, 2018).

Beyond its place in curricula, statistical literacy has also been an important topic in the area of research for about 20 years. Watson (1997), one of the pioneer researchers in statistical literacy, defined statistical literacy in a three-tiered hierarchical framework. These tiers are (i) understanding basic terminology about statistics and probability, (ii) understanding and interpreting those concepts in social and scientific contexts, and (iii) critically evaluating claims and arguments in real life. Hence, according to Watson (1997), statistical literacy is an ability consisting of understanding, interpreting and critically evaluating statistical messages that are encountered in the daily lives of individuals. In later years, using her framework, Watson conducted many research studies with her colleagues to understand students' statistical literacy (Watson & Moritz, 1999, 2000, 2000b; Watson & Callingham, 2003). In 2003, the three-tiered hierarchical framework was elaborated and a framework was developed consisting of six levels, which was named as the statistical literacy construct (Watson and Callingham, 2003). In 2006, depending on the obtained research results, Watson developed a model consisting of the major components of statistical literacy. One of those components is data and chance curriculum. This component involves data collection-sampling, data representation, chance, inference and average. Furthermore, there is another component in the model of Watson (2006), the importance of which is mentioned in the research by many authors (Cobb &Moore, 1997; Ben Zvi, 2004, Watson, 2006). This component is variation, which Cobb and Moore (1997) believe that it is "the omnipresence of variability" that cause the need for statistics (p. 801). Besides, Konold and Higgins (2003) explained that average is a signal in this noisy environment where the noise refers to variation. Furthermore, average and variation in the model of Watson (2006) are two important concepts that appear in the Turkish curriculum in the middle school years; hence; these are the concepts that are focused in this study.

Various studies regarding students' understanding of average and variation (Mokros & Russell, 1995; Watson & Moritz, 1999, 2000; Gal, Rotschild & Wagner, 1989; Watson & Callingham, 2003; Watson & Kelly, 2008; Shaugnessy, 2003b) exist in the related literature. In terms of the concept of average, it was concluded that students have difficulties in defining the concept of average (Watson & Moritz, 2000), do not know the representative nature of the concept of average (Mokros &Russell, 1995), cannot use average when comparing two data sets (Gal, Rotschild & Wagner, 1989; Watson &Moritz,1999) or cannot realize the effect of an outlier in the calculation of mean, a type of average, and hence cannot suggest the use of the median (Watson & Callingham, 2003). Findings of studies indicated that results were not different for the concept of variation. Most students have difficulty in defining and interpreting the concept of variation (Watson & Kelly, 2008) and they cannot use variation in comparing of two groups similar to the average (Shaugnessy, 2003b). On the other hand, Shaugnessy and Pfannkuch (2002) observed that when students draw a graph of a data presented to them in a table form, they can realize variation in the data. Similarly, Bright and Friel (1998) asserted that graphs can affect students' interpretation of the data presented to them. Furthermore, Enisoglu (2014) claimed that graphical representations could have an effect on students' interpretation of the average. To state it differently, graphical representations could have an influential role in students' interpretation of average and variation.

Moreover, Watson, Chick and Callingham (2014) asserted that each real life context does not reveal the same effect on students' understanding of average. In their study, they found that while students could easily interpret average appreciating variation in a weather context, they had difficulty in interpreting the meaning of the average of home prices. Similarly, McGatha, Cobb and Mc Clain (2002) observed that while most of the students could realize variation in a weather context, few of them noticed it in a basketball context. In other words, it is important to analyze students' understanding of average and variation in different real-life contexts since, as the researchers mentioned above asserted, results could show some differences from those of prior studies.

Therefore, the aim of this study is to investigate the statistical literacy of seventh grade students concentrating on the concepts of average and variation on graphical representations in different real life contexts. In line with the objective that students should be able to use mean and range, which are two measures related to average and variation, in comparing two data sets in real life situations (MoNE, 2018), statistical literacy of students regarding the concepts of average and variation will also be examined in comparative situations. Graphical representations that students learn up until seventh grade are bar and line graphs and pie charts. Since it is believed that bar and line graphs are appropriate for comparison of two data sets, the purpose of the present study is to analyze statistical literacy of seventh grade students regarding the concepts of average and variation on bar and line graphs.

1.1 Purpose of the Study

The purpose of the study is to analyze the statistical literacy of seventh grade students in terms of the concepts of "average" and "variation" on bar and line graphs related to the data obtained from social or scientific contexts. In this respect, the following research question and its sub-question directed the current study:

What are the statistical literacy levels of seventh grade students regarding the concepts of "average" and "variation" on bar and line graphs related to the data obtained from social or scientific contexts?

a. How do seventh grade students at different statistical literacy levels define, interpret and evaluate "average" and "variation" concepts on bar and line graphs related to the data obtained from social or scientific contexts?

1.2 Definitions of Important Terms

In this section, definitions of the main terms in this study are provided for the clarity of the research questions.

Statistical Literacy: According to Watson (1997), statistical literacy is an ability consisting of understanding, interpreting and critically evaluating statistical messages encountered in daily lives of individuals.

Furthermore, the statistical literacy level of a student is determined according to the modified version of the framework of Watson and Callingham (2003) by taking into consideration the answers provided by a student in the Statistical Literacy Test (SLT)

Average: Average is "A single number or measure that is representative of a larger collection of numbers" (Van de Wall, 2013, p. 446).

In this study, average refers to a measure that represents the data obtained from social or scientific contexts and given on bar and line graphs.

Variation: Variation refers to "A description or measurement of change" (Reading & Shaugnessy, 2004, p.202).

In this study, variation refers to a measure that represents the change in the data obtained from social or scientific contexts and given on bar and line graphs.

Bar Graph: A bar graph is "A diagram in which the numerical values of the variables are represented by height or length of lines or rectangles of equal width" (Oxford American Dictionary, 2006, p. 66).

Line Graph: A line graph is a representation to show the relation between two continuous data with a line drawn to connect the two points (Van de Walle, 2013).

Social or Scientific Contexts: Social or scientific contexts refer to written sources, such as newspapers, research reports, brochures or magazines, which are encountered in real life, especially in the media. This study includes the newspapers, brochures and magazines as social contexts and some research results as scientific contexts which were generally prepared by the researchers.

1.3 Significance of the Study

For effective participation into the social, political and cultural life or for personal satisfaction, being a literate person is significant (Is, 2003). According to the Organization for Economic Co-Operation and Development (OECD, 2001), literacy refers not only to the ability to read and write but also to the ability to use obtained knowledge in different real life situations. Taking into account that individuals make decisions using statistical information, which is found in almost every area in real life, it can be asserted that individuals should also be statistically literate. Statistical literacy is a key ability that people living in information societies should acquire (Gal, 2002) and an ability that students are expected to be equipped with in schools (GAISE, 2005; Gal, 2002). Therefore, it is important to investigate the statistical literacy of students.

In the related literature, many researchers express the importance of variation in statistics (Ben-Zvi, 2004; Cobb & Moore, 1997; Watson, Kelly, Callingham & Shaugnessy, 2003; Watson, 2006). It is the concept of variation which is the underlying reason for the existence of the statistics discipline (Watson, Kelly, Callingham & Shaugnessy, 2003). Data collection, data representation and calculation of averages are all essential to take variation under control (Watson, 2006). It is meaningless to calculate an average when there is no variation (Watson & Kelly, 2008). On the other hand, if variation does exist, the calculation of the average is essential to summarize a huge amount of information into a manageable form (Shaugnessy, 2007). In addition to summarizing and describing a data set, calculating the average is very useful for the comparison of two groups (Konold & Higgins, 2003; Mokros & Russell, 1995). Thus, it can be concluded that average and variation are two important concepts in statistics which are dependent on each other (Watson, 2006). Hence, it is essential that statistically literate individuals should be able to define, interpret and evaluate both concepts when they encounter with them in their daily lives. When such is the case, it is important to examine the statistical literacy of individuals regarding the concepts of average and variation. As aforementioned, since statistical literacy is an ability that schools are expected to equip students with (Gal, 2002; GAISE, 2005), and since the average is one of the mostly encountered statistics in real life which is dependent to the variation, it is essential to investigate the statistical literacy of students related to the concepts of average and variation.

There are various studies in the literature related to students' understanding of average and variation. Almost all of those studies showed that students' understanding of average and variation was not at the expected level (McGatha, Cobb & McClain, 2002, Mokros & Russell, 1995; Shaugnessy, 2003b; Watson & Kelly, 2008; Watson & Moritz, 1999a, 1999b, 2000). These researches revealed that students' definitions, interpretations and evaluations related to the concept of average and variation are inadequate. However, many curriculum documents including the Turkish curriculum states the importance of interpretation of statistics, including average and variation, in real life contexts (AEC, 1991; MoNE, 2018; NCTM, 2000). For example, these documents mention the appropriate use of three average types in real life contexts: mean, mode and median. It is explained that in addition to being able to calculate the three types of average, students should also be able to select the appropriate average type in different real life situations. On the other hand, several researchers asserted that different real life contexts do not show the same impact on students' understanding of average and variation (McGatha, Cobb & McClain, 2002; Watson, Chick & Callingham, 2014). For instance, while students could interpret average as mean in one context, they could understand it as the mode of the given data in another context. The current study examines interpretation and evaluation of the concepts of average and variation presenting students some alternative contexts. Hence, this study has the potential to make a significant contribution to the existing literature by having revealed students' interpretations and evaluations related to the concepts of average and variation in different contexts.

Furthermore, to decide correctly which of the three average types is appropriate indicates that students have conceptual understanding regarding the concept of average (Enisoglu, 2014). In the literature, there are several studies which concentrated on investigating whether students could be able to select median as the appropriate type of average when there is an outlier in the data (Watson &

Moritz, 1999b; Watson & Callingham, 2003). However, it was observed that studies which provided students a categorical data set, hence necessitates the use of the mode as an appropriate type of average, were limited. In this study, a context, which requires the use of mode as the appropriate average type since there is a categorical data, is presented to the students. Therefore, the current study is significant as it contributes to the existing literature valuable information about whether students are able to realize that the appropriate type of average to be used is the mode of the data when there is a categorical data set, which in turn can be useful in passing to the appropriate use of median where students had difficulty (Watson & Callingham, 2003; Watson & Moritz, 1999b).

Similar to the concepts of average and variation, various studies in the literature showed that students have some difficulties in interpreting different graphical representations, including bar and line graphs (Boote & Boote, 2017; Bright & Friel, 1998; Capraro, Kulm, Hammer, 2005; Kirsch, Jungeblatt &Mosenthal, 1988; Pereira-Mondeza & Mellor, 1991). The current study attempts to link the concepts of average and variation with graphical representations and analyzes the statistical literacy of students. Thus, the findings of this study could be important in providing information related to both of these areas and the intersection of these two areas.

Besides, Bright and Friel (1998) claimed that graphs can have different effects on students' interpretation of the data presented to them. Furthermore, Enisoglu (2014) stated that graphical representations could have an effect on students' interpretation of the average. In studies in the literature regarding the use of the median as the appropriate average type in case of an outlier, data were generally presented directly to the students. Different from those studies, in the present study, data were displayed on bar and line graphs. Thus, the current study is significant as it reveals the role of bar and line graphs on the realization of the outliers in the data set. Furthermore, Shaugnessy and Pfannkuch (2002) showed that while students could not realize variation in the data presented in a table form, most of them could easily realize it when they drew a graph and could make good predictions. Different from the studies in the literature, bar and line graphs were used in the current study; hence, the current study provides valuable information about the role of bar and line graphs on students' interpretation and evaluation of variation concept. In this way, if it is observed that bar and line graphs are useful in the interpretation and evaluation of the concepts of average and variation, they can be used in teaching the concepts of average and variation by teachers or teacher educators.

Moreover, Watson and Kelly (2005) suggested analyzing students' understanding of variation while comparing two graphs with equal averages but different variations. Several studies revealed that students do not use average and variation while comparing situations (Gal et al, 1989, Shaugnessy, 2003b, Watson & Moritz, 1999). Different from the studies in the literature, the present study initially asked students to calculate an average while comparing two graphs. Then, their understanding of variation was analyzed when they observed that averages are equal. In other words, this study is significant as it has observed whether students directly interpreted the variation of the presented data, which is the underlying reason for calculation of an average (Shaugnessy, 2007), while comparing two groups when they observed that the averages are the same. Hence, the findings of this study could be important in providing information to teachers, teacher educators or curriculum developers in that teaching of the concepts of average should be delayed until the students' understanding related to the variation concept complete.

In the context of Turkey, there are different studies analyzing students' understanding of the concept of average (Ucar & Akdogan, 2009; Enisoglu, 2014). They concluded that middle school students do not know the representative nature of the average, use different solution strategies in solving questions related to the mean, median and mode but generally did not depend on the graphical representations given. Moreover, many errors and difficulties in solving questions related to three measures of central tendency were observed (Enisoglu, 2014). Different from these studies, the current study investigated the interpretation and evaluation of the concept of average presenting students some alternative contexts which may have different effects (Watson, Chick & Callingham, 2014). Hence, this study is significant since it has the potential to make an important contribution to the existing Turkish literature by having revealed students' interpretations and evaluations related to the concepts of average and variation in alternative contexts.

On the other hand, there are not many comprehensive studies regarding the concept of variation in the accessible literature in Turkey. Furthermore, studies related to the statistical literacy of middle school students are limited, and there is a need to conduct further studies related to statistical literacy concentrating on specific concepts in the curriculum (Yolcu, 2012). Therefore, considering the fact that there is a limited number of studies related to variation and statistical literacy in the related Turkish literature, this study investigated statistical literacy of seventh grade students in terms of the variation concept. By also attaching two graphical representations, namely bar and line graphs, the results of such a study were able to provide distinctive and valuable information regarding whether students could interpret and evaluate variation on bar and line graphs and whether the statistical literacy of students related to the concepts of average and variation changed when data were presented on bar and line graphs.

Furthermore, in Turkey, the statistical literacy of students was examined generally through multiple choice tasks. However, to reveal students' understanding and questioning ability better, open-ended tasks should be used (Watson, 1997). Thus, this study is significant in that through open-ended tasks students' understanding and questioning ability related to the concepts of average and variation could be examined in detail.

Lastly, this study is significant in giving information to both pre-service and in-service teachers regarding what seventh grade students know and do not know related to not only the concepts of average and variation, but also bar and line graphs. Moreover, the present study could provide valuable information to teachers and teacher educators in the development of tasks that are necessary to raise statistically literate individuals. The results of this study could be important in the revision of current curricula and the development of textbooks taking into account statistical literacy and the specific concepts that this study focused on.

CHAPTER 2

LITERATURE REVIEW

The purpose of the study is to analyze seventh grade students' levels of statistical literacy in the concepts of average and variation on bar and line graphs related to the data obtained from social or scientific contexts. The review of literature is presented throughout this chapter.

The first part of the chapter begins with the concept of literacy and provides definitions of literacy. The second part of the chapter focuses on statistical literacy. Subsequently, theoretical frameworks and models of statistical literacy, and studies conducted on statistical literacy and specific concepts the study entails, namely average and variation, are addressed. Then follows a section on studies carried out in the context of Turkey. At the end of this chapter, a summary is presented to outline the underlying rationale of this research.

2.1. Literacy

"The ability to read and write" (Oxford Dictionaries, 2009, p.519) is defined as literacy in various dictionaries. For effective participation into the social, political and cultural life or for personal satisfaction, being literate is very significant (Is, 2003). Besides, the more individuals are literate, the more they are productive, so literacy is essential for economic welfare of a country (Metcalfe, Simpson, Todd & Toyn, 2013). However, just being able to read and write is not sufficient to participate actively into the society (Is, 2003). In addition to these skills, application and interpretation of the knowledge acquired in schools to the real life is necessary (OECD, 2001). In other words, The Organization for Economic Co-Operation and Development (OECD, 2001) presents a broad definition to the literacy concept beyond the existing definition. Literacy is the ability to use obtained knowledge in schools in different real life situations (OECD, 2001).

The responsibility of schools is to raise individuals who are literate in reading, mathematics and science (OECD, 2001). An individual who is literate in these areas should be able to understand and reflect on a given text (OECD, 2001). Moreover, he or she should be able to solve mathematical problems in real life and think scientifically in a world of technology and science (OECD, 2001). On the other hand, real life, especially media, is full of statistical messages (Ben Zvi & Garfield, 2004), which are biased or lack objectivity in many situations (Gal, 2004). Individuals should be able to understand and critically evaluate those biased or misleading messages (Gal, 2004; Watson, 2006). Therefore, besides the three domains mentioned by OECD, one of the goals of schools as stated by The Guidelines and Assessment for Statistics Education Report (GAISE, 2005) is to raise individuals who are statistically literate.

2.2 Statistical Literacy

Importance of statistical literacy is mentioned by many researchers and in many curriculum documents. However, there is no consensus on the definition of this concept (Shaugnessy, 2007, Sharma 2017). Therefore, initially, definitions made by different researchers will be presented in this section.

Definitions of statistical literacy can be traced back to Wallman's definition made in 1993. She observes statistical literacy as an important vehicle for prosperity of the society, and defines the concept as "the ability to understand and critically evaluate statistical results that permeate our daily lives-coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions." (p.1). The statement "public and private" in the definition of Wallman reveals that, besides its role for society, statistical literacy has an important role for individuals in making decisions depending on the information presented to them in their personal lives (Watson & Callingham, 2003).

Believing that statistical literacy is the intersection point of statistics and probability and everyday world (Watson, 2006), Watson (1997) defines the concept in a three-tiered hierarchical framework. These tiers include understanding basic terminology in statistics, understanding and interpreting the concepts of statistics in other contexts and questioning claims and arguments in real life. Hence, according to Watson (1997), statistical literacy is an ability consisting of understanding, interpreting and critically evaluating statistical messages that are encountered in the daily lives of individuals. This is the definition to be used in this study.

A similar definition to that of Watson was offered by Garfield in 1999. For Garfield (1999), statistical literacy is the understanding of statistical terms, symbols and words, understanding and interpretation of graphs and tables and making sense of statistics in real life. Questioning ability in the definition of Watson forms the difference between definition of Garfield (1999) and Watson (1997). On the other hand, Schield (1999) regards statistical literacy as a competency, just like reading and writing. It is being competent in critical thinking by which claims are supported with statistics (Schield, 1999). It also involves the understanding and interpretation of arguments presented (Schield, 1999). Moreover, asking good questions are an indispensable part of interpretation (Schield, 1999).

In another study, Gal (2002) describes statistical literacy for adults. For Gal (2002), adult statistical literacy is composed of two components. One of them is one's ability to interpret or critically evaluate the statistical information encountered in various contexts. The other is the ability to communicate one's opinions about certain statistical information, such as being able to explain his or her concerns about an implication drawn from a study. These are key skills which are necessary in this information age and expected educational outcomes in schools (Gal, 2002).

On the other hand, as a result of some discussions in several forums on statistical literacy, thinking and reasoning, statistical literacy was defined as a concept covering some basic skills to be used in understanding of statistical information or research results (Garfield, del Mas & Chance, 2003). Organizing of data, constructing data displays and dealing with different data representations constitute these skills. Comprehension of words, symbols or terms is also involved in the concept of statistical literacy (Garfield, del Mas & Chance, 2003). These researchers have not mentioned critical evaluation as an essential skill for statistical literacy. However, critical evaluation is regarded as a skill for statistical thinking as it is used to question the how and why of statistical investigations.

Watson and Callingham (2003) state that knowing only formulas or definitions is not sufficient to be a statistically literate person. Integration of these formulas or definitions with the contexts where some questions arise is necessary (Watson & Callingham, 2003). These researchers also describe statistical literacy as a construct consisting of six hierarchical levels, which will be explained in detail in the next section. In a further study, Towsend (2006) defined statistical literacy with four main skills: understanding and interpretation of statistical information, critical evaluation of the information involving statistical messages, application of the information to real life situations, communication of his or her concerns regarding the information to others. Apart from the application part, the definition provided by Towsend (2006) is very similar to the one made by Gal (2002).

To summarize, although researchers define the concept of statistical literacy in different ways, there are three common terms that could be identified in most of them. “Interpretation” and “critical evaluation” of statistical information are two terms generally emerging in definitions of statistical literacy. The last common term is “context”, which refers to real life or everyday world in the definitions. Taking into account that statistics makes sense only within a context (Scheaffer, 2006; Franklin et al., 2005), this last common emerging term should be inevitable in definitions of statistical literacy.

2.2.1 Theoretical Frameworks and Models of Statistical Literacy

In this section, main theoretical frameworks and models regarding statistical literacy will be summarized.

2.2.1.1 Watson's Theoretical Framework

In 1997, Watson described a framework to measure the statistical literacy level of students. Hierarchical in nature, this framework consists of three tiers, which does not take into account the explicit affective dimension, but only the cognitive dimension of statistical literacy (Yolcu, 2012).

The first tier is understanding of basic concepts in statistics and probability. In this tier, students only need to understand basic concepts like mean, random, percentage, graphing within a mathematical context. In addition, the calculation of a mean or a measure of spread not depending on any social or scientific context is a requirement of this tier.

In the second tier, beyond making definitions of the concepts or performing just computations, students need to understand and interpret the concepts in social or scientific contexts in order to reach conclusions and to make decisions.

The last tier involves the ability to question. In this tier, students do not believe every claim or argument presented to them in their daily lives. They should be able to critically evaluate claims that they encounter in a social or scientific context. For instance, when reading about the effect of a drug in an article, a statistically literate person should ask questions regarding the sample, such as the number of people in the sample and its representativeness of the population.

2.2.1.2 The Statistical Literacy Construct of Watson and Callingham

Conducting longitudinal research with about 3000 students whose grades varied between 3 and 9, Watson and Callingham (2003) validated that the concept of statistical literacy is hierarchical in nature. Rasch analysis revealed that the construct of statistical literacy involves understanding the concepts mentioned in a statistics and probability curriculum along with understanding the context that may involve bias or misinformation (Watson & Calllingham, 2003). Furthermore, analyses made it possible to identify and interpret hierarchical levels of the construct. Those levels are idiosyncratic, informal, inconsistent, consistent non-critical, critical and critical mathematical. Each level shows the characteristics of statistical literacy that students displayed in the tasks, and they can be seen in Table 2.1 below.

Table 2.1 Statistical literacy construct (Watson & Callingham, 2003, p.14)

Level	Brief characterization of step levels of tasks
6. Critical Mathematical	Task-steps at this level demand critical, questioning engagement with context, using proportional reasoning particularly in media or chance contexts, showing appreciation of the need for uncertainty in making predictions, and interpreting subtle aspects of language.
5. Critical	Task-steps require critical, questioning engagement in familiar and unfamiliar contexts that do not involve proportional reasoning, but which do involve appropriate use of terminology, qualitative interpretation of chance, and appreciation of variation.
4. Consistent Non-critical	Task-steps require appropriate but non-critical engagement with context, multiple aspects of terminology usage, appreciation of variation in chance settings only, and statistical skills associated with the mean, simple probabilities, and graph characteristics.
3. Inconsistent	Task-steps at this level, often in supportive formats, expect selective engagement with context, appropriate recognition of conclusions but without justification, and qualitative rather than quantitative use of statistical ideas.
2. Informal	Task-steps require only colloquial or informal engagement with context often reflecting intuitive non-statistical beliefs, single elements of complex terminology and settings, and basic one-step straightforward table, graph, and chance calculations.
1. Idiosyncratic	Task-steps at this level suggest idiosyncratic engagement with context, tautological use of terminology, and basic mathematical skills associated with one-to-one counting and reading cell values in tables.

In the *idiosyncratic* level, students make decisions based on personal beliefs and experience. Their engagement with context is idiosyncratic, and they use their daily language. For example, while drawing a male or a female name from a hat, students will decide that a female name will be drawn because the teacher is a female (Watson & Callingham, 2003).

In the second level, *informal*, students deal with context more than it is done so in the first level, but this engagement still depends on students' perceptions. Their answers may involve beliefs that are not statistical or do not include relevant aspects. For example, a student at this level answers a question about forming a sample like "ask everyone" or "ask the people I meet". He does not consider the aspect of representativeness of the population. Yet another example that could be given is that for the definition of the concept of average, they may use a single word like normal or only give an example to illustrate average (Watson & Callingham, 2003).

In the inconsistent level, characteristics observed begin to reflect the requirements of the second tier of Watson. In this level, the format of the items determines students' engagement with the context. If items provide additional support, students' engagement is more. For instance, while students use colloquial ideas in defining 'being average' in an open-ended task, when the question was given in a multiple-choice format, they could choose the appropriate alternative. Beyond single features, students at this level focus on more features; however, instead of using quantitative statistical ideas, students use qualitative ones; for example, students say that since there are more female names in the hat, choosing a female name is more likely. However, they do not express the probability of choosing a female name.

In the consistent non-critical level, requirements for Tier 2 clearly are observed. In this level, students engage with contexts, but this engagement does not involve critical questioning. Students can make criticisms, but these are not ones that are central. Definitions made at this stage involve multiple elements. At this stage, students suggest appropriate variations for chance tasks. They begin to use mathematical and statistical skills associated with mean, simple probabilities and graph characteristics. For example, they compute the mean of a small data set but do not observe the effect of an outlier.

The last two levels necessitate critical thinking skills just like in the third tier of Watson. In both of these levels, students critically engage with the context. They focus on central issues in the criticisms they make at level 5. Students at level 5 are more successful in concepts they are familiar with, just like in the school survey. They are able to compute the mean and median of a small data set. Students show that they appreciate variation with phrases like 'It will be close to half'. The difference between the two levels lies in the mathematical skills used. In the last level, *critical mathematical*, in the task of drawing a female or a male name from a hat, students give quantitative responses using proportional reasoning instead of qualitative ones like "the same" or "more". At this stage, students summarize the information given in graphs instead of merely reading data from the graphs. Students at this level know when to use the median as the appropriate measure of center.

This construct is closely related to the theoretical framework of Watson (Shaugnessy, 2007). As previously stated, the last two levels are similar to the third tier of Watson (1997), namely critical evaluation. While the objectives of the first tier are observed across different levels, requirements of Tier 2 begin to appear at level 3, *inconsistent*. Indeed, except for Tier 1, the tiers were separated into the two hierarchical parts.

2.2.1.3 Four-Stage Framework of Sharma and her colleagues

Believing in the need for assessment in a classroom setting, Sharma, Doyle, Shandil and Talakia'atu (2011) developed a four-stage framework for assessing the statistical literacy of students as a result of activities experimented with grade 9 students. The framework is a modified version of the statistical literacy construct proposed by Watson and Callingham (2003). The six levels in the statistical literacy were reduced to four in the framework of Sharma et al. (2011). Those stages are Informal/Idıosyncratic, Consistent/Non-Critical, Early Critical and Critical.

As the name suggests, the first stage, Informal/Idiosyncratic, is the combination of the first two levels of Watson and Callingham (2003). Therefore, this stage involves the performances of students for those levels, such as no engagement with context occurs in this stage or when there is engagement, it depends on students' personal beliefs.

Some of the requirements in the third level of Watson and Callingham (2003), Inconsistent, are located in the second stage and some are found in the third stage of Sharma and her colleagues' framework. For instance, while using qualitative statistical ideas on behalf of quantitative ideas means passing onto the third stage, reaching conclusions without any justification is a requirement of the second stage.

The last stage includes all the requirements of the Critical Mathematical level of Watson and Callingham (2003). In other words, students at this level engage in context with a questioning attitude. Moreover, they utilize complex mathematical and statistical skills in dealing with the context.

In the three sections above, some frameworks related to statistical literacy were presented. In the next two sections, some models for statistical literacy will be mentioned.

2.2.1.4 Two Elements the Model of Gal

In his paper, as previously mentioned at the beginning of Section 2.2, Gal (2002) defines statistical literacy as the capability of interpretation, critical evaluation and communication of statistical information and messages. These abilities are based on several knowledge bases which are related with each other and also on dispositions (Gal, 2002). In other words, the model of Gal assumes that statistical literacy of adults is composed of two interrelated elements. They are knowledge and disposition as can be seen in Table 2.2.

Table 2.2 A model of statistical literacy (Gal, 2002, p.4)

Knowledge elements	**Dispositional elements**
Literacy skills Statistical knowledge Mathematical knowledge Context knowledge Critical Questions	Beliefs and Attitudes Critical stance
↘ **Statistical Literacy** ↙	

While the knowledge component involves five cognitive elements, the dispositional component has two elements: (i) critical stance and (ii) belief and attitudes. All of these elements will be explained respectively. Gal (2002) notes that these components and elements are not independent of each other; in combination, they provide a statistically literate behavior. Moreover, he states that although they are necessary, an adult may not possess all of them to deal with in all contexts.

The knowledge component consists of literacy skills, statistical and mathematical knowledge, context knowledge and critical questions. To begin with, all statistical messages are sent within written or oral texts. Therefore, some literacy skills are inevitable to be a statistically literate adult. To comprehend the main text and the ones accompanying it, and to communicate ones' opinions, which are stated as text-processing skills by Gal (2002), are very significant to understand statistical messages. To express the relationship between literacy and statistical literacy, Gal (2002) mentions document literacy. He uses the document literacy concept of Kirsch and Mosenthal (1990). According to them, identifying, interpreting and using information given in lists, tables, charts, and graphs form the document literacy, which actually explains the clear connection between general literacy and statistical literacy.

The second knowledge element is statistical knowledge, which includes basic statistical and probabilistic concepts and procedures. This element shows a clear requirement for understanding statistical messages. According to the results of several research studies, Gal (2002) determined five important parts of this knowledge base. They are shown in Figure 2.1.

1. Knowing why data are needed and how data can be produced
2. Familiarity with basic terms and ideas related to descriptive statistics
3. Familiarity with basic terms and ideas related to graphical and tabular displays
4. Understanding basic notions of probability
5. Knowing how statistical conclusions or inferences are reached

Figure 2.1 Five parts of statistical knowledge base (Gal, 2002, p.10)

Figure 2.1 indicated that first of all adults should understand the necessity of data for research results, and also at least informally they should have information regarding data collection (Gal, 2002). Familiarity with the reason of conducting some commonly used designs in media, such as experiments or surveys, are also necessary for a statistically literate adult. Familiarity with some generally used basic concepts to send results, such as measures of center or percent and some graphs and tables to organize the collected data, comes after acknowledging the necessity of the data. Moreover, adults should interpret the language of chance and understand, at least intuitively, the variability underlying chance events. The last knowledge base is related to statistical inference. Gal (2002) believes that a statistically literate adult should have some basic ideas related to data analysis and the conclusions derived from it.

Beyond these knowledge elements, knowing the mathematical procedures behind statistics will allow for a more effective interpretation of them (Gal, 2002). However, Gal (2002) believes that it is not necessary to focus too much on the mathematics behind the statistics computed.

One of the other knowledge bases to understand statistical messages is the knowledge of context. According to the Gal (2002), the context facilitates the interpretation of the results obtained from the analysis. With the help of the context, sources of variation and error could be easily predicted. It is not easy for an adult to detect the mistakes in a study or to present alternative interpretations for results without understanding the context.

Critical skills in the framework of Gal (2002) is similar to the third tier in that of Watson (1997). Because of subjective report findings, conclusions or interpretations in the media, adults have to question everything presented to them. He states that adults should ask "worry questions" to examine the nature, validity and reliability of the messages sent. For example, when faced with a graph, a statistical literate adult could firstly ask and then be able to answer the question "Is a given graph drawn appropriately, or does it distort the trends in the data?

Evaluating critically statistical messages implies an action for adults (Gal, 2002). That is to say, in this process, adults are not in a passive situation. Gal (2002) believes that some dispositions are necessary for this action whether it is hidden or visible. As can be seen in Table 2.2, those dispositions are critical stance, and belief and attitudes. By critical stance, Gal (2002) means being prepared and dispositioned to ask his or her worry questions when encountered with data or statistical messages. To have a critical stance, adults should have a questioning attitude and a belief that it is not in conflict with taking a critical action. They have a right to be critical about statistical messages, and to ask "worry questions".

2.2.1.5 Watson's Statistical Literacy Model

In 2006, Watson presented a model for statistical literacy, which have some similar characteristics with Gal's model (2002). She formed her model based on the result of several studies which assessed students' level of statistical literacy after the three-tiered hierarchical framework of Watson (1997) and the statistical literacy construct of Watson and Callingham (2003) were described. The model reveals major components of statistical literacy and their relationship with each other. The components are context, mathematical/statistical skills, literacy skills, task motivation, task format and data and chance curriculum. The components and their relationships with each other can be seen in the Figure 2.2.

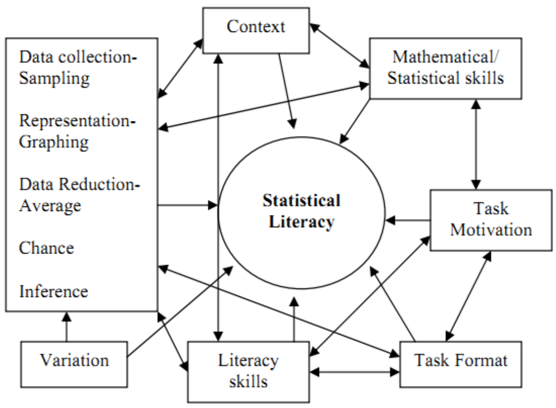


Figure 2.2 Relationship among components of statistical literacy (Watson, 2006, p.248)

The first component which is located on the left side of the model is data and chance curriculum. This component involves concepts in statistics and probability, namely data collection, data representation, data reduction, chance and inference. This component shows similarities with the statistical knowledge element of Gal (2002) but in the model of Watson (2006), the essential objectives for each concept is described in more detail. For instance, regarding the data representation concept, it is stated that students should be able to understand messages from the graphical displays, interpret them, and discover graphs that mislead individuals. Understanding the meaning of average in a given context, and interpreting the effectiveness of different types of measures of central tendencies are some of the requirements for the concept of average. For the concept of sampling, students should be able to understand the meaning of the concept of sample and the necessity of representative samples or evaluate possible bias in sampling. Analyzing graphical representations and making inferences from them, like comparing two groups and questioning the inferences made in the media are abilities needed for the concept of inference.

The second component which Watson (2006) believes is the main reason for statistics and has an effect on each of the concepts in the data and chance curriculum is variation. Regarding the concept of variation, students should be able to know what variation means, and how variation and sample are related. They should also be able to use variation while comparing two groups.

Mathematical and statistical skills form the third component of Watson's model. Mathematical skills involve the understanding of some concepts generally found in middle school curricula, such as proportions, part-whole relationships or percent. Students' lack of information about mathematics like fractions, decimals, and algebraic formulas block their understanding of statistical content (Ben Zvi & Garfield, 2004). By statistical skills, Watson (2006) means calculation of averages or basic probabilities. Definitions of terms also form an important part of mathematical and statistical skills.

Similar to Gal (2002), Watson (2006) regards context as a crucial component of the model. In the model, she mentions three types of contexts. They are isolated, familiar and unfamiliar contexts. While flipping a coin forms an example for isolated contexts, a school survey is a familiar context for students. Media extracts are good examples for unfamiliar contexts (Watson, 2006). She believes that unfamiliar contexts, among others, are the better ones for appreciation of the contexts and provide higher levels in statistical literacy.

Literacy skills is a clear component of statistical literacy since the concept itself includes the word literacy (Watson, 2006). As Gal (2002) makes a connection with the model of Kirsh and Mosenthal (1996) to reveal the connection between literacy and statistical literacy, Watson (2006) reveals this by making connections between The Four Resources Model of Luke and Freebody (1997). Their models are not hierarchical but necessitates a literate person to understand and interpret given texts and then to be able to form alternative meanings from them and lastly question the text appreciating some of them can include subjectivity. The consistency among the elements of literacy by Luke and Freebody (1997) and the tiers in Watson's theoretical framework (1997) reveals the relationship between literacy and statistical literacy.

Format of the task is another factor having a role in statistical literacy. Tasks can be either open-ended or multiple choice. Multiple choice tasks are more suitable when recognition of an answer is sufficient instead of constructing it, which is requested in some of the statistical literacy levels (Watson, 2006). However, open-ended tasks are the best if statistical understanding or questioning ability is to be assessed (Watson, 1997).

The last component of the model resembles the dispositional elements of Gal (2002). It is named as task motivation in the model. Although Watson mentioned only cognitive elements in her framework in 1997, she added an affective element to her model in 2006. With this component, she addresses all the dispositions of students to the task presented. These dispositions are a determining factor in reaching conclusions along with other components (Watson, 2006).

To conclude, there are several frameworks and models regarding statistical literacy in the literature and these models and frameworks have some similar and different characteristics. The aim of this study is to analyze statistical literacy of students regarding the concepts of average and variation on bar and line graphs. These concepts were chosen because they are the concepts in focus regarding statistics in the middle school curriculum in Turkey. Moreover, the necessary objectives for statistical literacy for the concepts of average and variation were mentioned in detail in the last model. Therefore, the last model and the framework of Watson and Callingham (2003), which is related to the model, were utilized in this study. How to use the framework to analyze statistical literacy levels of students will be explained in detail in the method chapter. The following section dwells on several research studies on statistical literacy.

2.3 Research on Statistical Literacy

Research on statistics began to gain increasing importance in accordance with the call for Shaugnessy in 1992. In a similar manner, research regarding statistical literacy of students has been conducted for about 20 years. In this section, results of some of these studies are presented. The aim of the current study is to analyze students' statistical literacy regarding the concepts of average and variation on bar and line graphs. Therefore, studies concentrating on the concepts of average, variation and data representation are the focus of this section. Furthermore, studies conducted not only in the context of statistical literacy but have important results regarding the concepts of the current study are presented.

2.3.1 Research on Students' Understanding of Average

Average is a power tool of statistics since it reduces a huge amount of information to a manageable form (Shaugnessy, 2007). The concept of average has been attracting the attention of researchers for a long time in the literature. Strauss and Bichler (1988) are some of those researchers. In their study in 1988, they focused on computational properties of the mean. For instance, they measured whether students aged between 8 and 14 were aware that mean was affected by extreme values or mean did not need to be a value in the data set. It was found that students could understand that some properties, such as mean, are affected by specific points in the data set or should be between extreme values. On the other hand, complicated properties, like including a zero value in the calculation of a mean or sum of deviations from the mean equals zero, were difficult for students (Strasuss & Bichler, 1988).

Conceptual studies regarding the concept of average began to be conducted in 1995. (Shaugnessy, 2007). Mokros and Russell (1995) analyzed fourth to eight grade students' understanding of representativeness by asking students to construct data sets. Furthermore, they examined the relationship between students' informal understanding of representativeness and their understanding of the mean that they had learnt in school. The application of two construction problems, one interpretation and one weighted means problem revealed five approaches that students used in solving problems. Although many of the students used more than one approach, researchers determined a predominant approach of every student. They separated these approaches mainly into two groups of approaches, namely recognizing and not recognizing representativeness. These approaches will be presented in detail in the following paragraphs.

The approaches of algorithm and average as mode were determined as the two approaches that did not embody an idea of representativeness. Students who recognized average as mode believed that average was the value that occured most in the data (Mokros & Russel, 1995). Therefore, constructing a data set for these students was regarded as easy since most of the data were placed into the mode. However, it was stated that to make their data realistic, all of the data were not placed into the mode, which is an indication that students also depended on their experience. For these students, the most in the data set was thought as the most typical one; therefore, the representative value of the average in the data set given was not recognized by these students (Mokros & Russel, 1995). Researchers asserted that even in questions with higher values of representativeness, those values were ignored and the focus was turned directly to the mode of the data. In addition to the mode, egocentric arguments were used by these students in their interpretations or constructions (Mokros & Russel, 1995). Furthermore, it was observed that the mean algorithm that students had learnt in school was not used by almost all of the students assuming the approach of recognizing average as mode. Researchers stated that even when students tried to use it, the algorithm they applied had little mathematical sense. Algorithmic approach was another approach that did not not recognize average as a value representing the data set as a whole (Mokros & Russel, 1995). It was concluded that these students knew the procedure to find the mean and directly apply this procedure when they encounter a problem. However, whether or not this procedure made sense was not considered; moreover, the data, the mean of which was computed was not taken into consideration either (Mokros & Russel, 1995). In other words, these students had the idea of what Mathew and Clark (2007) named as "process conception", meaning accepting the concept as a process, not as a measure of central tendency. It was observed that some of the students used some algorithms which were not meaningful. As in the study of Cai (1995), although students could compute the mean, they could not reverse the procedure to find total or specific data points.

On the other hand, approaches that recognize average as representative are average as reasonable, midpoint and mathematical point of balance (Mokros & Russel, 1995). What is typical in their own lives and what is mathematically reasonable were used as meanings for average for students using the reasonable approach. These students combined mathematics with real life and could use the mean algorithm (Mokros & Russel, 1995). However, it was observed that these students had a misconception that average is only an approximation if all of the data points are not given; therefore, the mean could not be computed by these students when sub-means were given. This is a similar result reported by Pollatsek, Lima and Well (1981). Pollatsek et al. (1981) observed that many undergraduate students could not compute weighted mean when sub-means were given. Like the students in the mode approach, these students also had egocentric arguments but they knew that average is something about the distribution of the data and some conceptions of representativeness, which is the trigger for future developments (Mokros & Russel, 1995).

There were some students, mostly eighth grade students, who constructed and interpreted their data according to the middle in the study of Mokros and Russel (1995). Researchers expressed that these students know that the middle value in the data represents the data given. Real life experiences and mean algorithm were used for checking their solutions (Mokros & Russel, 1995). One problem which the researchers observed with these students was that they were only successful with symmetrical distributions. They had difficulty in constructing non-symmetrical distributions (Mokros & Russel, 1995). The last approach revealed by Mokros and Russell (1995) was average as a mathematical point of balance. Mean is a balance point in the data set (Hardiman, Well & Pollatsek, 1984). In other words, total deviations above or below the mean are the same. Although there were no students in this study using this approach, some ideas were begun to be shown by two students, one sixth and one eight graders (Mokros & Russel, 1995). It was realized by the researchers that instead of equating deviations from the mean, these students balanced the totals below or above the means. Named by researchers as balancing total, the sum of the data points below or above the mean value was tried to be equated by these students. It seems that these students regard the mean as an equal sign. According to the researchers, although the ideas of these students were incorrect, they were looking for a balance point considering all of the data points. Moreover, these students were the ones who understood the data-total-mean relationship better (Mokros & Russel, 1995). Researchers asserted that the idea of representativeness is a foundation for the concept of average and it seems that students lose their informal knowledge of representativeness when they learn to find the mean. It was suggested that lessons should be designed so that they can connect new knowledge with their informal ideas of representativeness. If students mostly focus on procedures, they should be directed to summarizing and comparing data sets (Mokros & Russel, 1995).

Another conceptual study regarding to the concept of average similar to that of Mokros and Russell (1995) was designed by Watson and Moritz (2000). Questions like "Have you heard about the word average before?" and "What does an average of 3 hours of TV per day mean?" form the difference from the study of Mokros and Russell (1995). Even though it was not stated directly, in this research, the statistical literacy of students regarding the concept of average was analyzed according to the three-tiered framework, with specific focus on Tier 1 and Tier 2. Students' grades were between 3 and 9. Moreover, this study was a longitudinal one since the researchers wanted to examine students' understanding of average over time. Contexts in the study required students to use arithmetic mean in response to the results of the studies where students understood the average concept generally in terms of middle and most (Mokros & Russell, 1995). It was concluded that as students grew older, they began to realize the concept of average as a summary statistics representing a data set, which is a consistent result with that reported by Mokros and Russel (1995). In each grade, average was seldom described as representative. Instead, either colloquial language was used, like the average is "Okay..." or the ideas about middle or the most was mentioned in the questions asking for the meaning of average, as in the study of Mokros & Russell (1995) (Watson & Moritz, 2000). It was concluded that students realized the mean only as an add and divide algorithm, which is a consistent result with that reported by McGatha et al. (2002). However, most of the students beyond Grade 6 could not use the algorithm in a more complex problem solving task (Watson & Moritz, 2000). It was also observed that if students understand how to use the algorithm in a difficult setting, they can use that understanding in most of the situations without difficulty. Moreover, researchers believe that students' ideas of middle and the most about average could be a good starting point for classroom discussions. They also recommend delaying the teaching of mean until students develop strong informal understandings as with Backer and Gravemeijer (2004).

In another study, Watson and Moritz (1999b) examined students' statistical literacy related to the third tier of statistical literacy. They provided students with an authentic report from the media. In the report, a median house price and an average wage earner were the two statements that researchers focused on. First of all, students were asked for the meaning of average and median in the report (Watson & Moritz, 1999b). Then, the researchers asked the reason underlying the use of median related to the third tier of statistical literacy. While students described average the same way it was done so in the study of Watson & Moritz (2000) mentioned above, very few students could mention the effect of prices of cheap or expensive houses in cases where the mean was calculated (Watson & Moritz, 1999b). Researchers suggest explicit instruction regarding the effectiveness of different measures of central tendency in various contexts.

Conceptual studies about the concept of average continued in the year 2002 with the study of Konold and Pollatsek. Researchers offered four conceptual perspectives for the concept of average by extending the work of Mokros and Russell (1995). They are average as typical value, fair share, data reducer and signal amid noise. Appreciating variability as noise in a data set, researchers believed that average is a signal in this noisy environment, adding a new notion to the literature. This perspective was regarded as the most important among others since this conception is most useful in comparing data sets. Furthermore, comparing data sets is a good introduction for the concept of average (Konold & Pollatsek, 2002).

In the literature, there are also some studies to develop students' conceptions related to the concept of average. For instance, taking into consideration the above mentioned advice made by Mokros and Russell (1995) and by Watson and Moritz (2000) in delaying the teaching of the mean, Makar (2013) developed a learning trajectory to develop third grade students' conceptions of average. Students' conceptions of average were tried to be developed not with the comparison of data sets as suggested by Konold and Pollatsek (2002), but through informal inference.

It was observed that in answering the question “Is there a typical height for year 3 students? If so, what is it?”, students developed their conceptions starting from average as a reasonable approach stated in Mokros and Russell (1995). Students reached the idea of average as representative passing from the ideas average as mode and middle (mid-range) value, respectively.

The effect of different contexts on students’ understanding of average has also been investigated in recent years (Watson, Chick, Callingham, 2014). Concentrating on the second tier in the framework of Watson (1997), students were requested to interpret average in two different contexts, namely weather and home prices. The context of home prices was the one stated in the study of Watson and Moritz (1999b). One of the questions regarding the average number of children in a neighborhood asked for a specific numerical answer by using the procedure for finding the mean. Participants of the study were 247 students whose grades varied between 6 and 11. The performance of students showed variation across two different contexts which asked for the meaning of average. This showed that different contexts have different effects (Watson, Chick, Callingham, 2014). One of the expectations of researchers from this research was that there should be some kind of relationship between knowing the median and applying the mean. However, such a relationship could not be observed. Furthermore, it was concluded that performances of male students were higher in all of the questions than those of females but this difference was not statistically significant, which is contradictory with the results of Yolcu (2014) and Carmichael and Hay (2009).

The above mentioned studies were all about students’ understanding of average. There were also studies in the literature focusing on teachers with respect to the concept of average. It was found that teachers’ understanding of this concept was not much different from the understanding of the students (Jacobbe & Carvalho, 2011). Leavy and O’Loughlin (2006) observed that among teachers, the mean was the most common idea related to the concept of average. Furthermore, it was concluded that it was not easy for teachers to differentiate mean from median and mode (Leavy & O'Loughlin, 2006). Similarly, Hobden (2014) identified the statistical literacy levels of prospective teachers who were involved in an experiment for a new course in South Africa, Mathematical Literacy. Researcher tried to determine statistical literacy of the teachers regarding the concept of median within the context of HIV/AIDS survival times. It was found that the statistical literacy levels of teachers were low and more than half of the teachers did not have any idea regarding average or median. Some teachers interpreted median survival time as maximum survival time (Hobden, 2014). The results were surprising for the researcher since the teachers had completed a data handling module involving the concept of median just before the study. According to the researcher, these results reflect the gap between conceptual and procedural understanding of statistical concepts.

In conclusion, there are various studies regarding the concept of average in the literature. These studies revealed that students have some deficiencies regarding the concept; for instance, they do not know the representative nature of the average or they could not use average while comparing two data sets. However, contexts have a huge effect on students' understanding of average (Watson et al., 2014). Therefore, in the present study, students' statistical literacy, more specifically their interpretation and critical evaluation of the concept of average, was examined by presenting students some alternative contexts. Furthermore, when three measures of central tendency, namely mean, mode and median, are appropriate to be used, they are one of the foci of this study. In the literature, there are various studies examining whether students will realize the need to use the median when there is an outlier in the data. However, in the accessible literature, studies analyzing whether students could use the mode of the given data when the data were categorical was limited. Furthermore, in the studies regarding the use of the median in case of an outlier, data was generally presented directly to the students. In the present study, data were displayed on bar and line graphs, which in turn, could have an effect in realization of the outlier.

In the section that follows, some studies related to students' understanding of the concept of variation, which is another concept this study concentrated on, are summarized.

2.3.2 Research on Students' Understanding of Variation

Research on students' understanding of variation began to become more widespread when Shaugnessy, in his first review of statistics in 1992, stated the need for students' understanding of variability in addition to their understanding of measures of central tendency (Shaugnessy & Ciancetta, 2002). One of the earliest studies regarding the concept was assessing students' ideas of variation in a sampling environment (Shaugnessy, Watson, Moritz & Reading, 1999). The Gumball Problem in the National Assessment of Educational Progress (NAEP) exam in 1996 is a motivation for the researchers. The Gumball Problem can be observed in Figure 2.3.

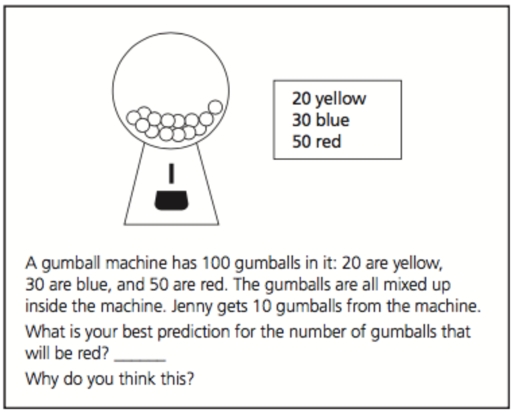


Figure 2.3 Gumball problem (NAEP, 1996)

It was observed that students give only one answer to this problem instead of suggesting an average. Shaugnessy et al. (1999) modified this problem into a similar task, namely the Lollie task, so that students could offer a range of possibilities. In this way, they could analyze students' thinking on variation in a sampling situation. Results showed that students offered lists with high, low, wide, narrow, and reasonable variation (Shaugnessy et al., 1999). It was stated that students presenting high variability believe that this is because there are more of the red ones. On the other hand, students, mostly 4th graders, believed that since there were more non-red ones, the list should involve less red ones (Shaugnessy et al., 1999). Furthermore, it was observed by the researchers that most 12th grade students presented a short list for choosing 10 lollies of five people such as 5, 5, 5, 5, 5 because 5 is the most likely outcome. These results suggested that students lose their intuitive ideas of variation in the school years just as they lose their ideas of representativeness when they learn the concept of mean (Mokros & Russell, 1995).

In later years, several research studies were conducted based on the Lollie task to analyze students' understanding of variation (Reading & Shaugnessy, 2000; Watson & Kelly, 2002). Reading and Shaugnessy (2000) developed a hierarchy based on students' description of variation as a result of the analysis of the interviews with 12 students by means of the Lollie task in 2000. The Description hierarchy can be seen in Table 2.3.

Table 2.3 Description hierarchy of Reading and Shaugnessy (Reading, 2004, p.87)

Levels	Focus of Responses
D1 - Concern with Either Middle Values or Extreme Values	Describe variation in terms of what is happening with either extreme values or middle values. *Extreme Values* are used to indicate data items that are at the uppermost or lowest end of the data, while *Middle Values* indicate those data items that are between the extremes.
D2 - Concern with Both Middle Values and Extreme Values	Describe variation using both the extreme values and what is happening with the values between the extremes.
D3 - Discuss Deviations from an Anchor	Describe variation in terms of deviations from some value but either the anchor for such deviations is not central, or not specifically identified as central.
D4 - Discuss Deviations from a Central Anchor	Describe variation by considering both a centre and what is happening about that centre.

Reading and Shaugnessy (2000) expressed that at the most basic level, students describe variation with either middle or extreme values. At the next level of sophistication, students begin to use both middle and extreme values (Reading & Shaugnessy, 2000). It was observed that while students start to look at deviations from a value which is not determined at the third level, at the last level they describe variation in the data using deviations from a center.

Different from the above studies, in 2002, students' understanding of variation was also examined in a probability environment by Shaugnessy and Ciancetta. Their motivation was the result of a question again in NAEP (1996), which can be seen in the figure below.

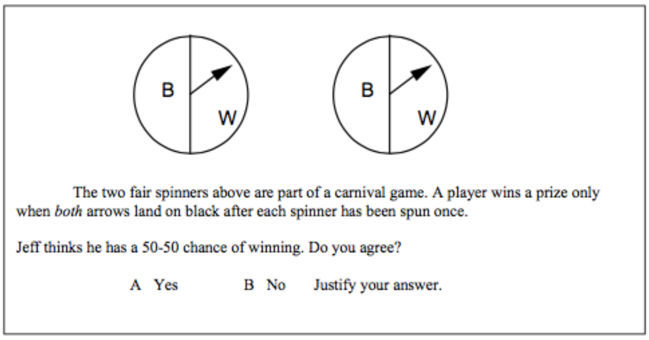


Figure 2.4 Spinner task (NAEP, 1996)

In this exam, it was observed that only 8% of the 12th grade students had given a correct response and justified their response correctly. Researchers considered that this could derive from the fact that students enrolled in the exam had not taken any mathematics lessons for several years. Subsequently, researchers decided to apply this task to students taking different levels of mathematics whose grades varied between 6 and 12. The results showed that when students' involvement in mathematics increased, their success in the spinner task also increased (Shaugnessy & Ciancetta, 2002). It was concluded that 97% of 11th and 12th grade students who had taken the calculus course correctly responded to the question in the spinner task by providing accurate justifications to their reasoning. It was observed that the performance of students who had taken intermediate level mathematics was just like that of the students in the NAEP (1996). They could not list the sample space for this probability task (Shaugnessy & Ciancetta, 2002). Researchers thought that playing the game may allow these students to observe variation and then provide the sample space. Therefore, they also investigated whether or not playing these spinners and gathering real data had an effect on students' understanding of variation. For this analysis, researchers interviewed with 28 students, who had not participated in the survey. It was found that while only 4 students could list the sample space before playing, this number increased to 12 out of 28 after playing with spinners (Shaugnessy & Ciancetta, 2002). From this result, researchers reached the conclusion that when students observed the variation in probability environments, they could compose the sample spaces themselves. Indeed, several research studies revealed that if students did their own experiments, either forming their own sample spaces as in the above study or trying to compare different groups, their appreciation of variation was developed much more compared to the beginning (Shaugnessy, Watson, Moritz & Reading, 1999; Watson, Skalicky, Fitzallen, Wright, 2009; Ben Zvi, 2004).

On the other hand, in 2003, Watson, Kelly, Callingham and Shaugnessy developed a survey covering the data and chance curriculum in which variation occurs automatically believing that to be able to understand students' understanding of variation, their understanding of change should be measured while measurements are made or events occur rather than assessing their understanding of standard deviation. They analyzed variation in chance, sampling and data handling contexts separately, and most of the questions were appropriate to the statistical literacy framework of Watson (1997). From the analysis of the results, four levels of understanding of variation, which fit well into the statistical literacy framework of Watson (1997) and that of Luke and Freebody (1997) were determined (Watson et al., 2003). The levels are presented in Table 2.4.

Table 2.4 Relationship of levels identified with the statistical literacy hierarchy (Watson et al.,2003, p.20)

Variable mapping (figure 1)	Luke and Freebody [41]	Statistical literacy hierarchy [40]
Level 4 – Critical aspects of variation: Employing complex justification or critical reasoning	Text analyst Critical practice	Tier 3 – Ability to question claims
Level 3 – Applications of variation: Consolidating and using ideas in context, inconsistent in picking salient features	Text user Pragmatic practice	Tier 2 – Application in context
Level 2 – Partial recognition of variation: Putting ideas in context, tendency to focus on single aspects and neglect others	Text participant Text-meaning practice	
Level 1 – Prerequisites for variation: Working out the environment, table/simple graph reading, intuitive reasoning for chance	Code-breaker Coding practice	Tier 1 – Language, definitions/processes

Levels in Table 2.4 were different from the levels in the description hierarchy of Reading and Shaugnessy (2000). It was stated that at the first level, students only recognize variation in simple contexts and they generally depend on their experience in the justification of the results. At the second level, students begin to use qualitative chance statements; however, their reasoning still does not reflect understanding of variation (Watson et al., 2003). Furthermore, it was observed that the word variation was familiar for students but they could not define the concept. Researchers expressed that they started to apply variation at Level 3. For instance, their offerings for 60 dice throws involved variation but that variation was not appropriate (Watson et al., 2003). Researchers observed that at the last level, students' understanding of variation was complete. In other words, they could present appropriate variation for spinner and dice tasks, and define the concept of variation correctly without depending on any example; in addition, they knew the significance of varying values in the computation of a mean (Watson et al., 2003). As previously stated, it was asserted by the researchers that levels identified a good fit into the framework of Watson (1997). Although Level 2 and Level 3 could match with Tier 2 and Level 4 with Tier 3, understandings of basic concepts in statistics and probability, which are requirements for Tier 1, began at Level 3 and were completed at Level 4. Therefore, it was thought that Tier 1 that matched with Level 1 in Table 2.4 can be placed at Level 3 and Level 4.

In addition to students' understanding of variation in a sampling or probability environment, their understanding was also assessed in a natural environment, namely the weather context, believing that weather is a context, in which everyone appreciates the variation (Watson & Kelly, 2005; Reading, 2004). Watson and Kelly (2005) interviewed 73 students from prep school to grade 9 using a Weather protocol for data collection. An average temperature of a city in Australia was given in the protocol and then some questions based on this average temperature were asked (Watson & Kelly, 2005). The questions in the protocol asked students to explain what the given average temperature tells about the temperature for that city, to suggest temperatures for different months in the year, to generate a graph for the given average and to analyze different graphs (Watson & Kelly, 2005). Older students were also asked to define variation. Questions in the protocol show that statistical literacy of students regarding the concepts of variation were analyzed in accordance with the statistical literacy framework of Watson (1997). Researchers presented the results of the study through four variables in accordance with the protocol. They are explanation, suggested data, graphing and definition (Watson & Kelly, 2005). It was observed by the researchers that the performance of students increased across the grades for all variables except for explanation. The results showed that students' explanations of variation when an average is given improves up to 7^{th} grade but then their performance drops (Watson & Kelly, 2005). Researchers believed that this reflects the inadequate connection between the concepts of average and variation in classrooms. Researchers asserted that the performance of students was better on graphing. Moreover, relationships between different pairs of variables were examined in the study, and although the association between pairs was not as much as expected, the strongest relationship was between the definition and graphing variables (Watson & Kelly, 2005). Furthermore, it was concluded by the researchers that if students could suggest appropriate varying data to a given average, their performance on graphing was better. Researchers stated that the difference in performance was not observed when students' understanding of variation was compared with their understanding of variation in a sampling or probability environment. At the end of their articles, researchers suggested that presenting students with a table showing daily temperatures of a month and asking them to summarize the information, or presenting students graphs with different variations about the same mean could be used to develop students' explanations of variation.

The weather context was also used by Reading (2004) to examine how students describe variation in an inference task. They gave students data related to weather, and asked them to suggest a month for a new celebration. Based on the hierarchy of Reading and Shaugnessy (2000) for description of variation in a sampling environment mentioned above, Reading (2004) examined students' answers and developed this hierarchy further (see Table 2.5). As can be observed in the Table 2.5, the results of the research showed that students' description of variation needs does not only quantitative but also qualitative explanations (Reading, 2004). According to the researcher, these qualitative descriptions are statistically less complicated than the first two levels of the description hierarchy of Reading and Shaugnessy (2000). Besides, the second cycle was very similar to the first two levels of the description hierarchy. Reading (2004) stated that the absence of more statistically complicated answers, that is, answers in the third and fourth levels in the hierarchy of Reading and Shaugnessy (2000), did not make it possible to refine those levels. That is why, Reading (2004) suggests refinements of these levels by studying with more advanced students. Developing students' description of variation using graphical representations and considering the influence of average when describing variation were among the other suggestions of the researcher.

Table 2.5 Refined version of the description hierarchy of Reading and Shaugnessy (Reading, 2004, p.97)

First Cycle Qualitative Responses	**element - qualitative feature of variation of data**
U1 - unistructural - one qualitative feature of variation	magnitude related - in an absolute sense to give indication of size of change, e.g., *pretty much consistent* or arrangement related - in a relative sense to give position, e.g., *spread out pretty evenly*
M1 - multistructural - more than one qualitative feature of variation	limiting related - set limits on the data values, e.g., *doesn't get too hot* and/or sequential related - deal with data item by item, e.g., *lots of dry days then a couple of wet days then a lot of dry days again*
R1 - relational - link qualitative features of variation	link the general limit with the discussion of blocks sequentially, e.g., *seems to fall pretty regularly but the amounts are not too much .. main pattern seems to be a short spell of dry days (3-5days) and then 1 or 2 wet days but rain is pretty light and not a large amount falls…*
Second Cycle Quantitative Responses	**element - quantitative feature of variation of data**
U2 - unistructural - one quantitative feature of variation	based on extreme values - discuss maximum, minimum, range or interior values - refer to blocks or patches of days
M2 - multistructural - more than one quantitative feature of variation	based on extreme values and/or interior values, e.g., refer to range but also to the rise and fall of temperatures throughout the month
R2 - relational - link quantitative features of variation	linking of extreme values and interior values may suggest immature notions of deviations, e.g., discussions including day-to-day deviations or 'averaged' deviations from day-to-day

The literacy aspect of the concept of variation was also analyzed by several researchers in the literature. For instance, Watson and Kelly (2008) analyzed personal concepts definitions of 3, 5, 7 and 9th grade students regarding the concept of variation as well as the concepts of random and sample. Researchers also provided students with instruction to examine whether their ability to define statistical vocabulary improved or not. The analysis of the results revealed that students' use of statistical vocabulary is very limited, showing their lack of interest in literacy skills in describing mathematical ideas in the middle school years (Watson &Kelly, 2008). Many students could not define the three terms and some of them confused the words with the words that they used in everyday language (Watson &Kelly, 2008). Nevertheless, researchers observed an improvement in the performance of students after they were provided with instruction. For instance, while students could not describe variation before instruction, they tried to define the same word as “a variation of something like food etc.” after the instruction (Watson &Kelly,2008). That is why, researchers stressed the importance of middle school years in the development of verbally expressing ideas and concepts. It was also shown that students could give examples of samples, random phenomena and things that vary, which could be a good starting point for classroom activities and the development of concepts.

Different from the above-mentioned studies, students’ understanding of variation was also measured in comparisons of groups (McGatha, Cobb & McClain, 2002; Shaugnessy, 2003b, Lewis, 2016). For instance, in a recent study in 2016, Lewis examined middle grade students’ understanding of variation by having them compare two box-plots. Giving students five-number summaries of distances of gummi bears from a catapult placed either in the back or belly, the researcher first asked the students to form box-plots and then compare the two plots to discover whether there is a significant difference or not. The results revealed that in comparing distributions, most of the students did not use the box-plots they had created nor any of the five-number summaries given to them (Lewis, 2016). The researcher observed that some of the students used ideas regarding variation using qualitative descriptions like “Both plots are equally consistent.”

In addition, the study of Gal, Rotschild and Vagner (1989) was one of the earlier studies regarding comparisons of groups. The researcher presented students, 3rd and 6th graders, with 9 comparison questions from two domains, namely outcomes of a frog jumping contest and scores on a school test. In the analysis, researchers focused on the strategies students used when making comparisons, not mentioning students’ use of variation. The findings revealed that most of the students had difficulty in comparing groups having an equal mean, mode and symmetry of distribution (Gal et al., 1989). Furthermore, the researchers asserted that comparisons involving different group sizes were the most difficult for students.

The researchers identified three strategies, which are statistical, proto-statistical and other/task-specific methods. In statistical strategies, students take into account all of the data points, not the specific ones and they make their decisions by summarizing the data in each group (Gal et al., 1989). It was found that in proto-statistical strategies, students do not use all the information they have; instead, they make their decisions by only taking into account various characteristics of the data (Gal et al., 1989). For example, researchers observed that in some instances, third graders made their comparisons only by looking at the mode; they did not consider the actual values of the data. Other/task specific strategies included strategies like addition (Gal et al., 1989). Researchers also encountered qualitative explanations as in the study of Lewis (2016) mentioned above. To illustrate, "The frogs in this team are less consistent because they are spread out more than the other team" was one example of a qualitative explanation. These explanations were placed into the category of other/task specific strategies (Gal et al., 1989). Though not stated by the researchers, these students had qualitative appreciation of variation among data sets as found in the hierarchy of Reading (2004). Researchers asserted that while proto-statistical strategies are the strategies that were used by many students, statistical strategies were used by very few students. This result for the third graders was not surprising for the researchers since third graders had not taken any statistical course. However, even though sixth graders had learnt some concepts related to average, they could not apply this knowledge either.

A similar study to that of Gal et al. (1989) was the research of Watson and Moritz (1999). The context was the comparison of two classes on a test. The grades of the students varied between 3 and 9. The comparison groups can be seen in Figure 2.6. As with Gal et al. (1989), Watson and Moritz (1999) also focused on the strategies the students resorted to. To compare graphs, students used numerical, totaling or finding average, and visual strategies (Watson & Moritz, 1999). It was realized that some students focused on individual values in making their comparisons; for instance, one student said, "Brown got a 7 and no one else did." (Watson & Moritz, 1999, p. 154). However, similar to the results which the study by Gal et al. (1989) yielded, the researchers stated that very few students used average or computed mean in comparing two graphs. Specifically, 10% and 54% of the students using them were in Grade 6 and 9, respectively (Watson & Moritz, 1999). Researchers believed that lack of students' experiences in the use of mean in comparison of data sets could be the reason underlying this result. Moreover, although the number of students in Pink and Black class (see Figure 2.5) were different, students did not realize this and concluded that Pink class is better (Watson & Moritz, 1999). When the researchers asked how they would compare the two graphs if their sizes were different, students resorted to visual strategies.

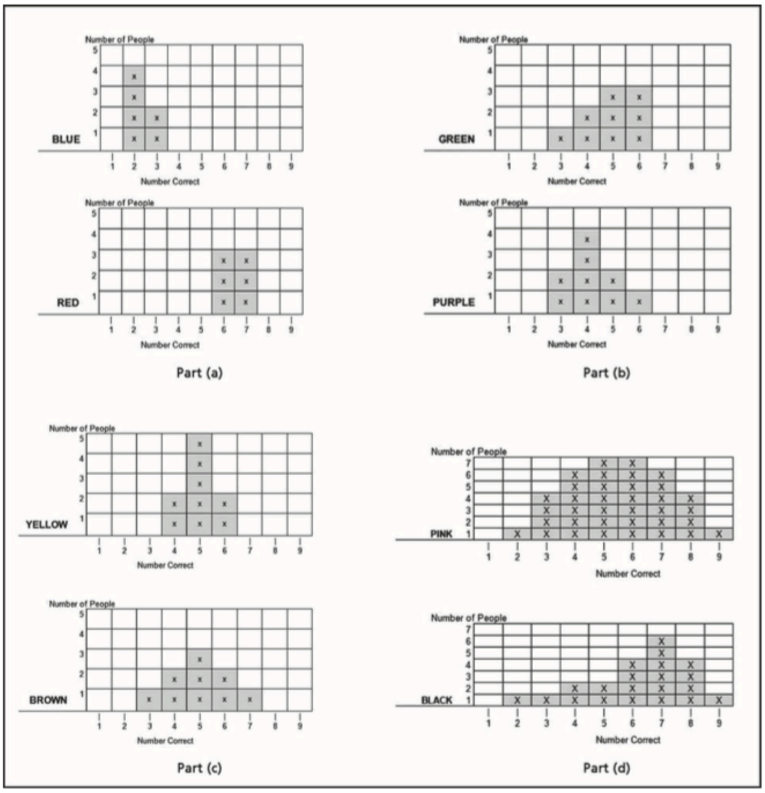


Figure 2.5 Comparison groups given to students in the study of Watson and Moritz (1999, p.151)

Different from the aforementioned studies of Gal et al. (1989) and Watson and Moritz (1999), Shaugnessy (2003b) and McGatha et al. (2002) concentrated on variation in comparing data sets. Shaugnessy (2003b) used comparisons of Yellow-Brown and Pink-Black classes (see Figure 2.6) with middle and high school students. Shaugnessy (2003b) stated that when comparing Yellow and Brown classes, students in both middle and high school used either average or variation in their comparisons; however, only 4 high school students could use both of them in their comparisons.

On the other hand, Shaugnessy (2003b) observed that students experienced difficulty in the Pink-Black class comparison as in the study of Watson and Moritz (1999). While most of the students decided that Pink class was more successful because of the more data values, some of them calculated the average or considered the distributional characteristics of the data, but they were few in number (Shaugnessy, 2003b).

Different from the aforementioned comparison studies, McGatha et al. (2002) provided students with two comparison tasks whose variations were different. One of the tasks was about choosing a basketball player for the final game of the year and the other was deciding on the month of a school trip when temperatures were given. In both situations, the researcher expected that students should be able to take consistency into consideration. It was observed that while three groups out of eight could consider variation in the basketball task, five out of seven could do so in the trip decision task. The other groups made their comparisons according to the totals or means. The study of McGatha et al. (2002) reveals the effect of context on students' consideration of variation in that context has different effects on students' understanding on average (Watson et al., 2014).

All of the studies mentioned above were related to the students' understanding of the concept of variation. In the literature, there are also some studies on the concept of variation in relation to the teacher. For example, Hammer and Rubin (2003) observed that in contrast to the students in the study of Watson and Moritz (1999), middle and secondary school teachers could use proportional reasoning when comparing two groups whose sizes were not equal. Furthermore, Canada (2004) conducted a qualitative study with thirty pre-service elementary mathematics teachers regarding their conceptions of variation and developed a framework including three elements: expecting, displaying and interpreting variation. Lastly, Makar and Confrey (2005) wondered whether pre-service teachers would only use mean when comparing two groups or they would also analyze variation and spread. The researchers observed that pre-service teachers used some statistical terms like mean, measure of spread, or they mentioned the shape of the data. It was also found that pre-service teachers used non-statistical language, such as clumps, chunks or spread out, which researchers believed could be the beginning of a more complete understanding.

To summarize, various studies are present in the literature regarding students' understanding of variation. Those studies found that students can describe variation both qualitatively and quantitatively, their understanding of variation can be improved with some experiments and they generally do not use variation when comparing two groups. However, in the study of McGatha et al. (2002), it was observed that different contexts have different effects on the consideration of variation in comparison situations. Therefore, in the present study, students' statistical literacy of variation, specifically interpretation and evaluation of the concept of variation will be investigated in different contexts. Furthermore, because of the positive effects of graphs on the realization of variation (Shaugness & Pfannkuch, 2002), students' understanding of variation was analyzed on graphical representations. Different from the studies in the literature, bar and line graphs were used in the current study. Lastly, taking into account the suggestion of Watson and Kelly (2005), the current study analyzed students' understanding of variation in comparison of two graphs with equal averages and different variations.

As stated previously, students' understanding of average and variation was examined on bar and line graphs in the present study. Therefore, the next section summarizes the results of some studies regarding graphs.

2.3.3 Research Studies on Graphs

Graphs are tools conveying important messages regarding data by using small amounts of place (Tortop, 2011) instead of huge descriptive writings (Özgün-Koca, 2001). With the use of graphs, some trends or irregularities are surfaced, which can be hidden in when data are presented in a table form, and help individuals in their predictions or decisions. This effect of graphs was clearly observed in the study of Shaugnessy and Pfannkuch (2002). In their study, graphing the data given in the table enabled students to discover the variation in the data and present good predictions. This feature of graphs makes them a popular tool in newspapers, textbooks or televisions (Shah, Freedman, & Vekiri, 2005); therefore, comprehending them is essential for individuals (Friel, Curcio & Bright, 2001; Shaugnessy, 2007). As such, graphs are given an important place in many curriculum documents, including the Turkish curriculum (NCTM, 2000; MNE, 2018)

Different from the other concepts in statistics, due to their vast amount of use, a lot of research exists regarding graphs (Shaugnessy, 2007). These research studies include reading and interpretation of different kinds of graphs, possible factors affecting graph comprehension, and difficulties and errors of students in graphs. For the purpose of this study, generally studies regarding bar and line graphs are mentioned in the following section, respectively.

One of the earlier research related to the the reading and interpretation of bar graphs was the study of Pereira- Mondeza and Mellor (1991). Researchers examined 4th and 7th grade students' ability in reading and interpreting bar graphs. They concluded that while students have a few problems regarding reading bar graphs, they could not make interpretations from the graphs given. Consistent with findings reported by other studies, it was observed that they could discover patterns in the graphs but could not make inferences beyond the data (Kirsch, Jungeblatt &Mosenthal, 1988; Bright &Friel, 1998, Capraro, Kulm, Hammer, 2005). On the other hand, Capraro, Kulm and Hummer (2005) provided students with data regarding the number and types of pets of 14 students, and they asked students to construct a graph which would enable them to find the typical number of pets in the data sets. Mostly bar graphs and pie charts were constructed by 46% and 35% of the students, respectively, and the others constructed line graphs (Capraro, Kulm & Hummer, 2005). However, researchers realized that students could not find the typical number of pets by interpreting their graph, which is an evidence indicating students' difficulties in interpretation. Furthermore, some students claimed that their graphs were histograms, but they were actually bar graphs (Capraro, Kulm & Hummer, 2005). Confusion with bar graphs and histograms was also observed in the study of Baker, Corbett and Koedinger (2002) and Whitaker and Jacobbe (2017). Whitaker and Jacobbe (2017) also indicated that 6 through 12 grade students had some misunderstandings regarding bar graphs and histograms. In 1998, Bright and Friel also investigated students' misunderstandings regarding bar graphs. Researchers observed that students did not use the height of the bars to determine the number of values; instead, they regarded them as a single value.

On the other hand, line graphs are representations revealing relationships between two continuous data with a line connecting two data points (Van de Wall, Karp, Bay-Williams, 2013). To show the general trend in the data, they can also be used in situations where two continuous data do not exist, but a discussion should be made with students regarding appropriateness of connecting data points (Friel & House, 2003). Few number of students constructed line graphs in the above mentioned study by Carparo et al. (2005). One of the reasons of this situation may be the type of the data presented. Organizing categorical data was revealed to be easier for students than organization of numerical data (Nisbet, Jones, Thornton, Langral & Mooney, 2003; Bright &Friel, 1998). Inability in distinguishing between categorical and numerical data could be another reason, which, in turn, leads to difficulty in choosing the appropriate type of graph for a given data set (Shaugnessy & Zawojewski, 1999; Friel, Curcio & Bright, 2001; Brasell & Rowe, 1993). Line graphs are one of the commonly used graphs in science classes; therefore, studies related to line graph comprehension in science also exist in the literature (Boote, 2012; Boote & Boote, 2017). For instance, Boote and Boote (2017) revealed that sixth grade students could not interpret line graphs related to science since they had difficulty in reading data in line graphs. Furthermore, it was shown that students' prior experience with other types of graphs, namely bar graphs, scatter plots and time-series, had both a positive and a negative effect on students' interpretation of line graphs.

Different from the above mentioned studies, students' statistical literacy regarding graphs were also examined in the literature. Watson and Moritz (1997a) investigated students' statistical literacy in graphs in accordance with the three-tiered framework by presenting them some graphs from newspapers. Results showed that only 10% of the students questioned a misleading graph. They used their personal experiences instead of the information in the graph when they were required to make a calculation. Another study of Watson (1997) revealed that while 9th grade students understood the information conveyed in a pie chart, 6th grade students had difficulty in doing so. A similar result to that reported by Watson and Moritz (1997a) was encountered in the study of Aoyama and Stephens (2003). The researchers found that although fifth and eight graders could understand the meaning conveyed through the graphs, they could not evaluate the information in the graphs critically.

All of the studies mentioned above were about students. In the literature, there are also several research investigating teachers' graphing abilities (Nisbet, 2001; Monteiro & Cainley, 2007). For example, Nisbet (2001) observed that similar to the students in the study of Nisbet et al. (2003), teachers also had difficulty in organization of continuous data. Furthermore, Monteiro and Cainley (2007) found that most of the pre-service teachers could not critically evaluate the graphs presented from media. Just like the students in the study of Watson and Moritz (1997a), pre-service teachers made their interpretations not only depending on the graphs given but also using their own experience.

To summarize, there are various studies in the literature regarding graphs. These research showed that students generally could read the data from the graphs easily, but they had difficulties in making interpretations. Several research indicated that students are more familiar with bar graphs and they had difficulties in choosing the appropriate graph type. On the other hand, Bright and Friel (1998) claimed that graphs can have different effects on students' interpretation of the data presented to them. Furthermore, several researchers asserted that graphical representations could have an effect on students' interpretation of average and variation. Therefore, the current study will investigate students' statistical literacy on graphical representations. The current study also examined statistical literacy of students regarding the concepts of average and variation in comparative situations. Graphical representations that students learn up until seventh grade are bar and line graphs and pie charts. Since it is believed that bar and line graphs are appropriate for comparison of two data sets, students' statistical literacy related to the concept of average and variation will be examined on bar and line graphs.

There are also studies in the Turkish literature related to the statistical literacy and the average and variation concepts that the current study concentrated on. Therefore, the next section summarizes the studies in the context of Turkey.

2.4 Research on Statistical Literacy in Turkey

In the Turkish middle school curriculum, there is a learning area named as data analysis, which includes objectives related to several statistical concepts. Some of those objectives are constructing and interpreting pie charts, and bar and line graphs, calculating and interpreting measures of central tendency and comparing two data sets. Furthermore, the Turkish curriculum gives high importance to interpreting statistical concepts in real-life contexts and making decisions based on these interpretations (MoNE,2018). However, there were not many studies regarding statistics until the call of Ulutas and Ubuz (2009). After their call regarding the necessity of studies regarding statistics and probability in the context of Turkey, studies related to statistics and statistical literacy began to improve.

In the Turkish literature, there are some studies focusing on the concept of average (Ucar & Akdoğan, 2009; Enisoğlu, 2014). Ucar and Akdoğan (2009) investigated middle school students' understanding of the concept of average by applying a test to a total of 18 students from grades six, seven and eight, 6 from each grade. It was found that while half of the students were aware of the representative value of the average, the other half interpreted it just as an algorithm, just like in other studies (Mokros & Russell, 1995; McGatha, 2002; Watson & Moritz, 2000, Enisoglu, 2014). However, the percentage of students appreciating the representative role of the average (50%) is high when compared to the findings of other studies.

On the other hand, Enisoglu (2014) concentrated on the three measures of central tendency: mean, mode and median. She analyzed solution strategies, errors and misinterpretations of seventh grade students in solving questions regarding measures of central tendency. Questions were asked through bar graphs. It was found that in finding a mean or constructing a data set when the mean was given, using the average formula was the most common strategy that seventh grade students used. It was also observed that some students used the balance model or the guess and check strategy. Enisoglu (2014) asserted that using numerical procedures was very common in finding the mode and median of a data set. When data were given in a graphical form, students depended on graphs, but in other times, they depended on numerical strategies (Enisoglu, 2014). Moreover, the researcher encountered many errors related to all concepts. For example, it was seen that some students only found the sum of the values and stated this sum as the mean of the data set as in the study of Watson and Moritz (2000). Besides, Enisoglu (2014) observed that generally seventh grade students considered the average to be equal to the mean of a data set. She believed that the reason of this result could be the name of the mean, i.e. arithmetic average, in the Turkish curriculum. It was also revealed that students had inadequate knowledge regarding when to compute and use the average of a data set. Students had the idea that when a datum is removed from the data set, the average will automatically decrease (Enisoglu, 2014). The researcher asserted that these results indicate that seventh grade students do not have a conceptual understanding regarding the concept of average.

There are also studies in Turkish literature regarding statistical literacy (Yolcu, 2012; Koparan, 2012; Koparan & Guven, 2014; Yolcu, 2014, Ozen, 2013; Gunduz, 2014). For instance, Yolcu (2012) identified the statistical literacy level of 1074 eighth grade students enrolled in public schools based on Watson's three tiered statistical literacy framework (1997). A Statistical Literacy Test was developed by the researcher. Items in the instrument was about all the concepts in the data and chance curriculum that Watson (2006) mentioned in her model. Furthermore, the items in the instrument were from different social contexts, such as health and media. The results of the study showed that the statistical literacy level of eighth grade students, who are very near graduation, is low (Yolcu, 2012). It was observed by the researcher that while students' performance was the highest in the second tier, their performance was low in the first and third tier, and lowest in the third one. This finding was similar to that reported by Watson and Callingham (2003). When Yolcu (2012) examined students' statistical literacy level in terms of the specified concepts, she reached conclusions similar to those reported in the literature. For example, students do not understand the concept of average as a representative value (Yolcu, 2012). Instead, they understand it as "add them up and divide algorithm" (Uçar & Akdoğan, 2009; Mokros & Russel, 1995, Watson & Moritz, 2000). It was observed in the research that students explained concepts of average and variation via measures of central tendency and spread, which the researcher interpreted as students having procedural understanding regarding these concepts. Among the other content domains, students' performance was relatively higher in graphs (Yolcu, 2012). The researcher assumed that this could derive from the fact that students in Turkey are very much exposed to graphs starting from pre-school to eighth grade. Related to the other purposes of the study, Yolcu (2012) concluded that students have positive attitudes towards statistics and there is a positive relationship between students' attitudes toward statistics and their statistical literacy level. Yolcu (2012) ended her study by recommending conducting research on specific concepts so as to make it possible to analyze them in more detail. Furthermore, conducting research to investigate the role of gender on the statistical literacy level of middle school students from grades six to eight, Yolcu (2014) showed that statistical literacy level of students did not change significantly by grade level, which is a conflicting result to that reported by Watson and Moritz (2000). In addition, it was found that there was no interaction between grade level and gender, but female students were generally better than male students (Yolcu, 2014).

Moreover, since several studies revealed that statistical literacy level of students was not as expected, Koparan (2012) designed an experimental research to develop the statistical literacy levels of students. A statistical literacy test was prepared by the author. Most of the questions were from the studies of Watson and Callingham (2003). While the project based learning approach was used in teaching of statistics concepts in the experimental group, the control group was taught via the traditional approach (Koparan, 2012). In the traditional approach group, some questions were solved after they were taught some basic concepts; on the other hand, in the experimental group, after the basic concepts were taught, the students conducted research, which entailed the steps from defining a problem to arriving at conclusions (Koparan, 2012). The researcher used the statistical literacy construct of Watson and Callingham (2003) to determine the statistical literacy level of the students. It was observed that almost for all components of the statistical literacy, the project based learning approach had a positive effect. For instance, for the component of data representation, there was a statistically significant difference between the scores of the experimental and control groups (Koparan, 2012). It was concluded by the researcher that while many of the students were at Level 2 in the experimental group before the experiment, they were at Level 3 after the experiment. Besides, after the experiment, there were more students in the fourth level. The researcher asserted that one of the benefits of the experiment was that students could interpret representations properly and could draw a conclusion from them after the experiment. Furthermore, after the experiment, students began to look critically at the data representations (Koparan, 2012). The researcher believed that this could derive from the fact that they drew many representations, and examined and questioned other groups' representations throughout the experiment. It was also observed that students did not use names for axes and did not include every information in their data. Despite students' familiarity with the bar graphs, they had difficulty in reading and interpreting pie charts and line graphs (Koparan, 2012). Similar to the results of Shaugnessy and Zawajeski (1999) and Tortop (2011), the researcher stated that students could not decide about the effectiveness of the graphs in different contexts.

Some of the above mentioned studies were among the studies in Turkey which were related to the graphs. There are also other research studies regarding graphs in the Turkish literature (Enisoglu, 2014; Erbilgin, Arıkan, Yabanlı, 2015; Memnun, 2013; Tortop, 2011). For example, Tortop (2011) examined errors and misconceptions of seventh grade students on graphs and whether there was an effect of regular instruction in schools on these errors and misconceptions. It was observed that instruction was not effective in eliminating students' errors and misconceptions. Furthermore, the knowledge of the teacher regarding students' possible errors and misconceptions were not adequate (Tortop, 2011). Therefore, the researcher recommended putting more emphasis on the education of in-service teachers. Different from the study of Tortop (2011), Erbilgin, Arıkan and Yabanlı (2015) concentrated only on line graphs and developed a measurement tool to assess middle school students' comprehension of line graphs and their ability to construct line graphs. Their measurement tool assessed graph comprehension abilities stated by Curcio (1987); in other words, abilities related to reading data, reading within the data and reading beyond the data. Seventh grade students' ability to read and construct line graphs was also analyzed by Memnun (2013). It was observed that while most of the students could read data given in the line graphs, they could not use those data in solving questions. Furthermore, it was realized that some of the students constructed a bar graph instead of a line graph (Memnun, 2013).

2.5 Summary of the Literature Review

In the previous section, the summary of the literature related to the purposes of the present study was presented. In accordance with the purposes of the present study, first of all, definitions of the concepts of literacy and statistical literacy were stated. Then, some theoretical frameworks and models regarding statistical literacy were reviewed. Afterwards, research related to students' understanding of average, variation and graphs were summarized. Finally, studies in the context of Turkey were presented.

Firstly, both researchers and curriculum documents agree on the necessity of literacy and statistical literacy for effective citizenship. Hence, various research studies were conducted on statistical literacy and the average and variation, which are two concepts that this study focused. These studies revealed that students' understanding of average and variation was not at the expected level. For example, students could not use average or variation while comparing groups (Shaugnessy, 2003b; McGatha et al., 2002). However, many curriculum documents including the Turkish curriculum states the importance of interpretation of statistics, including average and variation, in real life contexts (AEC, 1991; MoNE, 2018; NCTM, 2000). For example, these documents mention the appropriate use of three average types in real life contexts: mean, mode and median. It was asserted that students should be able to select the appropriate average type in different real life situations as well as calculating them. Nevertheless, different real life contexts do not have the same impact on students' understanding of average and variation (McGatha, Cobb & McClain, 2002; Watson, Chick & Callingham, 2014). For instance, while students could interpret average as mean in one context, they could understand it as the mode of the given data in another context. Hence, the current study examines interpretation and evaluation of the concepts of average and variation presenting students some alternative contexts. Hence, this study has the potential to make a significant contribution to the existing literature by having revealed students' interpretations and evaluations related to the concepts of average and variation in different contexts.

Furthermore, to decide the appropriate use of three average type is one of the foci of this study, which indicates students have conceptual understanding regarding the concept of average (Enisoglu, 2014). In the literature, there are some studies investigating students' realization of the use of median when there is an outlier (Watson & Callingham, 2003; Watson & Moritz, 1999b). However, in the accessible literature, studies which provided students with categorical data; hence, necessitates using of mode as appropriate type of average were limited. Therefore, this study presented students a context involving a categorical data set, so students were expected to use the mode of the given data as the appropriate type of average. By this way, the current study can contribute to the existing literature in providing the information about whether students made use of the mode when there were categorical data which in turn can be useful in passing to the appropriate use of median where students had difficulty (Watson & Callingham, 2003; Watson & Moritz, 1999b).

On the other hand, there were differences in students' interpretations when the data were presented in a graphical form (Bright & Friel, 1998) Furthermore, Enisoglu

(2014) stated that graphical representations could have an effect on students' interpretation of the average concept. In studies regarding the use of median in case of an outlier in the literature, data were generally presented directly to the students. Different from those studies, data were displayed on bar and line graphs in the present study, which, in turn, could have an effect in the realization of the outlier. Hence, this study is important because it shows the effect of bar and line graphs on the realization of the outliers in the data set. Graphs also have positive effects on the realization of variation in the data (Shaugnessy & Pfannkuch, 2002), so the present study examined students' understanding of variation on graphical representations. Different from the studies in the literature, bar and line graphs were used in the current study thus, the current study provides valuable information about the impact of bar and line graphs on students' interpretation and evaluation of variation concept. In this way, bar and line graphs can be used in teaching the concepts of average and variation by teachers or teacher educators.

Furthermore, taking into account the suggestion by Watson and Kelly (2005), the current study analyzed students' understanding of variation in comparison of two graphs with equal averages and different variations. Several studies revealed that students are reluctant to use average and variation in comparing situations (Gal et al, 1989, Shaugnessy 2003b, Watson & Moritz, 1999). Since it was shown in the literature that it would not be used when it was not asked, the present study initially asked students to calculate an average while comparing two graphs different from the studies in the literature. Then, their understanding of variation was analyzed when they observed that averages are equal. In other words, this study is significant as it has observed whether students directly interpreted the variation of the presented data, which is the underlying reason for calculation of an average (Shaugnessy, 2007), while comparing two groups when they observed that the averages are the same. Hence, the findings of this study could provide important information to teachers, teacher educators or curriculum developers in that teaching of the concepts of average should be delayed until the students' understanding related to the variation concept complete.

In the context of Turkey, there are some studies analyzing students' understanding of average concept in the Turkish literature (Ucar & Akdogan,2009; Enisoglu, 2014) however, different from these studies, the current study investigated the interpretation and evaluation of the concept of average providing students some alternative contexts, which can reveal different results. Hence, this study is significant since it has the potential to make an important contribution to the existing Turkish literature by having revealed students' interpretations and evaluations related to the concepts of average and variation in different contexts.

Besides, there is not many comprehensive studies regarding the concept of variation in the accessible literature in Turkey. Also, research on statistical literacy was scarce in Turkey and studies should be conducted on specific concepts so as to analyse them in detail (Yolcu, 2012). In this respect, there is a need to conduct further studies related to the concept of variation and statistical literacy. Thus, this study investigated statistical literacy of seventh grade students in terms of the variation concept. By also attaching two graphical representations, namely bar and line graphs, the results of such a study were able to provide valuable information regarding whether students could interpret and evaluate variation on bar and line graphs and whether the statistical literacy of students related to the concepts of average and variation changed when data were presented on bar and line graphs.

Furthermore, studies regarding statistical literacy of students in Turkey, such as the study by Yolcu (2012), measured students' statistical literacy generally through multiple choice tasks. However, to reveal students' understanding and questioning ability more effectively, open-ended tasks should be used (Watson, 1997). Thus, it is essential to analyze statistical literacy of students through open-ended tasks to examine students' understanding and questioning ability related to the concepts of average and variation better. Therefore, using open-ended tasks, this study aims to investigate the statistical literacy of seventh grade students about the concepts of average and variation on bar and line graphs related to the data obtained from social or scientific contexts.

CHAPTER 3

METHODOLOGY

This chapter aims to explain the methodology used in the current study. Therefore, this chapter includes information about the research design, population and sample, data collection instruments, the validity and reliability of the instruments, data collection procedure, analysis of data, assumptions and limitations. The internal and external validity of the study is presented at the end of the chapter.

3.1 Design of the Study

The purpose of the present study is to analyze seventh grade students' statistical literacy related to the concepts of average and variation on bar and line graphs related to the obtained from social or scientific contexts. In this respect, the following research question and its sub-question were addressed in the study:

What are the statistical literacy levels of seventh grade students regarding the concepts of "average" and "variation" on bar and line graphs related to the data obtained from social or scientific contexts?

a. How do seventh grade students at different statistical literacy levels define, interpret and evaluate "average" and "variation" concepts on bar and line graphs related to the data obtained from social or scientific contexts?

Fraenkel and Wallen (2006) stated that survey research designs are very useful in describing aspects or characteristics of a population, such as abilities or knowledge. Therefore, to examine statistical literacy of seventh grade students, the survey research design was used in the study. Particularly, this study requires collecting data from the selected sample from at one point of time; hence, the design of the current study is a cross-sectional survey (Fraenkel & Wallen, 2006). Furthermore, Jansen (2010) mentioned two types of survey, namely quantitative and qualitative. He stated that if the aim is to determine diversity of any topic in a population, then the survey type is qualitative. Since the current study aimed to investigate the diversity of the students' definitions, interpretations and critical evaluations regarding the concepts of average and variation, design of this study was qualitative survey study.

3.2 Sampling Procedure and Participants of the Study

The target population of the study is all seventh grade students in Ankara. However, since it is not possible to reach the target population, all seventh grade students in Cankaya and Akyurt districts of Ankara constituted the accessible population of the study. The convenience sampling method was used in the study because this sampling method selects the available group for the study (Fraenkel & Wallen, 2006). The reason of selecting this sampling procedure is that the researcher was a middle school mathematics teacher of one of the public schools in the Akyurt district and the other public school determined in Cankaya was very close to the university where the researcher was enrolled as an M.S. student. One of these schools had five seventh grade classes and the other had three classes. Thus, the Statistical Literacy Test (SLT) was applied to 164 seventh grade students in Akyurt and Cankaya, Ankara. Some characteristics of the sample are displayed in Table 3.1 below. Table 3.1 indicates that the total number of boys and girls was 82. While the average age for participants was 12.57, the average of their mathematics grade at the end of the last semester was 70.31. Lastly, while the students in the Akyurt district of Ankara had a moderate level of socio-economic status, the socio-economic status of students in Cankaya district was higher.

Table 3.1 Characteristics of the sample group by gender, age and mathematics grade

Classes	Sample Size(n)	Age(years) Average	Gender		Mathematics Grade Average
			Boys	Girls	
7/A	20	12.69	20 (12.2 %)	0 (0.0%)	74.64
7/B	19	12.50	19 (11.6 %)	0 (0.0%)	64.43
7/C	33	12.65	0 (0.0%)	33 (20.1%)	83.33
7/D	28	12.52	0 (0.0%)	28 (17.1%)	59.34
7/E	21	12.59	21 (12.8 %)	0 (0.0%)	58.86
7/A	13	12.53	7 (4.3%)	6 (3.7%)	63.59
7/B	12	12.48	5 (3.0%)	7 (4.3%)	75.81
7/C	18	12.57	10 (6.1%)	8 (4.8%)	82.47
Total	164	12.57	82 (50.0 %)	82 (50.0%)	70.31

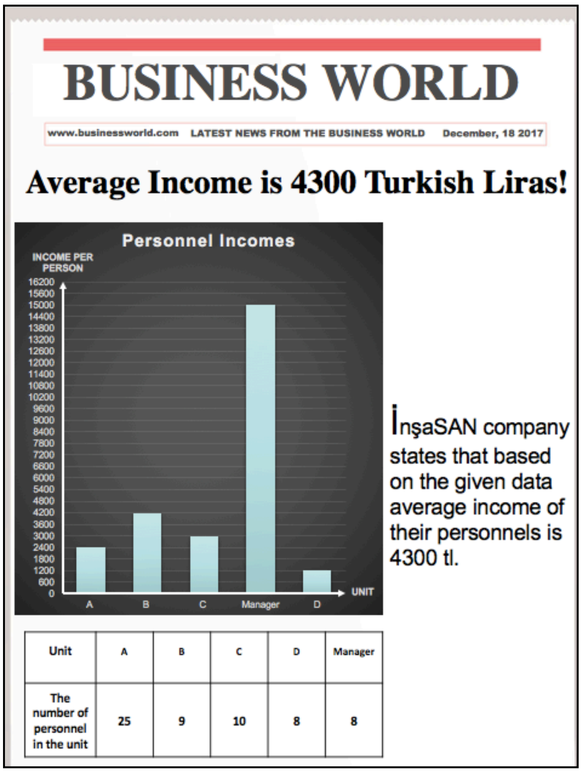
3.3 Data Collection Instruments

The present study intended to analyze the seventh grade students' statistical literacy in terms of the concepts of average and variation on bar and line graphs related to the data obtained from social or scientific contexts. The data of the study were collected through an instrument called the Statistical Literacy Test (SLT), developed by the researcher. In developing the instrument, three steps were followed. First of all, the objectives regarding the concepts of average and variation of fifth, sixth and seventh grade Turkish National Middle School Mathematics Education Curriculum were identified. The objectives related to bar and line graphs were also identified since statistical literacy of students regarding the specified concepts were tried to be determined on bar and line graphs. In accordance with the purpose of the study, Watson's Statistical Literacy Framework (1997) was taken into account in preparation of the instrument. Secondly, before developing the instrument, the related literature was reviewed and some questions were adapted from the literature. Lastly, the researcher prepared other questions so that there was at least one question for each of the objectives determined beforehand. Furthermore, the importance of context for statistical literacy was mentioned by many researchers, as it was stated in the previous chapter. Therefore, all of the questions were asked within a context. Before preparing the questions, appropriate contexts for the questions were investigated from the literature and several textbooks related to the concepts of the current study, and the contexts of the questions adapted from the literature were taken directly.

The SLT included 7 open-ended items. While questions 2 and 3 were adapted from the literature, the remaining questions were prepared by the researcher to analyze the research questions of the current study. In-depth information regarding all questions are provided below.

Question 1 (Q1), which was prepared by the researcher, aimed to analyze whether students could interpret the word 'average' in a context and evaluate the given average in the given context. In this question, students were given a piece of misleading information from a newspaper, business world. The news can be seen in Figure 3.1.

Fraenkel and Wallen (2006) suggest using median, the midpoint score in a data set, to find an average if there are extreme values in the data. There was also an extreme value, income of the manager, in the data set in the news given. However, the average income in the news was calculated using the mean, which is a type of average highly affected by extreme values (Fraenkel & Wallen, 2006). Therefore, this news was misleading and the question asked aimed at identifying whether students could evaluate the news and discover this situation.



Unit	A	B	C	D	Manager
The number of personnel in the unit	25	9	10	8	8

Figure 3.1 The news in Q1

Q1 involved 4 parts and they can be observed in Figure 3.2. The first part of the question required students to interpret the word ‘average’ in the news. To state it differently, this part assessed students’ interpretation of the word ‘average’ in a social context, a requirement for the second tier. The second part was added into the question after the pilot study. In the pilot study, it was observed that none of the students realized the high average because of the manager’s income in the third part of the question. Therefore, thinking of how the average in the news is calculated was thought to be useful for the third part of the question. The second part was also related with the second tier of the statistical literacy framework since it requires students to interpret the concept of average in the given context. Furthermore, since the first and second questions in Figure 3.2 were often used in the many studies of Watson to meet the requirements for the second tier, these questions were preferred to examine students’ interpretation of the average concept in different real life contexts.

a) What do you understand from the word **average** in the news?

b) How do you think the **average income** in the news was computed? Explain your answer clearly.

c) Suppose that you are working in a company where your income is 2900 TL. If you were placed in one of the units in the company mentioned in the news, would you give up your work taking into account the graph or average income and want to work in this company?

Yes/ No

- How did you reach this decision? Explain your answer clearly.

…………………………………………………………………………………………
…………………………………………………………………………………………
………………………………………………………………………………

d) Define the word "Average" which you use in mathematics class with your own words. Give an example of average.

Figure 3.2 Parts of Q1

The third part of the question was related to the third tier of Watson's hierarchical framework. In other words, this part examined whether or not students could critically evaluate the claim in the news, that the average income is 4300 TL. The question was prepared with the help of a question which is related to the third tier in the framework in the study of Watson (1997).

The last part of the first question asked students to define the word 'average' using their own words and wanted them to give an example related to the average. This question was a requirement for the first tier and was translated directly from the study of Watson and Moritz (2000). Although the three-tiered framework suggests Tier 1 before Tier 2 and Tier 3, the part related to Tier 1 was placed at the end in this question because understanding of the definitions is improved with their applications in various contexts (Watson, 2006).

Question 2 (Q2) asked students to investigate their interpretation of the concept of average and critical evaluation of the concept of variation. With this aim, students were required to compare two data sets whose averages were almost the same, but their variations were different. This question was adapted from the study of Mc Gatha et al. (2002). The original problem in Figure 3.3, named as Basketball All-Star, examined whether students took into account variation while comparing distributions in addition to the use of the mean.

One player will be selected from the Meigs basketball team to play in the all-star tournament. Below is a listing of the points scored by the top two candidates for the last eight games of the season. Based on this information, present an argument to support the selection of one of the players.

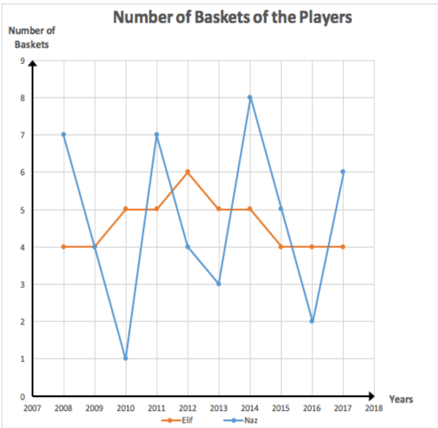
Player A: 11	*31*	*16*	*28*	*27*	*14*	*26*	*15*
Player B: 21	*17*	*22*	*19*	*18*	*21*	*22*	*20*

Figure 3.3 Original version of Q2 (McGatha et al., 2002)

In this problem, the player having a higher mean score, Player A, had higher inconsistency. In other words, there was a higher level of variation among his scores; therefore, he was not a suitable choice. Kazak (2016) translated this question directly into Turkish in her review on statistical reasoning. The Turkish version of the problem with some additional changes were used in the current study. Modifications for the problem were listed accordingly. First of all, the scores of the players were organized so that their means were almost the same. In this way, it could be possible to observe whether students would evaluate the variation while comparing two data sets. Moreover, several studies in the literature revealed that students generally do not use average in comparing distributions (McGatha, Cobb & McClain, 2002; Shaugnessy, 2003b). Therefore, in the first part of Q2, students were asked to calculate the average of two players and to observe that they were equal. By not asking students directly to compute the mean but the average, this part also investigated how students interpreted the concept of average in this basketball context. The second change was decreasing the number of baskets of the players in order not to trouble students so much with computation. Lastly, for the purpose of the study, the number of baskets was given on a line graph instead of in a table. The adapted version of the question is presented in the following figure:

Question 2:

The Trainer of young women's national team examines the scores of the players in the final games of the last 10 years to select players for the final game of this year. The following graph shows the scores of two players that the trainer was indecisive in their selection for the team.



a) Find the average number of baskets of two players in the last 10 final games. Show your work.

Average number of baskets of Elif:

Average number of baskets of Naz:

b) Trainer decided to select Naz for the final game of this year. Do you agree with the decision of the trainer?

Yes/No

- How did you reach this result? Explain your answer clearly.

....
....
....
....

Figure 3.4 Q2

Figure 3.4 shows that Q2 involves two parts. The purpose of the first part was explained in the previous paragraph. The second part of the question is related with the questioning ability of students to a decision given. A direct question regarding students' agreement with the decision of the trainer were asked to analyze whether students used variation while comparing data sets when the averages were almost the same. The reason of asking such a question was that such question structures generally were used in the interviews in the studies regarding statistical literacy to examine students' questioning ability.

Similar to Q2, Question 3 (Q3) was related to the concept of variation. In this question, students' interpretation of the variation concept was assessed in a natural context, namely weather, in which variation is appreciated by everyone (Reading, 2004). This question was adapted from the study of Watson and Kelly (2005). In that study, students' interpretation of the concept of variation was assessed by providing them with an average temperature for a city in Australia and asking various questions depending on this average, such as suggesting temperatures for 6 different days in a year or drawing a graph of the whole year. The original version of the question 3 is given in the following figure:

> Some students watched the news every night for a year, and recorded the daily maximum temperature in Hobart. They found that the average maximum temperature in Hobart was 17 C.
>
> How would you describe the temperature for Hobart over a year in a graph?

Figure 3.5 Original version of Q3 (Watson & Kelly, 2005)

Some changes were applied to this question. To begin with, the context of the question was changed into 'the temperature of Ankara', as it is believed that 'Ankara' is a familiar city for all the students participating in the study, and the question was given as a headline of a newspaper. Secondly, instead of providing them with an average of the whole year, an average of a month in a year, namely April, was given. The aim here was to eliminate possible confusions because of the variations within a month and variations among different months. Since it was observed that the temperature in the month of April fluctuates much in Ankara, this month was chosen for the context of the question. Lastly, students were provided with a graph paper and axes for them to draw a line graph. The adapted version of the question is portrayed in the following figure:

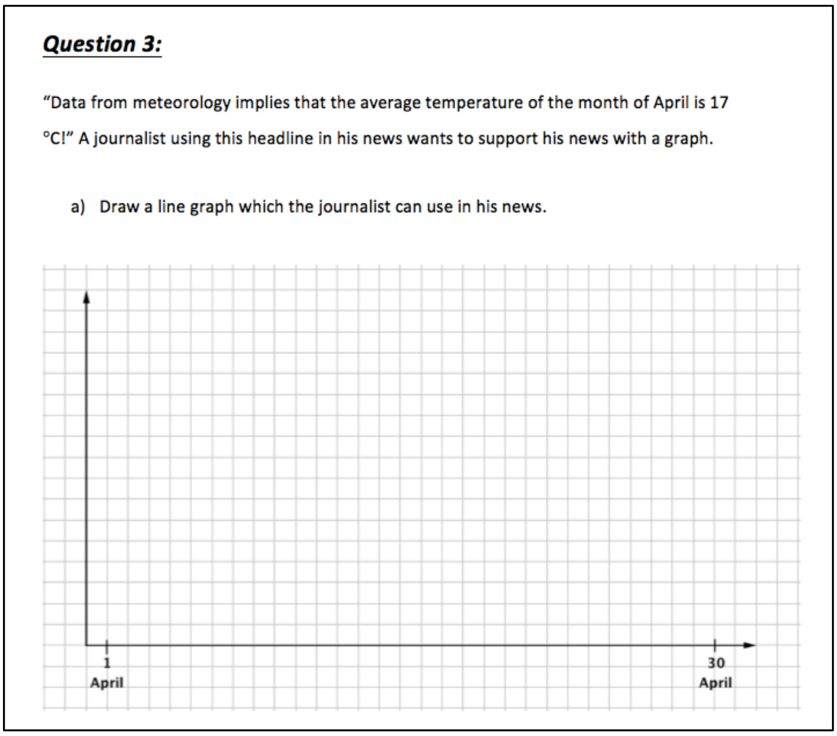
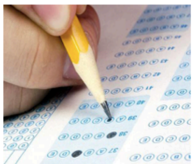


Figure 3.6 Q3

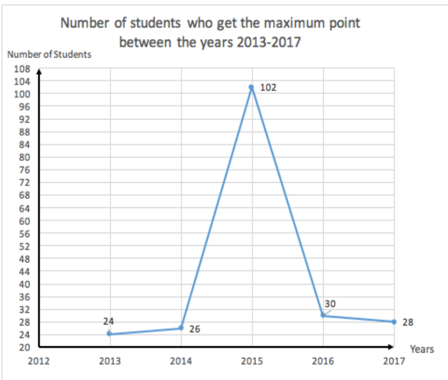
In the subsequent part, students were asked to interpret their graph, taking into consideration the given average temperature. The aim of this part was to understand how students draw graphs and, therefore, to examine their interpretation of variation to a given average. The last part of Q3 asks students to define the word 'range', which they learnt in mathematics class in their sixth grade. This part was to fulfil the requirement of the first tier related to the variation concept. Since the only objective related to the concept of variation was range, a measure of spread, in the middle school curriculum in Turkey, the definition of range instead of that of variation was asked. In other words, definition of the variation concept was examined with an indirect way by asking the definition of range concept.

The aim of Question 4 (Q4) was the same as that of Q1. In other words, this question analyzed students' interpretation of the concept of average in a different context. Furthermore, some misleading information regarding the average was presented through a brochure, and it was examined whether students could critically evaluate the given situation. The only difference between Q1 and Q4 was the type of the graphs. In Q4, data were presented through a line graph. This question also examined whether graphs have a role in the interpretation and critical evaluation of the concept of average. As can be observed in Figure 3.7, this question consists of three parts, and the purposes of these parts are the same with the those of the parts in Q1.

Question 4:



102 students get the maximum point in the foreign language exam in September! Average number in the last 5 years is 42 students! Come and succeed!



a) What do you understand from the word **average** in the brochure above?

b) How do you think **the average number of students** in the brochure was computed? Explain your answer clearly.

c) Suppose that you are now involved in a course that has an average number of 30 students getting the maximum point. If you knew that there was an equal number of students participating in each course in the given years, would you leave your own course <u>taking into account the broschure</u> and register to this course?

Yes/ No

- How did you reach this decision? Explain your answer clearly.

....
....
....
....

Figure 3.7 Q4

Question 5 (Q5), which was prepared by the researcher, was asked to the students in order to examine how students interpret and critically evaluate the concept of average. This question consists of two parts, and at the beginning, it presents students an interview from a magazine. The interview is provided in Figure 3.8.

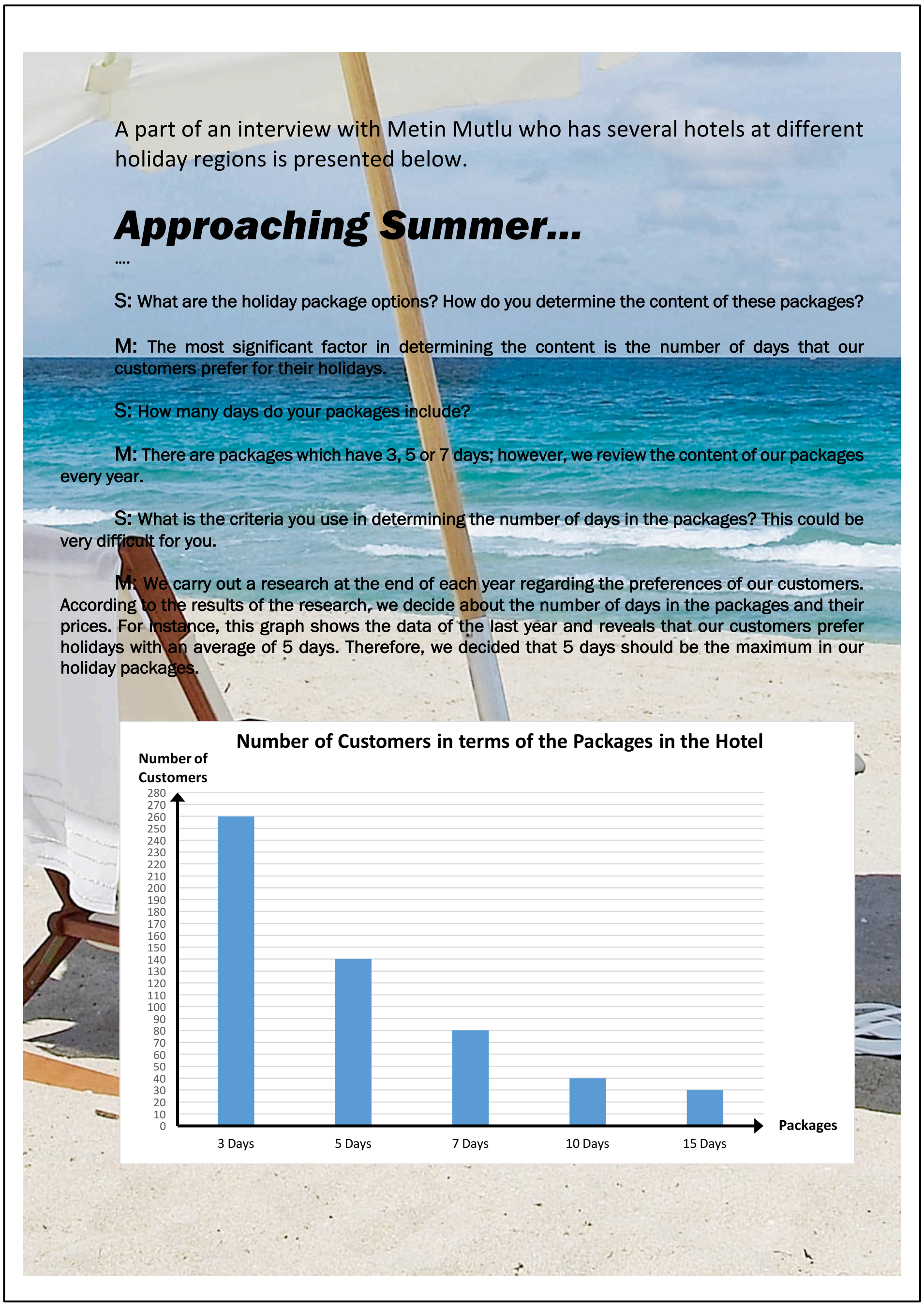


Figure 3.8 Part of an interview in Q5

Different from Q1 and Q4, the context of this question necessitates the use of mode as the type of average since data were presented through categories. Grawetter and Wallnau (2013) have stated that if the scale of measurement is nominal, that is, measurements are observed on a number of categories, the only average type to be used is mode. Moreover, if the number of days are computed for holiday packages by a travel agent, other measures will not be as useful as mode (Van de Wall, 2013). However, average in the interview was calculated using the mean. Q5 aimed to reveal whether or not students could critically evaluate this situation. The question entailed two parts. The first part asked students to think about how the average in the interview was calculated. The aim of this part was to help students in the following part regarding the third tier of statistical literacy. The second part investigated students' evaluation of a decision presented. Parts of Q5 are provided in the following figure:

a) How do you think the director has computed the average number of days as 5 by taking into consideration of the data in the graph? Explain your answer clearly.

b) Do you agree with the decision of the director regarding organizing a maximum of 5-day packages taking into consideration the above graph?

Yes/No

- How did you reach this result? Explain your answer clearly.

  ....
  ....
  ....

Figure 3.9 Parts of Q5

Similar to Q2, in Question 6 (Q6), students' critical evaluation of the concept of variation while comparing two data sets was assessed. The questions have the same purposes but the data in Q6 were given through a bar graph instead of a line graph. The aim here was to examine whether graphs have a role in the critical evaluation of the concept of variation. Besides, the importance of the consistency for the decision made was presented as a clue in the second part of Q6, stating that fluctuated wind speed increases maintenance cost. Q6 can be examined in Figure 3.10 below.

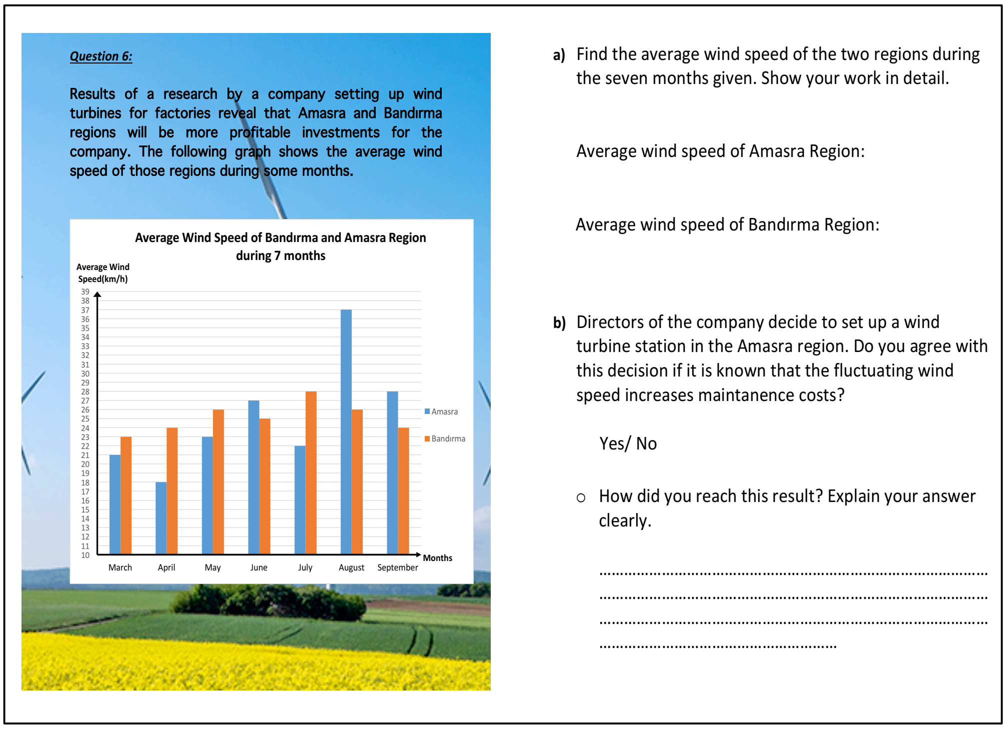


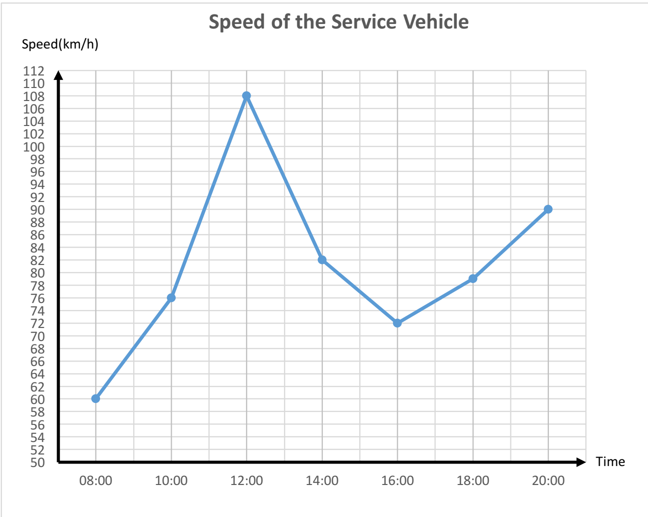
Figure 3.10 Q6

Lastly, the purpose of the seventh question (Q7) was to measure the critical evaluation of the concept of variation different from the comparative situations as in Q2 and Q6. With this aim, students were given a line graph showing the speed of a service vehicle at specific times of a day. The question was asked to identify whether students questioned the variation in the speed of the service. Not to influence students in drawing their graphs in Q3, the critical evaluation of the concept of variation was investigated through a different context. The seventh question is presented in Figure 3.11.

Question 7:

Determining the speed limit as 70 km/h for factory service vehicles, the Chamber of Drivers in Ankara declares imposing fines to the drivers not obeying the speed limit.

The graph below shows the speed of a service vehicle whose speed is measured at specific times in a day.



Depending on the above graph, the Chamber of Drivers did not impose any fines to the driver of this service vehicle. Do you agree with the decision of the chamber?

Yes / No

- How did you reach this result? Explain your answer clearly.

  ……
  ……
  ……………………………………………

Figure 3.11 Q7

In conclusion, students were given a test that included 7 open-ended questions with 16 sub-questions. The relevance of these questions with the objectives in the curriculum and the tiers in the framework of Watson (1997) is displayed in the following table of specifications.

Table 3.2 Table of specification with respect to the three tiers

	Tier 1	Tier 2	Tier 3
Average	1d	1a,1b,2a,4a, 4b,5a,6a	1c, 4c, 5b
Variation	3c	3a, 3b	2b, 6b, 7

Table 3.3 Table of specification with respect to the objectives in the curriculum

Objectives (MoNE,2018)	Grade Level	Related Questions
Students should be able to compute and interpret the mean of a data set.	6	2a,6a
Students should be able to compute range and interpret the range of a data set	6	3c
Students should be able to use mean and range in comparing two data sets.	6	2b, 6b
Students should be able to construct and interpret line graphs.	7	3a,3b
Students should be able to compute mean, median and mode of a data set and interpret them.	7	1a, 1b, 1c, 4a, 4b, 4c, 5a, 5b

Table 3.2 indicated that 11 sub-questions were related to the concept of average and 7 of them required students to interpret the concept of average in different social or scientific contexts; that is, 7 questions were related to the Tier 2. While 3 sub-questions were about the evaluation of the average concept, the question 1d

required students to define the concept of average so related to the Tier 1. On the other hand, 6 questions in SLT were about the concept of variation. While only question 3c was related to the Tier 1, 2 questions were asked for interpretation of the variation concept. Lastly, 3 questions were related to the Tier 3 of the variation concept. In other words, 3 questions required students to evaluate the variation concept in different social or scientific contexts. On the other hand, Table 3.3 indicated that the questions in SLT addressed some objectives in Turkish curriculum existing both in sixth and seventh grade. While there are 5 sub-questions addressing the sixth grade objectives, 10 sub-questions were about the objectives existing in the seventh grade.

3.4 Pilot Study

The pilot study was conducted by the researcher in the spring semester of the 2017-2018 academic year for several purposes: to check the validity and reliability of the SLT. In other words, to determine the appropriate duration for the test, and to check the clarity and comprehensibility of the questions in the test.

The pilot study was conducted with eighth grade students since these students had completed all the necessary objectives for the study by the same semester of the last academic year. 47 eighth grade students in a middle school in Akyurt district, Ankara, was selected for the pilot study based on the criteria of convenience. With 7 questions and 14 sub-questions, the Statistical Literacy Test was implemented by allowing the students 45 minutes to complete the test.

Some revisions were made in the data collection instrument in accordance with the feedback received from the students during the study and the detailed analysis of the students' answers. First of all, it was observed that students needed about 50 minutes to complete the test. Thus, students were given 55 minutes in the actual data collection stage. Secondly, the analysis of students' answers revealed that the second part of Q1 and Q4, which were related to the critical evaluation of the concept of average, did not reach the expected aim. Therefore, an additional component asking for the computation of average was added before these parts, with the belief that thinking regarding computation of average will be helpful in the realization of the misleading nature of the news or the brochure. Hence, although the pilot study consisted of 14 sub-questions, there were 16 sub-questions in the actual study. Moreover, it was observed that most of the students had given idiosyncratic responses to the parts related to the third tier in Q1, Q4 and Q5. In other words, answers were not based on the given news, brochure or the interview. Therefore, statements such as "taking into account the given graph or brochure" was added into these parts and underlined. In addition, since students had not shown how they computed the average in the first parts of Q2 and Q6, in the actual study, they were asked to show how they reached their answers. Another revision was related to the Yes/No part existing in all of the questions except in Q3. In the pilot study, some place was provided for both the Yes and No part. However, it was observed that some students thought that both places should be filled. Therefore, the format of these parts was changed by providing space only for them to state the reason of their decisions. Lastly, the pilot study revealed that the number of data in Q5 and Q7 challenged students in the computation of an average. Thus, data in both of the questions were decreased. The last form of the SLT is provided in Appendix C.

3.5 The Validity and Reliability of the Instrument

"The appropriateness, meaningfulness, correctness, and usefulness of the inferences a researcher makes" form the validity of an instrument (Fraenkel &Wallen, 2006, p.147). In other words, validity is related to the correct conclusions reached from the data collected (Fraenkel & Wallen, 2006). Fraenkel and Wallen (2006) state that three types of validity evidence can be gathered by the researchers. They are content, criterion and construct related evidence. The content and format of the data collection tools form the content related evidence of validity. To ensure the content validity of the instrument prepared, first of all, the objectives in the national middle school curriculum regarding the concepts of average, variation and graphs were analyzed. In accordance with the purposes of the study, the requirements for the three tiers in the statistical literacy framework of Watson (1997) were also written down. Subsequently, two tables of specifications, one for the objectives in the curriculum and the other for the tiers in the framework were prepared. The table of specifications was provided in Section 3.3.

Subsequently, the test was given to two experts from the elementary mathematics education department of two different public universities. The definition of the variable statistical literacy used in the current study and a brief explanation regarding the three tiers of Watson (1997) were also provided to the experts. The experts commented on the consistency of the questions with the tiers and objectives and the comprehensibility and clarity of the questions. In accordance with the comments of the experts, some revisions were done to the instrument. First of all, both of the experts stated that the line graphs in Q2 and Q7 were not appropriate for the contexts given; therefore, the contexts of those questions were changed. Secondly, the statement of "except for manager department" in the third part of Q1 was removed since it was stated that median will not be used in this case. Lastly, some minor wording changes were made in the Q4 and Q6 upon the experts' suggestions to clarify the items.

After the necessary revisions were made, the clarity and comprehensibility of the questions in the test were checked by piloting the preliminary version of the instrument with 47 eighth grade students who had completed all the objectives required for the study in the same semester of the last academic year. The final revisions were made to the instrument in accordance with the feedback received from the students during the pilot study and the detailed analysis of answers of the students. All of the revisions with their reasons were presented in Section 3.4.

As for reliability of an instrument, "The consistency of the scores obtained" is how the reliability of an instrument is defined (Fraenken & Wallen, 2006). To state it differently, an instrument is said to be reliable if individuals get similar scores at different administrations of the same instrument. To provide the reliability evidence, first of all, a graduate student in mathematics education was informed about the purposes and process of the current study. Subsequently, answers of 82 students, half of the participants in the actual study, were analyzed by the researcher and co-coder independently and students' statistical literacy levels were determined for each question according to the modified version of the framework of Watson and Callingham (2003). After the analysis, while statistical literacy levels of students which are determined same by both coders in each question were coded as "1", different statistical literacy levels were coded as "0" and Cronbach's alpha coefficient was calculated through SPSS program as 0.92. Fraenken and Wallen (2006) indicated that instruments having a reliability measure of at least 0.70 is considered as a reliable instrument; hence, it can be said that SLT is a reliable instrument.

3.6 Time Table and Data Collection Procedures

The present study aimed to investigate the statistical literacy of seventh grade students regarding the concepts of average and variation on bar and line graphs related to the data obtained from social or scientific contexts. In the fall semester of the 2017-2018 academic year and at the beginning of the spring semester, the instrument was developed by the researcher based on the related literature. The necessary revisions made in accordance with the feedback received from the experts were completed by the end of March. Then, before the data collection process, the necessary permissions from the Middle East Technical University Human Subjects Ethics Committee was taken (see Appendix A). Subsequently, official permissions required to implement the study in the determined schools were taken from the Ministry of National Education (see Appendix B). The data of the study were collected by the researcher during the spring semester of the 2017-2018 academic year. First of all, a pilot study was conducted in the month of April to check the validity and reliability of the instrument and the clarity and comprehensibility of the items in the test. Then, SLT was administered to 164 seventh grade students after the concepts related to the study were covered by the mathematics teachers of the specified schools. The instrument was implemented by the researcher during the students' regular class time, and the duration of the implementation was 55 minutes. At the beginning of the implementation, some information regarding the aim of the study was provided to all the classes by the researcher. Furthermore, the structure of the test was briefly explained to the students briefly. The students were told that there was a news or the results of a study in each question except in Q3. They were notified that the questions they were to answer were related to those news or results that they were provided with on the next page. It was also emphasized that the students should explain their thinking clearly and in detail. The students were also given information about confidentiality of their answers. The time schedule for the data collection procedure is presented in Table 3.4.

Table 3.4 Time schedule for the study

Date	Tasks
December 2017-March 2018	Development of the instrument
April 2018	Pilot study and revision of the instrument
May 2018	Data Collection
June-July 2018	Data Analysis

3.7 Analysis of Data

The aim of this study was to analyze seventh grade students' statistical literacy regarding the concept of average and variation on bar and line graphs related to the data obtained from social or scientific contexts. Particularly, statistical literacy levels of seventh grade students related to the concepts of average and variation on bar and line graphs related to the data obtained from social or scientific contexts were tried to be determined. To determine students' levels of statistical literacy, the framework of Watson and Callingham (2003), which was explained in Chapter 2, was used. As stated previously, this framework consists of six levels and is related to all concepts which Watson (2006) mentioned in her model. In other words, each level in the framework shows the characteristics of statistical literacy that students display in the tasks about the concepts of data collection-sampling, data representation, chance, inference, average and variation. Since this study focused only on the concepts of average and variation, initially, the characteristics in each level for the concepts of average and variation were specified from the framework, which are provided in Table 3.5 below.

Table 3.5 Characteristics shown in six levels related to the average and variation concepts in in the framework (Watson & Callingham, 2003)

	General Characteristics of the Levels	Characteristics related to the Average Concept	Characteristics related to the Variation Concept
Level 1 Idiosyncratic	-Personal Beliefs and Experience	- Use no term for average even in a colloquial sense	- Basic acknowledgement of change e.g. "won't look the same every day" - Idiosyncratic predictions of chance outcomes without justifications
Level 2 Informal	-Engagement with context but this engagement is intuitive, non-statistical or reflective of irrelevant aspect of the task context	- Single ideas related to average like okay, normal, same as others	- Too much, too little variations for spinner or die tasks or depended on strict probability
Level 3 Inconsistent	-More engagement with the context but this depends on the format of the items - Qualitative statistical ideas rather than quantitative - Appropriate conclusions but not suitable justifications	- Continuity of single ideas in open-ended items - Use of measures of central tendency or ideas related to representativeness in the tasks requiring recognition; that is, in multiple choice items	- Only recognition of chance, not variation in predicting repeated outcomes with a 50-50 spinner -A single aspect of the term variation e.g you get a choice
Level 4 Consistent-Non critical	-Appropriate contextual but non-critical engagement in various contexts - Use of mathematical and critical skills in straightforward settings	-Ideas related to mean, middle and most in description of average -Find mean of a small data set without recognition of the outlier	- A reason associated with variance in explaining differences in repeated samples of 50-50 spinner -Realistic variations for spinner and die tasks -Multiple elements related to the definition of variation e.g. Variation means to change something, The weather is going to vary over the next few days.

Table 3.5 (continued)

	General Characteristics of the Levels	Characteristics related to the Average Concept	Characteristics related to the Variation Concept
Level 5 Critical	-Requirement of critical thinking skills -Appropriate use of terminology	- Find mean or median of a small data set	- Words like "about" or "probably" in predicting the outcomes of spinning 50-50 spinner - An increase or change in the data over time or acknowledgement of variation explicitly in the visual appearance of the graphs
Level 6 Critical Mathematical	-Critical thinking skills together with mathematical skills -Quantitative statistical ideas	-Recognition of outliers in calculation of mean and suggesting the median as the appropriate measure	

Table 3.5 indicated the general characteristics and characteristics that students show related to the concepts of average and variation with respect to the six levels. These characteristics shown for the concept of average and variation will be explained briefly in the following sections.

At Level 1, while students use any term related to the concept of average, they acknowledge the change basically for the concept of variation. For instance, they answer the question of “Would the graph, which displays transportations that some students preferred in a day, look the same everyday?” as “It will not look the same everyday.”.

At Level 2, for the concept of average, students begin to use single ideas such as ‘normal’ or ‘okay’. On the other hand, in the tasks such as spinner or dice, students provide either too much or too little variations or they did not provide any variation in those tasks but depend on the strict probability.

At Level 3, while students continue using single ideas in describing the concept of average in an open-ended task, some ideas related to measures of central tendency begin to emerge when the tasks require recognition as in the multiple choice items. Students at Level 3 mention single aspects of the term variation. For instance, in defining the concept of variation, they state that you get a choice.

On the other hand, at Level 4, consistent non-critical, students engage with contexts, but this engagement does not involve critical questioning. Students at Level 4 describes the average concept with the ideas related to three measures of central tendency: mean, mode and median. Furthermore, students at Level 4 can compute the mean of a small data set but do not observe the effect of an outlier. On the other hand, students at this level suggest realistic variations to the chance tasks like spinner or dice.

At the last two levels, students show critical thinking skills just like in the third tier of Watson (1997). In both of these levels, students critically engage with the context. For the average concept, students at Level 5 can compute the mean and median of a small data set. Besides, Level 5 students appreciate variation and it can be understood with phrases like ‘It will be close to half’. At Level 5, students mention variation explicitly in the visual appearance of the graphs. Furthermore, appropriate terminology for both concepts are observed at Level 5. Besides, in the last level, *critical mathematical*, students recognize the effect of outliers in the calculation of mean and so suggest median as the appropriate measure of central tendency.

Moreover, as previously stated in Chapter 2, the framework of Watson and Callingham (2003) was related to Watson’s Statistical Literacy Framework (Watson, 1997) which was used to prepare SLT in the current study. While the objectives of the first tier are observed across different levels, requirements of Tier 2; that is, understanding and interpreting the concepts in social or scientific contexts begin to appear at Level 3, *inconsistent* and continue at Level 4, *consistent non-critical*. On the other hand, critical evaluation of the concepts in social or scientific contexts which is the requirement of the third tier was observed at the last two levels in the framework of Watson and Callingham (2003). Indeed, except for Tier 1, the tiers were separated into the two hierarchical parts.

When the framework in Table 3.5 was analyzed, it was realized that although it is useful for coding the answer of the students, more detailed characteristics were needed to analyze the answers of the students in the current study. Hence, some modifications were made to the framework in Table 3.5. To make these modifications, sections for the concepts of average and variation in the book of Watson (2006) and her related articles and article of Shaugnessy (2003b) were examined in detail, and some more characteristics were added to some levels by taking into consideration the general characteristics of the levels. The characteristics added can be observed in Table 3.6 below.

Table 3.6 Modified version of the framework of Watson and Callingham (2003)

	General Characteristics of the Levels	Characteristics related to the Average Concept	Characteristics related to the Variation Concept
Level 1 Idiosyncratic	-Personal Beliefs and Experience	- Use no term for average even in a colloquial sense	- Basic acknowledgement of change e.g. "won't look the same every day" - Idiosyncratic predictions of chance outcomes without justifications
Level 2 Informal	-Engagement with context but this engagement is intuitive, non-statistical or reflective of irrelevant aspect of the task context	- Single ideas related to average like okay, normal, same as others - Colloquial interpretations of the questions about the meaning and method of average - Sometimes refer to add up colloquially but not in a calculation sense	- Too much, too little variations for spinner or die tasks or depended on strict probability - Produce a graph with variation but not appropriate for the task -Focus on single values or use of the word "more" in comparison tasks
Level 3 Inconsistent	-More engagement with the context but this depends on the format of the items - Qualitative statistical ideas rather than quantitative - Appropriate conclusions but not suitable justifications	- Continuity of single ideas in open-ended items - Use of measures of central tendency or ideas related to representativeness in the tasks requiring recognition; that is, in multiple choice items -Sometimes appearance of the modal idea in interpretation of the contexts -Sometimes suggestion of mean without any discussion of why it would be appropriate	- Only recognition of chance, not variation in predicting repeated outcomes with a 50-50 spinner -A single aspect of the term variation e.g you get a choice -Realistic variations in drawing graphs but they are not complete -Visual strategies focusing on more than one values in comparison tasks

Table 3.6 (continued)

	General Characteristics of the Levels	Characteristics related to the Average Concept	Characteristics related to the Variation Concept
Level 4 Consistent-Non critical	-Appropriate contextual but non-critical engagement in various contexts - Use of mathematical and critical skills in straightforward settings	-Ideas related to mean, middle and most in description and interpretation of average -Finding of mean of a small data set without recognition of the outlier -Difficulty in computation of weighted mean	- A reason associated with variance in explaining differences in repeated samples of 50-50 spinner -Realistic variations for spinner and die tasks -Multiple elements related to the definition of variation e.g. Variation means to change something, The weather is going to vary over the next few days. -Realistic variation in complete graphs - Finding total or mean (numerical strategies) in comparison tasks sometimes with inclusion of individual values
Level 5 Critical	-Requirement of critical thinking skills -Appropriate use of terminology	-Find mean or median of a small data set -Ideas related to representative nature of average -Presentation of the idea of mode when data involve an outlier explaining the most would be more accurate	- Words like “about” or “probably” in predicting the outcomes of spinning 50-50 spinner - An increase or change in the data over time or acknowledgement of variation explicitly in the visual appearance of the graphs - Visual strategies taking into account variation in the graphs in comparison tasks
Level 6 Critical Mathematical	-Critical thinking skills together with mathematical skills e.g. use of proportions -Quantitative statistical ideas	-Recognition of outliers in calculation of mean and suggesting the median as the appropriate measure	- Integration of numerical and visual strategies taking into account variation in comparison tasks

Table 3.6 indicated that some modifications were made to all levels except for Level 1 for both average and variation concepts. In further sections, how these modifications were made explained for average and variation concept respectively.

Watson and Moritz (2000) stated that students provide colloquial interpretations to the questions asking about meaning and method of the average at the unistructural level which is the second level of SOLO Taxonomy. For example, students say that the average of 3 hours of TV in a day means around 3 hours of TV in a day. In other words, students use single ideas in their interpretation of average, which is a characteristics shown at Level 2 according to the framework in Table 3.4; therefore, colloquial interpretations regarding the meaning and method of the average concept were added to Level 2. It was also stated that at the same level, students can sometimes refer to adding up when they are asked how the average in the given context was calculated. However, they do not ignore the implications. This characteristic was also placed at Level 2. Researchers also asserted that at the same level, the modal idea can appear in the interpretation of the average concept such as most will watch 3 hours in a day and sometimes students could refer to the mean when the calculation of the average is asked. However, they do not know why it is appropriate to use the mean. Since these characteristics were believed to be better than the ones placed at Level 2, they were placed at Level 3 in the framework. In the same article, it was stated that at the next level, students use ideas related to measures of central tendency in describing and interpreting the average concept, sometimes combining two of them. They asserted that these students can calculate the mean of a data set but do not recognize the outlier and they have difficulty in calculating the weighted mean. Most of these characteristics were expressed in the framework at Level 4, except for the difficulty in the computation of weighted mean and interpreting the average concept using ideas related to three measures of central tendency; thus, these characteristics were also added to Level 4. Watson and Moritz (2000) also asserted that students use ideas related to the representative nature of the average concept at the highest level.

Since appropriate terminology for the concept of average was observed at Level 5 in the framework in Table 3.5, this characteristic, the use of ideas related to the representative nature of the average concept was placed into the Level 5. Furthermore, in a question entailing an outlier in the book of Statistical Literacy at School (Watson, 2006), it was stated that the students at the highest level could recognize outliers and either suggest the median or find the mean by excluding the outliers. The researcher stated that students before this level could find the mode of the given data set correctly and express the idea that if most agreed, it would be the average. Since recognizing the outlier exists at Level 6 in the framework, the other idea, the presentation of the mode when there is an outlier, was placed at Level 5.

On the other hand, for the variation questions, the article of Watson and Kelly (2005), from where Q3 was adapted, the inference and variation chapters in the book Statistical Literacy at School (Watson, 2006) and the article of Shaugnessy (2003b) were examined in detail. Watson and Kelly (2005) examined interpretations of the variation concept by requesting students to draw a graph. They observed that at the unistructural level, students begin to produce some graphs with variations which are not appropriate for the tasks set. As can be remembered, according to the framework in Table 3.5, students at Level 2 produce some variation in spinner or dice tasks but they are either too much or too little; that is, not appropriate for the given task. Since the characteristic mentioned in the article of Watson and Kelly (2005) was similar to the variation characteristic at Level 2 in Table 3.5, this characteristic was placed at Level 2. Then, Watson and Kelly (2005) asserted that after such answers, students could interpret the variation concept by firstly producing graphs with appropriate variation but are not complete; secondly, presenting complete graphs with realistic variations. Since in the framework, it was stated that students exhibit realistic variations in spinner or dice tasks at Level 4, characteristics of producing graphs with appropriate variation placed at Level 4. Characteristic of producing graphs with appropriate variation but are not complete were added to Level 3.

In the inference and variation chapters of the book Statistical Literacy at School, Watson (2006) states that students use individual values or only the word ‘more’ while comparing two data sets at the unistructural level. Since characteristics observed at the unistructural level generally exist at Level 2 and since students use single ideas related to the concept of average and variation at Level 2, the use of individual values in comparing two data sets characteristics was added into the Level 2. Then, Watson (2006) implies that at the multistructural level, students either use numerical strategies such as finding totals or means or uses some visual strategies focusing on more than one value. “There are more 6s and 5s than … over in the Purple class” shows an example of a student using visual strategies concentrating on more than one value in making his comparison (Watson, 2006, p.201). In the framework in Table 3.5, it was stated that a student can calculate mean of a data set at Level 4. Therefore, numerical strategies like finding mean were added to the Level 4. However, visual strategies focusing on more than one value was placed into the Level 3 thinking that these students still do not take into account all the data given. At the highest level, students use both numerical strategies and visual strategies taking into account the variation in the graphs in combination or make their comparison by only using visual strategies stating the variation in the graphs. Since one of the characteristics regarding the concept of variation at Level 5 in Table 3.5 is that students mention explicitly the variation in the graphs, the characteristic of using visual strategies stating the variation in the graphs in comparison tasks were placed at Level 5. On the other hand, Shaugnessy (2003b) stated in his article, in which he examined students’ strategies in two comparison tasks, students at the highest level use average and variation together to compare the two groups given. Hence, the characteristics of the use of numerical strategies and visual strategies taking account the variation in the graphs in combination was placed at the highest level, Level 6 in the framework.

Subsequently, the data of the study were started to be analyzed using the modified version of the framework of Watson and Callingham (2003) in Table 3.6, and the levels of the students were tried to be determined. During data analysis, all of the characteristics in the modified version of the framework was useful to determine statistical literacy levels of the students; however, it was realized that some students defined the concept of average using the ideas of mean, but could not reveal this understanding in interpretation of the questions, or they interpreted the concept of average in a social context using the idea of median by using the word 'middle'; however, no definition was provided for the concept of average. Nor could they interpret the average concept when asked questions regarding its computation. Since these students could not be placed at Level 4 because they could neither define nor interpret the concept of average, they were placed at Level 3 and one more characteristic was added to the framework (see Table 3.7): Ideas related to measures of central tendency in either definition or interpretation of the average concept. Moreover, it was observed that some of the Level 4 students used ideas related to the representative nature of the average concept. Therefore, that characteristics which was at Level 5 in Table 3.7 was taken into the Level 4. The last version of the framework is presented in Table 3.7.

Table 3.7 Modified version of the framework of Watson and Callingham (2003)

	General Characteristics of the Levels	Characteristic Related to the Average Concept	Characteristic Related to the Variation Concept
Level 1 Idiosyncratic	-Personal Beliefs and Experience	- Use no term for average even in a colloquial sense	- Basic acknowledgement of change e.g. "won't look the same every day" - Idiosyncratic predictions of chance outcomes without justifications
Level 2 Informal	-Engagement with context but this engagement is intuitive, non-statistical or reflective of irrelevant aspect of the task context	- Single ideas related to average like okay, normal, same as others - Colloquial interpretations of the questions about the meaning and method of average - Sometimes refer to add up colloquially but not in a calculation sense	- Too much, too little variations for spinner or die tasks or depended on strict probability - Produce a graph with variation but not appropriate for the task -Focus on single values or use of the word "more" in comparison tasks
Level 3 Inconsistent	-More engagement with the context but this depends on the format of the items - Qualitative statistical ideas rather than quantitative - Appropriate conclusions but not suitable justifications	- Continuity of single ideas in open-ended items - Use of measures of central tendency or ideas related to representativeness in the tasks requiring recognition; that is, in multiple choice items -Sometimes appearance of the modal idea in interpretation of the contexts -Sometimes suggestion of mean without any discussion of why it would be appropriate -Ideas related to measure of central tendency in either definition or interpretation of the average	- Only recognition of chance, not variation in predicting repeated outcomes with a 50-50 spinner -A single aspect of the term variation e.g you get a choice -Realistic variations in drawing graphs but they are not complete -Visual strategies focusing on more than one values in comparison tasks

Table 3.7 (continued)

	General Characteristics of the Levels	Characteristic Related to the Average Concept	Characteristic Related to the Variation Concept
Level 4 Consistent-Non critical	-Appropriate contextual but non-critical engagement in various contexts - Use of mathematical and critical skills in straightforward settings	-Ideas related to mean, middle and most in description and interpretation of average -Finding of mean of a small data set without recognition of the outlier -Difficulty in computation of weighted mean -Ideas related to representative nature of average	- A reason associated with variance in explaining differences in repeated samples of 50-50 spinner -Realistic variations for spinner and die tasks -Multiple elements related to the definition of variation e.g. Variation means to change something, The weather is going to vary over the next few days. -Realistic variation in complete graphs -Finding total or mean (numerical strategies) in comparison tasks sometimes with inclusion of individual values
Level 5 Critical	-Requirement of critical thinking skills -Appropriate use of terminology	-Find mean or median of a small data set -Presentation of the idea of mode when data involve an outlier explaining the most would be more accurate	- Words like “about” or “probably” in predicting the outcomes of spinning 50-50 spinner - An increase or change in the data over time or acknowledgement of variation explicitly in the visual appearance of the graphs - Visual strategies taking into account variation in the graphs in comparison tasks
Level 6 Critical Mathematical	-Critical thinking skills together with mathematical skills e.g. use of proportions -Quantitative statistical ideas	-Recognition of outliers in calculation of mean and suggesting the median as the appropriate measure	- Integration of numerical and visual strategies taking into account variation in comparison tasks

Lastly, the framework in Table 3.7 was used in the current study to determine statistical literacy levels of students about the concepts of average and variation on bar and line graphs related to the data obtained from social or scientific contexts. Their statistical literacy levels were examined for each question separately to observe the role of different contexts on the statistical literacy levels of students. First of all, students' definitions, interpretations and evaluations were coded in each question. Then, statistical literacy levels of students were determined by overall analysis of the codes in students' definitions, interpretations and evaluations related to the concepts of average and variation. This type of coding was named as upward coding by Jansen (2010). Moreover, statistical literacy levels of students were presented by combining the two levels in the framework such as Level 1-2 or Level 3-4. The reason of this is that as mentioned previously, SLT was prepared according to Watson's Statistical Literacy Framework (1997); in other words, students' definitions of the concepts of average and variation, their interpretations related to the concepts in a social or scientific contexts and their evaluations to the claims related to the average and variation concepts were examined. Furthermore, the framework used to determine statistical literacy levels of students were related to the Watson's Statistical Literacy Framework (1997). To state it differently, while the objectives related to the definitions of the concepts of average and variation are observed across different levels, requirements of Tier 2; that is, interpreting the average and variation concepts in social or scientific contexts begin to appear at Level 3 and continue at Level 4. On the other hand, critical evaluation of the average and variation concepts in social or scientific contexts which is the requirement of the third tier was observed at Level 5 and Level 6. For instance, statistical literacy level of a student who can evaluate the concept of average in a social or scientific context is either Level 5 or Level 6; hence, in this study, statistical literacy level of that student was stated as Level 5-6. Nevertheless, since characteristics shown at each level specified clearly in the framework, students' statistical literacy levels could also be determined as Level 5 and Level 6 and frequencies for each level were presented for each question in the next chapter.

Furthermore, in a question; for instance, related to the evaluation of the average concept, statistical literacy levels of the students are expected to be Level 5-6 according to the framework. However, since analysis were made with respect to the students' answers, their statistical literacy level can also be Level 1-2 or Level 3-4.

3.8 Assumptions and Limitations

There are some assumptions and limitations of the present study. To begin with, it was assumed that students' statistical literacy regarding the concepts of average and variation on bar and line graphs related to the data obtained from social or scientific contexts could be examined with the instrument developed. The second assumption is that students were willing and careful in answering all of the questions in the instrument. It was also assumed that their explanations were clear enough to reveal their thinking process.

The sample of the current study was selected via the convenience sampling method. Convenience samples do not represent the population selected (Fraenken & Wallen, 2006). Hence, this study is limited in the generalization of the findings to any seventh grade student. Fraenken and Wallen (2006) suggest providing demographic information in the use of convenience samples so that results can be generalized to samples in similar contexts. Demographic information was also provided in addition to the findings of the study. Therefore, findings can be generalized to the samples which have similar characteristics of the current study. Besides, the results reported in this study were limited to the questions in SLT. In other words, if different questions related to the concepts of average and variation were asked in different contexts, different results could be found.

3.9 The Internal and External Validity of the Study

Internal and external validity are two validity types that affect the validity of any study. Hence, some information regarding both validity types is provided in the following sections.

3.9.1 Internal Validity

Internal validity refers to any difference or relationship observed in the dependent variable are because of only independent variables determined but not by any other variable (Fraenken & Wallen, 2006). Internal validity threats differ depending on the type of the research design. Mortality, location and instrumentation are three main types of internal validity threats existing in survey research designs (Fraenken & Wallen, 2006).

To begin with, mortality is a threat when subjects are lost during the study. The current study was a cross sectional one; that is, data were gathered at one point of time; hence, mortality was not a possible thread. However, to provide maximum participation, the researcher got in contact with mathematics teachers of the specified schools. It was ensured that the students would not participate in any other activity on the administration day of the test, and mathematics teachers often reminded students that a test would be applied in the determined time.

The location threat occurs if the data collection location causes alternative explanations for the study (Fraenkel & Wallen, 2006). Location was not a possible threat for the current study since the instrument was implemented to all the students in their own classrooms with similar conditions.

The instrumentation threat involves three types of threats, namely instrumentation decay, data collector characteristics and data collector bias. First of all, if there is a change in the instrument or scoring, instrumentation decay is a possible threat for a study (Fraenken & Wallen, 2006). Nevertheless, this does not seem to be a threat for the present study since the instrument was administered just once and a framework was used to analyze the answers of the students. Data collector characteristics might cause a threat for studies in which data are gathered by different data collectors (Fraenken & Wallen, 2006). However, all of the data were collected by the researcher of the study; therefore, this threat was handled in the current study. Lastly, data collector bias occurs when the data of a study is changed either intentionally or unintentionally by the data collectors (Fraenken & Wallen, 2006). Data were collected by the researcher and no interaction, except for the explanations at the beginning of the administration, was allowed throughout the implementation. Therefore, this thread was taken under control.

3.9.2 External Validity

External validity refers to "the extent to which the results of a study can be generalized from a sample to a population" (Fraenkel & Wallen, 2006, p.107). Population and ecological generalizability are two types of external validity.

Population generalizability is related with the extent of the representativeness of a sample with the population selected. The sample should be representative of the interested population so as to ensure population generalizability of a study. In the current study, all seventh grade students in Ankara form the target population. The accessible population of the study is all seventh grade students in the Cankaya and Akyurt districts of Ankara. The SLT was administered to 164 seventh grade students who were selected through the convenience sampling method. Convenience samples do not represent the intended population (Fraenken & Wallen); therefore, results of the study could not be generalized to the population of interest but the results might be generalized to samples with similar conditions as explained below.

Fraenken and Wallen (2006) define ecological generalizability as "the degree to which the results of a study can be extended to other settings or conditions" (p.106). The present study was conducted with seventh grade students in public middle schools in Ankara. Some characteristics of students were provided in Section 3.2. Furthermore, in those schools, the same mathematics education curriculum is implemented. Furthermore, the students were using the mathematics textbook provided by MoNE. Hence, it was believed that the results of the study might be generalized to the public schools with similar settings and to students with similar characteristics.

CHAPTER 4

FINDINGS

The purpose of the current study was to analyze seventh grade students' statistical literacy in terms of the concepts of "average" and "variation" on bar and line graphs related to the data obtained from social or scientific contexts. More specifically, this study aimed to determine statistical literacy levels of seventh grade students and aimed to explain how students at different statistical literacy levels define, interpret and evaluate the concepts of "average" and "variation" on bar and line graphs related to the data obtained from social or scientific contexts.

In this chapter, the findings of the study are presented from four main aspects related to the concepts focused in this study. In other words, the findings for the questions related to the concepts of "average" and "variation" are presented, respectively. Since both concepts were analyzed on bar and line graphs separately, the findings for questions on bar and line graphs are presented in different sections.

In the first section, the findings for the questions related to the concept of average on bar graphs are presented. The second section examines the findings for the questions related to the concept of average on line graphs. In the subsequent sections, the findings for the questions related to the variation concept on line and bar graphs are presented, respectively. The organization of each section is in line with the research question of the current study. In other words, first of all, the statistical literacy levels of the students determined by the overall analysis of each question are presented in each section. Subsequently, to respond to the sub-question, students' definitions, interpretations and evaluations of the specified concepts are explained for each statistical literacy level determined in each question and supported with examples from students' answers. Furthermore, students' definitions, interpretations and evaluations of the specified concepts are presented comparatively in similar questions.

4.1 The Concept of Average on Bar Graphs

One of the aims of the present study was to analyze seventh grade students' levels of statistical literacy regarding the concept of average on bar graphs. With this aim, three questions, question 1 (Q1), question 5 (Q5) and the first part of question 6 (Q6a) were constructed. Q1 was related to the average income in a company, which was presented to the students through a news excerpt. In this question, an outlier, namely the income of the manager, was provided, but the average in the news was calculated by means of the mean, which is not an appropriate measure of average when there is an outlier. Whether or not students critically questioned the given average income for the company was examined. Furthermore, by asking students the meaning of the average in the news and how the average in the news was calculated, students' interpretation of the concept of average was investigated. On the other hand, Q5 was about the average number of days that customers preferred in a hotel and was presented to the students by means of the script of an interview. Although there were categorical data, the average number of days for the customers was calculated as 5 days based on the mean of the given data. Whether or not students realized that the average should be the mode of the given data was investigated. Moreover, as in Q1, by asking students how the average in the interview was calculated, the interpretation made of the concept of average was examined. Lastly, in Q6a, students were requested to calculate the average wind speed of two regions whose wind speeds throughout seven months were given on a double bar graph. By avoiding asking directly for the mean of the wind speeds, and instead asking for the average, the interpretation of the concept of average was investigated.

To determine students' statistical literacy levels for each question, the modified version of the framework of Watson and Callingham (2003) was used. The statistical literacy levels of the students were determined by means of the overall analysis of each question. In other words, the statistical literacy levels of the students were determined by taking into consideration their definitions and interpretations regarding the concept of average and their evaluations regarding the claims related to the concept of average in each question. The distribution of the percentage of students across the six levels for Q1, Q5 and Q6a are displayed in Table 4.1.

Table 4.1 The distribution of students across the six levels in the framework for Q1, Q5 and Q6a

	Questions related to Average on Bar Graphs		
Levels	**Question 1**	**Question 5**	**Question 6a**
1-2	44 (26.8%)	51 (31.1%)	57 (34.8%)
3-4	110 (62.1%)	68 (41.5%)	107 (65.2%)
5-6	10 (6.1%)	45 (27.4%)	0 (0.0%)
Total	164 (100.0%)	164 (100.0%)	164 (100.0%)

Table 4.1 revealed that when students' definitions, interpretations and evaluations of the concept of average were analyzed as a whole in Q1, most of the students (62.1%) performed at Level 3-4. On the other hand, while 26.8% of the students were identified to be at Level 1-2 in Q1, only 6.1% of the students performed at Level 5-6. Findings of the study were not different for Q5. That is, most of the students (41.5%) performed at Level 3-4. The number of students observed at Level 1-2 and Level 5-6 in Q5 was 31.1% and 27.4%, respectively. Lastly, Table

4.1 indicates that 65.2% of the students were at Level 3-4 in Q6a. The remaining 34.8% of the students performed at Level 1-2 and there were no students performing at Level 5-6. This was an expected result for Q6a since it was a question related to the interpretation of the concept of average concept only; hence, students were expected to be at either Level 1-2 or Level 3-4.

In the following sections, to answer the sub question of the present study, the answers of students at different statistical literacy levels are explained in detail for Q1, Q5 and Q6a providing examples from students' answers. To put it differently, how seventh grade students at different statistical literacy levels define, interpret and evaluate the concept of average on bar graphs is explained for Q1, Q5 and Q6a.

4.1.1 The Concept of Average on Bar Graphs in Question 1

In question 1 (Q1), the students were provided with a news excerpt regarding the average income in a company. In the question, the income of the personnel in different departments was presented on a bar graph and the average was calculated as 4300 TL by using the mean algorithm. However, the income of the manager was much higher than the income of the personnel in the other departments. To state it differently, there was an outlier in the given data and the students were expected to realize this outlier. Q1 entailed four sub-questions. In the first sub-question, Q1a, students were asked what they understood from the average given in the news. In this way, their interpretation of the concept of average in the given context was examined. In the second sub-question, Q1b, students' interpretation of the concept of average was investigated once more via a different question. Q1b asked students how the concept of average in the news was calculated. On the other hand, the third sub-question, Q1c, was about students' evaluation of the claim in the article regarding the average income. The students were asked the question of whether or not they would give up their current work which granted an income of 2900 TL by taking into account the information conveyed through the graph and the average income in the news. The last sub-question, Q1d, required defining the concept of average, which the students had learnt in their mathematics class. They were also required to give an example related to the concept of average.

The statistical literacy levels of the students were determined by an overall analysis of their answers to the four sub-questions. As presented in Table 4.1, 26.8% of the students performed at Level 1-2 in Q1. In the modified version of the framework of Watson and Callingham (2003), it was stated that any idea related to the concept of average is observed by Level 1 students while defining and interpreting the average. On the other hand, students at Level 2 express some single ideas about the concept of average, such as ‘normal’ or ‘the same as others’ while defining and interpreting the concept of average. Critical evaluation of the claims related to the concept of average in a social or scientific context is not an expected skill of Level 1-2 students since they cannot define or interpret the concept. In further sections, definitions, interpretations and evaluations of Level 1-2 students in Q1 are presented.

Findings of the present study showed that 10.9% of the students could not present any idea while defining the concept of average in Q1d; therefore, these students were placed at Level 1. However, some single ideas were begun to be used in defining the average by 5.4% of the students; hence, these students were placed at Level 2. The word *about* was the most preferred word. These students explained that, for them, average refers to the word *about*. Figure 4.1 shows the answer of such a student.

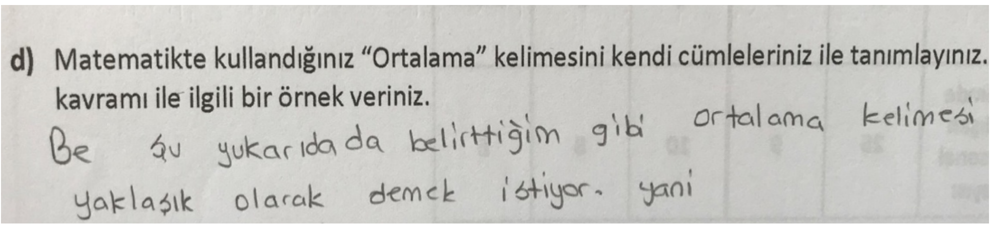
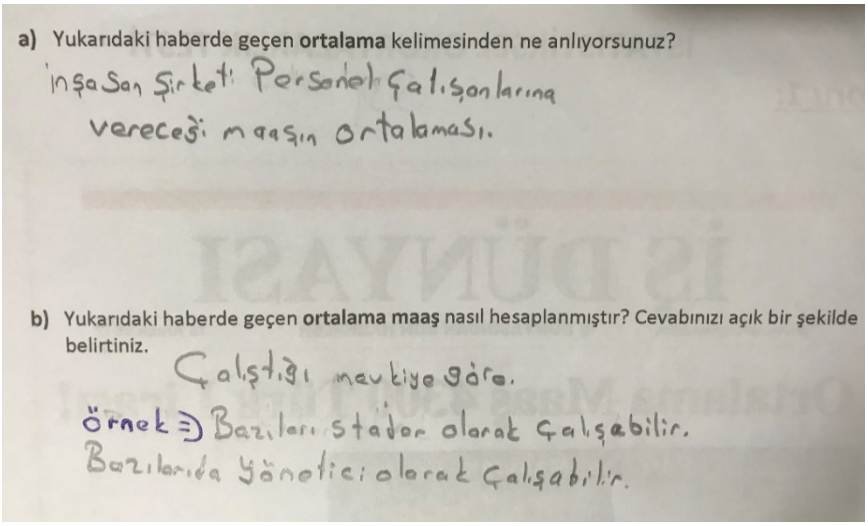


Figure 4.1 The definition of average by a Level 1-2 student in Q1d

Other single words or ideas preferred by students in defining the concept of average in Q1d were *almost*, *around*, *uncertain*, *rounding* and *estimation*.

On the other hand, when the answers of students whose statistical literacy levels were determined to be at Level 1-2 to the interpretation questions in Q1, it was observed that some of them either did not interpret the concept of average or presented some irrelevant responses. For example, S31 in Figure 4.2 made his interpretation presenting almost the same wording in the news given: "average income which will be given to personnel in the company". Moreover, he explained that average income was computed based on the department.



a) Yukarıdaki haberde geçen **ortalama** kelimesinden ne anlıyorsunuz?

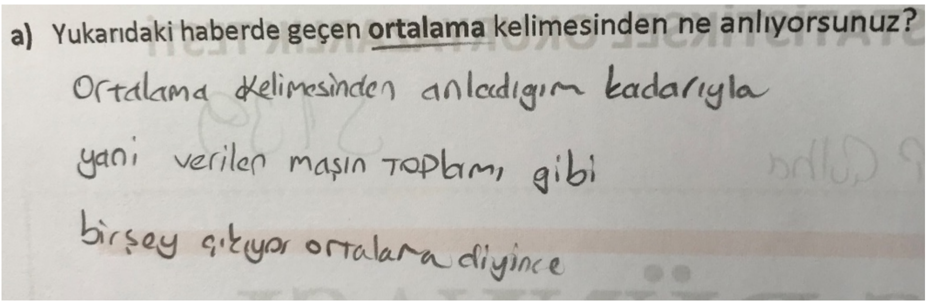
inşaSan Şirketi Personel Çalışanlarına vereceği maaşın ortalaması.

b) Yukarıdaki haberde geçen **ortalama maaş** nasıl hesaplanmıştır? Cevabınızı açık bir şekilde belirtiniz.

Çalıştığı mevkiye göre.
örnek ⇒ Bazıları stajer olarak çalışabilir.
Bazılarıda yönetici olarak çalışabilir.

Figure 4.2 Interpretations of average by a Level 1-2 student in Q1a and Q1b

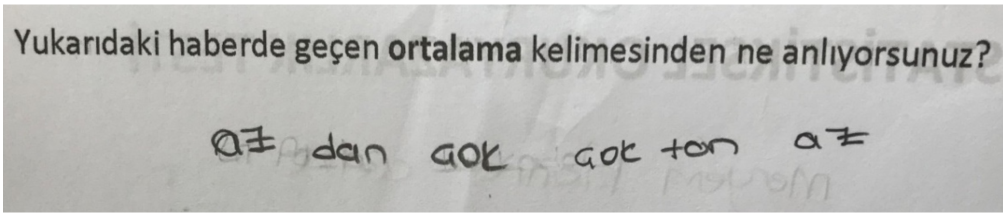
It was observed that some of the students interpreted the concept of average in Q1a and Q1b by using the same single ideas that they had used in defining the concept of average; thus, these students were placed at Level 2. Differently, the *sum* of the given salaries was used by some students to interpret the meaning of the concept of average in the given news in Q1a. A sample answer can be observed in Figure 4.3. The student stated that average is like the *sum* of the given incomes in the company.


a) Yukarıdaki haberde geçen **ortalama** kelimesinden ne anlıyorsunuz?

Ortalama kelimesinden anladığım kadarıyla yani verilen maşın toplamı gibi birşey çıkıyor ortalama diyince

Figure 4.3 Interpretation of average by a Level 1-2 student in Q1a

Besides, findings of the present study revealed that some of the students (7.3%) whose statistical literacy levels were determined to be at Level 2 could not define the concept of average in Q1d. However, all of them could interpret the concept of average when asked for its meaning in the given context in Q1a. This could be an indication of the effect of the context. Indeed, the student in Figure 4.4 could not define the concept of average in Q1d but interpreted it as *more than less and less than more,* which was coded as *more or less* as an interpretation in Q1a.


Yukarıdaki haberde geçen **ortalama** kelimesinden ne anlıyorsunuz?

az dan çok çok tan az

Figure 4.4 Interpretation of average by a Level 1-2 student in Q1a

However, the effect of the context was not observed when the answers of students at Level 2 were examined for the interpretation question regarding the computation of average in Q1b. It was found that most of the students could answer the question but they presented some irrelevant responses. Interestingly, these students could define the concept of average in Q1d or could make an interpretation regarding the meaning of average in the given news in Q1a but they had difficulty in interpreting the computation of average in the given news. For instance, S64 in Figure 4.5 both defined and interpreted the average concept using the word *about* in Q1d and Q1a, respectively; therefore, this student was placed at Level 2. However, as can be observed in Figure 4.5, she could not interpret the computation of the average in the given news but explained her own opinion regarding the salaries in the company: “The manager could increase or decrease income according to the person.”

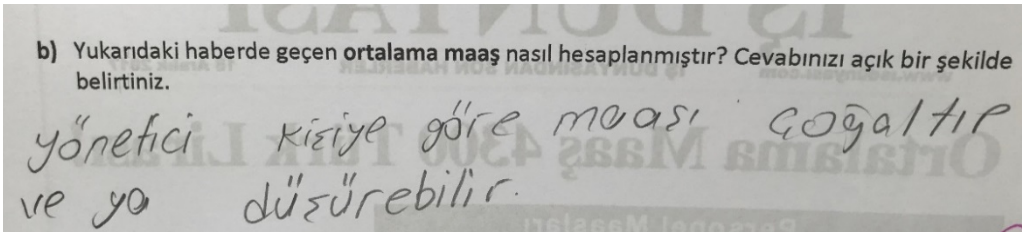
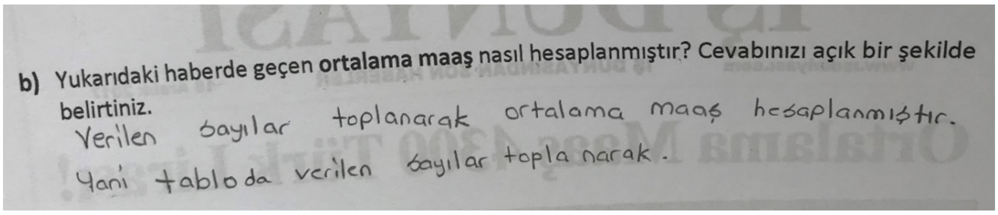


Figure 4.5 Interpretation of average by a Level 1-2 student in Q1b

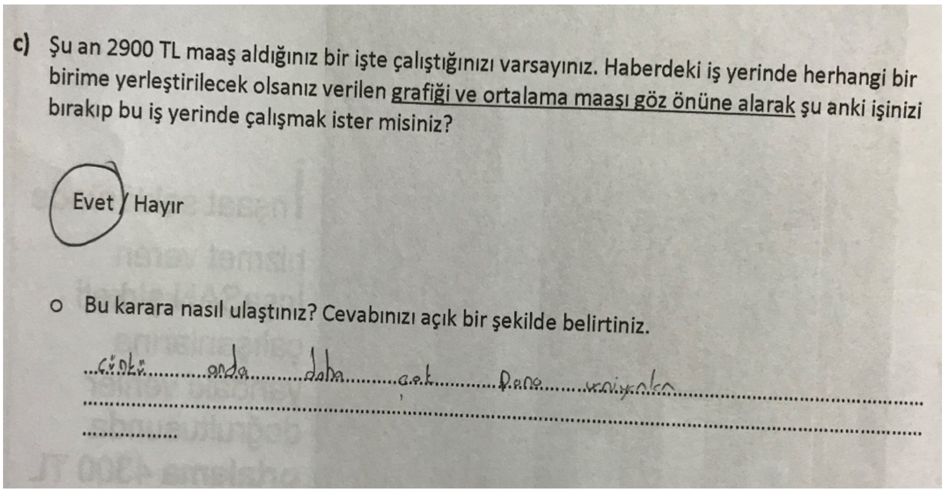
Similarly, the number of students whose statistical literacy levels was determined as Level 1-2 in Q1 and who used a single idea in interpreting the computation of average in the given news in Q1b was much less than the number of students who used a single idea interpreting the meaning of the average in Q1a. It was stated by Level 1-2 students that average was computed through *estimation* or *rounding* but these students did not show how the average income is obtained through *rounding* or *estimation*. There were two students who stated that average was calculated by adding all the given salaries like the student in Figure 4.6. Similar to S64 in Figure 4.5, S161 defined and interpreted the concept of average using the word *about*. However, she interpreted the computation of average in Q1b differently. She said that average in the news was calculated through the sum of the income of the personnel. Nevertheless, these students did not add the salaries of the given personnel and reached the result of 4300 TL.


b) Yukarıdaki haberde geçen **ortalama maaş** nasıl hesaplanmıştır? Cevabınızı açık bir şekilde belirtiniz.

Verilen sayılar toplanarak ortalama maaş hesaplanmıştır. Yani tabloda verilen sayılar topla narak.

Figure 4.6 Interpretation of average by a Level 1-2 student in Q1b

Findings of the present study indicated that none of the students whose statistical literacy levels were determined as Level 1-2 could critically evaluate the given average in the news in Q1c. While 10% of Level 1-2 students accepted the average income directly, some presented idiosyncratic responses to justify their decisions in Q1c. For example, a student in Figure 4.7 decided to give up his or her work by presenting the idea that average income in the news is more than their own income in the given context.


c) Şu an 2900 TL maaş aldığınız bir işte çalıştığınızı varsayınız. Haberdeki iş yerinde herhangi bir birime yerleştirilecek olsanız verilen grafiği ve ortalama maaşı göz önüne alarak şu anki işinizi bırakıp bu iş yerinde çalışmak ister misiniz?

Evet / Hayır

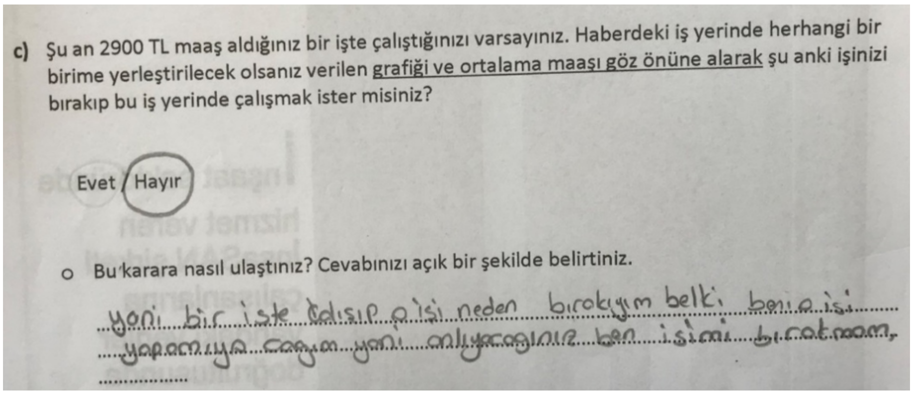
o Bu karara nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

çünkü orda daha çok para veriyorlar

Figure 4.7 Evaluation of average by a Level 1-2 student in Q1c

Figure 4.8 shows a sample of an idiosyncratic response made by S144 in Q1c. S144 states that there is no meaning in giving up her work while she is still working. She also thinks that there is a possibility that she is not suitable for that work; therefore, she decides not to give up her current work. Just as S144, 10.3%

of the students evaluated the news but did not focus on the critical features in the given news.



c) Şu an 2900 TL maaş aldığınız bir işte çalıştığınızı varsayınız. Haberdeki iş yerinde herhangi bir birime yerleştirilecek olsanız verilen grafiği ve ortalama maaşı göz önüne alarak şu anki işinizi bırakıp bu iş yerinde çalışmak ister misiniz?

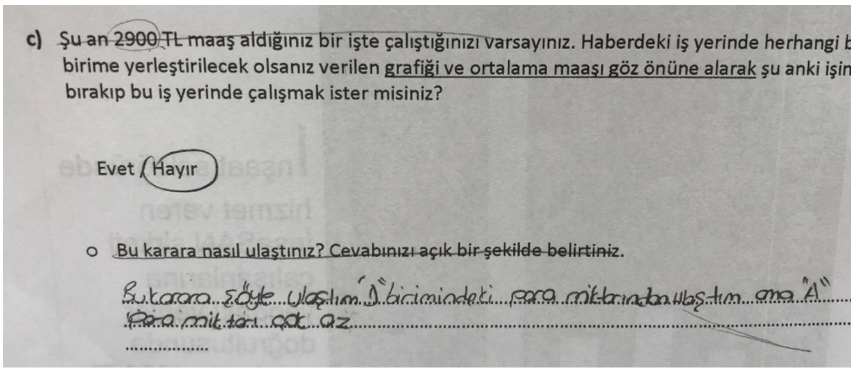
Evet / Hayır

o Bu karara nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

yani bir işte çalışıp o işi neden bırakayım belki benim o işi yapamıyacağım yani anlıyacağınız ben işimi bırakmam,

Figure 4.8 Evaluation of average by a Level 1-2 student in Q1c

Different from the students in Figure 4.7 and Figure 4.8, some students whose levels were determined as Level 1-2 made their evaluations focusing on only single ideas or some values in the given data. They did evaluate the given news in Q1c but did not question the average in the news and did not take into account the effect of all the data in the reported average. For instance, the student in Figure 4.9 stated that she reached her decision based on the salary of department D whose personnel get the minimum income in the company. She also stated that the salary of department A was much lower.



c) Şu an 2900 TL maaş aldığınız bir işte çalıştığınızı varsayınız. Haberdeki iş yerinde herhangi b birime yerleştirilecek olsanız verilen grafiği ve ortalama maaşı göz önüne alarak şu anki işin bırakıp bu iş yerinde çalışmak ister misiniz?

Evet / Hayır

o Bu karara nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

Bu karara şöyle ulaştım "D" birimindeki para miktarından ulaştım ama "A" para miktarı çok az

Figure 4.9 Evaluation of average by a Level 1-2 student in Q1c

On the other hand, the overall analysis of students' definitions, interpretations and evaluations of the concept of average in Q1 indicated that more than half of the students (62.1 %) performed at Level 3-4 in Q1. According to the modified version of the framework, students still use single ideas at Level 3 in defining or interpreting the average concept. However, students begin to use ideas related to measures of central tendency in either defining or interpreting the concept of average, especially in questions requiring the interpretation of the computation of average. Besides, at Level 4, students generally use some ideas related to measures of central tendency in defining and interpreting the concept of average. Some ideas related to the representative nature of the concept of average also begin at Level 4 in the definition and interpretation of the concept of average. It was stated in the framework that although students at Level 4 could compute the mean or median of a small data set, they could not compute weighted mean. Critical questioning of the concept of average in social or scientific context is still not observed at Level 3-4. As previously stated, findings of the present study showed that 62.1% of the students performed at Level 3-4 in Q1. Table 4.2 and 4.3 below reveals definitions and interpretations of students whose levels were determined as Level 3-4 in Q1. Since the characteristics observed at Level 3 and Level 4 are specified clearly in the framework, the levels of the students could be determined as Level 3 and Level 4 separately. More specifically, while 19.2% of the students performed at Level 3 in Q1, 46.8% of the students performed at Level 4. The definitions and interpretations made by the students whose levels were determined as Level 3 and Level 4 are presented in separate tables below. Moreover, since critical questioning of the concept of average was not observed in Level 3-4 students, the tables do not include an evaluation component.

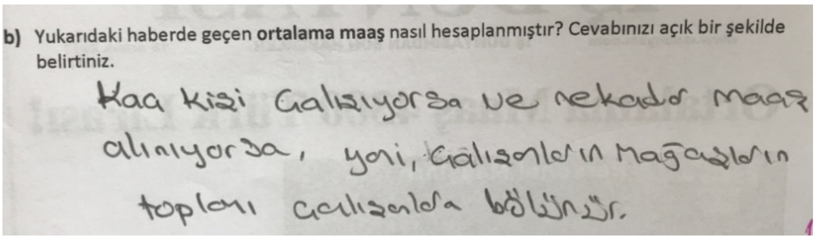
Table 4.2 The distribution of students at Level 3 in terms of their definitions and interpretations of the concept of average in Q1

	Definition of average	**Interpretation regarding the meaning of average**	**Interpretation regarding the computation of average**
Students at Level 3	Ideas related to measures of central tendency -Mean: 2 (1.2%) -Median: 2 (1.2%) -Mode: 0 (0.0%)	Ideas related to measures of central tendency -Mean: 1 (0.6%) -Median: 3 (1.8%) -Mode: 4 (2.4%)	Ideas related to measures of central tendency - Mean: 20 (7.9%) -Median: 0 (0.0%) -Mode: 0 (0.0%)
	Single Ideas 12 (7.2%)	Single Ideas 12 (7.2%)	Single Ideas 0 (0.0%)
	Irrelevant Responses 3 (1.8%)	Irrelevant Responses 10 (6.0%)	Irrelevant Responses 7 (4.2%)
	No Answer 13 (7.8%)	No Answer 2 (1.2%)	No Answer 5 (3.0%)
Total	32 (19.2%)	32 (19.2%)	32 (19.2%)

Table 4.2 indicated that 32 students (19.2%) performed at Level 3 in Q1. When the definitions of Level 3 students were examined, it was observed that 7.8% of the students could not present any definition related to the concept of average in Q1d. While 7.2% of the students used some single ideas in defining the concept of average, only 2.4% of the students used ideas related to measures of central tendency in defining the concept of average. There were also 3 students (1.8%) who provided some irrelevant responses while defining the average in Q1d. As can

be remembered, two questions related to the interpretation of the concept of average exist in Q1. While Q1a was related to the interpretation of the meaning of the given average in the news, Q1b was related to the interpretation of the computation of the average in the news. Table 4.2 showed that most of the students (7.2%) who were at Level 3 in Q1 interpreted the meaning of the concept of average in the given news by using single ideas. While 6.0% of the students presented some irrelevant responses in interpreting the meaning of the concept of average in Q1a, 4.8% of the students interpreted it using ideas related to three measures of central tendency. There were 2 students (1.2%) who did not provide any interpretation regarding the meaning of the concept of average. On the other hand, most of the students who performed at Level 3 (7.9%) interpreted the computation of the average in the given news by using ideas related to the measures of central tendency but the only idea observed was *mean*. While 7 (4.2%) of the students presented some irrelevant responses in interpreting the computation of the average in the given news in Q1b, 3.0% of the students could not provide any answer. In further sections, definitions, interpretations and evaluations of Level 3 students in Q1 are presented.

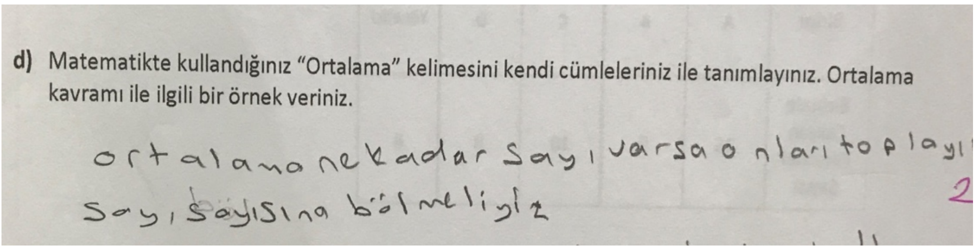
As indicated in Table 4.2, the findings of the present study revealed that 7.2% of the students whose statistical literacy levels were determined to be at Level 3 used some single ideas similar to those at Level 2 in defining the concept of average in Q1d or interpreting the meaning of the concept of average in Q1a. However, these students expressed the idea of *mean* in interpreting the calculation of the average in Q1b without any discussion of why it is appropriate to use that measure in calculation or made no effort to justify the given average using the mean algorithm. Hence, these students were placed at Level 3. For instance, the student in Figure 4.10 explained that the average in the given news was calculated by dividing the sum of the income of the personnel by the number of personnel in the company. Nevertheless, his suggestion for the calculation of the average was not appropriate to his definition or interpretation since this student defined the concept of average in Q1d as the *sum* of the given values. Similarly, he interpreted the meaning of the average income in Q1a as the sum of the given income.


b) Yukarıdaki haberde geçen **ortalama maaş** nasıl hesaplanmıştır? Cevabınızı açık bir şekilde belirtiniz.

Kaa kişi Galışıyorsa ve nekadar maaş alınıyorsa, yani, Galışanların mağaşların toplamı Galışanlara bölünür.

Figure 4.10 Interpretation of average by a Level 3-4 student in Q1b

Furthermore, it was observed that some students used ideas related to three measures of central tendency in either defining the concept of average in Q1d or interpreting its meaning in Q1a. However, they could not transfer their understandings to all of the definition or interpretation questions. Therefore, these students were placed at Level 3. For instance, the student in Figure 4.11 defined the concept of average in Q1d by using the idea of *mean*. She stated that the average was division of the sum of the existing numbers by the number of number. However, she could not apply her definition to her interpretation of the meaning of the concept of average in Q1a or her interpretation of the calculation of the concept of average in Q1b. While she stated that the meaning of the concept of average is *about* in Q1a, she provided her own opinion regarding the income in the company while interpreting the computation of the average in Q1b. Her answer was presented in Figure 4.5.


d) Matematikte kullandığınız "Ortalama" kelimesini kendi cümleleriniz ile tanımlayınız. Ortalama kavramı ile ilgili bir örnek veriniz.

ortalama nekadar sayı varsa onları toplayıp sayı sayısına bölmeliyiz

2

Figure 4.11 Definition of average by a Level 3-4 student in Q1d

Besides, as indicated in Table 4.2, 7.9 % of the students whose statistical literacy levels were determined as Level 3 used some ideas related to the *mean* in the interpretation question regarding the computation of average in Q1b. However, it was observed that some of those students provided only ideas related to addition and division as in the definition of the mean algorithm but did not relate it to the given context. S62 in Figure 4.12 answered the question of how the average income in the given news was calculated as first of all, all of the numbers were added, and then they are divided. The similarity between the average definition of S64 in Q1d in Figure 4.11 and that of S62 in Q1b is explicit.

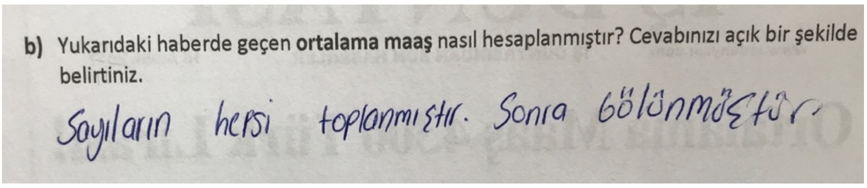


Figure 4.12 Interpretation of average by a Level 3-4 student in Q1b

Moreover, it was noticed that some students whose statistical literacy levels were determined as Level 3 and used some ideas related to the *mean* in the interpretation question regarding the computation of average in Q1b could not correctly mention the mean algorithm. For instance, the student in Figure 4.13 showed the calculation of the average in the given news by reversing the algorithm; in other words, by dividing the number of personnel by the income. Furthermore, the student did not state that whether the income is the total income of the all personnel or the total income of some personnel.

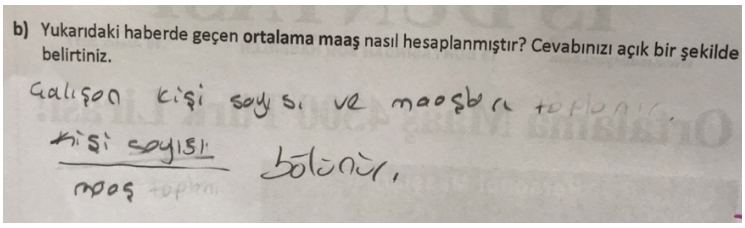
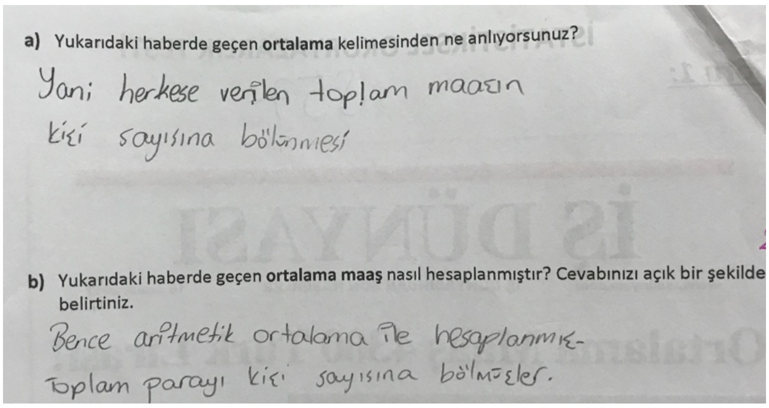


Figure 4.13 Interpretation of average by a Level 3-4 student in Q1b

On the other hand, the overall analysis of students' answers in four sub-questions in Q1 revealed that 78 students (46.8%) performed at Level 4 in Q1. Different from the students whose statistical literacy levels were determined as Level 3, students whose levels were determined to be at Level 4 used ideas related to measures of central tendency in at least two sub-questions in Q1. For instance, S57 in Figure 4.14 gave the signals of the mean in both interpretation questions in Q1. She interpreted the meaning of the concept of average as division of the sum of all of the given income by the number of people in Q1a. In Q1b, she said that the average in the given news was calculated through the mean; they divided the total money by the number of people. However, S57 could not define the average concept in Q1d; hence, the statistical literacy level of this student was determined as Level 4.


a) Yukarıdaki haberde geçen **ortalama** kelimesinden ne anlıyorsunuz?

Yani herkese verilen toplam maaşın kişi sayısına bölünmesi

b) Yukarıdaki haberde geçen **ortalama maaş** nasıl hesaplanmıştır? Cevabınızı açık bir şekilde belirtiniz.

Bence aritmetik ortalama ile hesaplanmış- toplam parayı kişi sayısına bölmüşler.

Figure 4.14 Interpretations of average by a Level 3-4 student in Q1a and Q1b

Table 4.3 below presents the definitions of the concept of average of Level 4 students in Q1d and their interpretations related to it in Q1a and Q1b.

Table 4.3 The distribution of students at Level 4 in terms of their definitions and interpretations of the concept of average in Q1

	Definition of average	**Interpretation regarding the meaning of average**	**Interpretation regarding the computation of average**
	Ideas related to representativeness 3 (1.8%)	Ideas related to representativeness 4 (2.4%)	Ideas related to representativeness 0 (0.0%)
	Ideas related to measures of central tendency -Mean: 46 (27.6%) -Median: 6 (3.6%) -Mode: 3 (1.8%)	Ideas related to measures of central tendency -Mean: 41 (24.6%) -Median: 9 (5.4%) -Mode: 7 (4.2%)	Ideas related to measures of central tendency -Mean: 72 (43.2%) -Median: 1 (0.6%) -Mode: 1 (0.6%)
Students at Level 4	Multiple definitions 3 (1.8%)	Multiple interpretations 2 (1.2%)	Multiple interpretations 2 (1.2%)
	Single Ideas 2 (1.2%)	Single Ideas 6 (3.0%)	Single Ideas 0 (0.0%)
	Irrelevant Responses 0 (0.0%)	Irrelevant Responses 6 (3.6%)	Irrelevant Responses 3 (1.8%)
	No Answer 15 (9.0%)	No Answer 3 (1.8%)	No Answer 1 (0.6%)
Total	78 (46.8%)	78 (46.8%)	78 (46.8%)

Table 4.3 revealed that 55 students out of 78 (33.0%) used ideas related to measures of central tendency in defining the concept of average in Q1d. *Mean* was the most preferred measure in defining the concept of average. While 9.0% of the students could not define the concept of average, 3 (1.8%) of them presented some ideas related to the representative nature of average while defining the concept of average in Q1d. On the other hand, in defining the concept of average in Q1d, 2 students (1.2%) still used some single ideas and 1.8% of the students used more than one idea. Ideas related to three measures of central tendency were most popular among the answers of Level 4 students in interpreting the meaning of the average in the given news in Q1a. Similar to the definition question, *mean* was the most preferred idea (24.6%) in interpreting the meaning of the concept of average. It was realized that 3.0% of the students provided either some single ideas or irrelevant responses in interpreting the meaning of the concept of average in Q1a. Similar to the definition question, 2.4 % of the students presented some ideas related to the representative nature of the concept of average in interpreting its meaning in the given news. While 3 students (1.8%) could not interpret the meaning of the concept of average in Q1a, 2 of them (1.2%) provided multiple interpretations. Lastly, when the students' answers were analyzed in the interpretation question related to the computation of the average in the given news in Q1b, it was noticed that again students chose ideas related to three measures of central tendency. However, almost no other measure than *mean* was encountered in the students' interpretation of the calculation of the average in Q1b. 43.2% of the students used the idea of *mean* when interpreting the calculation of the average. The remaining 4 students (2.4%) either could not interpret the average concept or presented some irrelevant responses in interpreting the computation of the average in Q1b. In further sections, definitions and interpretations of Level 4 students are examined in detail by providing examples from students' answers.

It can be inferred from Table 4.3 that mean was the most preferred measure of central tendency by Level 4 students in defining the concept of average in Q1.

Interestingly, 15 students (9.0%) whose levels were determined as Level 4 could not present any definition in Q1d. The number is much higher than it is in the previous level, Level 3. However, it was observed that 9 of those 15 students that did not define the concept in Q1d provided some examples that showed the application of the mean algorithm in a different context, just as the student in Figure 4.15.

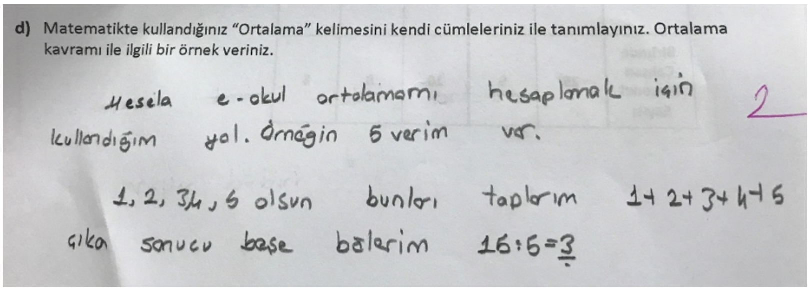


Figure 4.15 Example of average by a Level 3-4 student in Q1d

S160 whose statistical literacy level was determined as Level 4 in Figure 4.15 tried to define the concept of average in Q1d by presenting an example related to her average grade point. In other words, this student needed a context to show his/her understanding of the concept of average.

On the other hand, students whose statistical literacy levels were determined to be at Level 4 used some ideas related to other measures of central tendencies in defining the concept of average. However, the number was much less when it was compared with the number of students who used the idea of *mean* in defining the concept of average. For instance, the student in Figure 4.16 defined it as the *middle value* in the data set in Q1d. He said that average is the middle of anything. Although he did not use the word *median* explicitly, he explained his idea clearly by providing an example, which can be examined in Figure 4.16.

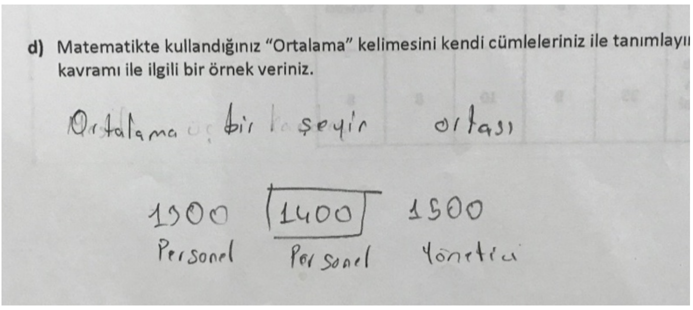


Figure 4.16 Definition of average by a Level 3-4 student in Q1d

Similarly, students whose statistical literacy levels were determined as Level 4 did not use the exact word of *mode* but they used some words which referred to the idea of mode, such as *generally* or *majority* in defining the concept of average in Q1d. Only 3 students (1.8%) presented the idea of mode in defining the concept of average. Furthermore, it was observed that 1.8% of the definitions made by Level 4 students reflected the representative nature of the concept of average. A sample answer of such a student is shown in the following figure.

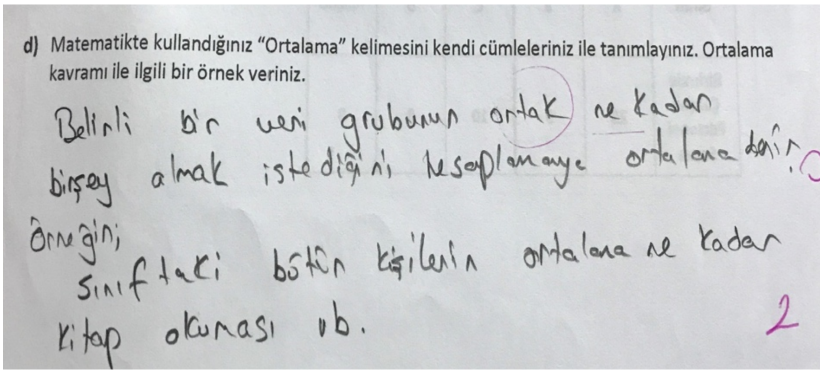


Figure 4.17 Definition of average by a Level 3-4 student in Q1d

S95 in Figure 4.17 stated that average is something about calculation and finding a common value for a data set. Although this student did not mention representativeness explicitly in her definition, it was thought that this student is aware that average is a measure related to all of the data in the given data set. It was also noticed that there were a few Level 4 students who presented multiple definitions in Q1d, just as the student in Figure 4.18.

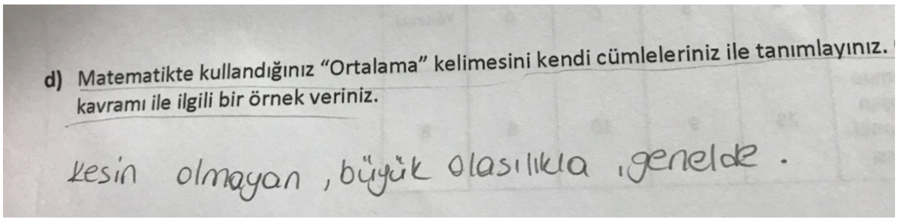


Figure 4.18 Definition of average by a Level 3-4 student in Q1d

S41 in Figure 4.18 defined the concept of average in Q1d using three words: *uncertain*, *quite likely* and *generally*. S41 believes that average does not give a certain value but it reflects the general situation. To state it differently, this student defined the concept by using both a single idea, *uncertain,* and the idea of mode by using the word *generally*. Different from S41, 2 students (1.2%) defined the concept by using two measures of central tendency. Figure 4.19 displays the definition of these students.

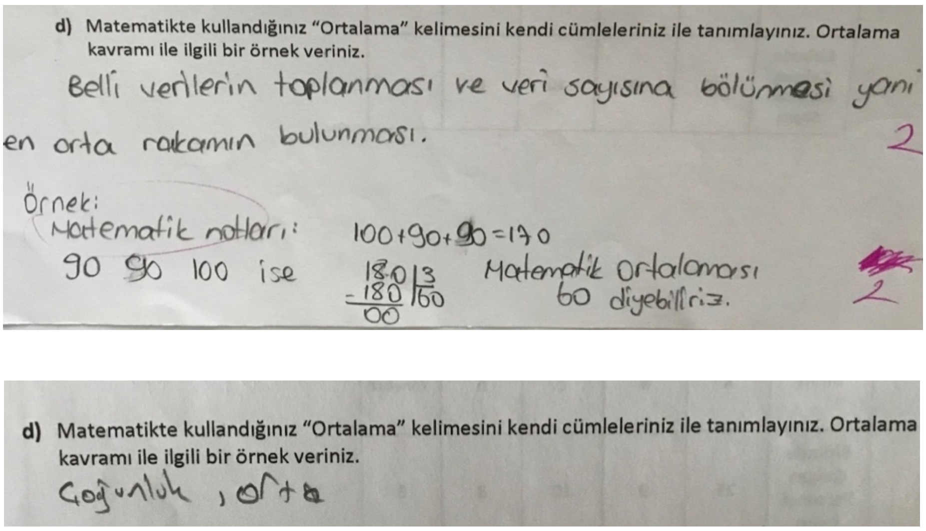


Figure 4.19 Definition of average by two Level 3-4 students in Q1d

S34 in Figure 4.19 said that the average was the division of the sum of certain data by the number of data, that is, she continues, finding the *middle number*. In other words, S34 defined the concept of average by using the ideas of *mean* and *median* in Q1d. On the other hand, S89 defined it by using the ideas of *mode,* using the word *majority,* and *median*. Moreover, although S34 believed that with the application of *mean*, the *median* of the data set is obtained, it can be observed from her example in Figure 4.19 that she did not realize that 60 is not the middle value in her data set.

When the answers of the students whose statistical literacy level was determined as Level 4 to the interpretation question in Q1 were examined, it was realized that mean was the most preferred measure of central tendency, as in the definition of the concept of average in Q1d. As at Level 3, the usage of *mean* increased much more, reaching 44.4%, in the interpretation question regarding the computation of average in Q1b. Even though some students defined or interpreted the concept of average as *median* or *mode* in the definition question in Q1d or in the interpretation of the meaning of the concept of average in Q1a, they used the idea of *mean* in their interpretation of the calculation of the given average in the given news in Q1b. For example, S89 in Figure 4.19 above defined the concept by using the ideas of *median* and *mode*. Moreover, he interpreted the concept of average again as *majority* in the interpretation question which asked for the meaning of the average in the given news in Q1a, which can be seen in Figure 4.20. However, as can be observed in Figure 4.20, he stated that to obtain the *majority*, the sum of all could be divided by the number of data in Q1b.

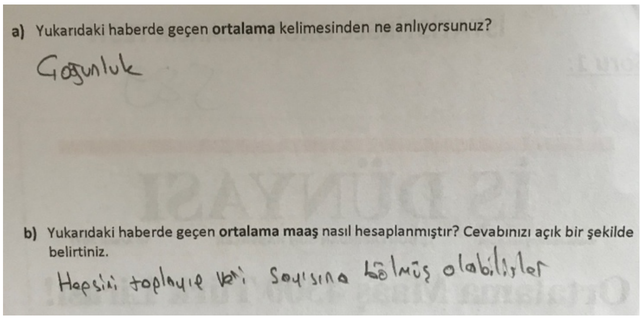
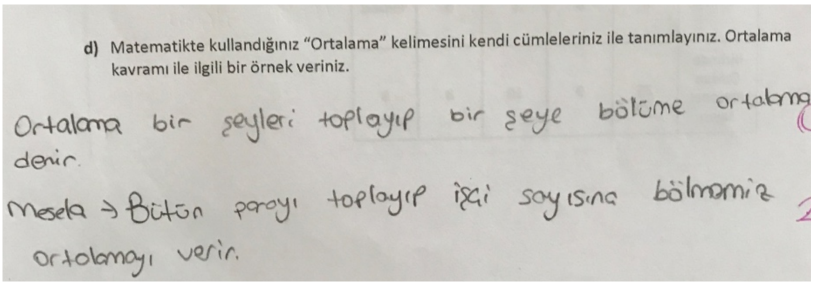


Figure 4.20 Interpretations of average by a Level 3-4 student in Q1a and Q1b

3.0% of the students whose statistical literacy levels were determined as Level 4 used some single ideas in interpreting the meaning of the concept of average in Q1a. However, it was realized that these students defined the concept of average by using the idea of *mean* in Q1d. For instance, S44 in Figure 4.21 defined it in Q1d as dividing the sum of something by something else. However, she interpreted the meaning of the average income in the given news in Q1a as *about*. Since S44 interpreted the calculation of the concept of average in the given news in Q1b through the *mean* of the given data, she was placed at Level 4, but the differences between the definition and interpretation provided by S44 reveals the gap between the student's own understanding of the concept of average and what she had learnt in the school.



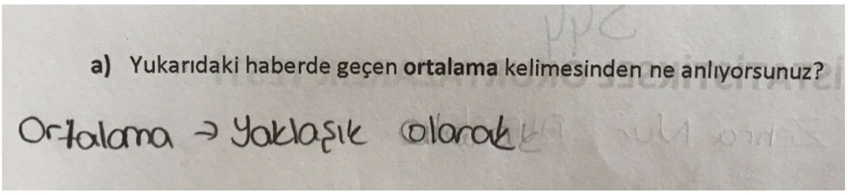


Figure 4.21 Definition and interpretation of average by a Level 3-4 student in Q1d and Q1a, respectively

The interpretation of the meaning of the average concept of one of the Level 4 students exemplifies the above situation once more. The interpretation of the student in Q1a could be observed in Figure 4.22. S96 in Figure 4.22 provided two interpretations related to the meaning of the concept of average in Q1a, one for the context of mathematics and the other as used in the public. While she interpreted the meaning of the concept of average as the mean in mathematics, she stated that average is used to mean general in the daily language of the public.

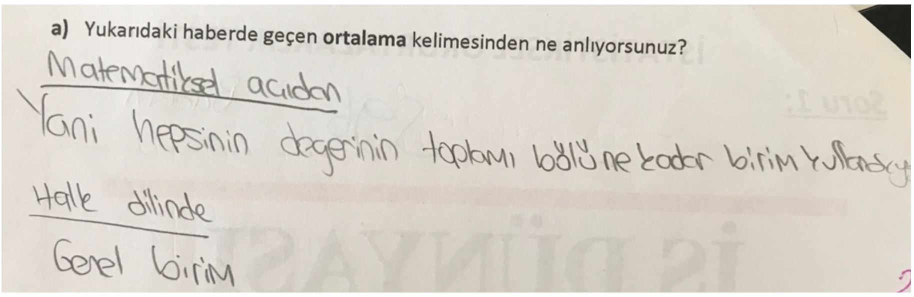


Figure 4.22 Interpretations of average by a Level 3-4 student in Q1a

Furthermore, it was realized that although 43.2% of the students whose statistical literacy levels were determined to be at Level 4 used the idea of *mean* in the interpretation question related to computation of average in Q1b, most of them did not state or apply the algorithm correctly as the students whose statistical literacy levels were determined as Level 3. For example, the student in Figure 4.23 below thought of using the *mean* algorithm to find the average income in the company. However, instead of dividing the sum of incomes with the total number of personnel in the company, he suggested to dividing it by 5, which is the number of departments in the company. To put it differently, S32 did not consider computing the weighted mean in Q1b.

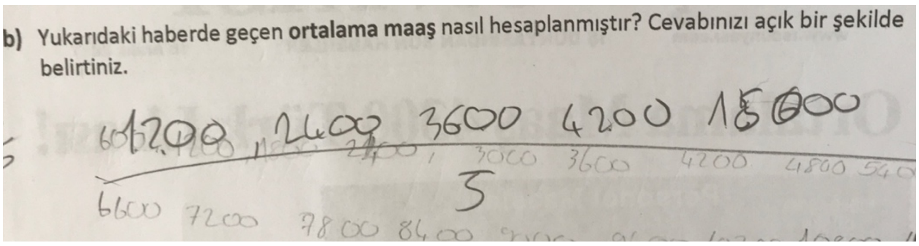


Figure 4.23 Interpretation of average by a Level 3-4 student in Q1b

Only 2 students (1.2%) among the other Level 4 students in Q1b suggested finding the *weighted mean* to find the given average as the student in Figure 4.24, but none of them found the average by applying their suggestion. This was expected of Level 4 students since, in the framework, it was stated that students could not apply the algorithm in complex situations, such as in the calculation of the weighted mean.

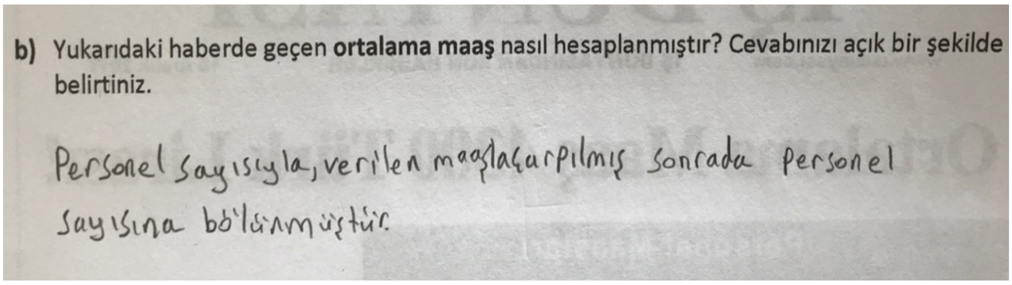


Figure 4.24 Interpretation of average by a Level 3-4 student in Q1b

Different from the above students, some students whose levels were determined as Level 4 mentioned finding the *mean* by reversing the algorithm in the interpretation question related to the calculation of the average in the given news in Q1b. To state it differently, these students, as the student in Figure 4.13, suggested finding the *mean* by dividing the total number of data by the sum obtained from adding all the numbers in the data.

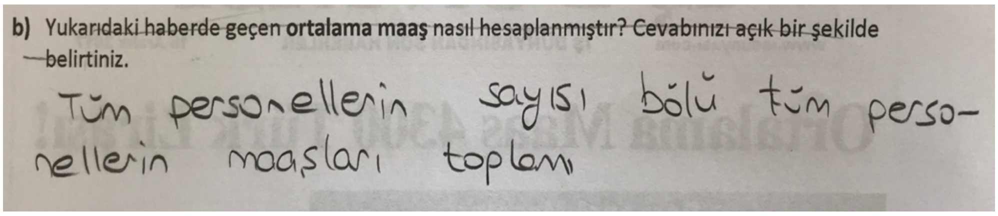


Figure 4.26 Interpretation of average by a Level 3-4 student in Q1b

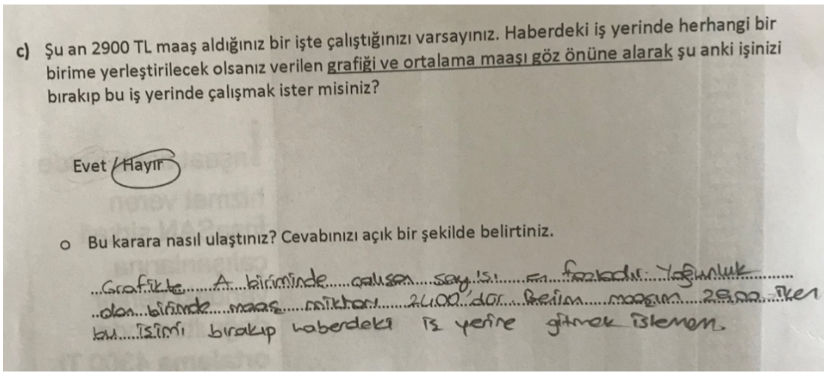
On the other hand, the overall analysis of the students' answers in the four sub-questions in Q1 revealed that 10 students (6.1%) could reach Level 5-6 in Q1. In the modified version of the framework, it was asserted that in addition to defining and interpreting the average concept correctly, students critically evaluate the claims regarding the concept of average observed in social or scientific contexts at Level 5-6. It is stated that at Level 6, students recognize the effect of outliers in the calculation of mean and, therefore, suggest the median as the appropriate measure of central tendency. At Level 5, students do not realize the effect of an outlier, but can focus on the mode of the given data and state that if most agree, it will be the average. Findings of the present study showed that none of the students recognized the outlier in the data set given in Q1; therefore, no students performed at Level 6. However, it was observed that some students evaluated the given average focusing on the mode of the given data set and on the uncertainty of the departments in Q1. Therefore, these students were placed at Level 5. Table 4.4 below displays the definitions, interpretations and evaluations of the concept of average of the students whose statistical literacy levels were determined to be at Level 5.

Table 4.4 indicates that most of the students (5.4%) at Level 5 generally used ideas related to measures of central tendency in defining the concept of average in Q1d. Ideas related to the representative nature of the concept of average was also used by one of the students in defining it in Q1d. When the interpretations of Level 5 students related to the meaning of the concept of average in Q1a were examined, it was realized that students interpreted its meaning by using ideas related to measures of central tendency. The only emerging measure was the *mean*. Furthermore, only 1 student (0.6%) presented multiple interpretations related to the meaning of the concept of average in Q1a. Similarly, all of the students (6.1%) used the idea of *mean* in their interpretation related to the computation of the average in the given news. Lastly, it was observed that 8 students (4.8%) used ideas related to the *mode* of the given data set when making their evaluations of the claim related to the average in the given news in Q1c. The remaining 2 students (1.2%) mentioned ideas related to *uncertainty of departments* when making their evaluations in Q1c. As can be easily realized, the only difference between Level 4 and Level 5 students were their evaluations of the claim related to the concept of average in Q1c. Level 5 students used similar ideas with those of Level 4 students in defining the average in Q1d and interpreting the average in Q1a and Q1b. Therefore, further sections do not present definitions and interpretations of students whose statistical literacy levels were determined as Level 5. Only evaluations of Level 5 students regarding the claim related to the given average in the given news in Q1c are provided with supporting examples from students' answers.

Table 4.4 The distribution of students at Level 5 in terms of their definition, interpretation and evaluation of the concept of average in Q1

	Definition of the average	**Interpretation regarding the meaning of the average**	**Interpretation regarding the computation of the average**	**Evaluation of the average**
Students at Level 5	Ideas related to representativeness 1 (0.6%)	Ideas related to representativeness 0 (0.0%)	Ideas related to representativeness 0 (0.0%)	Ideas related to representativeness 0 (0.0%)
	Ideas related to measures of central tendency - Mean: 7 (4.2%) - Median: 2 (1.2%) -Mode: 0 (0.0%)	Ideas related to measures of central tendency - Mean: 9 (5.4%) - Median: 0 (0.0%) -Mode: 0 (0.0%)	Ideas related to measures of central tendency - Mean: 10 (6.1%) - Median: 0 (0.0%)- Mode: 0 (0.0%)	Ideas related to measures of central tendency -Mode: 8 (4.8%) - Median: 0 (0.0%) -Mode: 0 (0.0%)
	Multiple Definitions 0 (0.0%)	Multiple Interpretations 1 (0.6%)	Multiple Interpretations 0 (0.0%)	Ideas related to uncertainty of department 2 (1.2%)
Total	10 (6.1%)	10 (6.1%)	10 (6.1%)	10 (6.1%)

It was previously stated that many of the students (4.8%) whose statistical literacy levels were determined as Level 5 evaluated the claim related to the given average in the news in Q1c by using the idea of mode. In other words, these students used the income of the maximum number of personnel in the company to evaluate the given average. For example, S70 in Figure 4.27 asserted that department A has the maximum number of personnel, and the amount of income in that department is 2400 TL; therefore, she decided not to give up her current work, where she gets 2900 TL. To state it differently, although S70 did not clearly mention it, she believed that the income of most of the personnel would be the average income in the company; that is 2400 TL.


c) Şu an 2900 TL maaş aldığınız bir işte çalıştığınızı varsayınız. Haberdeki iş yerinde herhangi bir birime yerleştirilecek olsanız verilen grafiği ve ortalama maaşı göz önüne alarak şu anki işinizi bırakıp bu iş yerinde çalışmak ister misiniz?

Evet / Hayır

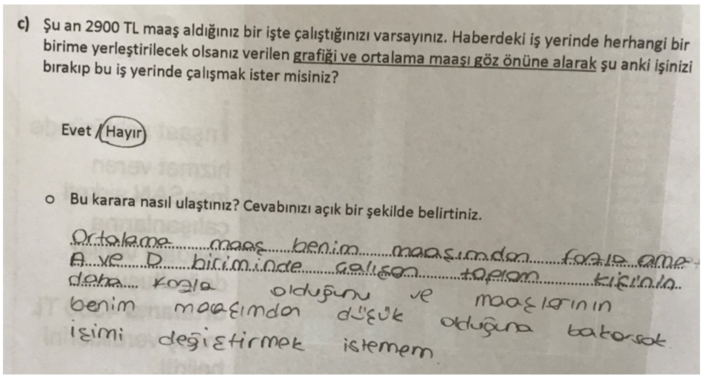
o Bu karara nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

Grafikte A biriminde çalışan sayısı en fazladır. Yoğunluk olan biriminde maaş miktarı 2400'dür. Benim maaşım 2900 iken bu işimi bırakıp haberdeki iş yerine gitmek istemem.

Figure 4.27 Evaluation of average by a Level 5-6 student in Q1c

Some Level 5 students, such as in Figure 4.28, also considered the number of personnel in department D together with the number of personnel in department A while presenting her evaluation of the claim related to the concept of average in Q1c. S43 in Figure 4.28 states that the average income in the news is higher than her income; however, if she takes into consideration that the sum of the personnel in departments A and D is high and their income is less than hers, she does not want to change his/her work.

It can be understood from the evaluation of S43 that these students still do not suspect the given average in the news. Nevertheless, it was believed that they were beginning to realize that the higher the number of personnel, the more effect it will have on the average income; therefore, these students were placed at Level 5.


c) Şu an 2900 TL maaş aldığınız bir işte çalıştığınızı varsayınız. Haberdeki iş yerinde herhangi bir birime yerleştirilecek olsanız verilen grafiği ve ortalama maaşı göz önüne alarak şu anki işinizi bırakıp bu iş yerinde çalışmak ister misiniz?

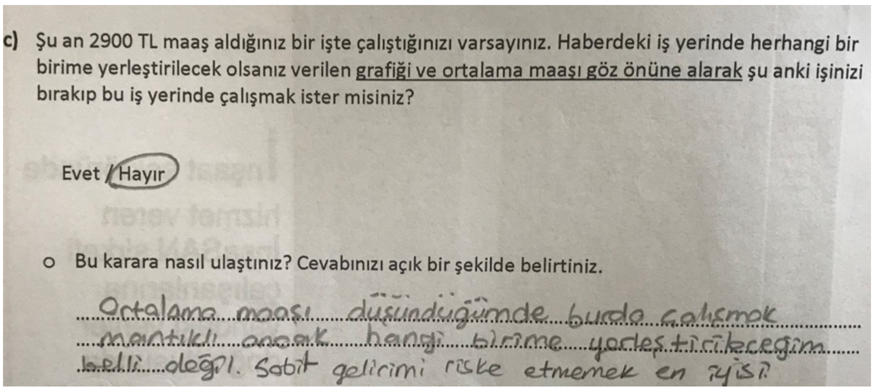
Evet / Hayır

o Bu karara nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

Ortalama maaş benim maaşımdan fazla ama A ve D biriminde çalışan toplam kişinin daha fazla olduğunu ve maaşlarının benim maaşımdan düşük olduğuna bakarsak işimi değiştirmek istemem

Figure 4.28 Evaluation of average by a Level 5-6 student in Q1c

Moreover, similar to the above students, 2 students (1.2%) made their evaluations using ideas related to the *uncertainty of departments*. Figure 4.29 displays the answer of such a student. S79 in the figure below said that it was logical to work in the company mentioned in the question when the average income was taken into account, but as it was not certain in which department he would be placed in, he thought it was better to take no risk. Similar the students whose statistical literacy levels were determined as Level 5 and who used the idea of *mode* in their evaluations in Q1c, these students still do not suspect the given average in the news. However, since these students focused on the critical features in the given news and since it was believed that these opinions could be the beginning for the recognition of the outlier, these students were placed at Level 5.


c) Şu an 2900 TL maaş aldığınız bir işte çalıştığınızı varsayınız. Haberdeki iş yerinde herhangi bir birime yerleştirilecek olsanız verilen grafiği ve ortalama maaşı göz önüne alarak şu anki işinizi bırakıp bu iş yerinde çalışmak ister misiniz?

Evet / Hayır

o Bu karara nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

Ortalama maaşı düşündüğümde burda çalışmak mantıklı ancak hangi birime yerleştirileceğim belli değil. Sabit gelirimi riske etmemek en iyisi

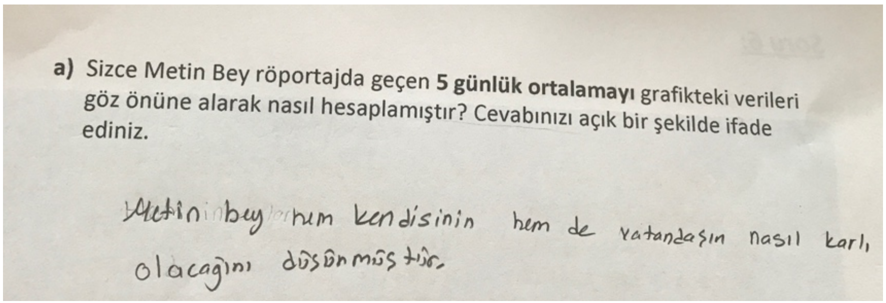
Figure 4.29 Evaluation of average by a Level 5-6 student in Q1c

4.1.2 The Concept of Average on Bar Graphs in Question 5

Question 5 (Q5) had similar purposes with Q1, but the contexts were different. In Q5, students were given an interview in which they were provided with a bar graph displaying the number of customers in terms of different package types in the hotel. In the interview, the average was calculated using the weighted mean and found that the average number of days that customers preferred was 5 days. However, the data presented in the interview was categorical; hence, the mean is not an appropriate measure of average type. Students were expected to realize the categorical data and suggest the use of the mode of the given data as the appropriate type of average. Q5 involved 2 sub-questions. In the first sub-question, Q5a, the students were asked how the average in the interview was calculated. In this way, their interpretation of the concept of average in the given context was examined. In the second sub-question, Q5b, students were required to evaluate the given average asking the question of whether or not they agreed with the decision of the director regarding the organizing packages of a maximum of 5 days by taking into account the bar graph given.

The statistical literacy levels of students were determined by the overall analysis of the students' answers to the two sub-questions in Q5 and their average definitions in Q1d. As Table 4.1 indicated, just as in Q1, most of the students (41.5%) performed at Level 3-4 in Q5. The percentage of students at Level 1-2 and Level 5-6 in Q5 were 31.1% and 27.4%, respectively. These percentages are higher than the those of students performing at Level 1-2 and Level 3-4 in Q1. Since the frequencies of students' definitions of average at all levels were almost the same in Q1, further sections explain the students' interpretations related to the computation of the average in Q5a and their evaluations of the claim related to the given average in the interview in Q5b. The findings for Level 1-2, Level 3-4 and Level 5-6 students are presented in the following paragraphs by providing examples from students' answers.

First of all, 51 students (31.1%) performed at Level 1-2 in Q5. The students' interpretations and evaluations at Level 1-2 regarding the average concept showed similar characteristics with those of Level 1-2 students in Q1. To state it differently, at Level 1, students could not interpret the given average in the interview. They either left the interpretation question blank or provided irrelevant responses as the student in Figure 4.30. S36 in Figure 4.30 explained that Mr. Metin thought about how both he and his customers would be profitable. She did not mention any idea regarding the computation of the average with the given data; therefore, this student was placed into Level 1.

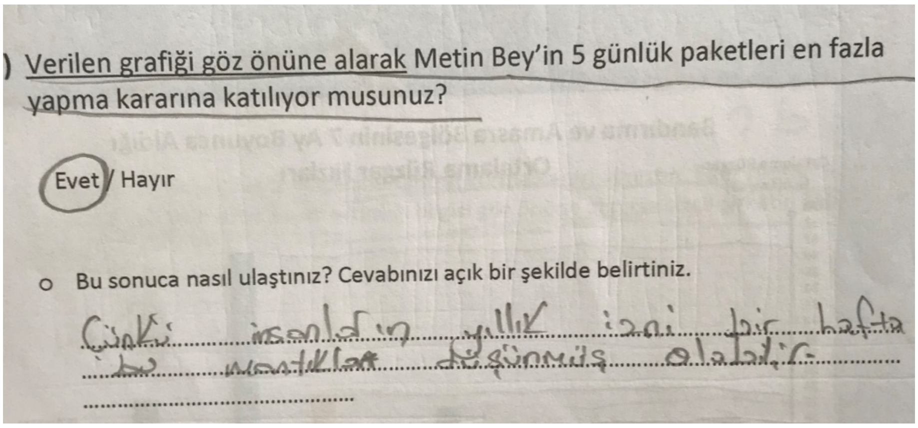

a) Sizce Metin Bey röportajda geçen **5 günlük ortalamayı** grafikteki verileri göz önüne alarak nasıl hesaplamıştır? Cevabınızı açık bir şekilde ifade ediniz.

Metin bey hem kendisinin hem de vatandaşın nasıl karlı olacağını düşünmüştür.

Figure 4.30 Interpretation of average of a Level 1-2 student in Q5a

On the other hand, it was observed that students whose statistical literacy levels were determined as Level 2 in Q5 still have difficulties in interpreting the calculation of the average in Q5a. These students were placed at Level 2 because they have used some single ideas, like *about* in defining the concept of average in Q1d.

When the evaluations of Level 1-2 students in Q5 related to the given average in the interview was examined, evaluations similar to those in Q1 were encountered. In other words, these students could not evaluate the given average in the interview in Q5b, but either accepted the given average directly or presented idiosyncratic responses, just as S114 in Figure 4.31. S114 agreed with the decision given because she stated that Mr. Metin thought that people's annual holiday was one week.


) Verilen grafiği göz önüne alarak Metin Bey'in 5 günlük paketleri en fazla yapma kararına katılıyor musunuz?

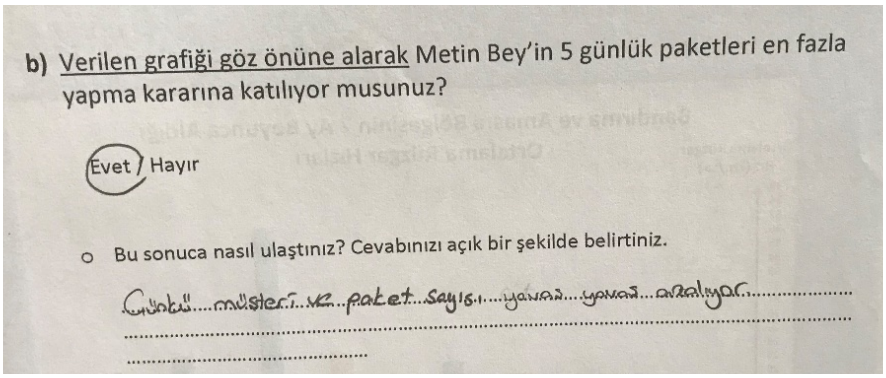
Evet / Hayır

o Bu sonuca nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

Çünkü insanların yıllık izni bir hafta bu mantıkla düşünmüş olabilir.

Figure 4.31 Evaluation of average by a Level 1-2 student in Q5b

Different from Q1, some students at Level 1-2 in Q5b depended on the given graph in making their evaluations related to the average given in the interview, but they only read the given data or mentioned the variation in the data. For instance, S141 in Figure 4.32 explained that she agreed with the decision of the director since the number of customers and packages were decreasing slowly. However, S141 did not explain how she used variation in making her evaluations in Q5b.


b) Verilen grafiği göz önüne alarak Metin Bey'in 5 günlük paketleri en fazla yapma kararına katılıyor musunuz?

Evet / Hayır

o Bu sonuca nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

Çünkü müşteri ve paket sayısı yavaş yavaş azalıyor.

Figure 4.32 Evaluation of average by a Level 1-2 student in Q5b

On the other hand, the overall analysis of students' interpretations and evaluations of the average concept in Q5 and their definitions of the average concept in Q1d indicated that almost half of the students (41.5%) performed at Level 3-4 in Q5. Since the characteristics observed at Level 3 and Level 4 were specified clearly in the framework, the students' levels could be determined as Level 3 and Level 4 separately in Q5. More specifically, while 22.2% of the students performed at Level 3 in Q5, 16.8 % of the students performed at Level 4. As previously mentioned, the frequencies of the definitions made by Level 3 and Level 4 students related to the concept of average was almost the same in Q1. Furthermore, as stated in the framework, critical questioning of the concept of average was still not observed at Level 3 and Level 4. It was revealed in the current study that Level 3-4 students made their evaluations related to the concept of average in the interview in Q5b, just as the students at Level 1-2 in Q5. Thus, the following table displays only the interpretations of students whose statistical literacy levels were determined to be at Level 3 and Level 4 in Q5 regarding the computation of the average in the given interview in Q5a.

Table 4.5 indicates that 22.2% of the students performed at Level 3 in Q5. Most of these students (7.8%) interpreted the computation of the given average in the interview using ideas related to the measure of central tendency in Q5a. Just as in Q1, the idea of the *mean* was the most preferred one (7.8%). While 12 students (7.2%) provided an irrelevant response in interpreting the computation of the average, 4.2% of the students could not present any interpretation in Q5a. These students were placed at Level 3 in Q5 because they used some ideas related to the measure of central tendency in defining the concept of average in Q1d. On the other hand, 16.8% of the students' statistical literacy levels were determined as Level 4 in Q5. Different from the students whose statistical literacy levels were determined as Level 3, students whose levels were determined as Level 4 used ideas related to measures of central tendency in both interpreting the concept of average in Q5a and defining the average in Q1d. It was observed that all of the Level 4 students (16.8%) in Q5 interpreted the calculation of the average in the given interview by using ideas related to the measure of central tendency. Only 2 of

28 (1.2%) students used some ideas related to the *mode* of the given data while interpreting the calculation of the average in Q5a. The remaining students (15.6%) used the idea of the *mean* in their interpretations.

Table 4.5 The distribution of Level 3 and Level 4 students in terms of their interpretations regarding the computation of average in Q5

	Students at Level 3	**Students at Level 4**
Interpretation regarding the computation of average	Ideas related to measures of central tendency - Mean: 13 (7.8%) -Median: 0 (0.0%) -Mode: 5 (3.0%) Irrelevant Responses 12 (7.2%) No Answer 7 (4,2%)	Ideas related to measures of central tendency - Mean: 26 (15.6%) -Median: 0 (0.0%) -Mode: 2 (1.2%) Irrelevant Responses 0 (0.0%) No Answer 0 (0.0%)
Total	37 (22.2%)	28 (16.8%)

Not much difference is observed when Table 4.5 is compared with Table 4.2 and Table 4.3, which display the definitions and interpretations of Level 3 and Level 4 students in Q1. In other words, students at both Level 3 and Level 4 generally interpreted the calculation of the average by using the ideas of the *mean* in Q1b and Q5b. However, the number of students using the idea of the *mean* at Level 4 in Q5b (15.6%) was a little higher than the number of students using the idea of the *mean* at Level 4 in Q1b. For example, S9 in Figure 4.33 found the *mean* of the data by dividing the total number of customers by 5, which is the number of different packages in the hotel. However, the answer of 90 did not bother the student. As this example reveals, although the idea of the *mean* was the most mentioned measure of central tendency in the calculation of the average, the students seemed to experience some difficulties in the application of the mean as in Q1.

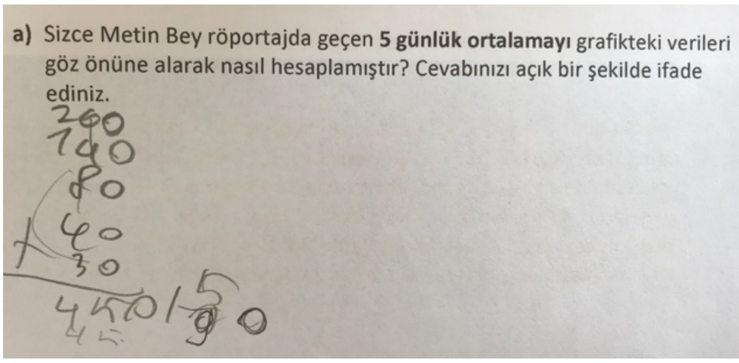


Figure 4.33 Interpretation of average by a Level 3-4 student in Q5a

It was realized that different from Q1, the idea of the mode was used more frequently while interpreting the calculation of the average in Q5a by both Level 3 and Level 4 students. For instance, S59 in Figure 4.34, whose statistical literacy level was determined as Level 4 in Q5, stated that the average was calculated by taking into account the day selected by many of the customers. However, this student did not feel the necessity of checking out her opinion from the data given.

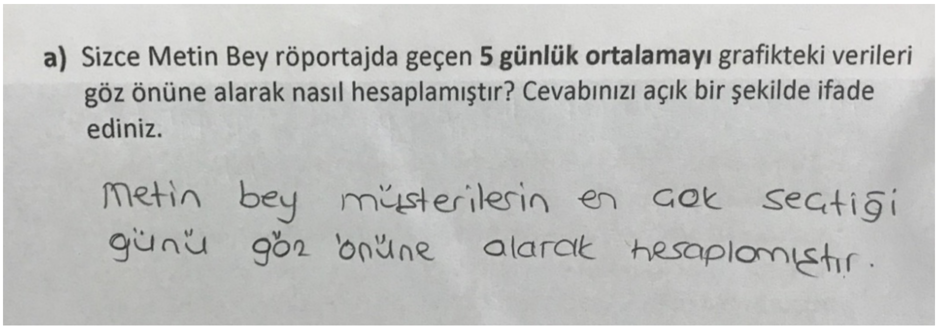
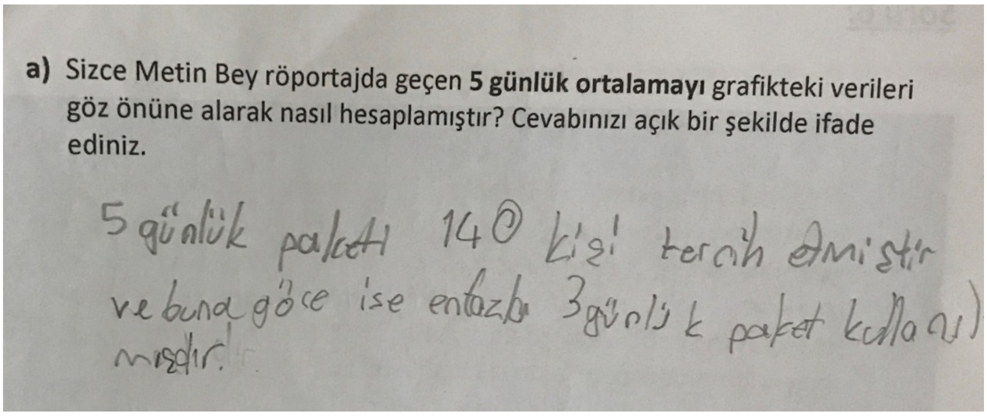


Figure 4.34 Interpretation of average by a Level 3-4 student in Q5a

It was noticed that only one Level 4 student in Q5, who interpreted the calculation of the average in the given interview as the *mode,* checked out his opinion from the given data. S30 in Figure 4.35 stated that 140 customers preferred 5-day packages and based on this, a maximum of 3-day packages were used. In Q5b, which required the evaluation of the average in the interview, he reached the conclusion that 3-day packages should be used since they were the most preferred ones by the customers. The answer of the student can be examined in Figure 4.35.


a) Sizce Metin Bey röportajda geçen **5 günlük ortalamayı** grafikteki verileri göz önüne alarak nasıl hesaplamıştır? Cevabınızı açık bir şekilde ifade ediniz.

5 günlük paketi 140 kişi tercih etmiştir ve buna göre ise en fazla 3 günlük paket kullanılmıştır.

Figure 4.35 Interpretation of average by a Level 3-4 student in Q5a

On the other hand, the overall analysis of the students' answers in the two sub-questions in Q5 and their definitions of average in Q1d revealed that 45 students (27.4%) could reach Level 5-6 in Q5. The number of students performing at Level 5-6 in Q5 was much more when compared to Q1. It was noticed that Level 5-6 students in Q5 realized that the average mentioned in the interview was not appropriate and expressed that 3-day packages should be used instead of 5-day packages. These students were placed into Level 5, not Level 6, since they had some deficiencies in either defining the concept of average in Q1d or interpreting the calculation of the average in Q5a. For instance, as previously stated, S30 in Figure 4.35 used the idea of the *mode* in both the interpretation and evaluation part of the fifth question. However, he could not present any definition for the concept of average in Q1d. As aforementioned, the frequencies of the definitions made by Level 5 students related to the concept of average almost the same as those in Q1. Therefore, the following table only displays the interpretations and evaluations of the students whose statistical literacy levels were determined as Level 5 in Q5.

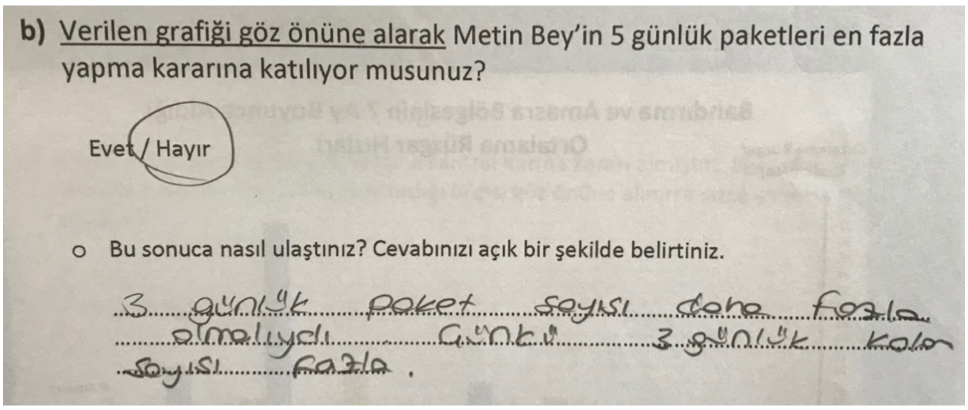
Table 4.6 The distribution of Level 5 students in terms of their definitions, interpretations and evaluations of the concept of average in Q5

	Interpretation regarding computation of the average	**Evaluation of the average**
Students at Level 5	Ideas related to measures of central tendency - Mean: 24 (14.6%) - Median: 2 (1.2%) -Mode: 1 (0.6%) Irrelevant Responses 14 (8.5%) No Answer 4 (2.4%)	Ideas related to measures of central tendency -Mean: 1 (0.6%) -Middle:1 (0.6%) -Mode: 43 (26.2%)
Total	45 (27.4%)	45 (27.4%)

Table 4.6 indicated that most students (16.4%) used ideas related to the measure of central tendency in interpreting the computation of the average in Q5a. Similar to Q1, the idea of the *mean* was preferred by most of the students (14.6%) in making their interpretations related to the calculation of the average in the given interview. There were only 3 students who used ideas of the *median* and *mode* during their interpretation of the calculation of the average in Q5a. While 14 students (8.5%) presented some irrelevant responses, 2.4% of the students could not provide any interpretation related to the computation of the average in Q5a. Besides, when the

evaluations of Level 5 students regarding the average number of days in the interview were examined, similar to Q1, it was noticed that almost all of them (26.2%) used ideas of the *mode* in evaluating the average in Q5b. There were 2 students (1.2%) who evaluated the given average in the interview by using the ideas of the *median* and *mean*. When Table 4.5 and Table 4.6 were compared, it was realized that Level 5 students used ideas similar to those of Level 4 students while interpreting the calculation of the average in Q5a. It seems that the only difference between Level 4 and Level 5 students were their evaluations of the claim related to the concept of average in Q5b. Therefore, the following sections do not present interpretations of students whose statistical literacy levels were determined as Level 5. Only the evaluations of Level 5 students regarding the claim related to the given average in the interview in Q5b are provided with supporting examples from students' answers.

As Table 4.6 indicated, 26.2% of the students used the idea of the *mode* in the evaluation of the concept of average in Q5b. These students did not use the word *mode* explicitly, but they stated that they did not agree with the decision of the director since 3-day packages were preferred by most of the customers. Figure 4.36 displays the answer of such a student. S43 said that the number of 3-day packages should be more since there are more people that stay 3 days in the hotel.


b) Verilen grafiği göz önüne alarak Metin Bey'in 5 günlük paketleri en fazla yapma kararına katılıyor musunuz?

Evet / Hayır

o Bu sonuca nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

3 günlük paket sayısı daha fazla olmalıydı çünkü 3 günlük kalan sayısı fazla.

Figure 4.36 Evaluation of average by a Level 5-6 student in Q5b

It was realized that 2 students (1.2%) in Q5b evaluated the given average in the interview by using the ideas of the median and mode. Nevertheless, both students could not correctly apply the measure they had chosen. For example, S81 in Figure 4.37 tried to calculate the mean of the given data in the interview. However, the student did not divide the total number of days by the total number of customers, but divided the total number of customers, 550, by the total number of days in the packages, 40. Since the average found, 13.7, is not 5, he did not agree with the decision given. Since it was believed that these students suspect the given average and tried to evaluate the concept by focusing on critical features, these students were placed into Level 5.

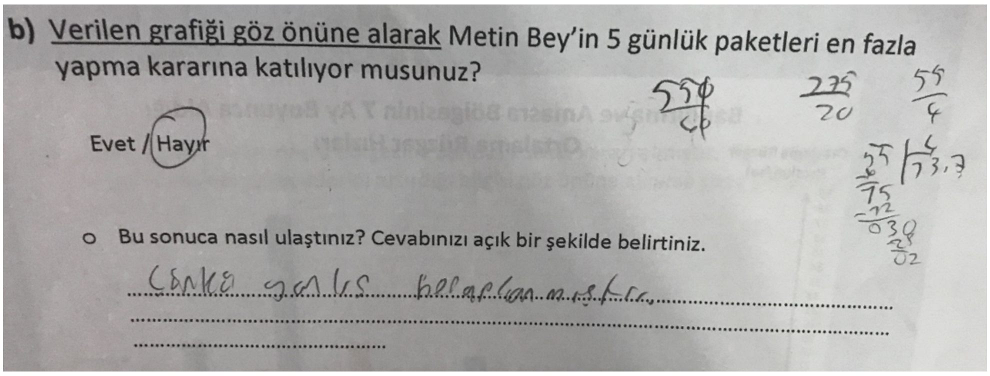


Figure 4.37 Evaluation of average by a Level 5-6 student in Q5b

4.1.3 The Concept of Average on Bar Graphs in Question 6

The last question, which was related to the concept of average on bar graphs was the first part of question 6 (Q6a). In this question, students were given a double bar graph which displayed the speed of the wind in the regions of Amasra and Bandırma throughout the seven months. The question requested the students to find the average wind speed of both regions. By not asking the students the computation of the mean directly, but the average, students' interpretation of the concept of average was investigated.

The statistical literacy levels of the students were determined by the overall analysis of the students' answers in Q6a and their definitions of average in Q1d. Table 4.1 indicated that most of the students (65.2%) performed at Level 3-4 in Q6a. While 34.8% of the students performed at Level 1-2, there were no students performing at Level 5-6. This was an expected result for Q6a since it was a question related only to interpretation of the concept of average; hence, students were expected to be at either Level 1-2 or Level 3-4. Since the frequencies of students' definitions of average at all levels were almost the same as those in Q1, the following sections explain the interpretations of students related to the concept of average in Q6a. The findings for Level 1-2 and Level 3-4 students are presented with examples from students' answers.

Firstly, 57 students (34.8%) performed at Level 1-2 in Q6a. Students at these levels could not interpret the concept of average as the students at Level 1-2 in Q1 and Q5. At Level 1, most of the students could not present any answer. The remaining students either read all of the data in the given graph, as S29 in Figure 4.38, or they presented any value in the graph for both of the regions.

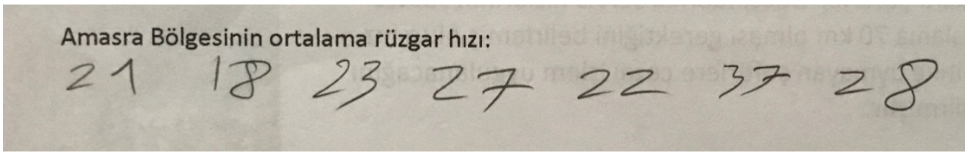


Figure 4.38 Interpretation of average by a Level 1-2 student in Q6a

On the other hand, different from Q1 and Q5, the sum of the given data was frequently preferred (11.6%) by Level 2 students to interpret the concept of average in Q6a. Figure 4.39 displays the answer of such a student. S53 added all of the given data for both regions to find the average wind speed. Furthermore, she made a calculation error in finding the sum for the Bandırma region.

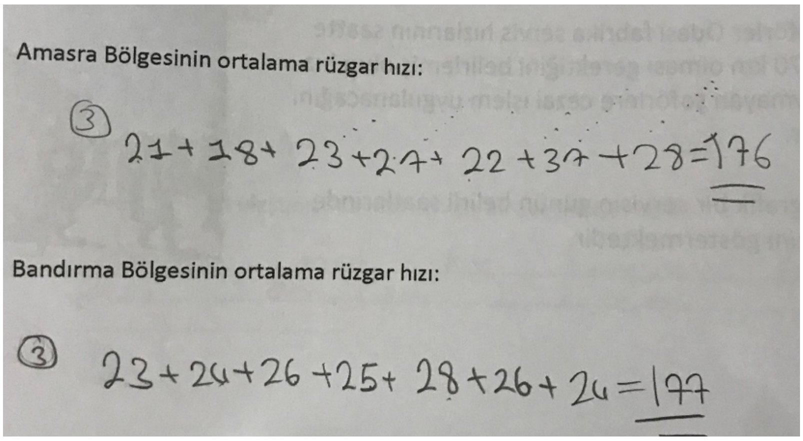


Figure 4.39 Interpretation of average by a Level 1-2 student in Q6a

On the other hand, the analysis of the students' interpretations of the concept of average in Q6a and their definitions of the average concept in Q1d as a whole indicated that more than half of the students (65.2%) performed at Level 3-4 in Q6a. Since the characteristics observed at Level 3 and Level 4 were specified clearly in the framework, the levels of the students could be determined as Level 3 and Level 4 separately in Q6a. More specifically, while 28.8% of the students performed at Level 3 in Q6a, 37.2 % of the students performed at Level 4. The percentages were higher when compared to those in Q5. As previously mentioned, the frequencies of the definitions made by Level 3 and Level 4 students related to the concept of average was almost the same in Q1. To state it differently, single ideas were used frequently at Level 3 in defining the concept but some ideas related to measures of central tendency began to emerge. However, at Level 4, students generally defined the concept using the ideas of measures of central tendency. As there is no question in Q6a related to the critical questioning of the average concept, the following table only displays the interpretations of students whose statistical literacy levels were determined as Level 3 and Level 4 regarding the calculation of the average in Q6a.

Table 4.7 The distribution of Level 3 and Level 4 students in terms of their interpretations in Q6a

	Students at Level 3	Students at Level 4
Interpretation regarding the computation of the average	Ideas related to measures of central tendency - Mean: 30 (18.3%) - Median: 2 (2.4%) -Mode: 1 (0.6%)	Ideas related to measures of central tendency - Mean: 61 (37.2%) - Median: 0 (0.0%) -Mode: 0 (0.0%)
	Ideas related to the sum of the given data 5 (3.0%)	Ideas related to the sum of the given data 0 (0.0%)
	Ideas related to the place of the average 2 (1.2%)	Ideas related to the place of the average 0 (0.0%)
	Irrelevant Responses 3 (1.8%)	Irrelevant Responses 0 (0.0%)
	No Answer 3 (1.8%)	No Answer 0 (0.0%)
Total	48 (28.8%)	61 (37.2%)

Table 4.7 indicated that 28.8% of the students performed at Level 3 in Q6a. Most of these students (21.3%) interpreted the computation of the average by using ideas related to the measure of central tendency in Q6a. The idea of the *mean* was the most preferred (18.3%), just as in Q1 and Q5. Different from Q1 and Q5, there were 5 students (3.0%) who found the sum of the given data as the average. While

3 students (1.8%) provided an irrelevant response in interpreting the computation of the average, 1.8% of the students could not present any interpretation in Q6a. These students were placed at Level 3 in Q6a because they used some ideas related to the measure of central tendency in defining the concept of average in Q1d. Lastly, 2 students (1.2%) presented some ideas related to the place of the average to interpret the calculation of the average in Q6a. On the other hand, 37.2 % of the students' statistical literacy levels were determined as Level 4 in Q6a. Different from students whose statistical literacy levels were determined as Level 3, Level 4 students used ideas related to measures of central tendency in both interpreting the concept of average in Q6a and defining the average in Q1d. It was observed that all of the Level 4 students (37.2%) in Q6a interpreted the calculation of the average by using ideas related to the measure of central tendency, and all of them used the idea of the *mean* in their interpretations.

Some differences are observed when Table 4.7 is compared with Tables 4.2, 4.3 and 4.5. As in Q1 and Q5, the *mean* was the most preferred measure of central tendency at Level 3 and Level 4; however, the usage of the *mean* was observed frequently especially by Level 3 students in Q6a. For example, the student whose statistical literacy level was determined as Level 3 in Q6a in Figure 4.40 interpreted the average as the mean of the given data and could correctly apply the algorithm.

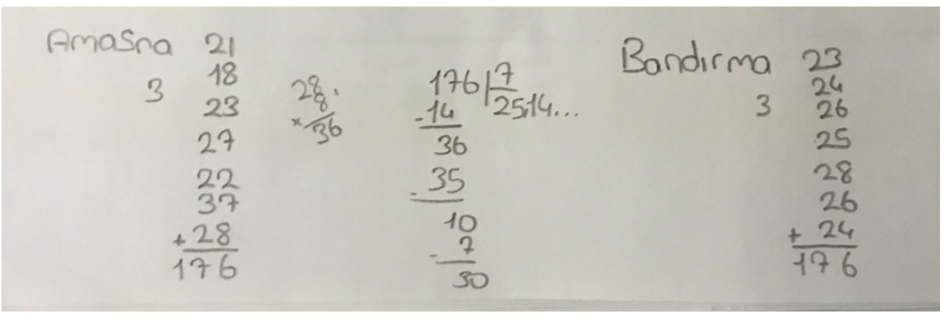


Figure 4.40 Interpretation of average by a Level 3-4 student in Q6a

However, it was observed that calculation errors were very common in finding the *mean*. For instance, while S103 in Figure 4.41 found the mean for the Bandırma region to be approximately 29 km/h, he found the mean for the Amasra region to be approximately 2011 km/h, and it seems that this student did not care much about the results he found.

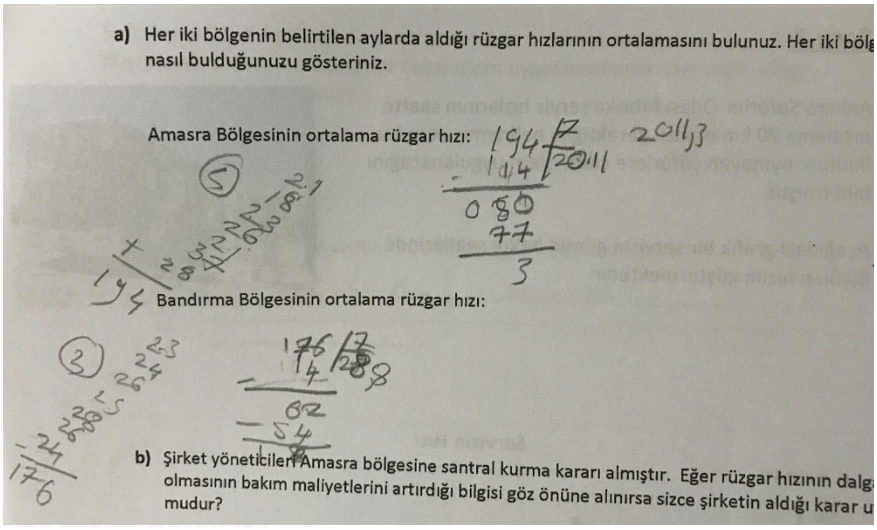


Figure 4.41 Interpretation of average by a Level 3-4 student in Q6a

In addition to interpreting the average as the *mean* and *sum* of the given data, 1.2% of the students whose statistical literacy levels were determined as Level 3 in Q6a mentioned some ideas related to the *place of the avera*ge. For example, S15 in Figure 4.42 stated that the average for the Amasra region would be between 18 and 37, and the average for the Bandırma region would be between 23 and 28. The values presented by the student were the maximum and minimum values for both of the regions. Since it was believed that these students could at least consider that the average could not be less than the minimum value and more than the maximum value, they were placed into Level 3.

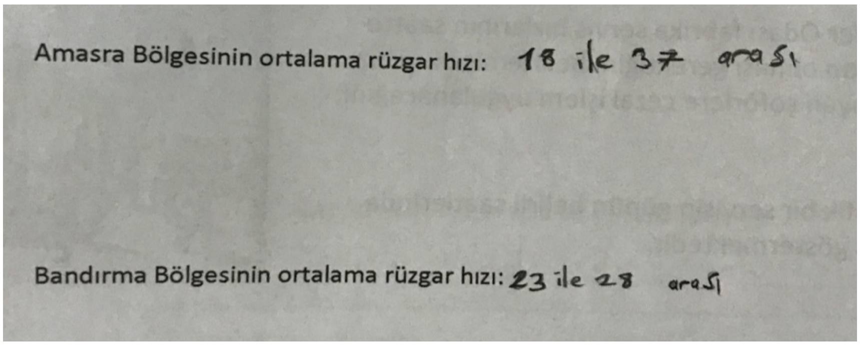


Figure 4.42 Interpretation of average by a Level 3-4 student in Q6a

On the other hand, the overall analysis of the students' answers in Q6a and their average definitions in Q1d revealed that 61 students (37.2%) could reach Level 4 in Q6a. The number of students performing at Level 4 in Q6a was more than the number of students performing at Level 4 in Q5, but less in Q1. As Table 4.7 indicated, different from the students whose statistical literacy levels were determined as Level 4 in Q1 and Q5, no ideas other than the *mean* was encountered by Level 4 students in Q6a. To state it differently, when there was a need to calculate the average, the only idea that Level 4 students came up with was the *mean* algorithm.

Table 4.8 in below summarizes the findings of the present study related to the concept of average on bar graphs. The findings of the present study revealed that in general, the statistical literacy levels of the students in the questions related to the concept of average on bar graphs were determined as Level 3-4. There were not many students reaching Level 5-6 in all of the questions. To put it differently, while students could interpret the average concept on bar graphs, most of them could not evaluate the presented average in the questions critically. Lastly, the *mean* was the most preferred measure of central tendency in interpreting the concept of average on bar graphs.

Table 4.8 The distribution of students across the six levels in the framework in the questions related to concept of average on bar graphs

The Concept of Average on Bar Graphs			**Statistical Literacy Level**		
Question	**Skill**	**Focus**	**Level 1-2**	**Level 3-4**	**Level 5-6**
Q1	Definition Interpretation Evaluation	Median	44 (26.8%)	**110** **(62.1%)**	10 (6.1%)
Q5	Interpretation Evaluation	Mode	51 (31.1%)	**68** **(41.5%)**	45 (27.4%)
Q6	Interpretation	Calculation	57 (34.8%)	**107** **(65.2%)**	0 (0.0%)

In this section, findings related to the average concepts on bar graphs were presented. The other aim of this study was first of all to determine statistical literacy levels of seventh grade students regarding the concept average on line graphs and then to examine how students at different statistical literacy levels define, interpret and evaluate the average concept on line graphs related to the data obtained from social or scientific contexts. Hence, in the next section, the findings of the present study related to the concept of average on line graphs are presented.

4.2 The Concept of Average on Line Graphs

To examine students' statistical literacy about the concept of average on line graphs, two questions were prepared similar to the ones in Section 4.1. Those questions were the first part of question 2 (Q2a) and question 4 (Q4). In Q2a, students were requested to calculate the average number of baskets of two players whose numbers of baskets in the last ten years were given on a line graph. By not directly asking for the mean, but for the average of the number of baskets, the interpretation of the concept of average was investigated. On the other hand, Q4 was related to the success of a course about which information in the last 5 years was provided through a brochure. In this question, an outlier, the number of students in 2015, was provided, but the average in the brochure was calculated by means of the mean, which is not an appropriate measure of average when there is an outlier. It was examined whether students critically questioned the average number of students for the course. Furthermore, asking students the meaning of the average in the brochure and how the average in the brochure was calculated, students' interpretation of the concept of average was investigated.

To determine students' levels for each question, the modified version of the framework of Watson and Callingham (2003) was used. The students' statistical literacy levels were determined by the overall analysis of each question. In other words, the students' statistical literacy levels were determined by taking into account their definitions and interpretations regarding the concept of average and their evaluations regarding the claims related to the concept of average in each question. The distribution of students across six levels for Q2a, and Q4 are displayed in Table 4.9.

Table 4.9 The distribution of students across the six levels in the framework for Q2a and Q4

	Questions related to the Average on Line Graphs	
Levels	**Question 2a**	**Question 4**
1-2	48 (29.3%)	54 (32.9%)
3-4	116 (70.7%)	110 (67.1%)
5-6	0 (0.0%)	0 (0.0%)
Total	164 (100.0%)	164 (100.0%)

Table 4.9 revealed that when students' definitions, interpretations and evaluations of the concept of average were analyzed as a whole in Q2a, most of the students (70.7%) performed at Level 3-4. On the other hand, while 29.3% of the students were at Level 1-2 in Q2a, there were no students performing at Level 5-6. This was an expected result for Q2a since it was a question related only to the interpretation of the concept of average; hence, students were expected to be at either Level 1-2 or Level 3-4. The findings of the study were not different for Q4. To state it differently, most of the students (67.1%) performed at Level 3-4. The number of students observed at Level 1-2 was 32.9%. Lastly, Table 4.8 indicates that any student could reach Level 5-6 in Q4 although it consisted of an evaluation of the concept of average.

In the following sections, to respond to the sub-question of the present study, the answers of students at different statistical literacy levels will be explained in detail for Q2a and Q4 by providing examples from students' answers. To put it differently, how seventh grade students at different statistical literacy levels define, interpret and evaluate the concept of average on line graphs is explained for Q2a and Q4.

4.2.1 The Concept of Average on Line Graphs in Question 2a

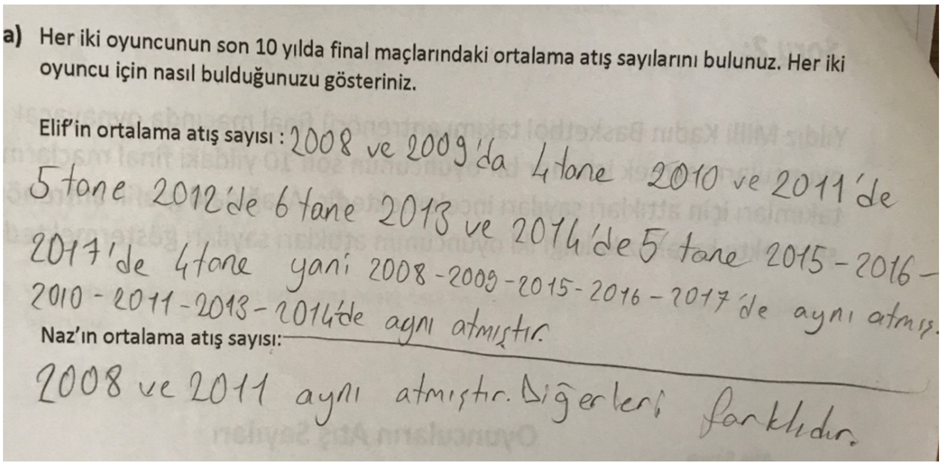
The first question, which was related to the concept of average on line graphs was the first part of question 2 (Q2a). This question was similar to Q6a. In this question, students were provided with a double line graph, which displayed the number of baskets of two basketball players in the final game of the last ten years. The question requested students to find the average number of baskets of the two players in the last ten years. By not directly asking students for the computation of mean, but the average, students' interpretation of the concept of average was investigated.

Students' statistical literacy levels were determined by an overall analysis of students' answers in Q2a and their definition of average in Q1d, the only definition question related to the average concept in SLT. As revealed in Table 4.9, most of the students (70.7%) performed at Level 3-4 in Q2a. The percentage was higher than the that of students performing at Level 3-4 in Q6a. While 29.3% of the students were performed at Level 1-2, there were no students performing at Level 5-6. This was an expected result for Q2a since it was a question related only to the interpretation of the concept of average; hence, students were expected to be at either Level 1-2 or Level 3-4. Therefore, the following sections explain the definitions made by students in Q1d and their interpretations in Q6a. The findings are presented for Level 1-2 and Level 3-4 students by providing examples from students' answers.

In the modified version of the framework of Watson and Callingham (2003), it was stated that any idea related to the concept of average are observed by Level 1 students in their definitions and interpretations of the concept of average. On the other hand, students at Level 2 express some single ideas about the concept of average, such as *normal* or *the same as others* in defining and interpreting the concept of average. Critical evaluation of the claims related to the concept of average in a social or scientific context is not an expected skill for Level 1-2 students since they cannot define or interpret the concept at this level. Since Q2a did not require students to evaluate the average concept, the definitions and interpretations of Level 1-2 students in Q2a are presented in further sections.

The findings of the present study showed that 12.8% of the students could not present any idea in defining the concept of average in Q1d or interpreting it in Q2a; therefore, these students were placed at Level 1. Even though these students tried to make some interpretations to find the average number of baskets of the two players, they either read all of the data given to them, such as the student in Figure 4.43, or presented any value from the given data as the average of the players. S63 stated that average number of baskets of Elif, one of the players, was 4 in 2008 and

2009, 5 in 2012, 6 in 2013, 5 in 2014 and 4 in 2015, 2016 and 2017; that is, the score of Elif was the same in the years 2008, 2009, 2015, 2016 and 2017 and was the same in the years 2010, 2011, 2013 and 2014. For the average number of baskets for Naz, the other player, S63 stated that her score was the same in the years 2008 and 2011; and her other scores were different. Such a finding was also obtained in Q6a.


a) Her iki oyuncunun son 10 yılda final maçlarındaki ortalama atış sayılarını bulunuz. Her iki oyuncu için nasıl bulduğunuzu gösteriniz.

Elif'in ortalama atış sayısı : 2008 ve 2009'da 4 tane 2010 ve 2011'de 5 tane 2012'de 6 tane 2013 ve 2014'de 5 tane 2015-2016-2017'de 4 tane yani 2008-2009-2015-2016-2017'de aynı atmış. 2010-2011-2013-2014'de aynı atmıştır.

Naz'ın ortalama atış sayısı: 2008 ve 2011 aynı atmıştır. Diğerleri farklıdır.

Figure 4.43 Interpretation of average by a Level 1-2 student in Q2a

However, some single ideas started to be used in defining the concept of average by 16.5% of the students; hence, these students were placed at Level 2. Most of the Level 2 students preferred the *sum* of the given data in their interpretation of the concept of average in Q2a, just as Level 2 students in Q6a. To state it differently, these students added all the given values in the given graph to find the average of the baskets of the two players. Figure 4.44 presents a sample answer. S56 in the following figure added the 6 and 9 values from the given graph instead of adding the 10 values given in the graph to find the average of the baskets of the two players, which implies that students have problems in reading data.

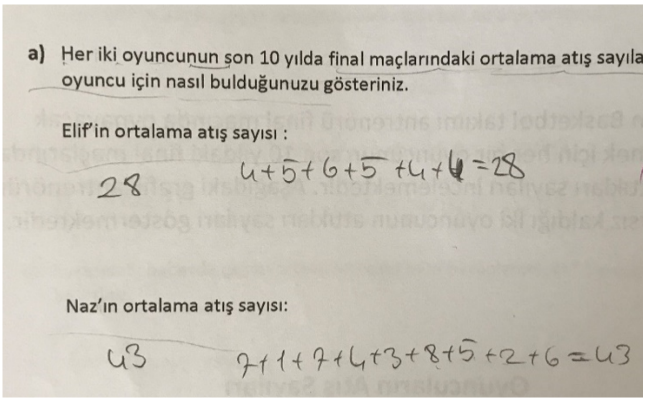


Figure 4.44 Interpretation of average by a Level 1-2 student in Q2a

On the other hand, the analysis of students' interpretations of the average concept in Q2a and their definitions of the average concept in Q1d as a whole indicated that more than half of the students (70.7%) performed at Level 3-4 in Q2a. According to the modified version of the framework, students still use single ideas at Level 3 in defining or interpreting the concept of average. However, students begin to use ideas related to measures of central tendency in either defining or interpreting the concept of average, especially in questions requiring the interpretation of the computation of average. Besides, at Level 4, students generally use some ideas related to measures of central tendency in defining and interpreting the concept of average. Some ideas related to the representative nature of the average concept also begin at Level 4 while defining and interpreting it. It was stated in the framework that although students at Level 4 could compute the mean of a small data set, they could not compute its weighted mean. Critical questioning of the concept of average in social or scientific contexts is still not observed at Level 3-4. As previously stated, the findings of the present study showed that 70.7% of the students performed at Level 3-4 in Q2a. Tables 4.9 and 4.10 below present the definitions and interpretations of students whose levels were determined as 3-4 in Q2a. Since the characteristics observed at Level 3 and Level 4 are specified clearly in the framework, the levels of students could be determined as Level 3 and Level 4 seperately. More specifically, while 33.5% of the students performed at Level 3 in Q2a, 37.2% of the students performed at Level 4. The percentages were higher when compared to the number of students who performed at Level 3 and Level 4 in Q6a. The definitions and interpretations of the students who were determined to be at Level 3 and Level 4 are presented in different tables below. Moreover, since critical questioning of the concept of average was not an expected skill in Q2a, the tables do not include an evaluation part.

Table 4.10 The distribution of Level 3 students in terms of their definition and interpretation of the concept of average in Q1d and Q2a, respectively

	Definition of average	**Interpretation regarding the computation of average**
Students at Level 3	Ideas related to measures of central tendency -Mean: 8 (4.9%) -Median: 4 (2.4%) -Mode: 0 (0.0%)	Ideas related to measures of central tendency - Mean: 38 (22.8%) -Median: 4 (2.4%) -Mode: 4 (2.4%)
	Ideas related to the sum of the given data 0 (0.0%)	Ideas related to the sum of the given data 0 (0.0%)
	Single Ideas 15 (9.1%)	Single Ideas 0 (0.0%)
	Irrelevant Responses 3 (1.8%)	Irrelevant Responses 0 (0.0%)
	No Answer 25 (15.2%)	No Answer 3 (1.8%)
Total	55 (33.5%)	55 (33.5%)

Table 4.10 indicated that 55 students (33.5%) performed at Level 3 in Q2a. When the definitions of Level 3 students were examined, it was observed that 15.2% of the students could not present any definition related to the concept of average in Q1d. While 9.1% of the students used some single ideas in defining the concept of average, 7.3% of the students used ideas related to measures of central tendency in defining the average concept. There were also 3 students (1.8%) who provided some irrelevant responses while defining the average in Q1d. As can be remembered, Q2a, related to the interpretation of the concept of average, asked for the average number of baskets of the two basketball players. Table 4.10 shows that most of the the students who performed at Level 3 (25.2%) interpreted the computation of the average in Q2a by using ideas related to the measure of central tendency. The most preferred idea observed (22.8%) was the *mean*, the percentage of which was higher when compared to the other questions. As in Q6a, there were 10 students (6.1%) who found the *sum* of the given data as the average. Three students (1.8%) provided an irrelevant response in interpreting the computation of the concept of average, but these students were placed at Level 3 in Q2a because they used some ideas related to the measure of central tendency in defining the concept of average in Q1d.

It was realized that in Q2a, as previously mentioned, although 15.2% of Level 3 students could not define the concept of average in Q1d and 9.1% of them presented some single ideas in defining it, most of them (22.8%) could compute the average number of baskets of the two players with ideas related to measures of central tendency, particularly by means of the *mean*. For example, S90 in Figure 4.45 could not define the average in Q1d, but she could find the mean number of baskets correctly for both players.

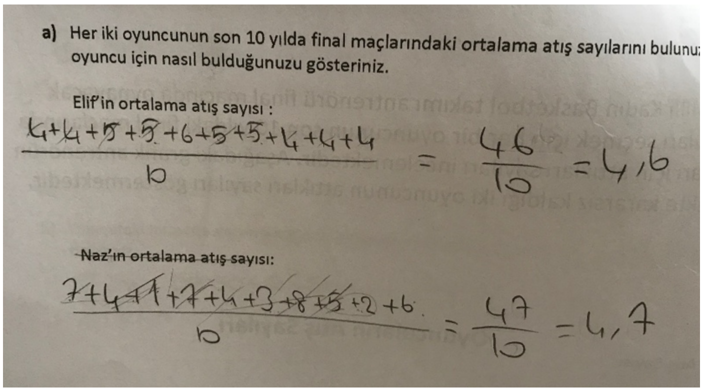


Figure 4.45 Interpretation of average by a Level 3-4 student in Q2a

Interestingly, it was noticed that even though students whose statistical literacy levels were determined as Level 3 in Q2a defined the concept of average by using ideas related to measures of central tendency in Q1d, they could use the *sum* of the given data to find the average number of baskets of the two players as in Q6a. For instance, S118 in Figure 4.46 defines the average in Q1d as the *mean* of the given data set; however, as can be examined in the figure, he found the *sum* of the number of baskets of the two players to find the average number of baskets. This student also seemed to experience difficulties in reading data as S56 in Figure 4.44.

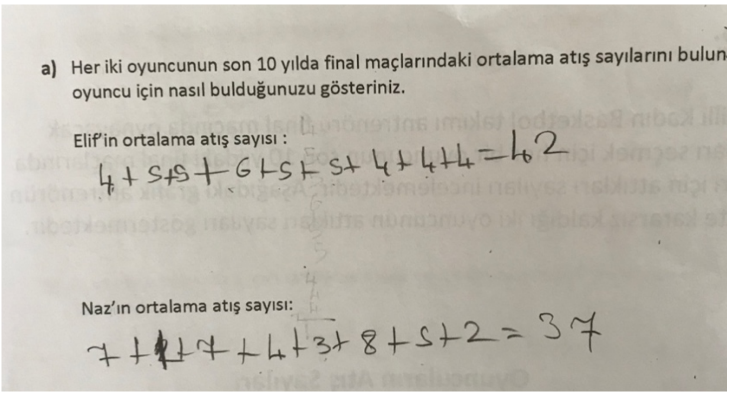


Figure 4.46 Interpretation of average by a Level 3-4 student in Q2a

On the other hand, the overall analysis of students' answers in Q2a and their definitions of average in Q1d revealed that 61 students (37.2%) could reach Level

4 in Q2a. The number of students performing at Level 4 in Q2a was the same as the number of students performing at Level 4 in Q6a. Different from the students whose statistical literacy levels were determined as Level 3, students whose levels were determined as Level 4 used ideas related to measures of central tendency in both defining and interpreting the concept of average in Q1d and Q2a, respectively. Table 4.11 below presents Level 4 students' definitions and interpretations of the concept of average in Q1d and Q2a, respectively.

Table 4.11 The distribution of Level 4 students in terms of their definitions and interpretations of the concept of average in Q1d and Q2a, respectively

	Definitions of average	**Interpretations regarding the computation of average**
Students at Level 4	Ideas related to representativeness 4 (2.4%) Ideas related to measures of central tendency -Mean: 44 (26.8%) -Median: 6 (3.6%) -Mode: 3 (1.8%) Multiple definitions 0 (0.0%)	Ideas related to representativeness 0 (0.0%) Ideas related to measures of central tendency -Mean: 59 (36.0%) -Median: 2 (1.2%) -Mode: 0 (0.0%) Multiple interpretations 0 (0.0%)
Total	61 (37.2%)	61 (37.2%)

Table 4.11 revealed that 53 students out of 61 (32.2%) used ideas related to measures of central tendency in defining the concept of average in Q1d. The *mean* was the most preferred measure in defining the concept of average. While 4 students (2.4%) presented some ideas related to the representative nature of the

average in defining the average in Q1d, 2.4% of the students used more than one idea in defining the concept of average. Similarly, ideas related to three measures of central tendency were the most frequent answers among Level 4 students in interpreting the computation of the average in Q2a. Similar to the definition question, the *mean* was the most preferred idea in interpreting the average concept with 36.0% in Q2a. Only 2 students (1.2%) used the idea of the *median* in finding the average number of baskets of the two players. Since the ideas that emerged in the responses of Level 4 students while defining and interpreting the concept of average in Q1d and Q2a, respectively, were previously supported with many examples (see Figure 4.14, Figure 4.17 and Figure 4.45), the following sections do not present Level 4 students' definitions and interpretations for Q2a once more. However, it was realized that, as indicated in Table 4.10, almost no ideas other than the *mean* was encountered among the responses provided by Level 4 students in Q2a, just as in Q6a. To state it differently, when there is a need to calculate the average, the only idea that Level 4 students put forward was the *mean* algorithm.

4.2.2 The Concept of Average on Line Graphs in Question 4

In Question 4 (Q4), the students were provided with a brochure regarding the average number of students in a course. In the question, the number of students who got the maximum point in a foreign language exam in the last 5 years was presented on a line graph, and the average was calculated as 42 by using the mean algorithm. However, the number of students receiving the maximum point in the foreign language exam in 2015 was much higher than it was in the other years. To state it differently, there was an outlier in the given data and students were expected to realize this outlier. Q4 included three sub-questions. In the first sub-question, Q4a, students were asked what they understood from the given average in the brochure. In this way, their interpretation of the concept of average in the given context was examined. In the second sub-question, Q4b, students' interpretation of the concept of average was investigated once more with a different question. Q4b asked students how the average in the brochure was calculated. On the other hand, the third sub-question, Q4c, was about students' evaluations of the claim in the article, the average number of students. The students were asked the question of whether or not they would give up their current course in which the average number of students getting the maximum point in a year was 30 by taking into account the graph and the average number of students in the brochure.

The students' statistical literacy levels were determined by an overall analysis of the students' answers in three sub-questions in Q4 and their definitions of average in Q1d. As indicated in Table 4.9, most of the students (67.1%) performed at Level 3-4 in Q4. 32.9% of the students were at Level 1-2, but there were no students performing at Level 5-6 although Q4c required students to evaluate the average in the brochure. Since the frequencies of the definitions of average made by the students at Level 1-2 and Level 3-4 were almost the same as those reported in the previous section, the following sections explain the interpretations of students related to the concept of average in Q4a and Q4b and their evaluations of the claim related to the given average in the brochure in Q4c. The findings are presented for Level 1-2 and Level 3-4 by providing examples from students' answers. Furthermore, this question was very similar to Q1 in Section 4.1.1. The only difference was that the data were displayed on a line graph in Q4; therefore, the findings are presented in line with the findings of Q1.

The findings of the present study showed that 32.9% of the student performed at Level 1-2 in Q4. The percentage was higher when compared to that of students at Level 1-2 in Q1. When the students' answers to the interpretation questions in Q4 were examined, it was observed that some of them either did not interpret the average concept in the brochure or presented some irrelevant responses. For example, S60 in Figure 4.47 answered the interpretation question related to the computation of average in Q4b as the maximum number of students. S60 interpreted the meaning of average in Q4a as the maximum number in the given data; however, she did not mention any idea regarding the calculation of the average in Q4b.

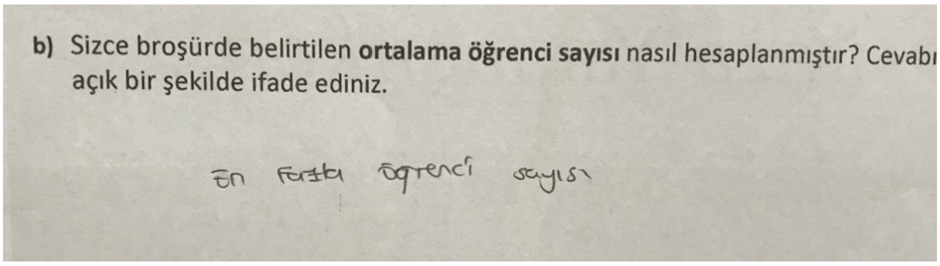
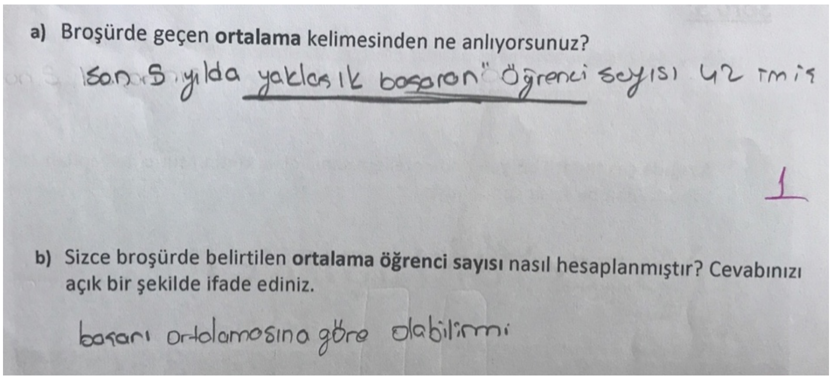


Figure 4.47 Interpretation of average by a Level 1-2 student in Q4b

It was observed that some of the students interpreted the meaning of the concept of average in Q4a by using the same single ideas that they have used in their interpretation of the concept in Q1a; thus, these students were placed at Level 2 in Q4. As in Q1, the word *about* was the most preferred (5.0%) single idea. However, it was noticed that the students could not show the same performance in the interpretation question regarding the computation of average in Q4b. It was found that most of the students at Level 2 could answer the question but they presented some irrelevant responses as the student in Figure 4.48. S58 in Figure 4.48 interpreted the meaning of average in Q4a by explaining that the number of successful students in the last 5 years was *about* 42; therefore, this student was placed at Level 2. However, as can be seen in Figure 4.48, she could not interpret the computation of the average in the given brochure in Q4b; instead, she responded to the question with another question: “Can it be according to the average of success?”

The same finding as the above one was also obtained in Q1. Hence, it can be concluded that even though students could define or interpret the meaning of the concept of average in a given context, they seemed to possess some difficulties in interpreting the computation of the average concept.



a) Broşürde geçen **ortalama** kelimesinden ne anlıyorsunuz?

Son 5 yılda yaklaşık başarılı öğrenci sayısı 42 imiş

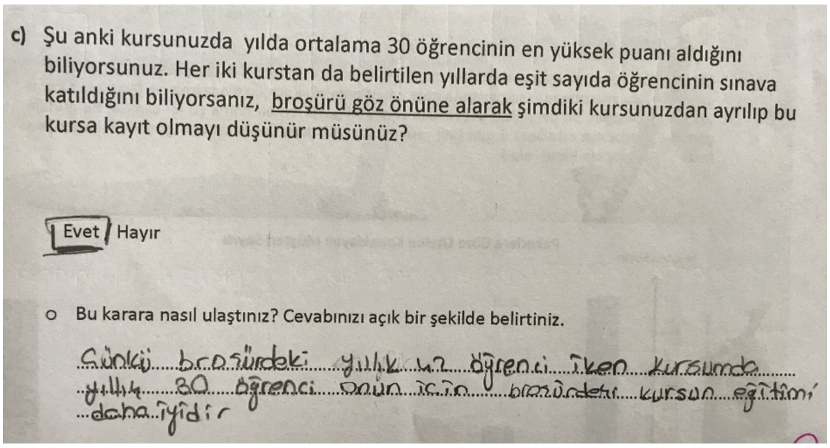
1

b) Sizce broşürde belirtilen **ortalama öğrenci sayısı** nasıl hesaplanmıştır? Cevabınızı açık bir şekilde ifade ediniz.

başarı ortalamasına göre olabilir mi

Figure 4.48 Interpretations of average by a Level 1-2 student in Q4a and Q4b

The findings of the present study indicated that none of the students whose statistical literacy levels were determined as Level 1-2 could evaluate the given average in the brochure critically in Q4c. As can be remembered, Q4c required students to evaluate the given average with the question asking whether or not students would give up their current course where the average number of students receiving the maximum point in a year was 30. It was noticed that while 6.7% of the Level 1-2 students directly accepted the average number of students in the brochure, some of them presented idiosyncratic responses to justify their decisions in Q4c. For example, S58 in Figure 4.49 decided to give up her current course with the justification that the education of the course mentioned in the brochure was better as there were 42 students in a year in course mentioned in the brochure, while her course had 30 students in a year. As can easily be understood from the answer of S58, these students did not show any signs of doubt about the given average number in the brochure.


c) Şu anki kursunuzda yılda ortalama 30 öğrencinin en yüksek puanı aldığını biliyorsunuz. Her iki kurstan da belirtilen yıllarda eşit sayıda öğrencinin sınava katıldığını biliyorsanız, broşürü göz önüne alarak şimdiki kursunuzdan ayrılıp bu kursa kayıt olmayı düşünür müsünüz?

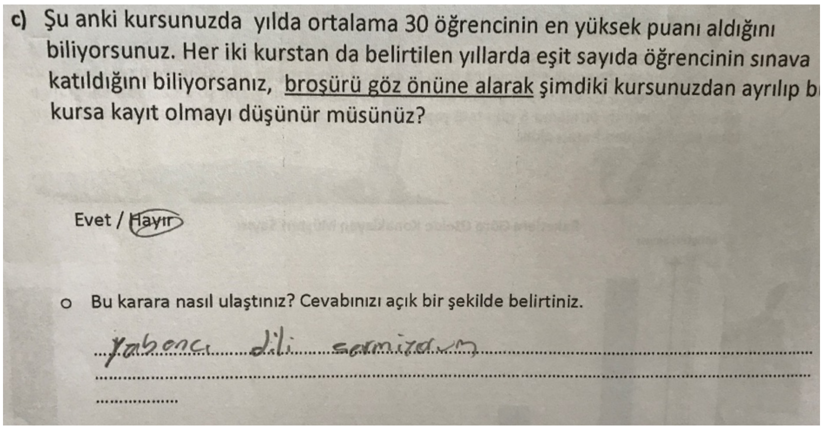
Evet / Hayır

- Bu karara nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

Çünkü broşürdeki yıllık 42 öğrenci iken kursumda yıllık 30 öğrenci onun için broşürdeki kursun eğitimi daha iyidir

Figure 4.49 Evaluation of average by a Level 1-2 student in Q4c

Figure 4.50 below shows the idiosyncratic response of S24 in Q4c. S24 did not consider giving up his current course since he stated that he did not like foreign language. As S24, 17.6% of the students tried to evaluate the news but did not focus on the critical features given in the brochure.


c) Şu anki kursunuzda yılda ortalama 30 öğrencinin en yüksek puanı aldığını biliyorsunuz. Her iki kurstan da belirtilen yıllarda eşit sayıda öğrencinin sınava katıldığını biliyorsanız, broşürü göz önüne alarak şimdiki kursunuzdan ayrılıp b
kursa kayıt olmayı düşünür müsünüz?

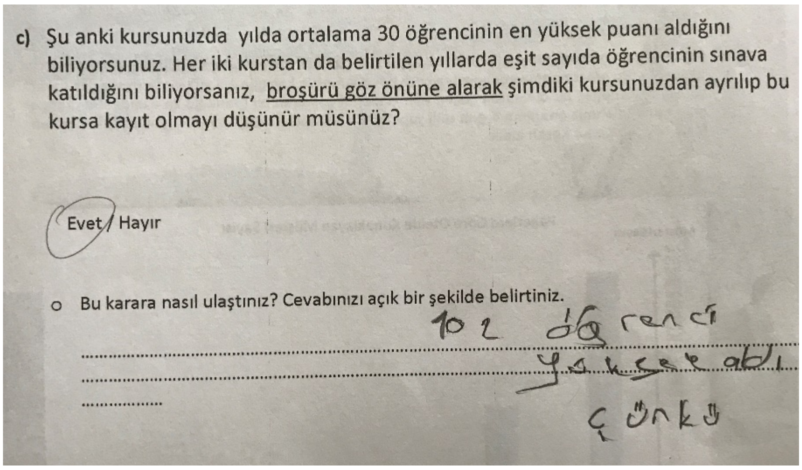
Evet / Hayır

- Bu karara nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

yabancı dili sevmiyorum

Figure 4.50 Evaluation of average by a Level 1-2 student in Q4c

Different from the students in Figure 4.49 and Figure 4.50, some students whose levels were determined as Level 1-2 in Q4 made their evaluations focusing on only single or some values in the given data. They did evaluate the given brochure in Q4c, but did not question the average in the brochure and did not take into account the effect of all the data in the reported average. For instance, S72 in Figure 4.51 gave his response as 102 students in the year 2015 by concentrating solely on the outlier in the given data. He decided to attend the course mentioned in the brochure because 102 students got maximum points in that course. It was concluded that these students could actually realize the outlier in the given data; however, they did not consider its effect on the average number of students.


c) Şu anki kursunuzda yılda ortalama 30 öğrencinin en yüksek puanı aldığını biliyorsunuz. Her iki kurstan da belirtilen yıllarda eşit sayıda öğrencinin sınava katıldığını biliyorsanız, broşürü göz önüne alarak şimdiki kursunuzdan ayrılıp bu kursa kayıt olmayı düşünür müsünüz?

Evet / Hayır

o Bu karara nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

102 öğrenci yüksek aldı çünkü

Figure 4.51 Evaluation of average by a Level 1-2 student in Q4c

On the other hand, the overall analysis of students' interpretations and evaluations of the concept of average in Q4 and their definitions of the concept of average in Q1d indicated that more than half of the students (67.1%) performed at Level 3-4 in Q4. Since the characteristics observed at Level 3 and Level 4 were clearly specified in the framework, the levels of students could be determined as Level 3 and Level 4 separately in Q4. More specifically, while 21.6% of the students performed at Level 3 in Q4, 45.1% of the students performed at Level 4. As aforementioned, the frequencies of the definitions made by Level 3 and Level 4 students of the concept of average were almost the same as those in Q2a. Moreover, as stated in the framework, critical questioning of the average concept was still not observed at Level 3 and Level 4. It was revealed in the current study that Level 3-4 students made their evaluations related to the concept of average in the brochure in Q4c just as the students at Level 1-2 in Q4. Thus, the following tables display only interpretations of students whose statistical literacy levels were determined as Level 3 and Level 4 in Q4 regarding the meaning of the average concept in Q4a and the computation of the average in the given brochure in Q4b.

Table 4.12 The distribution of Level 3 students in terms of their interpretations of the meaning of the average concept in Q4a and its computation in Q4b

	Interpretation regarding the meaning of average	**Interpretation regarding the computation of average**
Students at Level 3	Ideas related to measures of central tendency -Mean: 1 (0.6%) -Median: 0 (0.0%) -Mode: 1 (0.6%)	Ideas related to measures of central tendency - Mean: 22 (13.2%) -Median: 0 (0.0%) -Mode: 0 (0.0%)
	Single Ideas 9 (5.4%)	Single Ideas 0 (5.4%)
	Multiple interpretations 2 (1.2%)	Multiple interpretations 0 (1.2%)
	Irrelevant Responses 23 (13.8%)	Irrelevant Responses 9 (5.4%)
	No Answer 5 (3.0%)	No Answer 5 (3.0%)
Total	36 (21.6%)	36 (21.6%)

Table 4.12 indicated that 36 students (21.6%) performed at Level 3 in Q4. As can be remembered, there were two questions related to the interpretation of the

concept of average in Q4. While Q4a was related to the interpretation of the meaning of the given average in the brochure, Q4b was related to the interpretation of the computation of the average in the brochure. Table 4.12 shows that most students (13.8%) at Level 3 in Q4a interpreted the meaning of the average concept in the given brochure by providing some irrelevant responses. The percentage is higher when compared to Q1, which can be an indication of the role of context. While 5.4% of the students presented some single ideas in interpreting the meaning of the average concept in Q4a, 1.2% of the students interpreted the concept of average by using ideas related to three measures of central tendency. There were also 2 students who provided multiple interpretations regarding the meaning of the average concept in Q4a. On the other hand, most students who performed at Level 3 (13.2%) interpreted the computation of the average in the given brochure by using ideas related to the measure of central tendency, but the sole idea observed was the *mean*, just as in Q1. While 9 (5.4%) students presented some irrelevant responses in interpreting the computation of the average in the given news in Q4b, 3.0% of the students could not provide any answer at all.

It was realized that although most students at Level 3 continued to use single ideas or presented irrelevant responses in interpreting the meaning of the average concept in Q4a, they stated that the average in the brochure was calculated through the *mean* of the given data; therefore, these students were placed at Level 3. Furthermore, it was observed that some of them applied the algorithm correctly and reached the average in the brochure. For example, S132 in Figure 4.52 stated almost the same wording in the brochure to interpret the meaning of the average concept in Q4a: the average number of students in 2013, 2014, 2015, 2016 and 2017. However, S132 interpreted the computation of the average as the *mean* of the given data set and he could correctly apply the mean algorithm as can be seen in Figure 4.52.

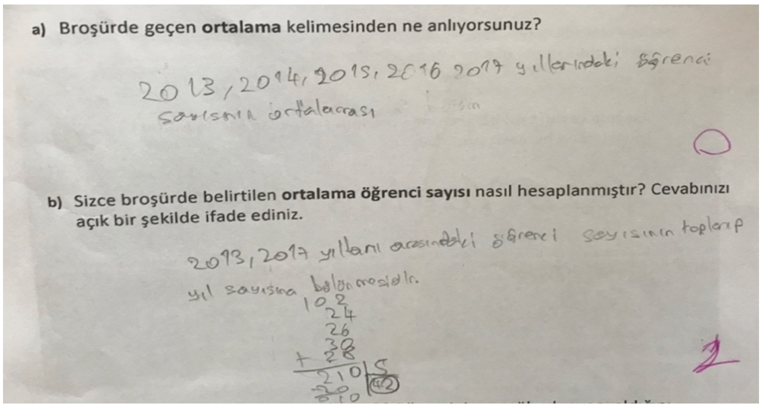


Figure 4.52 Interpretations of average by a Level 3-4 student in Q4a and Q4b respectively

Moreover, as observed in Q1, some students whose statistical literacy levels were determined as Level 3 interpreted the meaning of the average as the *sum* of the given data set. However, they did not find the sum of the given values to calculate the average in Q4b but they used the *mean* algorithm to reach the average number of students. Figure 4.53 shows a sample answer. Although S61 interpreted the meaning of the average concept in Q4a as the *sum* of the students in all the years, she could correctly reach the average number of students in Q4b by using the *mean* algorithm.

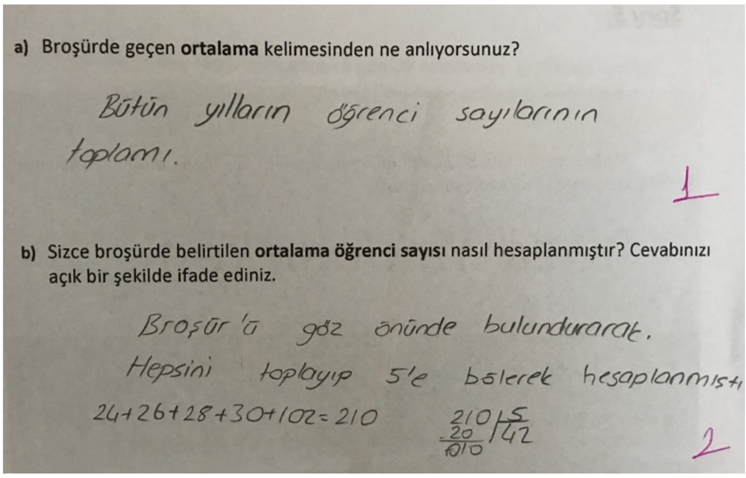


Figure 4.53 Interpretations of average by a Level 3-4 student in Q4a and Q4b

On the other hand, the overall analysis of the students' answers to the three sub-questions in Q4 and their definitions of average in Q1d revealed that 74 students (45.1%) performed at Level 4 in Q4. The percentage was almost the same as that of Level 4 students in Q1. Different from the students whose statistical literacy levels were determined as Level 3, students whose levels were determined as Level 4 used ideas related to measures of central tendency in at least two of the four parts examined for Q4. For instance, S5 in Figure 4.54 gave the signals of the *mean* in both interpretation questions in Q4. She interpreted the meaning of the concept of average in Q4a as the division of the sum of the years 2013, 2014, 2015, 2016 and 2017 by 5. In Q4b, she stated that the average in the given brochure was calculated by means of the *mean*; she divided the total number of students by the total number of years and obtained 42. However, S5 could not define the concept of average in Q1d; hence, the statistical literacy level of this student was determined as Level 4.

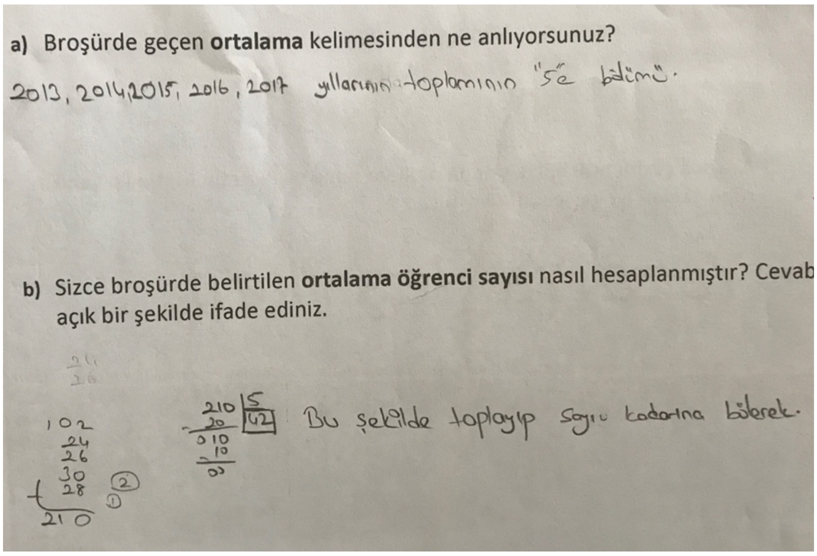


Figure 4.54 Interpretations of average by a Level 3-4 student in Q4a and Q4b

Table 4.13 below presents the interpretations of the concept of average of Level 4 students in Q4a and Q4b.

Table 4.13 The distributions of Level 4 students in Q4 in terms of their interpretations in Q4a and Q4b

	Interpretation regarding the meaning of average	**Interpretation regarding the computation of average**
Students at Level 4	Ideas related to representativeness 1 (0.6%)	Ideas related to representativeness 1 (0.6%)
	Ideas related to measures of central tendency -Mean: 37 (22.6%) -Median: 5 (3.0%) -Mode: 9 (5.4%)	Ideas related to measures of central tendency -Mean: 72 (43.9%) -Median: 0 (0.0%) -Mode: 0 (0.0%)
	Multiple interpretations 0 (0.0%)	Multiple interpretations 0 (0.0%)
	Single Ideas 4 (2.4%)	Single Ideas 4 (2.4%)
	Irrelevant Responses 16 (9.8%)	Irrelevant Responses 2 (1.2%)
	No Answer 1 (0.6%)	No Answer 1 (0.6%)
Total	74 (45.1%)	74 (45.1%)

Table 4.13 revealed that ideas related to the three measures of central tendency were the most frequent answers among Level 4 students in interpreting the meaning of the average in the given brochure in Q4a. Similar to Q1, the *mean* was the most preferred idea (22.6%) in interpreting the meaning of the concept of

average. It was realized that 9.8% of the students provided some irrelevant responses to exhibit their interpretation in Q4a. There were no Level 4 students in Q1 interpreting the meaning of the average concept through irrelevant responses. This reveals once more that students experienced more difficulties in the context of Q4. Moreover, 4 students (2.4%) presented some single ideas in interpreting the meaning of the average concept in Q4a. There was one student who presented some ideas related to the representative nature of the average concept, provided multiple interpretations, and could not interpret the meaning of the average concept in the given brochure. Lastly, when the students' answers were analyzed in the interpretation question related to the computation of the average concept in the given brochure, it was noticed that students once again chose ideas related to three measures of central tendency. However, almost no other measure than the *mean* was encountered in the students' responses of interpreting the calculation of the average concept in Q4b. 43.9% of the students used the idea of the *mean* when interpreting the calculation of the average concept. The remaining 2 students (1.2%) presented some irrelevant responses in interpreting the computation of the average in Q4b. Since the ideas that emerged among the responses of Level 4 students in interpreting the concept of average in Q4a and Q4b were supported with many examples previously, (see Figure 4.54, Figure 4.20 and Figure 4.23), the following sections do not present interpretations of the students whose statistical literacy levels were determined as Level 4 in Q4 once more. Nevertheless, as observed in Q1, it was noticed that although some students mentioned other measures of central tendency in the interpretation question asking for the meaning of the average in the brochure in Q4a, they did not use this understanding in the interpretation question regarding the computation of the average. In the question requiring the interpretation of the computation of the given average, they stated that the average was calculated through the *mean*. For example, S112 in Figure 4.55 stated that for him average means the middle of all, but he did not find the average by searching for the middle value. Instead, he stated that to find the average, he needed to add all the numbers and divide the sum by the number of years there were.

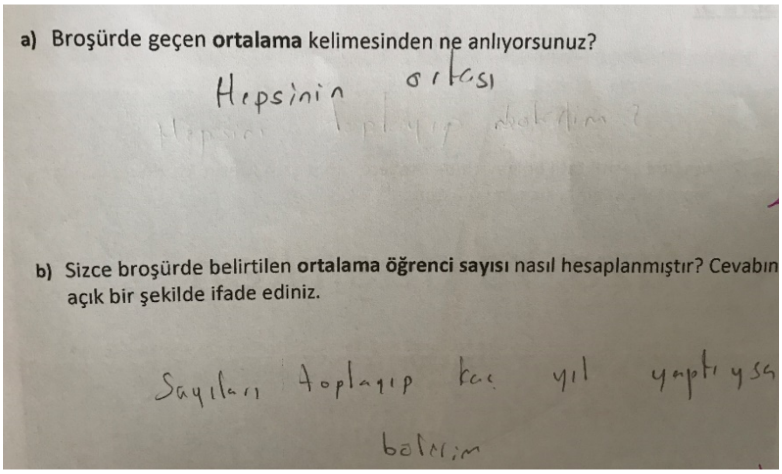


Figure 4.55 Interpretations of average by a Level 3-4 student in Q4a and Q4b

Table 4.14 in below summarizes the findings related to the concept of average on line graphs. The findings of the present study revealed that in general, the students' statistical literacy levels in the questions related to the concept of average on line graphs were determined as Level 3-4. There were no students reaching Level 5-6 in the line graph questions related to the average. To put it differently, while students could interpret the average concept on line graphs, not all of them could critically evaluate the presented average in the questions. Lastly, the *mean* was the most preferred measure of central tendency in interpreting the concept of average on line graphs as students used it frequently in interpreting the average concept on bar graphs.

In this section, findings related to the average concepts on bar and line graphs were presented. The other concept that this study focused was variation. In other words, the current study analyzed statistical literacy of seventh grade students about the concept of variation on bar and line graphs related to the data obtained from social or scientific contexts. Hence, the next two sections dwell on the findings of the current study related to the variation concept on line and bar graphs, respectively.

Table 4.14 The distribution of students across the six levels in the framework in the questions related to concept of average on line graphs

The Concept of Average on Line Graphs			Statistical Literacy Level		
Question	**Skill**	**Focus**	**Level 1-2**	**Level 3-4**	**Level 5-6**
Q2	Interpretation	Calculation	48 (29.3%)	**116 (70.7%)**	0 (0.0%)
Q4	Interpretation Evaluation	Median	54 (32.9%)	**110 (67.1%)**	0 (0.0%)

4.3 The Concept of Variation on Line Graphs

One of the purposes of the study was to analyze students' statistical literacy regarding the concept of variation on line graphs. Three questions, namely the second part of question 2 (Q2b), question 3 (Q3) and question 7 (Q7) were prepared with this aim. Q2 was related to the selection of a basketball player for the final game of the year. In this question, the students were provided with the number of baskets of two players in the final games of the last ten years on a line graph. While there was a high variation in the scores of one of the players, the others' scores were more consistent. However, the average scores of the players were almost the same. Of the two basketball players, the trainer of the players chose the one with varying scores. The students were requested to evaluate the decision of the trainer. By means of this question, whether or not students could critically evaluate the variation in the scores of the players was measured. On the other hand, Q3 aimed to measure students' interpretation of the variation concept. In this question, the students were provided with information regarding the average temperature of Ankara in the month of April, and they were requested to draw a line graph. In this way, students' interpretations of variation with a given average were examined. Furthermore, the aim of requesting the definition of range in Q3 was to examine students' knowledge related to the definition of variation. Lastly, students' ability to critically evaluate the variation concept, different from the comparison situation in Q2b, was examined through Q7. Q7 presented students the decision of the chamber of drivers about the speed of a service vehicle, the speed of which varied considerably. Whether or not students could critically evaluate the variation in the speed of the service was looked into.

To determine the levels of students for each question, the adapted and modified version of the framework of Watson and Callingham (2003) was used similar to the questions related to the the concept of average. In other words, the statistical literacy levels of students were determined by taking into consideration their definitions of range, and their interpretations and evaluations regarding the concept of variation in each question. The distribution of the students across six levels for Q2b, Q3 and Q7 are displayed in Table 4.15 below.

Table 4.15 The distribution of students across the six levels in the framework for Q2, Q3 and Q7

	Questions related to Variation on Line Graphs		
Levels	**Question 2b**	**Question 3**	**Question 7**
1-2	44 (26.8%)	119 (72.5%)	136 (71.0%)
3-4	97 (59.1%)	45 (27.4%)	26 (27.9%)
5-6	23 (14.0%)	0 (0.0%)	2 (1.2%)
Total	164 (100.0%)	164 (100.0%)	164 (100.0%)

As presented in Table 4.15, when the students' definitions, interpretations and evaluations of the variation concept were analyzed as a whole in Q2b, it was revealed that most of the students (59.1%) performed at Level 3-4. On the other hand, while 26.8% of the students were at Level 1-2 in Q2b, only 14.0% of the students performed at Level 5-6. As indicated in Table 4.15, in Q3, 72.5% of the students were found to be at Level 1-2. The remaining 27.4% of the students performed at Level 3-4, and there were no students performing at Level 5-6. This was an expected result for Q3 since it was a question related solely to the interpretation of the variation concept; hence, the students were expected to be at either Level 1-2 or Level 3-4. The findings were not different for Q7. To put it differently, it was found that most students (71.0%) performed at Level 1-2. While 27.9% of the students performed at Level 3-4, only 2 students (1.2%) were found to be at Level 5-6 in Q7.

To respond to the sub-question of the research question in the present study, the following sections present detailed explanations on the answers of the students at different statistical literacy levels for Q2b, Q3 and Q7 by providing examples from students' answers. In other words, how seventh grade students at different statistical literacy levels define, interpret and evaluate the variation concept on line graphs is explained for Q2b, Q3 and Q7.

4.3.1 The Concept of Variation on Line Graphs in Question 2b

Q2 was related to the selection of a basketball player for the final game of the year. In the question, the trainer of the basketball players needed to make a decision between the two players whose number of baskets in the final games of the last ten years were displayed on a double line graph. Although the average number of baskets of each of the two players was almost same, the variation in the number of baskets of the two players was different. While there was a wide variation among the scores of one of the players, Naz, the scores of the other player, Elif, remained more consistent throughout the ten years. As previously mentioned in Section 4.2.1, in the first part of Q2 (Q2a), students were requested to find the average number of baskets for the two players. On the other hand, in the second part of Q2

(Q2b), the students were given the information that the trainer decided to select Naz, who was the player with a wide variation in scores. The students were asked to evaluate the decision of the trainer. In this way, whether or not the students could critically evaluate the variation in the scores of the two players after they realized that their average scores were the same was examined.

The statistical literacy levels of the students were determined by an overall analysis of the students' evaluations of variation in Q2b and their definitions of range in Q3 since there was no question related to the definition of variation in Q2. However, in the analysis, it was noticed that even though some students could not define the concept of range in Q3, they could evaluate the variation in the scores of the players in Q2b. For example, S79 in Figure 4.56 could not present any definition for the concept of range. However, she could evaluate the variation in the scores of the two players stating that while the scores of Elif were almost the same in each year, there was a fluctuation in the scores of Naz in every year. Therefore, to determine statistical literacy levels of students in this question, the students' definitions of range in Q3 were not taken into consideration; only their evaluations related to the concept of variation were taken into account in Q2b. However, the frequencies of students who could define and could not define the concept of range were determined for each level. As presented in Table 4.15, the study revealed that most students (59.1%) performed at Level 3-4 in Q2b. While 26.8% of the students performed at Level 1-2, 14.0% of the students performed at Level 5-6 in Q2b. The sections that follow explain the students' evaluations related to the concept of variation in Q2b. The findings are presented for Level 1-2, Level 3-4 and Level 5-6 students by providing examples from students' answers. Moreover, the number of students who could and could not define the concept of range at each level is presented at the end of each level.

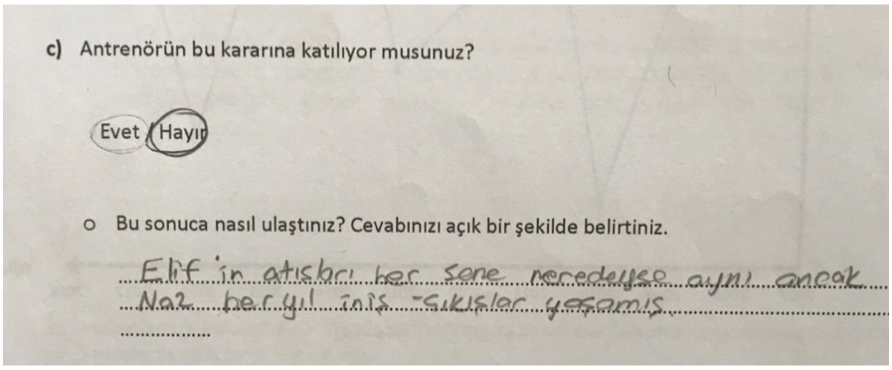


Figure 4.56 Evaluation of variation by a Level 5-6 student in Q2b

As can be observed in Table 4.15, 26.8% of the students performed at Level 1-2 in Q2b. In the modified version of the framework of Watson and Callingham (2003), it was stated that critical evaluation of the variation concept in a comparison situation was not an expected skill for Level 1-2 students. It was stated that while students depend on idiosyncratic beliefs in making comparisons at Level 1, students at Level 2 make their comparisons concentrating on only a single value in the given data sets. Since the characteristics observed at Level 1 and Level 2 are specified clearly in the framework, the levels of students could be determined as Level 1 and Level 2 separately. More specifically, while 12.8% of the students performed at Level 1 in Q2b, 14.0 % of the students performed at Level 2. In the following sections, initially, evaluations of variation made by Level 1 and Level 2 students in Q2b are presented by providing examples from students' answers. Subsequently, the number of Level 1-2 students who could and could not define the concept of range in Q3 is provided.

The findings of the present study revealed that 12.8% of the students either could not answer Q2b or provided some idiosyncratic responses. Therefore, these students were placed at Level 1. For example, S48 in Figure 4.57 agreed with the decision of the trainer. S48 stated that Naz was successful, smart and she made use of tactics in the game. She also mentioned some personality traits of Naz. This student evaluated the decision somehow, but her focus was not on the given data but on the personal traits of the player.

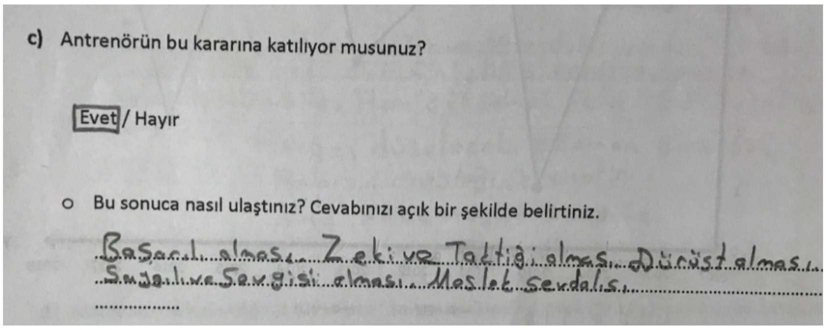


Figure 4.57 Evaluation of variation by a Level 1-2 student in Q2b

Moreover, it was observed that 14.0 % of the students used some single values when comparing the scores of the players; therefore, they were placed at Level 2. Figure 4.58 displays the answer of such a student. S16 in the below figure stated that while the minimum score of Elif was 4, the minimum score of Naz could be 1. Thus, he did not agree with the decision of the trainer. To state it differently, when making their decisions, these students did not concentrate on all the data, but only some specific values in the given data.

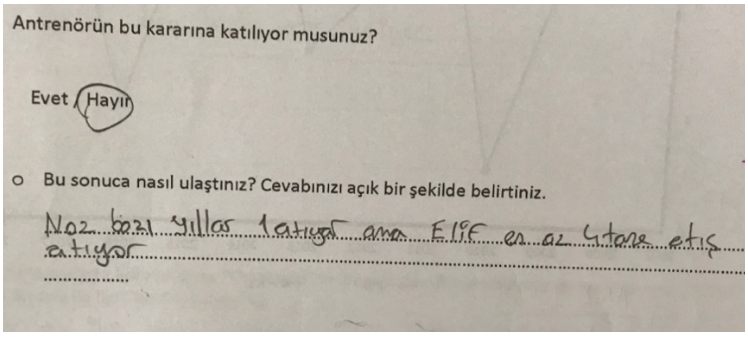


Figure 4.58 Evaluation of variation by a Level 1-2 student in Q2b

Surprisingly, S16 in Figure 4.58 above interpreted the average as the mean of the data in Q2a. In other words, he computed the mean scores of each player in Q2a. Because the total score he found for Naz was wrong, the mean score of Elif turned out to be greater than mean score of Naz. However, he did not use this information in making his evaluation. Instead, he concentrated on some values in the given data. It was realized that 13 of 23 (7.9%) students at Level 2 could calculate an average value in Q2a, but did not use that value in making their evaluations.

Besides, when Level 1 and Level 2 students' definitions of range were examined in Q3, it was observed that most students could not define the concept of range or presented some irrelevant responses: 9.1% and 11.6% of Level 1 and Level 2 students, respectively. Only 3 students at Level 1-2 could define the range as the difference between the maximum and minimum value in a data set. It was also realized that a few students presented some ideas regarding the difference, but could not present a complete definition. For example, the student in Figure 4.59 explained the range as the difference between one number and another and showed the gap between the numbers 10 and 99 as the range; however, she did not explain which numbers the difference should be applied to. Whether those numbers were in the data set or not was also left unexplained by the student.

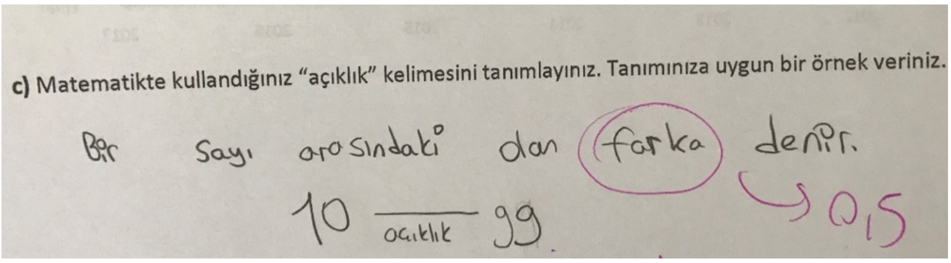
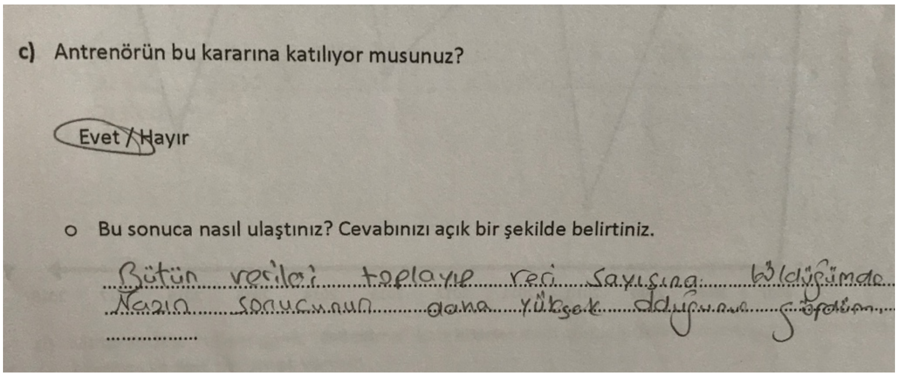


Figure 4.59 Definition of range by a Level 1-2 student in Q3

On the other hand, the analysis of students' evaluations of the variation concept in Q2b indicated that more than half of the students (59.1%) performed at Level 3-4 in Q2b. As aforementioned, in Q2b, only the evaluations of variations made by the students were taken into account to determine their statistical literacy levels since it was observed that although the students could evaluate the variation in the scores of the players, they could not define the range concept in Q3. In the modified version of the framework of Watson and Callingham (2003), it was stated that instead of only depending on single values in their comparisons, they make their comparisons by concentrating on more than one value in the given data sets at Level 3. However, at Level 4, students use numerical strategies to compare two data sets. To state it differently, they make their comparisons by either computing the sum or a measure of central tendency of the two data sets. Since the characteristics observed at Level 3 and Level 4 are specified clearly in the framework, the levels of the students could be determined as Level 3 and Level 4 separately; however, it was observed that none of the students performed at Level 3. More specifically, 59.1% of the students used a numerical strategy to compare the scores of the two players. It was realized that these students did not consider the variation in the scores of the players. While 37.8% of the students used the mean of the scores of the players to evaluate the decision of the trainer, 18.9% of the players used the sum of the scores in making their comparisons. There were also a few students calculating the median or mode of the given data sets and used them in their comparisons. In the sections that follow, first of all, variation evaluations of Level 4 students in Q2b are presented by providing examples from students' answers. Subsequently, the number of Level 4 students who could and could not define the concept of range in Q3 is provided.

As it was stated, 59.1% of the students' levels were determined as Level 4 since they used a numerical strategy to compare the two data sets. To state it differently, to evaluate the decision of the trainer, most students used the average score of the two players that they had computed in Q2a. For example, S74 in Figure 4.60 calculated the mean score of each player correctly in Q2a and found the scores of 4.7 and 4.6 for the players Naz and Elif, respectively. Furthermore, S74 stated that the score of Naz was greater when the sum of the data was divided by the number of data. It can be concluded that even a 0.1 difference in the mean scores was very important for this student.


c) Antrenörün bu kararına katılıyor musunuz?

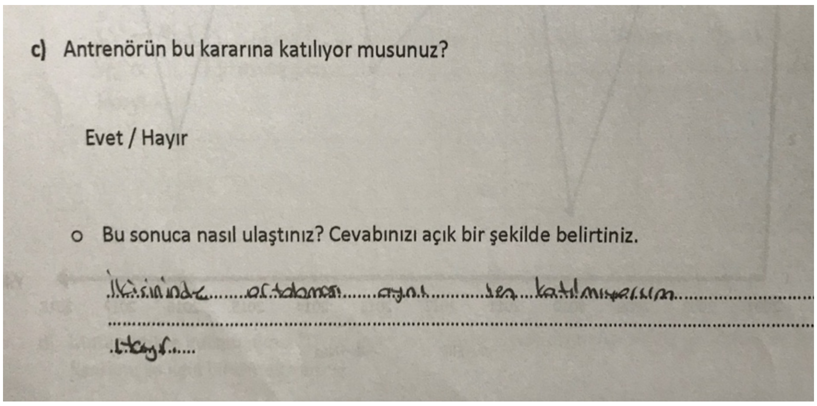
Evet / Hayır

o Bu sonuca nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

Bütün verileri toplayıp veri sayısına böldüğümde Nazlı sonucunun daha yüksek olduğunu gördüm.

Figure 4.60 Evaluation of variation by a Level 3-4 student in Q2b

Moreover, it was observed that some Level 4 students in Q2b computed the mean scores of the two players, but found them equal because of calculation errors. However, these students, such as S154 in Figure 4.61, stated that they did not agree with the decision of the trainer since the mean scores of the two players were equal; they did not mention any further idea regarding which players to select.


c) Antrenörün bu kararına katılıyor musunuz?

Evet / Hayır

o Bu sonuca nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

İkisinin de ortalaması aynı ben katılmıyorum

Hayır

Figure 4.61 Evaluation of variation by a Level 3-4 student in Q2b

Besides, as mentioned previously, about 20.0% of the students whose statistical literacy levels were determined as Level 4 used the sum of the players in making their decision. Interestingly, it was realized that although some students calculated the mean scores of the two players in Q2a, which asked for the average of the two players, they calculated the sum of the players again to compare the players. To state it differently, they did not use the mean that they had calculated in Q2a.

When the definitions of range made by the students whose statistical literacy levels were determined as Level 4 in Q2b were examined, it was observed that more students (19.5%) could define the concept of range when compared to the number of students whose statistical literacy levels were determined as Level 1 and Level 2. However, there were still 47 students (28.7%) who could not provide any definition of range. Furthermore, 6.1% of the students stated that range is a concept related to the difference, but could not provide a complete definition.

Lastly, the analysis of students' evaluations of the variation concept in Q2b indicated that 14.0% of the students performed at Level 5-6. In the framework, it is stated that a critical evaluation of the variation concept in a situation of comparison was an expected skill for Level 5-6 students. It is stated that while students make their comparisons using visual strategies including the variation concept at Level 5, at Level 6, they use numerical strategies and visual strategies including the variation concept together to make their comparisons. Since the characteristics observed at Level 5 and Level 6 were specified clearly in the framework, the levels of the students could be determined as Level 5 and Level 6 separately. More specifically, while 12.1% of the students performed at Level 5, only 3 students (1.8%) performed at Level 6 in Q2b. In the following sections, firstly, evaluations of variation made by students whose statistical literacy levels were determined as Level 5 and Level 6 in Q2b are presented by providing examples from students' answers. Then the frequencies of Level 5-6 students who could and could not define the concept of range in Q3 will be provided.

The findings of the present study indicated that 12.1% of the students only used some visual strategies, implying appreciation of variation in the scores of the players to make their decision related to the decision of the trainer. Therefore, these students were placed at Level 5. For instance, S79 in Figure 4.62 stated that she did not agree with the decision of the trainer since while the scores of Elif were almost the same in each year, the scores of Naz fluctuated in every year. Even though S79 did not use the exact word of variation, it was clear that she realized that there was a higher variation in the scores of Naz than there was in the scores of Elif by examining scores of the two players in the given graph. Furthermore, it was observed that even though S79 had computed average score of the two players in Q2a, she did not make use of those averages in making her comparison but only depended on the variation in the scores of the players. It was observed that 19 of 23 Level 5 students (11.7%) did not use the averages that they had calculated in Q2a.

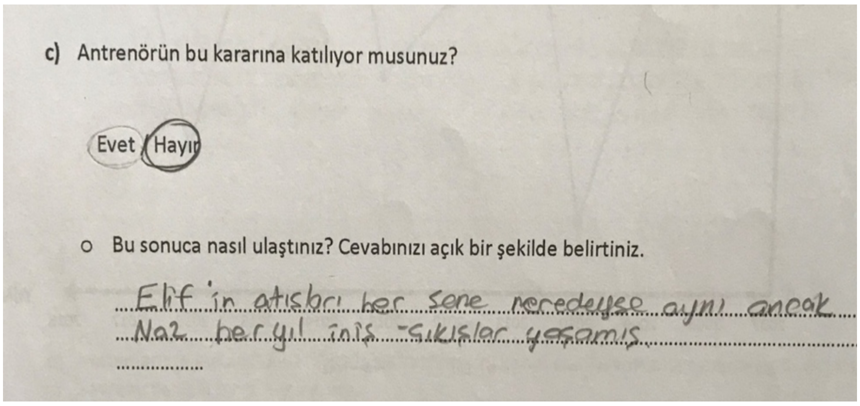
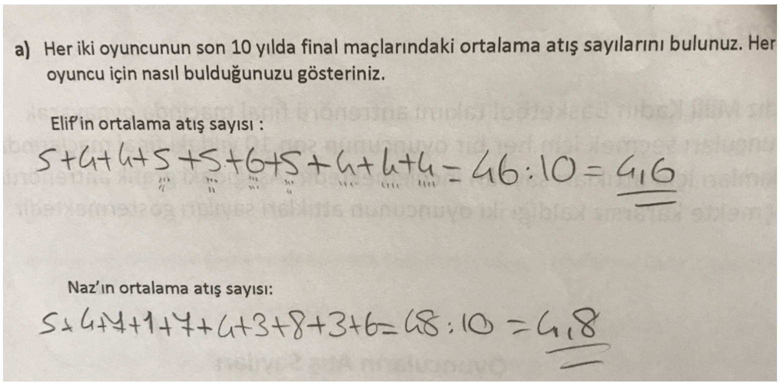


Figure 4.62 Evaluation of variation by a Level 5-6 student in Q2b

On the other hand, only 3 students (1.8%) used both a numerical strategy and a visual strategy, implying the appreciation of variation in the scores of the players, to make their comparisons in Q2b. Thus, these students were placed at Level 6. For example, S84 in Figure 4.63 calculated the mean scores of the two players in Q2a, and in the evaluation of the decision of the trainer in Q2b, S84 stated that although the average of Naz's scores was greater than that of Elif's, there were some dramatic drops in Naz's scores; therefore, to prefer Elif was more reliable.



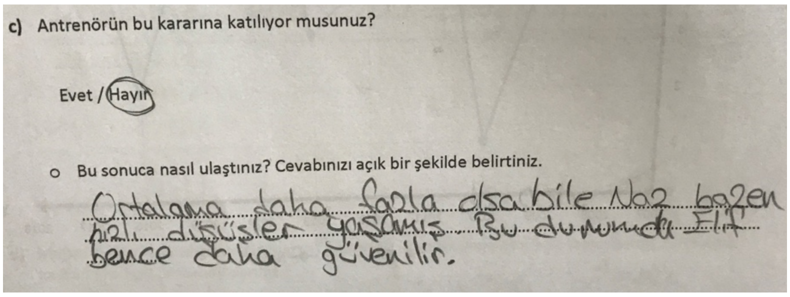
a) Her iki oyuncunun son 10 yılda final maçlarındaki ortalama atış sayılarını bulunuz. Her oyuncu için nasıl bulduğunuzu gösteriniz.

Elif'in ortalama atış sayısı :

5+4+4+5+5+6+5+4+4+4=46:10=4,6

Naz'ın ortalama atış sayısı:

5+4+4+1+7+4+3+8+3+6=48:10=4,8



c) Antrenörün bu kararına katılıyor musunuz?

Evet / Hayır

o Bu sonuca nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

Ortalama daha fazla olsa bile Naz bazen hızlı düşüşler yaşamış. Bu durumda Elif bence daha güvenilir.

Figure 4.63 Interpretation of average and evaluation of variation by a Level 5-6 student in Q2a and Q2b, respectively

Likewise, S45 in Figure 4.64, whose statistical literacy level was determined as Level 6, calculated the sum of the scores of the two players, but she found the sum of Elif's scores to be more than the sum of Naz's scores in Q2a, which asked for the average of the two players. In the evaluation of the decision of the trainer in Q2b, S45 stated that the score of Elif was higher than that of Naz; moreover, considering the graph, it can be observed that there are fluctuations in the scores of Naz. She continued stating that Naz had high scores in one year, while she had low scores in another year, but no difference was observed among the scores of Elif. To state it differently, S45 tried to explain that while the scores of Elif was consistent, the scores of Naz was inconsistent, which implies that she took into consideration the variation in the scores of the two players. Furthermore, since she expressed that Elif's score was higher than that of Naz, she also used a numerical strategy to make her comparison.

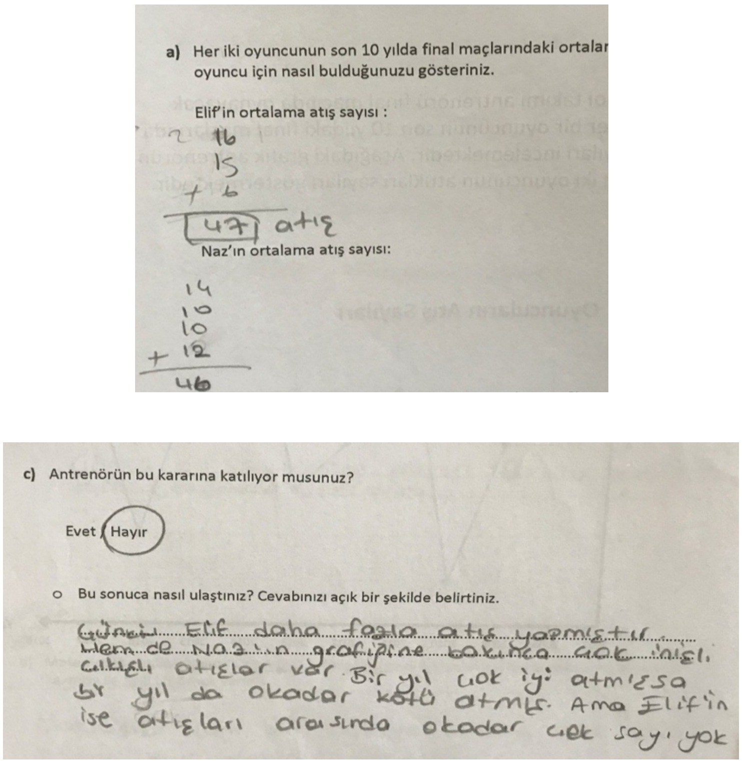


Figure 4.64 Interpretation of average and evaluation of variation by a Level 5-6 student in Q2a and Q2b, respectively

Moreover, when the definitions of range made by students whose statistical literacy levels were determined as Level 5-6 in Q2b were analyzed, similar results were obtained from the responses of Level 1-2 and Level 3-4 students. It was observed that 9 of 23 (5.4%) Level 5-6 students could not present any definition. Only 2 students (1.2%) among 23 students whose statistical literacy levels were determined as Level 5-6 could define the concept of range as the difference between the maximum and minimum value in a data set.

4.3.2 The Concept of Variation on Line Graphs in Question 3

Question 3 (Q3) was another question which was asked to investigate the statistical literacy of students related to the variation concept. In the first part of Q3 (Q3a), the students were provided with information regarding the average temperature of

Ankara in the month of April, and they were requested to draw a line graph which reflected the given average temperature. In the second part of Q3 (Q3b), the students were asked to interpret their graph by taking into account the given average temperature. In this way, students' interpretation of the variation concept to a given average could be analyzed. The last part of Q3 (Q3c) requested students to define the concept of range, which is a measure related to the concept of variation. In other words, the definition of the variation concept was examined indirectly by asking the definition of range.

The statistical literacy levels of the students were determined by an overall analysis of the graph which the students had drawn in Q3a, their interpretations in Q3b and their definitions of range in Q3c. However, in the analysis, it was noticed that even though some students could not define the concept of range in Q3c, they could present a graph that reflected an appropriate variation as in Q2b above. For example, S74 in Figure 4.65 could not present any definition for the concept of range; however, as can be examined in the figure, she reflected the appropriate variation through her graph: Her graph included temperatures between 4 and 30°C. S74 stated that to construct her graph, she computed the average of each of the two data as 17 °C. Thus, to determine statistical literacy levels of students in this question, their definitions of range in Q3c were not taken into consideration; only their graphs and interpretations related to the concept of variation were taken into account in Q3a and Q3b. However, the frequencies of the students who could and could not define the concept of range in Q3c were determined for each level. As indicated in Table 4.15, most students (72.5%) performed at Level 1-2 in Q3. The percentage was much higher when compared with that of Level 1-2 students in Q2b mentioned above. While 27.4% of the students performed at Level 3-4, there were no students performing at Level 5-6 in Q3. This was an expected result for Q3 since it was a question related only to the interpretation of the variation concept; hence, the students were expected to be at either Level 1-2 or Level 3-4. The following sections display students' graphs and their interpretations related to the concept of variation. The findings are presented for Level 1-2 and Level 3-4 students by providing examples from students' answers. Moreover, the number of students who could and could not define the concept of range at each level is presented at the end of each level.

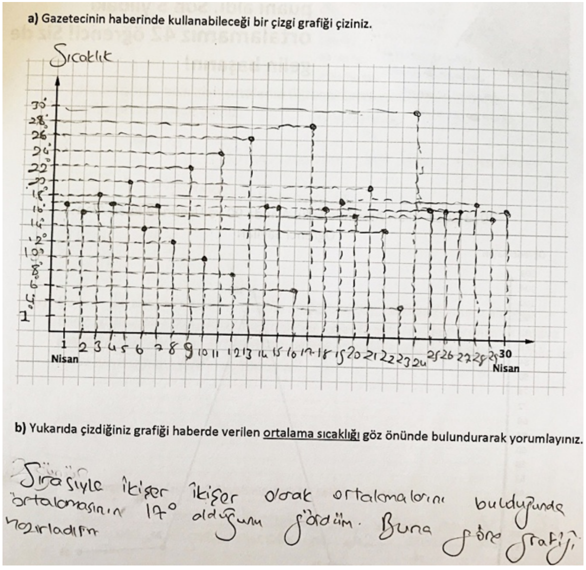


Figure 4.65 Graph of a Level 3-4 student in Q3a and her interpretation of the graph in Q3b

The findings of the current study revealed that different from the result obtained for Q2b, most students (72.5%) performed at Level 1-2 in Q3. In the framework, it is asserted that while students cannot draw a graph reflecting variation at Level 1, at Level 2, they begin to produce a graph with variation, but this variation is not appropriate for the task given. Since the characteristics observed at Level 1 and Level 2 are specified clearly in the framework, the students' levels could be determined as Level 1 and Level 2 separately. More specifically, while 51.8% of the students performed at Level 1 in Q3, 20.7 % of the students performed at Level 2. In the sections that follow, initially, the graphs and interpretations produced by Level 1 and Level 2 students in Q2b are presented by providing examples from their answers. Then, the number of Level 1-2 students who could and could not define the concept of range in Q3c is provided.

As aforementioned, more than half of the students (51.8%) performed at Level 1 in Q3. These students were placed at Level 1 since most of them (36.6%) had not drawn any graph. Furthermore, 9.8% of the students either only interpreted the given average temperature but did not draw any graph or drew a graph but only showed the temperature of a single day as 17 °C. These students were also placed at Level 1. For example, S5 in Figure 4.66 did not draw any graph in Q3a and stated that the temperature increased up to 17 °C in 30 days. From her explanation, it can be understood that she thought the maximum temperature in the month of April was 17°C. However, since she did not mention the temperatures of any of the other days, her interpretation of variation could not be understood, so this student was placed at Level 1.

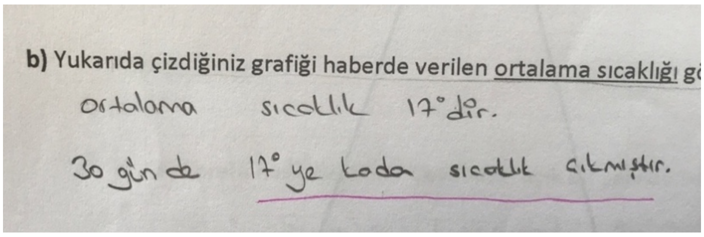


Figure 4.66 Interpretation of the graph by a Level 1-2 student in Q3b

On the other hand, it was realized that some students did not draw a complete graph, but only showed the temperature of a single day as the average temperature. In other words, for these students, for the average to be 17 °C, the temperature of one of the days being 17 °C is sufficient. These students did not consider the temperatures of the other days; that is, they did not think of the variation in the temperature. Therefore, these students were placed at Level 1. Figure 4.67 displays the answer of such a student.

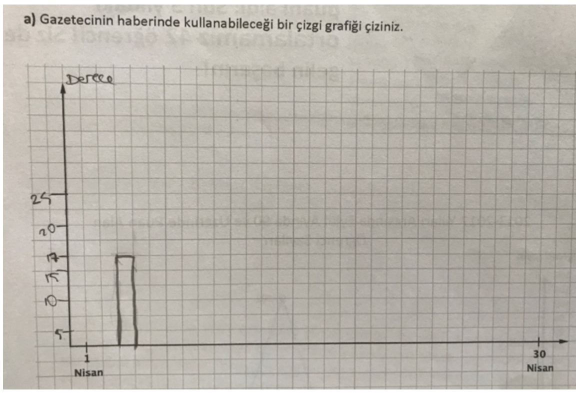


Figure 4.67 Graph of a Level 1-2 student in Q3a

Level 1 also included students that did not consider the variation in the weather and believed that the temperatures of all the days would be the same, i.e. 17 °C. For example, Figure 4.68 displays the graph of such a student. S12 stated that the temperature was 17 °C throughout the month and did not include any temperature other than 17 °C in her graph.

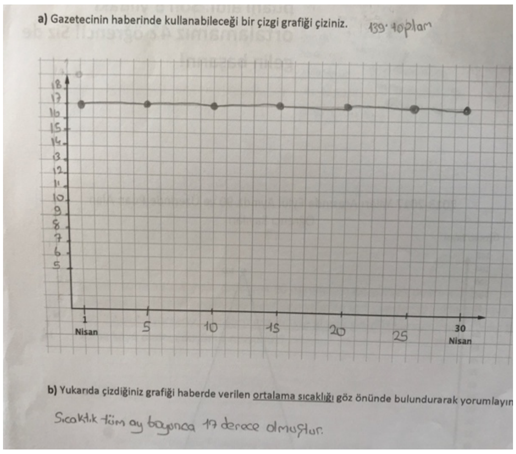


Figure 4.68 Graph of a Level 1-2 student in Q3a and her interpretation of the graph in Q3b

On the other hand, the findings of the present study revealed that 20.7% of the students started to include variation in their graphs; however, these were determined as inappropriate since the graphs of the students did not reflect the real-world situation. However, since variation was included, students were placed at Level 2, not Level 1. It was observed that these students either included too wide or too small variations in their graphs or their graphs were similar to the that of linear equations. For instance, S78 in Figure 4.69 drew his graph using the idea of the mode. He showed the temperatures of most days as 17 °C. He also stated that the temperature remained constant during the 26th and 27th day of the month and the mode was 17 °C. S78 could also use the mean algorithm since he multiplied 18 by 8 in his graph as can be observed in the figure. Nevertheless, his graph did not include any temperature other than 16, 17 and 18 °C, which actually did not reflect the real-world situation; hence, this student was placed at Level 2.

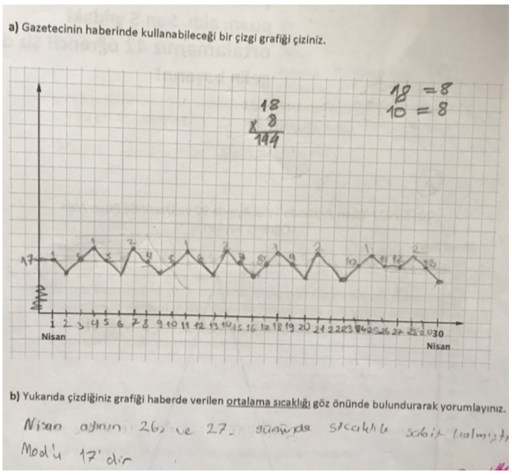


Figure 4.69 Graph of a Level 1-2 student in Q3a and his interpretation of the graph in Q3b

Likewise, S90 in Figure 4.70 stated that, in general, temperatures were between 10 and 20 °C; however, her graph included a day whose temperature was 50°. In other words, she produced a graph with a highly broad variation. Therefore, this student was placed at Level 2, just as the student in Figure 4.69.

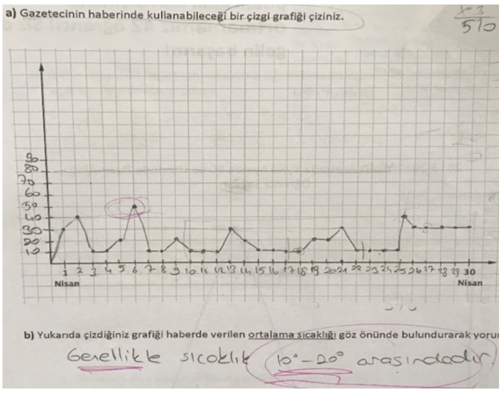


Figure 4.70 Graph of a Level 1-2 student in Q3a and his interpretation of the graph in Q3b

Furthermore, the graph of some Level 2 students were like that of linear equations. Figure 4.71 displays such a graph. S21 asserted that the temperature increased 1° C every day and then remained at 17°C. Since such a consistent increase and stability is not expected in the weather in the month of April, the graphs of these students were regarded as including inappropriate variation, so these students were placed at Level 2.

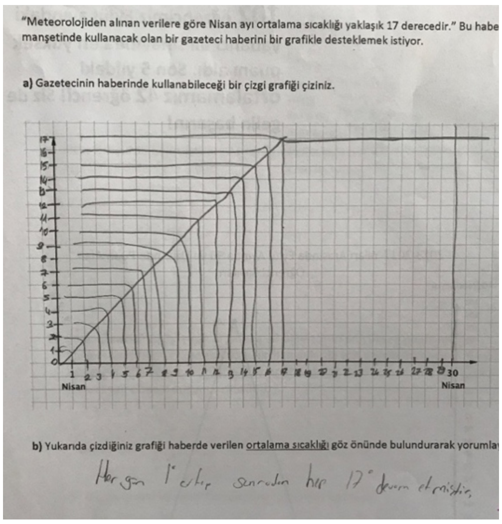


Figure 4.71 Graph of a Level 1-2 student in Q3a and his interpretation of the graph in Q3b

Besides, when the definitions of range made by students whose statistical literacy levels were determined as Level 1-2 in Q3 were examined, it was revealed that the findings were not different from the definitions of both Level 1 and Level 2 students in Q2b. Particularly, most students could not define the concept of range or presented some irrelevant responses: 36.6% and 11.4% of Level 1 and Level 2 students, respectively. It was observed that the number of students defining the concept of range correctly was higher when compared to Q2b: 11.0% and 6.1% of Level 1 and Level 2 students, respectively. There were also some students who presented some ideas regarding the difference, but could not present a complete definition.

It was realized that the definition of variation had no impact on students' interpretation of variation at all levels; therefore, students' range definitions were not taken into account in determining their statistical literacy levels. For example, S12 in Figure 4.68 whose statistical literacy level was determined as Level 1 could define the concept correctly; however, as previously stated, her graph did not include any variation; she showed the temperature of all the days as 17 °C . To state it differently, this student could not use her information related to the range concept in interpretation of the concept of variation.

On the other hand, the analysis of students' graphs in Q3a and their interpretations related to the concept of variation in Q3b indicated that 27.4% of the students performed at Level 3-4 in Q3. The percentage was much lower when compared to the that of students at Level 3-4 in Q2b. As aforementioned, in Q3, only the graphs and interpretations related to the concept of variation of the students were taken into account to determine their statistical literacy levels since it was observed that although the students could interpret the variation in the temperature, they could not define the range concept in Q3c. In the modified version of the framework of Watson and Callingham (2003), it is stated that students at these levels can produce graphs reflecting an appropriate variation. In other words, students at Level 3-4 can interpret the variation in the temperature. The difference between the two levels is that while students cannot draw a complete graph at Level 3, at Level 4, students reflect the temperature of all the days to show the given average and to interpret the variation in the temperature. In other words, Level 4 students believe that the temperature of all the days is important to determine the average temperature of the month. Since the characteristics observed at Level 3 and Level 4 are specified clearly in the framework, the levels of the students could be determined as Level 3 and Level 4 separately in the current study. More specifically, while 8.5% of the students performed at Level 3 in Q3, 18.9% of the students performed at Level 4. In the following sections, first of all, graphs and interpretations of Level 3 and Level 4 students in Q3 are presented by providing examples from students' answers. Then, the number of Level 3-4 students' who could and could not define the concept of range in Q3c is provided.

The findings of the present study revealed that the graphs of 8.5% of the students included an appropriate variation, but they did not complete their graphs. For example, S131 in Figure 4.72 showed only the temperature of 9 days. The temperatures of the days varied between 10 and 24. Since such a range in the temperature is expected in the month of April, the graphs of these students were identified as including an appropriate variation, but since the graph was not completed, these students were placed at Level 3.

Furthermore, 18.9% of the students did draw a complete graph and also included an appropriate variation in their graphs; hence, the levels of these students were determined as Level 4. For example, S74 in Figure 4.73 included temperatures between 4 and 30°C. S74 also stated while interpreting her graph in Q3b that the average of each two data was 17 °C. As can clearly be understood, this student initially determined different pairs of temperatures whose mean was 17 °C, such as 18 and 16 °C or 28 and 6 °C. Then, placed these data into her graph. Since S74 did not determine only one pair but included various different pairs in her graph, S74 was believed to have provided an appropriate variation. Furthermore, she showed the temperatures of all the days; therefore, she was placed at Level 4, the maximum level expected in Q3 since there was no question related to the evaluation of the variation concept.

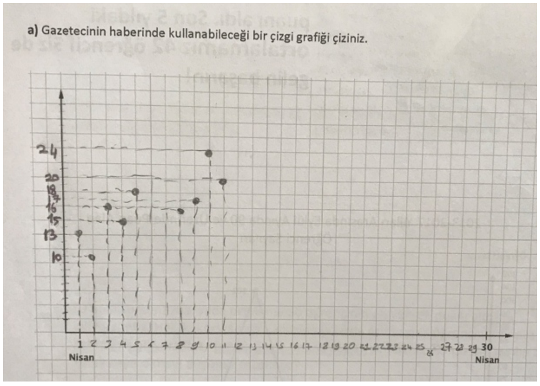


Figure 4.72 Graph of a Level 3-4 student in Q3a

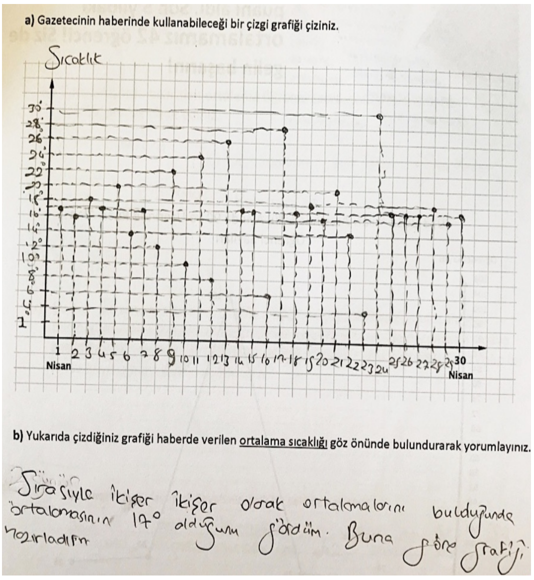


Figure 4.73 Graph of a Level 3-4 student in Q3a and her interpretation of graph in Q3b

Similarly, another student in Figure 4.74 drew his graph taking into account every fifth day of the month. Similar to the above student, graph of S24 in the Figure included five different temperatures whose mean is 17 °C; therefore, placed at Level 4.

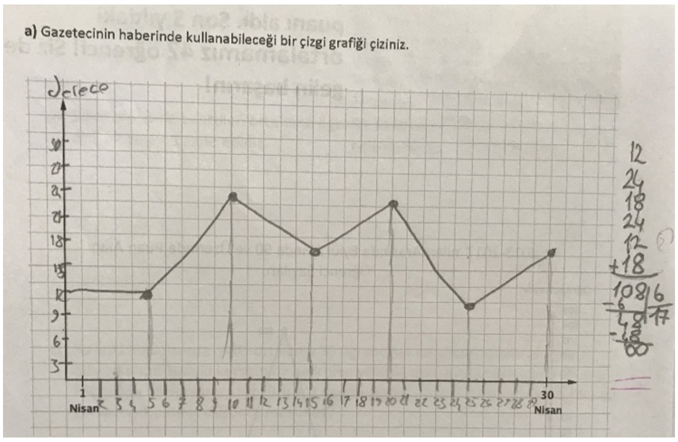


Figure 4.74 Graph of a Level 3-4 student in Q3a

Lastly, the definitions of range made by the students whose statistical literacy levels were determined as Level 3-4 in Q3 were examined. The findings were not different from the definitions of both Level 3-4 students in Q2b and of Level 1-2 students in Q3. Particularly, most students could not define the concept of range or presented some irrelevant responses: 7.9% and 7.3% of Level 3 and Level 4 students, respectively. It was observed that the percentage of students defining the concept of range correctly was lower when compared to Q2b: 1.2% and 7.9% of Level 3 and Level 4 students, respectively. There were 3 (1.8%) Level 4 students who presented some ideas regarding the difference, but could not present a complete definition.

4.3.3 The Concept of Variation on Line Graphs in Question 7

The last question related with the variation concept on line graphs was Question 7 (Q7). Q7 provided students with a line graph which displayed the speed of a factory service at specific times in a day. In the context, there was a wide variation in the speed of the service throughout the day, and its average speed was higher than the speed limit, 70 km/h. However, the Chamber of Drivers did not impose any fines to the driver of the service. The students were requested to evaluate the Chamber in related to the service driver. By means of this question, whether or not students could critically evaluate the variation in the speed of the service was examined.

The statistical literacy levels of the students were determined by an overall analysis of students' evaluations of variation in Q7 and their definitions of range in Q3c since there was no question related to the definition of variation in Q7. However, in the analysis, no effect of the definition of range on students' evaluation of variation was observed as in Q2b and Q3. To state it differently, even though some students could not define the concept of range in Q3c, they could evaluate the variation in the speed of the service in Q7. Therefore, to determine the statistical literacy levels of students in this question, their definitions of range in Q3c were not taken into consideration as in Q2b and Q3; only their evaluations related to the concept of variation in Q7 were taken into account. However, the frequencies of the students who could and could not define the concept of range were determined for each level. Table 4.15 revealed that while most students (71.0%) performed at Level 1-2 as in Q3, 27.9% of the students performed at Level 3-4 in Q7. Q7 required the evaluation of the variation concept so most students were expected to be at Level 5-6. However, only 2 students (1.2%) performed at Level 5-6 in Q7. The following sections explain students' evaluations related to the concept of variation in Q7. The findings are presented for Level 1-2, Level 3-4 and Level 5-6 students by providing examples from students' answers. Nevertheless, the percentage of students who could and could not define the concept of range at each level are not presented at the end of each level for Q7 because the frequencies obtained for each level were very similar to the ones obtained in Q2b and Q3. Furthermore, this question was similar to Q2b in Section 4.3.1. Both questions examined whether or not the students could critically evaluate the variation concept; therefore, the findings will be presented in line with the findings of Q2b.

The findings of the present study revealed that 71.0% of the students performed at Level 1-2 in Q7 which is a similar result obtained in Q3. As stated previously, the framework asserts that if students make their evaluations based on any explanation or idiosyncratic responses, they are placed at Level 1. On the other hand, if they use only a single value in their evaluations, their levels are determined as Level 2. In the current study, 36.2 % of the students either did not provide any answer or provided some idiosyncratic responses; therefore, their statistical literacy levels were determined as Level 1 for Q7. For instance, S21 in Figure 4.75 stated that if he went at 70 km/h, it lasted 2 days to reach a place; however, if he went at 120 km/h, it lasted 1 day. Since S21 used his own opinion regarding the speed of the service, but not the given graph, the answer of this student was determined as idiosyncratic and his statistical literacy level for Q7 was determined as Level 1.

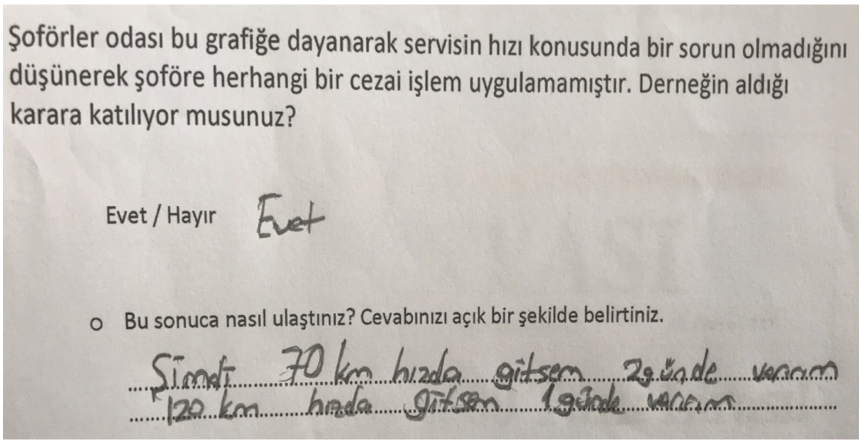
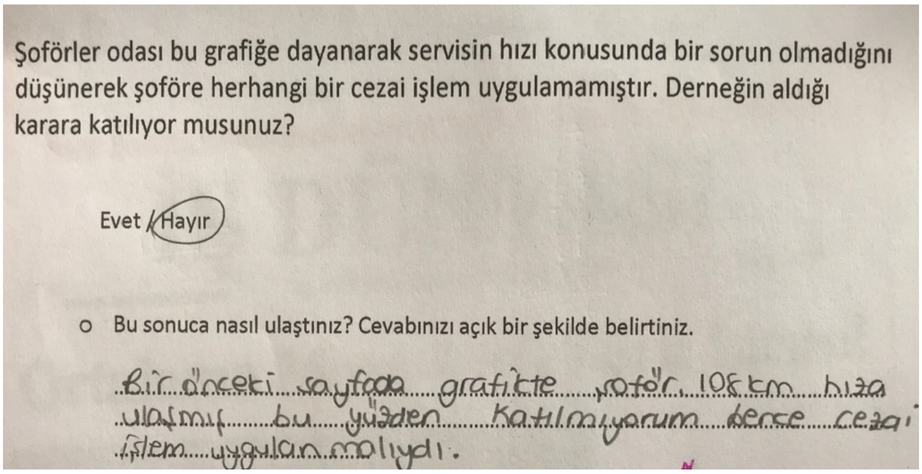


Figure 4.75 Evaluation of variation by a Level 1-2 student in Q7

Moreover, it was observed that 34.8% of the students depended on only a single value in the given graph and compared that value with the average speed, 70 km/h, to make their evaluations. Therefore, these students were placed at Level 2. It was observed that the students generally focused on the maximum speed that the service reached at 12:00, 108 km/h. For example, S5 in Figure 4.76 said that in the graph, the driver reached the speed of 108 km/h; therefore, the fine should have been imposed. Since S5 concentrated on only one value but not all the given values in the graph, this student was placed at Level 2.



Şoförler odası bu grafiğe dayanarak servisin hızı konusunda bir sorun olmadığını düşünerek şoföre herhangi bir cezai işlem uygulamamıştır. Derneğin aldığı karara katılıyor musunuz?

Evet / Hayır

- Bu sonuca nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

Bir önceki sayfada grafikte şoför 108 km hıza ulaşmış bu yüzden katılmıyorum bence cezai işlem uygulanmalıydı.

Figure 4.76 Evaluation of variation by a Level 1-2 student in Q7

On the other hand, the analysis of students' evaluations of the variation concept in Q7 indicated that 27.9% of the students performed at Level 3-4 in Q7. In the modified version of the framework of Watson and Callingham (2003), it is stated that at Level 3, students not only depend only a single value in their comparisons but also concentrate on more than one value in the given data sets. However, at Level 4, students use numerical strategies while comparing two data sets. To state it differently, they make their comparisons by either computing the sum or a measure of the central tendency of the two data sets. Since the characteristics observed at Level 3 and Level 4 are specified clearly in the framework, the levels of the students could be determined as Level 3 and Level 4 seperately. More specifically, while 12.0% of the students performed at Level 3, 15.9% of the students used a numerical strategy to make their evaluations; therefore, they were placed at Level 4. It was realized that these students did not consider the variation in the speed of the service. In the following sections, the evaluations of variation made by Level 3 and Level 4 students in Q7 are presented by providing examples from students' answers.

As mentioned previously, different from Q2b, 12.0% of the students' levels were determined as Level 3 since they had used more than one value in the given graph to make their evaluations in Q7. For instance, S22 in Figure 4.77 did not only focus on 108 km/h as S5 in Figure 4.76. Instead, S22 drew a line for the average speed and stated that the driver exceeded the speed limit 6 times; therefore, fines should be imposed. In other words, this student concentrated on almost all the values in the given data but did not calculate an average value; hence, he was placed at Level 3 in Q7.

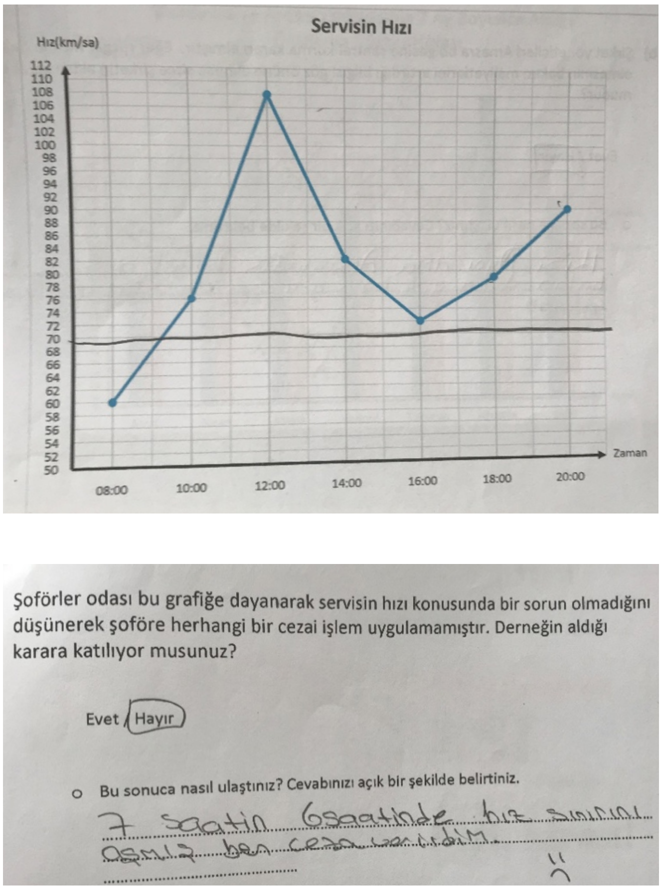


Figure 4.77 Evaluation of variation by a Level 3-4 student in Q7

On the other hand, 15.9% of the students used a numerical strategy to evaluate the decision of the Chamber of the Drivers. Different from Q2b, no student calculated the sum of the speeds in Q7, but all of them computed the mean of the speed of the service and made their evaluations accordingly. Since these students used a numerical strategy as in Q2b and thus have taken into account all the given values, these students were placed at Level 4. Figure 4.78 displays the answer of such a student. S34 in the figure calculated the mean of the given data as 81 km/h, and since it is higher than the average speed, she did not agree with the decision of the Chamber. However, it was realized that S34 also stated that the speed of the service was very high at specific times of the day. In other words, even though this student could calculate the mean of the given data set considering all of the given values, she still depended on some single values in making her evaluation.

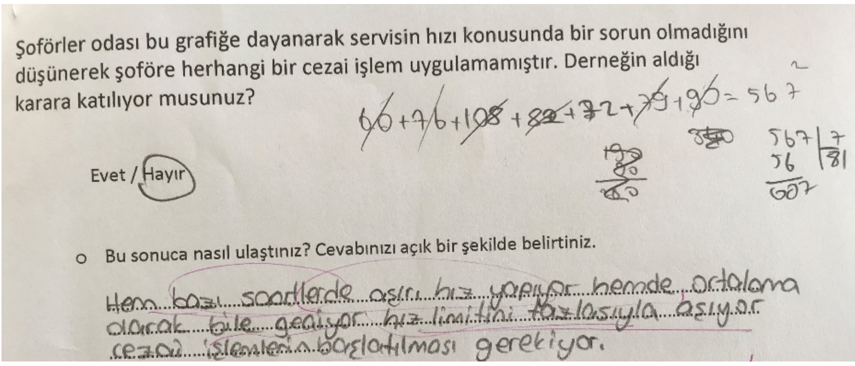
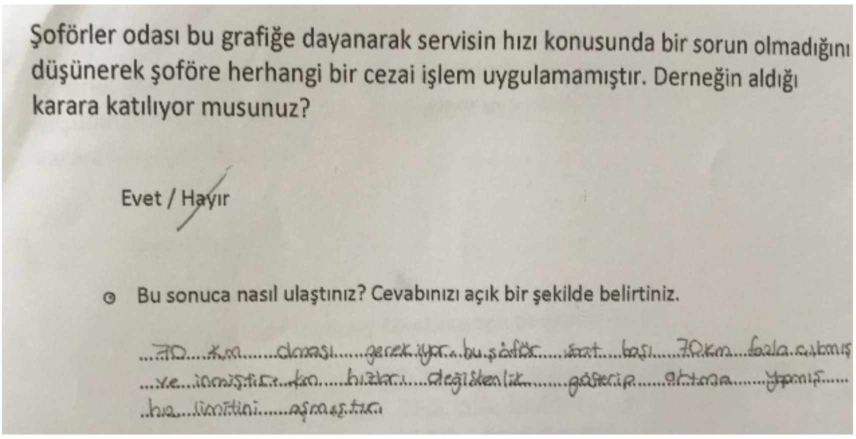


Figure 4.78 Evaluation of variation by a Level 3-4 student in Q7

Lastly, the analysis of students' evaluations of the variation concept in Q7 indicated that only 2 students (1.2 %) performed at Level 5-6. In the framework, it is stated that critical evaluation of the variation concept is an expected skill for Level 5-6 students. It is explained that while students make their evaluations using visual strategies including the variation concept at Level 5, at Level 6, they use numerical strategies and visual strategies including the variation concept together to make their evaluations. It was observed that 2 students reaching the Level 5-6 used only visual strategies including the variation concept; hence, these students were placed at Level 5. A sample answer of one of these students is provided in the section that follows.

One of the students using a visual strategy appreciating the variation in the speed of the service was S42. As can be observed in Figure 4.79, S42 stated that that speed of the service should be 70 km/h; however, the speed of this driver increased and decreased from 70 km/h at every hour. She maintained that the speed of the service showed some variation and exceeded the speed limit. This student stated the variation in the speed of the service explicitly but did not calculate an average value; therefore, the statistical literacy level of the student was determined as Level 5.


Şoförler odası bu grafiğe dayanarak servisin hızı konusunda bir sorun olmadığını düşünerek şoföre herhangi bir cezai işlem uygulamamıştır. Derneğin aldığı karara katılıyor musunuz?

Evet / Hayır

Bu sonuca nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

70 km olması gerekiyor bu şoför saat başı 70km fazla çıkmış ve inmiştir km hızları değişkenlik gösterip artma yapmış hız limitini aşmıştır

Figure 4.79 Evaluation of variation by a Level 5-6 student in Q7

Table 4.16 in below summarizes the findings related to the concept of variation on line graphs. The findings of the present study revealed that in general, the statistical literacy levels of the students in the questions related to the variation concept on line graphs were determined as Level 1-2. There were not many students reaching Level 5-6 in all of the questions. To put it differently, most students could neither interpret the variation concept on line graphs nor evaluate the variation in the questions critically. Generally, they could not answer the questions or depended on idiosyncratic responses.

Table 4.16 The distribution of students across the six levels in the framework in the questions related to concept of variation on line graphs

The Concept of Variation on Line Graphs		**Statistical Literacy Level**		
Question	**Skill**	**Level 1-2**	**Level 3-4**	**Level 5-6**
Q2	Evaluation	44 (26.8%)	**97 (59.1%)**	23 (14.0%)
Q3	Interpretation	**119 (72.5%)**	45 (27.4%)	0 (0.0%)
Q7	Evaluation	**136 (71.0%)**	26 (27.9%)	2 (1.2%)

In this section, findings related to the variation concept on line graphs were presented The other aim of this study was first of all to examine statistical literacy of seventh grade students regarding the concept variation on bar graphs related to the data obtained from social or scientific contexts. Hence, in the next section, the findings of the present study related to the concept of variation concept on bar graphs are presented.

4.4 The Concept of Variation on Bar Graphs

To measure statistical literacy levels of students regarding the concept of variation on bar graphs, one more question was prepared, similar to Q2b in section 4.3. The question was the second part of question 6 (Q6b). In the question, the students were given a double bar graph displaying the wind speed of two regions, Amasra and Bandırma, throughout seven months. Although the average wind speed of the two regions were equal, the variation in the speed of the wind in the Amasra region was greater. While the wind speed of the Amasra region varied substantially, the wind speed of the Bandırma region remained more consistent throughout the seven months. In the first part of Q6 (Q6a), the students were requested to find the average wind speeds of the two regions. On the other hand, in the second part of Q6 (Q6b), the students were told that the company had decided to select the Amasra region where the speed of the wind varied substantially, and the students were asked to evaluate the decision of the company. By means of this question, whether or not students could critically evaluate the variation in the speeds of the winds, after they realized that their average speeds were the same, was examined. Different from Q2b, the importance of variation was also expressed in Q6b by stating that a fluctuation in wind speed increases maintenance cost.

To determine the levels of the students for each question, the modified version of the framework of Watson and Callingham (2003) was used, just as in the other variation questions explained above. In other words, the statistical literacy levels of the students were determined by taking into consideration their definitions of range in Q3 and their evaluations regarding the concept of variation in Q6b. The distribution of students across six levels for Q6b are displayed in Table 4.17 below.

Table 4.17 The distribution of students across the six levels in the framework for Q6b

	Question related to Variation on Bar Graphs
Levels	**Question 6b**
1-2	68 (41.4%)
3-4	61 (37.2%)
5-6	35 (21.3%)
Total	164 (100.0%)

As indicated in Table 4.17, when students' definitions, and evaluations of the variation concept were analyzed as a whole in Q6b, it was revealed that, different from Q2b, while almost half of the students (41.4%) performed at Level 1-2 in Q6b, 37.2% of the students performed at Level 3-4. Moreover, 21.3% of the students performed at Level 5-6 in Q6b. The percentage was higher when compared to the number of students who performed at Level 5-6 in Q2b.

To address the sub-question related to the research question of the present study, the following sections dwell on detailed explanations of the answers of the students at different statistical literacy levels for Q6b with examples from students' answers. In other words, how seventh grade students at different statistical literacy levels define and evaluate the variation concept on bar graphs is explained for Q6b.

As aforementioned, the statistical literacy levels of the students were determined by an overall analysis of students' evaluations of variation in Q6b and their definition of range in Q3 since there was no question related to the definition of variation in Q6 and there was no interpretation question related to the concept of variation in Q6. However, in the analysis, it was noticed that even though some students could not define the concept of range in Q3, they could evaluate the variation in the wind speeds in Q6b, just as in the other variation questions. Therefore, to determine the statistical literacy levels of the students in this question, their definitions of range in Q3 were not taken into consideration; only their evaluations related to the concept of variation in Q6b were taken into account. However, the frequencies of the students who could and could not define the concept of range was determined for each level. As indicated in the Table 4.17 above, almost half of the students (41.4%) performed at Level 1-2 in Q6b. While 37.2% of the students performed at Level 3-4, 21.3% of the students performed at Level 5-6 in Q6b. The following sections explain students' evaluations related to the concept of variation in Q2b. The findings are presented for Level 1-2, Level 3-4 and Level 5-6 students by providing examples from students' answers. Nevertheless, the number of students who could and could not define the concept of range at each level is not presented at the end of each level for Q6b as in Q7 because the frequencies obtained for each level were very similar to the ones obtained in Q2b and Q3. Furthermore, this question was similar to Q2b in Section 4.3.1. Both questions examined the ability to critically evaluate the variation concept in a comparison situation; therefore, the findings are presented in line with the findings of Q2b.

Although most students were expected to be at Level 5-6 in Q6b since it was a question related to the evaluation of the variation concept, as can be observed in Table 4.17, most students (41.4%) performed at Level 1-2 in Q6b. The percentage was much higher when compared to that of students performing at Level 1-2 in Q2b. In the modified version of the framework of Watson and Callingham (2003), it is stated that critically evaluating the variation concept in a comparison situation was not an expected skill for Level 1-2 students. It is stated that while students depend on idiosyncratic beliefs in making comparisons at Level 1, students at Level 2 make their comparisons concentrating on only a single value in the given data sets. Since the characteristics observed at Level 1 and Level 2 are specified clearly in the framework, the levels of students could be determined as Level 1 and Level 2 seperately. More specifically, while 28.0% of the students performed at Level 1 in Q2b, 13.4% of the students performed at Level 2. In the following sections, initially, the evaluations of variation made by Level 1 and Level 2 students in Q6b are presented by providing examples from students' answers.

The findings of the present study showed that 28.0% of the students either could not answer Q2b or provided some idiosyncratic responses. Therefore, these students were placed at Level 1. For example, S104 in Figure 4.80 explained that it is a discrimination to set up wind turbines only in the Amasra and Bandırma regions. This student evaluated the decision somehow but his focus was not on the given data but on his own opinion.

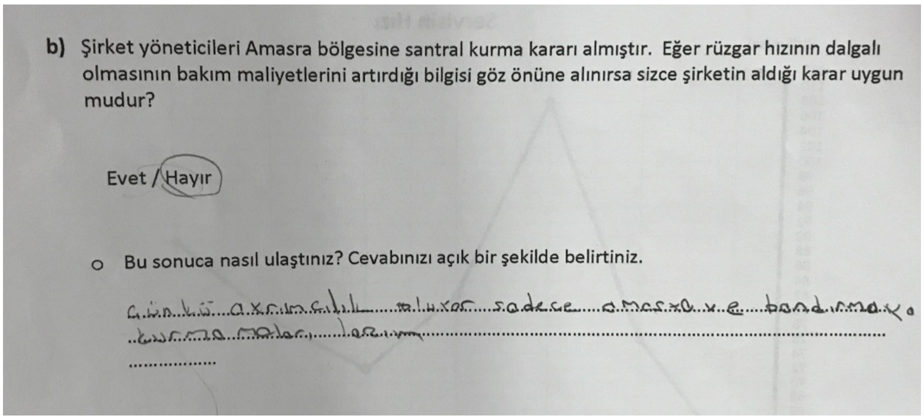


Figure 4.80 Evaluation of variation by a Level 1-2 student in Q6b

Furthermore, it was realized that 13.4% of the students used some single values when comparing the wind speeds of the two regions; therefore, they were placed at Level 2. Figure 4.81 displays the answer of such a student. S141 in the figure below calculated the average of the two regions as being equal by using the mean of the given wind speeds in Q6a. However, when making her evaluation, she stated that although they were equal, the Amasra region had higher wind speeds. It is believed that this student is most probably concentrating on the 37 km/h wind speed that the Amasra region had in the month of August. Similar to Q2b, although S141 calculated the average wind speed of the two regions as being equal, she did not direct her attention to the variation in the wind speeds, but only to the high values.

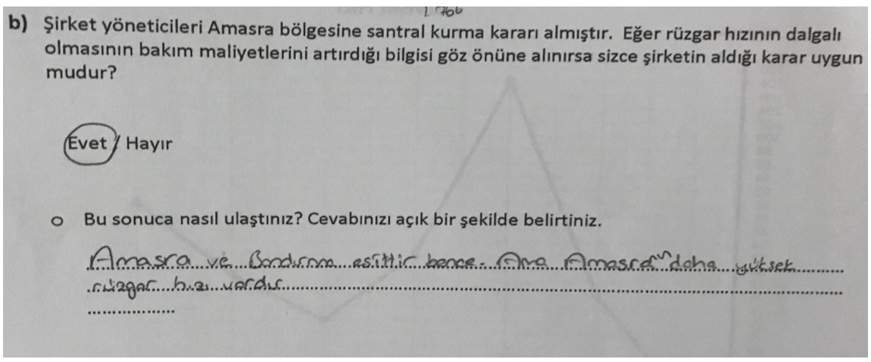
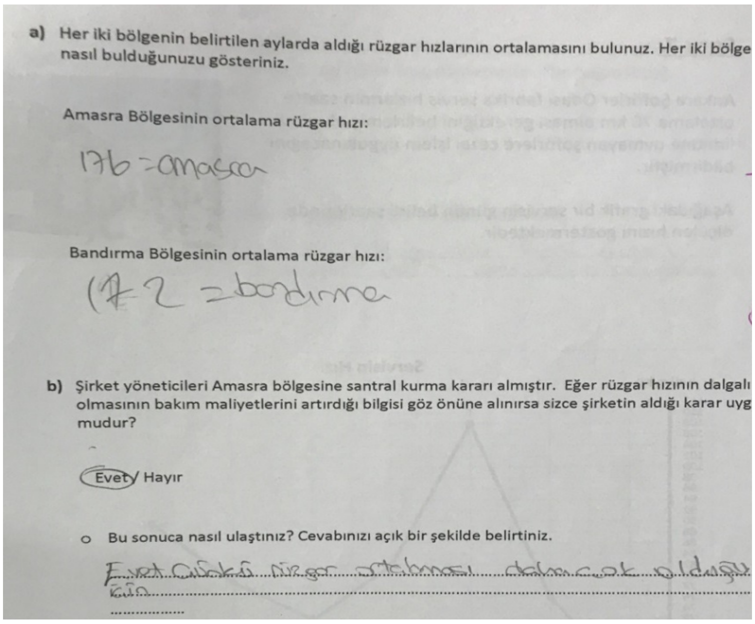


Figure 4.81 Evaluation of variation by a Level 1-2 student in Q6b

On the other hand, the analysis of the students' evaluations of the variation concept in Q6b indicated that 37.2% of the students performed at Level 3-4 in Q6b. As aforementioned, in Q2b, only the evaluations of variation of the students were taken into account to determine their statistical literacy levels since it was observed that although the students could evaluate the variation in the scores of the players, they could not define the range concept in Q3. In the modified version of the framework of Watson and Callingham (2003), it is stated that instead of only depending on single values in their comparisons, students at Level 3 make their comparisons by concentrating on more than one value in the given data sets. However, at Level 4, students use numerical strategies in comparison of two data sets. To state it differently, they make their comparisons by either computing the sum or a measure of central tendency of the two data sets. Since the characteristics observed at Level 3 and Level 4 are specified clearly in the framework, the levels of the students could be determined as Level 3 and Level 4 separately, but it was observed that none of the students performed at Level 3 as in Q2b. More specifically, 37.2% of the students used a numerical strategy to compare the wind speeds of the two regions. It was realized that these students did not consider the variation in the wind speeds of the two regions. While 26.2% of the students used the mean of the wind speeds to evaluate the decision of the company, 10.4% of the students used the sum of the wind speeds of the two regions in making their comparisons. The percentage was lower when compared to that of the students using the mean and the sum in Q2b. There were also a few students calculating the median or mode of the given data sets and used them in their comparisons. In the following sections, the evaluations of variation made by Level 4 students in Q6b are presented by providing examples from students' answers.

As it was stated, 37.2% of the students' levels were determined as Level 4 since they used a numerical strategy while comparing two data sets. To state it differently, to evaluate the decision of the company, most students used the average wind speeds of the two regions that they had computed in Q6a. For example, S140 in Figure 4.82 calculated the sum of the wind speeds of the two regions but found the wind speed of the Amasra region to be greater because of calculation errors. Furthermore, S140 stated that she agrees with the decision of the company since the average of wind speed in the Amasra region is higher.



a) Her iki bölgenin belirtilen aylarda aldığı rüzgar hızlarının ortalamasını bulunuz. Her iki bölge nasıl bulduğunuzu gösteriniz.

Amasra Bölgesinin ortalama rüzgar hızı:

176 = amasra

Bandırma Bölgesinin ortalama rüzgar hızı:

172 = bandırma

b) Şirket yöneticileri Amasra bölgesine santral kurma kararı almıştır. Eğer rüzgar hızının dalgalı olmasının bakım maliyetlerini artırdığı bilgisi göz önüne alınırsa sizce şirketin aldığı karar uyg mudur?

Evet / Hayır

o Bu sonuca nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

Evet çünkü rüzgar ortalaması daha çok olduğu için

Figure 4.82 Interpretation of average by a Level 3-4 student in Q6a and her evaluation of variation in Q6b

Moreover, it was observed that some Level 4 students in Q6b computed the mean scores of the two regions and correctly found them to be equal. However, these students stated that both regions are suitable since the mean of the wind speeds of the two regions are equal or, just as in Q2b, they stated no further idea regarding in which places to set up the wind turbines. Figure 4.83 displays answer of such a student. S96 stated that both of them is appropriate since both are the same.

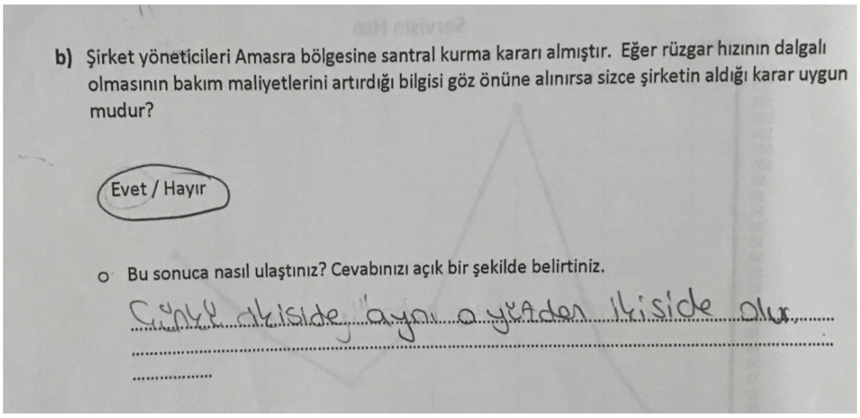
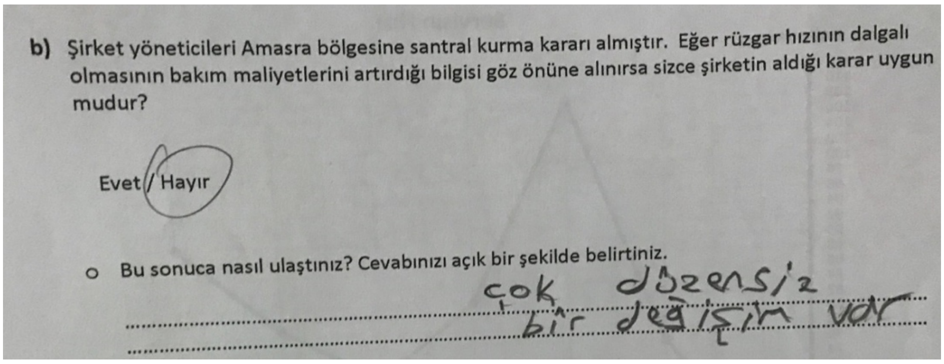


Figure 4.83 Evaluation of variation by a Level 3-4 student in Q6b

Besides, as mentioned previously, about 10.4% of the students whose statistical literacy levels were determined as Level 4 used the sum of the given wind speeds in making their decision. Interestingly, it was realized that although some students calculated the mean wind speed of the two regions in Q6a, which asked for the average wind speeds of the two regions, they used the sum of the wind speeds of the two regions to compare the two regions. To state it differently, they did not use the mean that they had calculated in Q6a. However, the percentage was lower than that of the students calculating the mean in Q2a, but not using it to compare the two players in Q2b.

Lastly, the analysis of the students' evaluations of the variation concept in Q6b indicated that 21.3% of the students performed at Level 5-6. The percentage was higher when compared to that of Level 5-6 students in Q2b. In the framework, it is stated that the ability to critically evaluate the variation concept in a comparison situation was an expected skill of Level 5-6 students. It was stated that while students make their comparisons using visual strategies including the variation concept at Level 5, at Level 6, they use numerical strategies and visual strategies including the variation concept together to make their comparisons. Since the characteristics observed at Level 5 and Level 6 are specified clearly in the framework, the students could be determined to be at Level 5 and Level 6 separately. Particularly, 17.7% of the students performed at Level 5, while 6 students (3.6%) performed at Level 6 in Q6b. In the following sections, the evaluations of variation of students whose statistical literacy levels were determined as Level 5 and Level 6 in Q6b are presented by providing examples from students' answers.

The findings of the present study indicated that 17.7% of the students used only some visual strategies, implying the appreciation of variation in the wind speeds of the two regions in making their decisions related to the decision of the company. Therefore, these students were placed at Level 5. For instance, S72 in Figure 4.84 stated that she did not agree with the decision of the company since the variation in the wind speeds in the Amasra region is so irregular. S72 explicitly mentioned the variation in the wind speeds of the Amasra region. Furthermore, it was observed that even though S72 computed the average wind speed of the two regions in Q6a, she did not use those averages in making her comparison, but only depended on the variation in the wind speeds of the two regions. It was observed that 18 of 29 Level 5 students (11%) did not use the averages that they had calculated in Q6a.


b) Şirket yöneticileri Amasra bölgesine santral kurma kararı almıştır. Eğer rüzgar hızının dalgalı olmasının bakım maliyetlerini artırdığı bilgisi göz önüne alınırsa sizce şirketin aldığı karar uygun mudur?

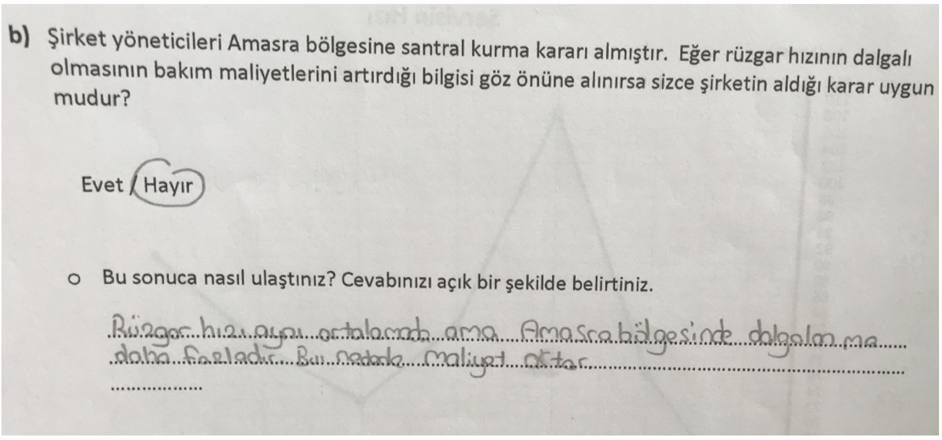
Evet / Hayır

- Bu sonuca nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

çok düzensiz bir değişim var

Figure 4.84 Evaluation of variation by a Level 5-6 student in Q6b

On the other hand, only 6 students (3.6%) used both a numerical strategy and a visual strategy implying an appreciation of variation in the wind speeds of the two regions to make their comparisons in Q6b. According to the modified version of the framework in Table 3.7, the statistical literacy level of these students is Level 6. The percentage was higher when compared to that of Level 6 students in Q2b. For example, S12 in Figure 4.85 calculated the mean of the wind speeds of the two regions in Q6a as being equal, and in the evaluation of the decision of the company in Q6b, S12 stated that the average wind speeds of the two regions were the same, but that the wind fluctuated considerably in the Amasra region. Therefore, the maintenance cost would increase. Even though S12 did not use the exact word of variation as S72 in Figure 4.84 above, it was clear that she had realized the variation in the wind speeds of the Amasra region by addressing the statement in the question, 'fluctuated wind speed increases maintenance cost'.



b) Şirket yöneticileri Amasra bölgesine santral kurma kararı almıştır. Eğer rüzgar hızının dalgalı olmasının bakım maliyetlerini artırdığı bilgisi göz önüne alınırsa sizce şirketin aldığı karar uygun mudur?

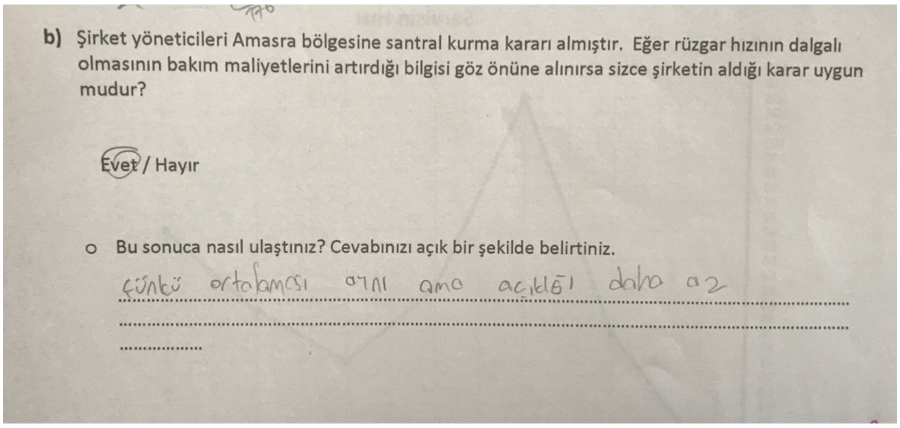
Evet / Hayır

o Bu sonuca nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

Rüzgar hızı aynı ortalamada ama Amasra bölgesinde dalgalanma daha fazladır. Bu nedenle maliyet artar.

Figure 4.85 Evaluation of variation by a Level 5-6 student in Q6b

Furthermore, only 1 student (0.6%) mentioned range, a measure of variation, in addition to the average he had calculated in Q6a to evaluate the decision of the company. As can be examined in Figure 4.86, S68 explained that although the average of the two regions was same, there was a much narrower range in the Amasra region. Although S68 calculated range of the two regions wrongly, he was aware that a narrow range was better for them.



b) Şirket yöneticileri Amasra bölgesine santral kurma kararı almıştır. Eğer rüzgar hızının dalgalı olmasının bakım maliyetlerini artırdığı bilgisi göz önüne alınırsa sizce şirketin aldığı karar uygun mudur?

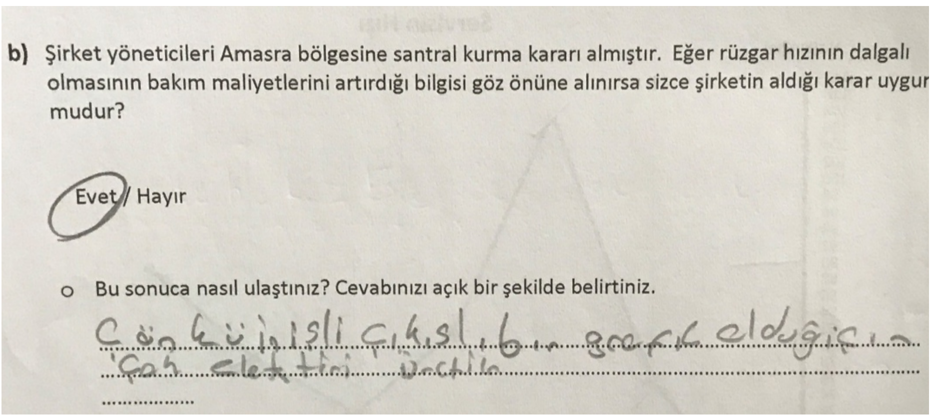
Evet / Hayır

o Bu sonuca nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

çünkü ortalaması aynı ama açıklığı daha az

Figure 4.86 Evaluation of variation by a Level 5-6 student in Q6b

It was also realized that some students whose statistical literacy levels were determined as Level 5 and thus used visual strategies taking into account variation believed that, opposite to the opinion of S68 in Figure 4.86, it was better to have greater variation. For instance, S101 in Figure 4.87 agreed with the decision of the company and he explained that since the graph displayed fluctuations, more electricity would be produced. As can be understood from the answer of the student, even though he mentioned the variation in the wind speed of the Amasra region implicitly, he believed that it would enable the company to produce more electricity.



b) Şirket yöneticileri Amasra bölgesine santral kurma kararı almıştır. Eğer rüzgar hızının dalgalı olmasının bakım maliyetlerini artırdığı bilgisi göz önüne alınırsa sizce şirketin aldığı karar uygun mudur?

Evet / Hayır

o Bu sonuca nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

Çünkü inişli çıkışlı bir grafik olduğu için çok elektrik üretilir

Figure 4.87 Evaluation of variation by a Level 5-6 student in Q6b

Table 4.18 in below summarizes the findings related to the concept of variation on bar graphs. The findings of the present study revealed that in general, the statistical literacy levels of the students in the questions related to the variation concept on bar graphs were determined as Level 1-2, just as it turned out to be for the variation questions on line graphs. However, more students reached the Level 5-6 in Q6b when compared to the percentage of students reaching Level 5-6 in the questions related to the variation concept on line graphs.

Table 4.18 The distribution of students across the six levels in the framework in the questions related to concept of variation on bar graphs

The Concept of Variation on Line Graphs		**Statistical Literacy Level**		
Question	**Skill**	**Level 1-2**	**Level 3-4**	**Level 5-6**
Q6	Evaluation	**68 (41.4%)**	61 (37.2%)	35 (21.3%)

4.5 Summary of the Findings

The purpose of the current study was to analyze seventh grade students' statistical literacy related to the concepts of average and variation on bar and line graphs related to the data obtained from social or scientific contexts. More specifically, initially, this study tried to determine the statistical literacy levels of seventh grade students regarding the concepts of "average" and "variation" on bar and line graphs related to the data obtained from social or scientific contexts. Then, it investigated how seventh grade students at different statistical literacy levels define, interpret and evaluate "average" and "variation" concepts on bar and line graphs related to the data obtained from social or scientific contexts. The findings of the present study revealed that the statistical literacy levels related to the concept of average of most seventh grade students was Level 3-4 for both bar and line graphs. There were not many students reaching Level 5-6 in average questions on both bar and line graphs. In other words, while the students could interpret the concept of average on bar and line graphs related to the data from social or scientific contexts, most of them experienced difficulty in critically evaluating the concept of average on bar and line graphs, particularly on line graphs. Moreover, the *mean* was the most preferred measure of central tendency in interpreting the concept of average on bar and line graphs. On the other hand, it can be concluded that the statistical literacy levels of most of the seventh grade students related to the variation concept were determined as Level 1-2 for both bar and line graphs. Only in one of the questions related to line graphs, most of the students could reach Level 3-4. There were not many students reaching Level 5-6 in variation questions on both bar and line graphs; but the percentage of Level 5-6 students was higher in the questions related to the bar graphs. In other words, most of the seventh grade students experienced difficulty in interpreting and critically evaluating the variation concept on bar and line graphs related to the data obtained from social or scientific contexts. It was found that students generally could not answer the questions related to the variation concept or they presented some idiosyncratic responses.

CHAPTER 5

DISCUSSION, IMPLICATIONS AND RECOMMENDATIONS

The aim of this study was to analyze seventh grade students' statistical literacy in terms of the concepts of "average" and "variation" on bar and line graphs related to the data obtained from social or scientific contexts. With this aim, a qualitative survey research design was implemented to determine the statistical literacy levels of seventh grade students and to analyze how students at different statistical literacy levels define, interpret and evaluate the average and variation concepts on bar and line graphs related to the data obtained from social or scientific contexts. In this chapter, findings of the current study are discussed and compared with previous research studies. First of all, the findings regarding the statistical literacy of students in terms of the concept of average on bar and line graphs related to the data obtained from social or scientific contexts are discussed. Then, the findings regarding the statistical literacy of students in terms of the concept of variation on bar and line graphs related to the data obtained from social or scientific contexts are summarized. Lastly, the implications of the present study and recommendations for further research studies are provided.

5.1 Students' Statistical Literacy in terms of The Concept of Average

One of the aims of the present study was to analyze the statistical literacy of seventh grade students regarding the concept of average. In this respect, the students were asked five different questions related to the concept of average, Q1, Q2, Q4, Q5 and Q6. First of all, the statistical literacy levels of the students were determined by an overall analysis of each question. The findings of the present study revealed that in general, the statistical literacy levels of the students in the questions related to the concept of average on bar and line graphs were Level 3-4. Although Chick and Pierce (2011) stated that schools aim to raise individuals whose statistical literacy level is critical mathematical; i.e., the highest level in the framework, there were not many students reaching this level in all of the questions including bar and line graphs. To state it differently, while students could interpret the average concept on bar and line graphs, most of them could not critically evaluate the average presented to them in almost all of the questions. Furthermore, almost all the students did not mention the representative nature of the concept of average while defining the concept.

This finding was consistent with that reported by several research studies in the literature indicating that more difficulties were observed in the definition and evaluation of the statistical concepts when compared to the interpretation of them among the students at different grade levels (Watson, 2006; Watson & Callingham, 2003; Yolcu, 2012). There might be several reasons underlying such a finding. One of the reasons can be that the Turkish curriculum does not focus on the critical evaluation of statistical concepts as much as it does on their interpretation. In other words, since students becomes familiar with the interpretation of the concepts but not with their critical evaluations in social or scientific contexts, they could have difficulties in the critical evaluation of the concept of average in different social or scientific contexts.

Another reason could be the teachers of the students. In her study with pre-service teachers, Ozen (2013) concluded that pre-service teachers could not evaluate several statistical claims critically. Furthermore, several researchers stated that teachers have a huge role in students' success related to the critical evaluation of the given data (Chick &Pierce, 2011; Watson, 2006). Researchers asserted that if teachers do not have a critical attitude towards the given data, their students will not attain such an attitude either. Thus, teachers' attitudes towards the statistical messages that they have encountered in the real life could be the reason underlying students' inadequacy in critical evaluation of the average concept in the present study.

On the other hand, teachers' inappropriate use of mathematical language in defining the concept of average might be the reason why students experience difficulties in the definition of the concept of average (Yolcu, 2012). Indeed, Yesildere (2010) observed that pre-service teachers do not use mathematical language correctly, and Miller (1993) stated that mathematics teachers generally use daily language in their teaching instead of using the accurate terminology. Hence, the findings in the current study is not surprising if teachers of the students participating in the current study do not use the appropriate language in the teaching of the concept of the average.

In the present study, after the students' statistical literacy levels were determined for each question, their answers to the definition, interpretation and evaluation questions related to the concept of average were analyzed in detail to address the sub question related to the research question of the study. The following sections initially present discussions related to the students' definitions and interpretations related to the concept of average. In the analysis of the answers of the students, the definitions of the students were generally compared with their interpretations related to the concept of average. Hence, in the following sections, the definitions and interpretations of the students are generally discussed together. Moreover, when the students' interpretations of the concept of average on bar and line graphs related to the data obtained from social or scientific contexts were examined in detail, not much difference was observed in students' interpretations of the concept of average on bar and line graphs. Therefore, students' interpretations related to the concept of average are not discussed separately for each question, instead, the findings that are common in all the questions are discussed. Subsequently, since students' evaluations related to the concept of average varied in different questions, each question which required the evaluation of the concept of average was discussed separately following the definition and interpretation of the concept.

When the students' definitions and interpretations were analyzed, it was observed that some students defined and interpreted the concept of average using single ideas such as *normal* or *okay*. However, many students generally defined the concept of average or interpreted its meaning in the given social or scientific contexts through the use of three measures of central tendency and the most used measure of central tendency was the *mean*. This was a consistent finding with those reported by many other research studies in the literature (Brown & Silver, 1989; Enısoglu, 2014; Ucar &Akdogan, 2009; Watson, 2006; Yolcu, 2012). On the other hand, it was observed that the *median* and the *mode*, which are other two measures of central tendency, were used by a few students. This finding could be attributed to the fact that teachers focus solely on the mean algorithm in teaching the concept of average since the *mean* is the most common idea related to the concept of average among teachers as stated by Leavy and O'Loughlin (2006). Yolcu (2012) also expressed the same opinion based on their findings. Furthermore, another reason underlying such a finding could be the similarity between the words *mean* and *average* in the Turkish language (Enisoglu, 2014). In the Turkish language, mean is named as "arithmetic average" and so the first opinion that emerges in students' mind when they hear the word *average* is arithmetic average; i.e, mean. As an alternative point of view, this finding could also be attributed to the curriculum and textbooks in Turkey. In one of the objectives in the Turkish curriculum, it is stated that "Students should be able to find and interpret average, median and mode of a data set" (MoNE, 2018, p. 70). It is also stated that the effectiveness of these three concepts should be discussed in different real life contexts (MoNE, 2018). In other words, the concept of average is used in the curriculum instead of the concept of mean, which indicates the possible reason underlying some findings obtained in the current study. Furthermore, when some textbooks in Turkey were analyzed, it was observed that some of them use the term average when students are requested to find the mean of a data set. Thus, it is not surprising for the students in this study to think of the concept of mean directly when they hear the word average.

Although the concept of mean is preferred much by seventh grade students in defining or interpreting the concept of average, the findings of the current study indicated that students seemed to have a procedural, not a conceptual, understanding regarding the concept of mean. One of those findings was the incorrect application of the mean algorithm, which is a consistent finding with those reported by many studies in the literature (Cai, 2000; Maverach, 1983; Pollatsek et al. 1981; Watson & Moritz, 1999). It was observed in this study that some students used the reversed version of the algorithm; i.e. they divided the number of data by the sum of the given data, or they had difficulty in the computation of the weighted mean, which is a consistent finding with that reported by the study of Watson and Moritz (2000). These findings observed in the present study clearly reflected that conceptual understanding of the students in this study regarding the concept of mean was inadequate since if students had a conceptual understanding related to any concept, they would apply an algorithm related to this concept correctly in every situation they encountered (Kilpatrick, Swafford & Findell, 2001). However, the students in the current study did not apply the mean algorithm in different situations, and they added some numbers and divided the sum by another number, but it was clear that they did not know which numbers they needed to add and divide. This is a clear indication of procedural understanding about the mean concept (Cai, 2000); hence, it can be inferred that the students in this study had a procedural understanding related to the concept of mean.

Another finding of this study showing that students had procedural understanding regarding the mean concept was the inconsistency between students' definitions and interpretations. For instance, some students defined the concept of average and interpreted its meaning in the given social or scientific contexts with the use of single ideas such as *about* or *more or less* but most of these students stated that average in the given contexts was calculated using the mean or sum of the data set. However, students in the present study did not care whether their opinion regarding the calculation of the average concept was consistent with their definition or interpretation of the concept, which is a finding parallel to that reported in the study of Mokros and Russell (1995). There were also students in this study who could define average as the mean concept but could not transfer their definitions to the interpretation of the average concept in the given social or scientific contexts. In other words, they could not use the idea of mean in interpreting the concept of average in different contexts. The underlying reason could be students having only procedural understanding related to the concept of average. To state it differently, they only know the addition and division algorithm but do not know what the obtained value represents (Konold & Higgins, 2003; Mokros & Russell, 1995). Therefore, the students in the present study either could not interpret the concept in a given context using the idea of mean or they used their own understandings to interpret the concept of average.

Different from the students who defined or interpreted the concept of average as the mean of a data set, when students' answers to the interpretation questions were analyzed to be able to answer the sub-question of the current study, it was observed that some students interpreted the concept of average as the maximum value in the given data sets. In other words, these students believed that any data greater than the average value did not exist in the given data. This finding was parallel to that reported in a study of Enisoglu (2014) and also of Hobden (2014), who realized that some teachers interpreted the median value as the maximum value in the given data set. The reason of such a finding could be due to students' and teachers' lack of conceptual understanding related to the concept of average (Cai, 2000; Hobden, 2014). If students in the current study could interpret average as balance point in a data set, they had the information that there were also smaller values than the average in the given data set.

Findings of the present study indicated that there are many students who had procedural understanding regarding the mean and average concept since they do not know what the obtained value represent when they do addition and division or they could interpret the average concept as the maximum value in a data set. Nevertheless, it was observed that some students in this study did not calculate an average value in the questions which required to calculate an average but gave information regarding the place of the average. These students stated that average should be between the minimum and maximum value in the given data. This was a different finding obtained from the studies in the literature. Since these students thought that average cannot be smaller than the minimum value and larger than the maximum value but should be in between them, it was believed that these students have conceptual understanding related to the concept of average. This study contribute to the literature in this sense since the ideas of these students could be used in teaching the place of the average in a data which in turn can help students to interpret average as balance point in a data set. Discussions about the questions such as "When do you think the average will be close to the minimum value or in the middle of the data?" can be useful in reaching this aim.

As previously mentioned, when students' definitions and interpretations related to the concept of average were analyzed, it was found that there is an inconsistency between definitions and interpretations of the students which indicated that students in this study had a procedural understanding regarding the concept of mean. Another conclusion that can be made in relation to this finding is that there is a gap between students' own understanding of the average concept and what they have learnt in school. Kilpatrick et al. (2001) stated that if students do not have a conceptual understanding about the concept, this gap is inevitable. This conclusion was clearly observed in the answer of some students in the current study. For instance, a student expressed two different meanings for the concept of average: one for public language and the other for mathematics. While the student expressed that average means generally in the public language, he mentioned add and divide algorithm for the meaning of the average concept in mathematics. On the other hand, another student in the present study interpreted the average concept as median but stated that average was calculated with the mean algorithm. To state it differently, that student did not regard the median as a calculation related to the concept of average and believed that mean algorithm is essential to find the median of the given data. These findings implied that teaching of the average concept in the schools do not follow the students' own understanding and students only learn procedural knowledge regarding the concept of average and cannot link their own understanding of the average concept with what they have learnt in the school. Also, these findings support the idea of Mokros and Russell (1995) who indicated that students lose their ideas of representativeness related to the average concept when they learn the mean algorithm in the schools. Researchers suggested that lessons should be designed so that students can connect new knowledge with their informal ideas of representativeness and Watson and Moritz (2000) believed that students' ideas of middle and most about average could be a good starting point for classroom discussions.

When students' definitions and interpretations related to the concept of average was examined to be able to answer the sub-question of the present study, it was also observed that some contexts were useful for students in interpretation of the concept of average. Even in the higher levels of statistical literacy, there were students that had difficulties in defining the concept of average. However, it was observed that even though these students could not define the concept, they could interpret its meaning in the given social or scientific contexts or they presented some contextual examples to define the concept. This interesting finding arising from the study confirms the study of Watson and Moritz (2000) in terms of the fact that contexts help students in interpretation of the concept of average. Nevertheless, benefit of the contexts was not observed in the interpretation questions regarding the computation of the average concept. In other words, it was observed that some students, mostly performed at Level 2, could both define the concept of average and interpret its meaning in the given social or scientific contexts through single ideas but they could not answer the question of how the average in the given context was calculated. Parallel with the study of Watson and Moritz (2000), in the present study, students had difficulties in the questions related to the computation of the average concept.

On the other hand, it was observed in the current study that some contexts were not as useful as the others in helping the students to interpret the meaning of the concept of average in the given social or scientific contexts. For instance, a higher number of students could not interpret the meaning of the concept of average in the given brochure or provided some irrelevant responses in Q4. The reason underlying such a finding could be the context of Q4 since Gal (2002) and Watson (2006) stated that interpretation of a statistical concept could not be expected from students if they did not understand the context of the given question. On the other hand, in Q5, where students were provided with a bar graph displaying the preferences of the customers in a hotel, the number of students interpreting the concept of average through the idea of mode was higher when compared to the other interpretation questions. This finding indicates the usefulness of the context of Q5 in interpreting the concept of average as the mode of the given data.

To address the sub question of the present study related to the critical evaluation of the concept of average in different social or scientific contexts, three questions, Q1, Q4 and Q5, were analyzed in detail. A general observation in these three questions was that most of the students could not evaluate the claims related to the concept of average in social or scientific contexts. Moreover, most of them did not display any doubt about the given average and either accepted the given average directly or provided some idiosyncratic responses, which is a finding parallel to that reported in the study of Ozen (2013), who found that even pre-service teachers could not evaluate the popular media texts, but either accepted the given information directly or presented their own opinions.

Another common observation among the three evaluation questions was that there were some students who tried to evaluate the given average but it was observed that they made their evaluations focusing on only some values in the given data sets. To put it differently, the students did not think of the effect of all the data on the given average since they did not have any information regarding the representative nature of the concept of average. In this study, it was observed that only a few students could begin to understand the representative nature of the concept of average, which is a consistent finding with that reported in the other research studies in the literature (Enısoglu, 2014; Mokros & Russell, 1995; Watson & Moritz, 2000; Ucar & Akdogan, 2009; Yolcu, 2012). If students knew that average is a summary statistics representing all data in the given contexts (Mokros & Russell, 1995), they would not concentrate on only some values in the given data set as the students in the current study but on all of the data to make their evaluations.

Beyond these common observations in the evaluation questions, it was realized that the contexts of the questions, that the students were provided, led them to make different evaluations. Hence, each of these questions is discussed separately in the following sections.

Firstly, the students were asked two questions, Q1 and Q4, which included an outlier and so necessitated the use of the median as an appropriate type of average. While data were presented on bar graphs in Q1, data were displayed on a line graph in Q4. The purpose of these questions was to reveal whether the students could critically evaluate that mean is not an appropriate type of average when there is an outlier in the data. It was expected that graphs can be helpful in realizing the outlier in the data; however, the finding obtained in this study was similar to the ones obtained by several researchers (Watson & Callingham, 2003; Watson & Moritz, 1999b). That is, even though the data were presented on bar and line graphs, students in the current study could not realize the outlier in the given data and so could not critically evaluate the average in the given social or scientific contexts. The reason underlying such a result could be that students were not exposed to an instruction where they discussed the effectiveness of the three measures of central tendency since several researchers stated that without explicit instruction about the effectiveness of different measures of central tendency in various contexts, most students could not realize their effectiveness themselves and so they could not reach the higher levels of statistical literacy (Hobden, 2014; Watson, 2006; Watson & Callingham, 2003; Sharma, 2017). Although it is stressed in the Turkish curriculum that the focus of the activities should be the effectiveness of the three measures of central tendency in different situations, the findings of the current study indicated that students might not be engaged in such activities before. Furthermore, it was believed that the interpretation question related to the computation of the concept of average could help students in recognizing the effect of the outlier in the computation of the mean. However, the findings of the present study showed that although there are many students stating that average was calculated through the mean of the given data, they did not feel the necessity of computing the mean in almost all of the questions. Just in Q4, some students found the average number of students using the mean algorithm but this did not enable the students to realize the effect of the outlier in the calculation of the mean. Hence, this finding supports the view of the above researchers once more that which measures of central tendency is effective in different situations should be discussed explicitly with students.

On the other hand, although none of the students could critically evaluate the average presented in Q1 and Q4 mentioned above, it was observed that in Q4, some students gave the signals that they could actually recognize the outlier in the given data sets. In Q4, where the data were presented on line graphs, most of the students evaluated the given average by using only a single value in the data sets; however, that single value which the students focused on was the outlier in the given data sets. To state it differently, some students could actually recognize the outlier in the data; however, they did not consider the effect of this outlier on the given average. Although these students did not show any doubt about the given average and realized the outlier, the idea of these students in Q4 could be a good beginning for the discussions regarding the effect of the outlier on the given average as suggested by many researchers and also observed in the current study (Hobden, 2014; Watson, 2006; Watson & Callingham, 2003; Sharma, 2017).

Q5 was another question in which students' critical evaluations related to the concept of average was examined, but in this case, there were categorical data displayed on bar graphs; hence, the mode of the data was the appropriate measure of average. It was realized that the number of students who could critically evaluate the given average in Q5 was higher when compared to the number of students providing a critical evaluation in Q1 and Q4. The findings of the present study indicated that even though the students did not explicitly say that average should be the mode of the given data, that is, 3-day packages, they did not agree with the mentioned average in the context and explained that the most preferred packages by the customers should be used. Therefore, it can be concluded that the context provided in Q5 was helpful in the evaluation of the concept of the mode, one of the three types of the average concept.

Lastly, one of the interesting findings of the current study was that some of the students could make the above evaluations related to the concept of average but they had difficulties in defining or interpreting the average concept. As can be remembered, there were also some students in the current study who could interpret the average in the given contexts but could not provide a definition. Since they could evaluate the average concept critically in the given social or scientific contexts, their statistical literacy levels were determined as Level 5-6, which is the highest level in the framework of Watson and Callingham (2003). Although it was stated in the framework that the appropriate usage of terminology is observed at Level 5, some of the students could not define the concept. Even though the students could define the concept of average, their answers implied that their definitions were procedural not conceptual. In other words, students interpret average as the concept of mean and they know that some numbers are added and divided by the others; however, they do not know what the obtained value represent. Therefore, they could not use the average concept in making their interpretations or evaluations. Instead, with the help of the context of the questions, they could either interpret or evaluate the average concept. However, in the framework, any characteristics indicating that definitions should be conceptual does not exist. In other words, any definition is accepted in the framework regardless of procedural or conceptual. Hence, it can be concluded that the framework needs some modifications indicating that interpretations and evaluations related to the concept of average should be based on the conceptual understanding of the students in addition to the use of the context.

To summarize, it was observed that some contexts provided to the students were useful for the interpretation and evaluation of the concept of average. However, most of the findings related to the average concept in the present study supported the idea of Shaugnessy (2007), who stated that instead of concentrating on the conceptual understanding of the statistical concepts, schools mostly focus on procedures and computations of these concepts.

In this section, findings related to the average concepts on bar and line graphs were summarized and discussed. The other concept that this study focused on was variation. In other words, the current study analyzed the statistical literacy of seventh grade students in terms of the concept of variation on bar and line graphs related to the data obtained from social or scientific contexts. Hence, in the next section the findings of the current study related to the variation concept on line and bar graphs are discussed in order to address the the two research questions of this study.

5.2 Students' Statistical Literacy in terms of The Concept of Variation

One of the aims of the present study was to analyze the statistical literacy of seventh grade students regarding the concept of variation. In this respect, the students were asked four different questions related to the concept of variation, namely Q2, Q3, Q6 and Q7. First of all, the statistical literacy levels of the students were determined by an overall analysis of each question. The findings of the present study revealed that in general, the statistical literacy levels of the students in the questions related to the concept of variation on bar and line graphs were determined as Level 1-2. Only in one of the questions related to the line graphs, most students could reach Level 3-4. In all of the questions including bar and line graphs, there were not many students reaching the maximum level: Level 5-6. In other words, most of the seventh grade students had difficulty in interpreting and critically evaluating the variation concept on bar and line graphs related to the data obtained from social or scientific contexts. Furthermore, most of the students could not define the concept of range, which is a measure related to variation.

This finding was consistent with those reported in several research studies in the literature, indicating that more difficulties were observed in the definition and evaluation of the statistical concepts when compared to their interpretation among the students at different grade levels (Watson, 2006; Watson & Callingham, 2003; Yolcu, 2012). However, the current study revealed that seventh grade students also had difficulties in interpreting the variation concept on bar and line graphs related to the data obtained from social or scientific contexts. There might be several reasons underlying such a finding. One of the reasons can be the curriculum in Turkey. As stated previously, students can be familiar with the interpretation of statistical concepts since the Turkish curriculum focuses much on the interpretation of statistical concepts, but not on their critical evaluation in different social or scientific contexts. Therefore, difficulties that were observed regarding the evaluation of the variation concept was an expected finding for the current study. However, it was revealed that students also showed difficulties in the interpretation of the concept of variation. This finding indicates that the Turkish curriculum does not focus on the interpretation of the variation concept as much as it does on the average concept. Indeed, the only objective related to the interpretation of the variation concept in the Turkish curriculum is the interpretation of the concept of range, a measure of variation. There is no objective related to the interpretation of the variation concept in a natural context or in a probability and sampling environment. Hence, such a finding related to the interpretation of the variation concept is not surprising for the present study.

In the present study, after the students' statistical literacy levels were determined for each question, the answers of students to the definition, interpretation and evaluation questions related to the concept of variation were analyzed in detail to respond to the sub-question related to the research question of the present study. The following sections provide discussions of those findings related to this analysis for the definitions, interpretations and critical evaluations of the concept of variation.

First of all, when the students' definitions related to the concept of range in the current study were analyzed, it was observed that many students could not define the concept of range regardless of their identified statistical literacy level. This is a consistent result with the findings of Watson and Kelly (2008), which indicates that students at different grade levels cannot define various statistical concepts correctly. However, Yolcu (2012) reached a contradictory finding. In her study, Yolcu (2012) found that students could define the concept of variation by means of the range concept and they could define the range correctly. From this finding, she concluded that students have procedural understanding regarding the concept of variation since the students in the study of Yolcu (2012) described the variation concept with a procedure but did not mention any idea related to the change. However, in the current study, most of the students, even in the higher statistical literacy levels, could not present a definition of the range concept. This finding can be an indication of the fact that students' procedural understanding regarding the concept of variation was inadequate even though one of the objectives in the sixth grade is related to the range concept in the Turkish curriculum. The reason underlying such a finding could be the lack of attention to this objective by the teacher of the students. If teachers only gave the definition of the concept of range in the sixth grade but did not discuss with the students what the value, which is obtained from the difference of the maximum and minimum values, represents or when it is useful to use the concept, then the inadeaquacy of the seventh grade students in the current study in the defining the concept of range is inevitable.

On the other hand, when students' answers to the question related to the interpretation of the variation concept was examined to be able to answer the sub-question, it was realized that the findings did not differ much from the those of the definition question related to the variation concept. To state it differently, most students could not interpret the variation in the weather context in Q3. In this question, students were requested to draw a line graph reflecting the average temperature of Ankara in the month of April but most of the students did not provide any graph. Therefore, it was believed that these students could not interpret the variation in the weather. This was a contradictory finding with that reported in the study of Watson and Kelly (2005), who observed that students' performance was better in graphing variation than other variables that they analyzed in their study, such as explanation of variation. As such, this finding observed in the current study could be attributed to the students' difficulties in drawing line graphs, which was also observed in the study of Capraro et al. (2005). The current study showed that students experienced some difficulties in drawing line graphs because, as can be examined in the graphs of students in Section 4.3.2, there were students who could interpret the variation in the weather but provided either a bar graph or a picture consisting of only points just like a scatter plot. This finding supports the findings of several other research studies in the literature in which it was reported that students confused different graph types (Baker et al., 2002; Boote & Boote, 2017; Memnun, 2013; Whitaker & Jacobbe, 2017). It was also observed that in this study, a few students were careful about the name of the axes or title of their graph. The conclusion that can be made from all of these findings is that it seems that students in the current study did not deal much with drawing their own graphs but generally presented graphs prepared by someone else.

One of the interesting findings obtained from Q3, which was a question related to the interpretation of the variation concept, was that for some students, the temperature of a single day was sufficient to be considered as the average temperature, reflecting the average temperature of the month. On the other hand, for some others, the temperature of all of the days should be 17 °C, which was the given average temperature of the month. That is, these students either only showed the temperature of one of the days as 17 °C or the temperature of all of the days as 17 °C. In either case, their graphs did not reflect any variation in the temperature. One of the reasons of this finding could be that these students either interpreted the average concept as representing only one value in the given data sets, or they interpreted the concept as all of the values in the data should be the same value. The second interpretation was different from the interpretations in the questions related to the average. However, it was believed that this reason is not so possible since in the study of Watson and Kelly (2005), students even at the lowest levels, could either state the temperature of more than one day or answer the question, "Are the temperature of all days as 17 C?" with a negative response. Another possible reason could be that these students did not want to produce a graph but drew such graphs since they were easy for them.

Furthermore, it was realized that in Q3, a higher number of students interpreted the concept of average as the maximum value in the given data set as it was observed in the other average questions since they drew a graph where the maximum temperature was 17 °C. As previously mentioned, the reason of such a finding could be students' lack of conceptual understanding related to the concept of average as also stated by Cai (2000). If students in the current study could interpret average as the balance point in a data set, they had the information that there were also higher values than the average in the given data set.

Obtained from the comparative analysis of the students' definitions of range and their interpretations of the concept of variation, the findings indicated that some students could define the concept of range or could state that range is a concept related to the difference; however, they did not know which numbers needed to be subtracted or whether those numbers were in the data set or not. However, these students could not show the same performance in the interpretation of the concept of variation. Parallel to the study of Yolcu (2012), these findings indicated that these students only had procedural understanding regarding the concept of variation. In other words, these students only knew how to calculate the range of a data set, by means of the difference of some numbers, but did not have any idea related to the what the obtained value represented as several researchers observed for the concept of average (Konold & Higgins 2003; Mokros & Russell, 1995). These students do not know that range is a measure of variation that can be used in making their interpretations. The conclusion that can be obtained from this finding can be that in schools, teachers focus on the computation of the range concept but not on its interpretation although the interpretation of the concept of range was one of the objectives in the Turkish curriculum.

To address the sub-question of the present study related to the critical evaluation of the variation concept in different social or scientific contexts, three questions, Q2, Q6 and Q7, were analyzed in detail. A general observation in these three questions was that a higher number of students could evaluate the variation concept in comparison situations when compared to the number of students who evaluated the concept of average; however, only a few students could evaluate the variation in the question which required the evaluation of the variation concept in a context different from the comparison situation. In other words, parallel to the study of Shaugnessy and Pffankuch (2002), graphs were helpful to discover the variation in the data sets given when there were two data sets, but their benefits were not observed when there was only one data set. The reason of this finding can be attributed to the fact that comparing the values in the two data sets could direct the students to the variations within the data sets. In other words, students might discover the consistency or inconsistency within the data while analyzing the variation between the data sets. Another reason of this finding could be the context of the question, Q7, which required the evaluation of the variation concept in a context different from the comparison situation. The critical evaluation of the variation concept is not expected from the students in the current study if they did not understand the given context as also stated by Gal (2004) and Watson (2006).

Another common observation realized in all of the three questions related to the evaluation of the variation concept was that many students used only a single value or some values in the given data to make their evaluations, which was also observed in the studies of several other researchers (Gal et al., 1989; Shaugnessy, 2003b; Watson & Moritz, 1999). These students did not concentrate on all the data to evaluate the question, just as they did not focus on all the data in the evaluation of the concept of average. However, Watson (2006) stated that using single or some values is the beginning for evaluating the concept of variation. Hence, it can be concluded that students in this study might begin to evaluate the variation in the data critically since they used a single value or some values in making their evaluations.

Beyond these general observations in the evaluation questions related to the variation concept, it was realized that some important findings were obtained in the present study in the questions related to evaluation of the variation concept in the comparison of two data sets. Therefore, those findings will be discussed in the followings sections.

First of all, two questions, Q2 and Q6, required students to evaluate the concept of variation in the comparison of two data sets. While data were displayed on line graphs in Q2, the students were presented the data through bar graphs in Q6. It was realized that the number of students evaluating the variation in the data sets was higher in Q6 than in Q2. One of the reasons of this finding might be the statement that indicated the importance of the variation concept in Q6. In the question, it was stated that the fluctuating wind speed increased the maintenance cost. This statement may have directed students to the given graphs and to realize the variation in the wind speeds. Another reason could be the graph type. Capraro et al. (2005) stated that students are more familiar with bar graphs. Hence, the familiarity of students with bar graphs in the current study could enable them to be better at evaluating the variation concept on bar graphs than on the line graphs. Yet another reason might be an objective found in the sixth grade in the Turkish curriculum. It is stated in the objective that students should be able to use the concepts of mean and range in comparing two graphs, and the graph type used in the comparisons are the bar graphs. Hence, students' familiarity with such tasks might enable them to evaluate the variation in the data sets in Q6. However, it was realized that only one student mentioned the range in the wind speeds in Q6. The remaining students expressed the variation in the data qualitatively just like in the description hierarchy of Reading (2004). In other words, students in the present study did not use the concept of range when making their evaluations of the variation concept just as they did not use it in interpretation of the variation. This could be because students only know how to calculate the range of a data set through the difference of some numbers, but do not have any idea related to the what the obtained value represents.

Although a higher number of students could evaluate the variation in the given data sets in Q6, in other words, there were more students whose statistical literacy level was determined as Level 5-6, most of the students performed at Level 1-2, which are the first two levels in the framework of Watson and Callingham (2003) in Q6. However, in Q2, in which students' were requested to evaluate the variation in the scores of the two basketball players, most students' statistical levels were determined as Level 3-4. Parallel to the thoughts of Mc Gatha et al. (2002), this finding could be because of the role of the basketball context in Q2, i.e. students' familiarity with the basketball game might result in fewer number of idiosyncratic responses in Q2 and, in turn, few number of students at Level 1-2.

On the other hand, evaluation questions related to the concept of variation in comparison of two groups also provided some findings for the concept of average. Different from the studies related to the comparison of two groups in the literature, in the present study, students were initially asked to calculate the average of the two data sets since it was shown in the literature that many students do not use the average while comparing two data sets (Gal et al., 1989; Shaugnessy, 2003b; Watson & Moritz, 1999). However, it was observed that even though the average was asked, students did not use it to make their comparisons; most of them concentrated on single values in the data. This was a consistent finding with those reported in many studies in the literature (Gal et al., 1989; Konold & Pollatsek, 2002; Shaugnessy, 2003b; Watson & Moritz, 1999). One of the reasons of this finding could be that students do not feel the need for using the average to compare the two data sets as also stated by Gal (2005). Furthermore, parallel to the opinions of Konold and Pollatsek (2002), this finding could be attributed to the fact that students do not know the representative nature of the average concept. If students in the present study had the information that average is a summary statistics representing all the data, they would use it in comparing two groups.

Differently, although fewer in number, there were some students who used the average they had calculated in making their comparisons. They calculated either the sum or mean of the given values as average and made their comparisons using those averages. However, it was observed that even a 0.1 point difference in the mean scores of the data sets or 1 point difference in the sum of the scores was very significant for these students; therefore, they did not direct their attention to the variation in the data sets. Even though they found the average of the data sets to be equal, most of them did not evaluate the variation in the data sets to make their comparisons. Thus, it can be concluded that how much difference is necessary to regard the found difference as significant and what can be done if the difference in the average scores is not much should be explicitly discussed with the students. These discussions can help students to direct their attention to the evaluation of variation in the data sets in addition to the evaluation of the concept of average.

Lastly, a similar finding with the average questions was also observed in the questions related to the evaluation of the variation concept. The findings of the present study indicated that some of the students could make the above evaluations related to the concept of variation but they had difficulties in defining or interpreting the variation concept just like the students who could interpret the variation in the given contexts but could not provide a definition. Since they could evaluate the variation concept critically in the given social or scientific contexts, their statistical literacy levels were determined as Level 5-6, which is the highest level in the framework of Watson and Callingham (2003). Although it was stated in the framework that the appropriate usage of terminology is observed at Level 5, most of the students at either Level 5 or Level 6 could not define the concept of range. Even if students could define the range, their answers implied that their definitions are procedural not conceptual. In other words, students know that range is the difference between the maximum and minimum value in a data set; however, they do not know what the obtained value represent. Therefore, they could not use the range concept in making their interpretations or evaluations. Instead, with the help of the context of the questions and graphs, they could either interpret or evaluate the variation concept. However, in the framework, any characteristic indicating that definitions should be conceptual does not exist. In other words, any definition is accepted in the framework regardless of whether it is procedural or conceptual. Hence, it can be concluded that the framework needs some modifications indicating that interpretations and evaluations related to the concept of variation should be based on the conceptual understanding of the students in addition to the use of the context.

To summarize, it was observed that the comparison of two groups were useful for students to evaluate the concept of variation critically in social or scientific contexts. However, findings related to the variation concept in the present study supported the idea of Shaughnessy (2007) once more: Schools mostly focus on procedures and computations of statistical concepts instead of concentrating on students' conceptual understanding of them.

In the next section, some implications for stakeholders are presented based on the findings of the current study.

5.3 Implications

In the current study, the statistical literacy of seventh grade students in terms of the concepts of "average" and "variation" on bar and line graphs related to the data obtained from social or scientific contexts were examined. In this section, some possible implications for teachers, textbook writers, teacher educators and curriculum developers are presented based on the findings of this study.

Statistical literacy is an ability that schools are expected to equip students with; however, this study revealed that there were not many seventh grade students reaching the higher levels of statistical literacy in the framework of Watson and Callingam (2003) regarding the concepts of average and variation on bar and line graphs related to the data obtained from social or scientific contexts. More specifically, it was observed that students had difficulties in defining, interpreting and evaluating the concepts of average and variation. Actually, there were students who could present a definition for both of the two concepts but their definitions were procedural, not conceptual. In other words, they could provide the definitions of mean or range; however, they did not know what these measures represent related to the data presented to them. When such was the case, they could not use their definitions in the process of interpreting and evaluating the average or variation concepts in different social or scientific contexts. This could be because teachers and textbooks merely give the definitions of the concepts and focus solely on the calculations by giving data sets without any context. Hence, teachers could focus more on the context and can use and develop tasks like those given in the current study to enhance the interpretation and evaluation of the average and variation concepts in different contexts rather than giving students the definitions of the average and variation concepts directly. Similarly, textbook writers should not give the definition of the average and variation concepts directly but include questions with different contexts where students need to make interpretations and evaluations of the average and variation concepts. In this way, students' understanding of definitions of the related concepts will also be improved since as Watson (2006) stated, students' understanding of definitions develop as they apply the definitions in various contexts. In other words, teachers and textbook writers could use contextual questions to teach definition, interpretation and evaluation of the average and variation concept together but not separately.

The findings of the present study related to the concept of average revealed that a higher number of students could evaluate the average concept critically when the mode is the appropriate type of average; that is, when there is a categorical data. Hence, teachers can start by making use of the contexts including categorical data for critical evaluation of the average concept. In this way, students begin to attain a critical attitude to the statistical messages in the given contexts. They can start to learn that statistical messages presented to them are not always correct but sometimes could mislead them (Gal, 2004). Then, teachers can use contexts where the median is the appropriate type of average in which students have more difficulties in making critical evaluations, which was shown by several studies including the current study (Watson & Callingham, 2003; Watson & Moritz, 1999b).

On the other hand, findings related to the variation concept indicated that more students could critically evaluate the variation in comparing two groups when compared to the number of students who critically evaluated the concept of average in different social or scientific contexts. In the question that did not include a comparison situation, most of the students used a single value or some values, which implies that students in the present study begin to evaluate the concept of variation critically as stated by Watson (2006). To state it differently, it was found that many seventh grade students began to realize variation in the given data sets especially in the contexts which are familiar to them such as basketball. Therefore, teachers could begin their instructions by using the tasks which involve familiar contexts to the students and which attract students' attention to the variation in the given data sets. After students feel the need to calculate a summary statistics in the enviroment where varying data values exist, which was stated by means of a context of noisy enviroment by Konold and Pollatsek (2002), teachers could direct students' attention to the concept of average. In other words, the teaching of the mean, mode and median concepts can might be delayed until the concept of variation is understood completely by students as also suggested by Bakker and Frederickson (2005).

English (2013) stated that statistical literacy is not an ability that is attained quickly, but needs a very long time to develop; teaching to raise statistical literate individuals should start from the early years of schooling. In line with the findings of the current study and as suggested by English (2013), curricula can be modified to raise statistically literate individuals. As previously stated, not many objectives exist regarding the critical evaluation of statistical concepts in the Turkish curriculum, but rather, interpretation of statistical concepts has a significant place in the curriculum. So as to make critical evaluation of the statistical concept in the classrooms more explicit, some objectives can be added to the curriculum, such that students should be able to critically evaluate the variation in the data sets which are provided to the students in different real life contexts. Furthermore, only two objectives related to the concept of variation exist in the middle school mathematics curriculum in Turkey and they were related to the concept of range. However, Watson et al. (2009) observed that with the use of appropriate tasks, even first grade students could realize the variation in the data sets, which is the reason underlying the calculation of an average (Shaugnessy, 2007). Furthermore, Watson and Moritz (1999) stated that comparison activities were very motivating for students and these kinds of activities should begin from the third grade. If students become engaged in activities where they can appreciate the variation beginning from their early years in the school, they might feel that a summary statistics, an average, is necessary when they get to middle school.

On the other hand, Watson et al. (2014) expressed that curriculum documents play a significant role in students' ability to interpret the average concept as the concept of mean since average is still understood as the concept of mean in many curricula. The findings reported by Watson et al. (2014) were clearly observed in the Turkish curriculum since the objective 7.4.1.2 states that "Students should be able to find and interpret the average, median and mode of a data set." It was also asserted that the effectiveness of these three concepts should be discussed in different real life contexts (MoNE, 2018, p.70). To put it differently, the curriculum did not state the concepts of mean, median and mode as the three average types but used the concept of average instead of the concept of mean. This might be the possible reason underlying some findings obtained in the current study. Hence, this objective can be modified so that anyone reading the objective should be able to understand clearly that mean, median and mode are the three different types of the concept of average.

Lastly, the findings of the present study provide some implications for teacher educators. With the SLT prepared, pre-service teachers' statistical literacy related to the concepts of average and variation can be examined, and similar tasks in the test can be used in methodology courses. Hence, pre-service teachers attain familiarity to such tasks and then they might use similar tasks in their classrooms when they become teachers to raise statistically literate students.

In this section, some implications of the present study for teachers, teacher educators and curriculum developers were presented. The following section presents some recommendations for further research studies.

5.4 Recommendations for Further Research Studies

The present study focused on the seventh grade students' statistical literacy in terms of the concepts of "average" and "variation" on bar and line graphs related to the data obtained from social or scientific contexts. Some recommendations for further research studies can be suggested grounded on the findings of this study.

To begin with, the participants of the current study were selected by means of the convenience sampling method, which included 164 7th grade students in two different public schools in Ankara. Similar studies, in which the random sampling method is used, could be conducted so as to generalize findings of the study to the other seventh grade students. Furthermore, it can be useful to conduct longitudinal studies to examine whether or not the statistical literacy of students in terms of the concepts of average and variation on bar and line graphs related to the data obtained from social or scientific contexts develop over time. Moreover, private school students' statistical literacy related to the concepts of average and variation can be examined, which can reveal the role of different school types on statistical literacy levels of the students.

On the other hand, the findings of the present study were limited with the questions in SLT. More research could be conducted with some different tests or SLT can be developed further with the inclusion of different contexts or different graph types. Moreover, the questions related to the interpretation of the variation concept were limited in number and many students could not provide a graph. Hence, it can be advised that SLT can be developed further by inclusion of the interpretation questions related to the concept of variation as in the study of Watson and Kelly (2005). In addition, asking for the definition of the concept of variation directly instead of the definition of range can be suggested. Findings of such a study can reveal whether or not students have the information that variation is a concept related to the change in the data sets. It can also show students' own understandings, which in turn, can help the teaching of the variation concept.

Besides, interviews can be useful to make a more in-depth analysis of the thinking processes of the students. Through interviews, how students interpret and evaluate the concepts of average and variation can be examined in more detail and the findings of such a study can reveal the possible factors in students' inadequacy in the interpretation and evaluation of the average and variation concepts. Moreover, in the current study, it was decided that many students could not appreciate variation to a given average since they did not provide any graph. In the interviews, students might be requested to try to draw a graph giving more and then students' interpretation of variation could be examined. Furthermore, in the interviews, students might be requested to compare their definitions with their interpretations which in turn could provide students to realize the inconsistency between their definitions and interpretations.

Finally, some recommendations can be made related to the framework used for data analysis in the current study. The findings of the present study revealed that students do not make their interpretations or evaluations using their conceptual understanding related to definitions of the average and variation concepts. To put it differently, students just know how to calculate the mean and range of a data set but they do not know why they calculate these values. Therefore, they could not use the mean or range concept in making their interpretations or evaluations. Instead, with the help of the context of the questions, they could either interpret or evaluate the average and variation concepts. However, in the framework, any characteristic indicating that definitions should be conceptual does not exist. In other words, any definition is accepted in the framework regardless of whether it is procedural or conceptual. Hence, it can be concluded that the framework needs some modifications indicating that interpretations and evaluations related to the concepts of average and variation should be based on the conceptual understanding of the students related to the definitions of the average and variation concepts in addition to the use of the context.

REFERENCES

Aoyama, K., & Stephens, M. (2003). Graph Interpretation Aspect of Statistical Literacy: A Japanese Perspective. *Mathematics Education Research Journal*, *15*(3), 207-225.

Australian Education Council. (1991). A national statement on mathematics for Australian schools. Carlton, Vic.: Author.

Bakkar, A., & Gravemeijer, K. P. E. (2004). Learning to reason about distribution. In J. Garfield & D. Ben Zvi (Eds.), The challenge of developing statistical literacy, reasoning and thinking (pp.147-168). Dordrecht, The Netherlands: Kluwer.

Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2001). Toward a model of learning data representations. *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society* (pp. 45-50). Mahwah, NJ: Erlbaum.

Ben-Zvi, D. (2004). Reasoning about variability in comparing distributions, *Statistics* Education Research Journal, 3(2), 42-63.

Ben-Zvi, D., & Garfield, J. (2004). Statistical Literacy, Reasoning and Thinking: Goals, Definitions and Challenges. in D. Ben-Zvi, & J. Garfield, The Challenge of Developing Statistical Literacy, Reasoning and Thinking (pp. 3-15). Dordrecht: Kluwer Academic Publishers.

Biehler, R. (1997). Students' difficulties in practicing computer-supported data analysis: Some hypothetical generalizations from results of two exploratory studies. In J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics: Proceedings of the 1996 IASE*

Round Table Conference (pp. 169-190). Voorburg, The Netherlands: International Statistical Institute.

Boote, S. K. (2012). Assessing and understanding line graph interpretations using a scoring rubric of organized cited factors. *Journal of Science Teacher Education, 25*(3), 333–354. doi: 10.1007/s10972-012-9318-8

Boote, S. K., & Boote, D. N. (2017). Leaping from discrete to continuous independent variables: Sixth graders' science line graph interpretations. *The Elementary School Journal, 117*(3), 455–484.

Brasell, H. M. & Rowe, M. B. (1993). Graphing skills among high school students. *School Science and Mathematics, 93*(2), p. 63-70.

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.

Carmichael, C., & Hay, I. (2009). Gender differences in middle school students' interests in a statistical literacy context. In R. Hunter, B. Bicknell, & T. Burgess (Eds.), Crossing divides: *Proceedings of the 32nd annual conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 97–104). Palmerston North: MERGA.

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. (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.

Chick, H. L. & Pierce, R. (2011). Teaching for statistical literacy: Utilising affordances in real-world data. *International Journal of Science and Mathematics Education, 10*(2), 339-362. doi: 10.1007/s10763-011-9303-2

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *American Mathematical Monthly, 104*(9), 801-823.

Curcio, F.R. (1987). Comprehension of mathematical relationships experienced in graphs. *Journal for Research in Mathematics Education, 18,* 382-393.

English, L. (2013). Promoting statistical literacy through data modelling in the early school years. In E. J. Chernoff & B. S. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 441-457). New York, NY: Springer.

Enisoglu, D. (2014). *Seventh grade students' possible solution strategies, errors and misinterpretations regarding the concepts of mean, median and mode given in bar graph representations.* Unpublished masters' thesis, Middle East Technical University, Ankara.

Erbilgin, E., Arikan, S. & Yabanli, H. (2015). Çizgi grafiğini yorumlama ve oluşturma becerilerinin ölçülmesi. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi (KEFAD), 16*(2), 43-61.

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

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2005). *Guidelines for assessment and instruction in statistics education (GAISE) report: A Pre-K–12 curriculum framework*. Alexandria, VA: American Statistical Association.

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

Friel, S. N. & House, P. A. (Eds.). (2003). Navigating Through Data Analysis in Grades 6 – 8. Reston, VA: NCTM.

Halpern, D. F. (1997). *Critical thinking across the curriculum*. Mahwah, NJ: Lawrence Erlbaum.

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

Gal, I. (2002). Adult's statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, *70*(1), 1-25.

Gal, I. (2004). Statistical Literacy: Meanings, Components, Responsibilities. In D. Ben-Zvi, & J. Garfield, The challenge of developing statistical literacy, reasoning and thinking (pp. 47-78). Dordrecht: Kluwer Academic Publishers.

Garfield, J. B., Ben-Zvi, D., Chance, B., Medina, E., Roseth, C., & Zieffler, A. (2008). Learning to reason about center. In J. B. Garfield and D. Ben-Zvi (Eds.), *Developing students' statistical reasoning: Connecting research and teaching practice* (pp. 187–200). Dordrecht, The Netherlands: Springer.

Garfield, J., delMas, R., & Chance, B. (2003). *Web-based assessment resource tools for improving statistical thinking.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Gravetter, F. J., & Wallnau, L. B. (2016). *Statistics for the behavioral sciences* (9th ed.). Belmont, CA: Cengage Learning.

Gündüz, N. (2014). *İlköğretim matematik öğretmen adaylarının istatistiksel okuryazarlıkları ile istatistiğe yönelik tutumları arasındaki ilişkinin incelenmesi* (Unpublished master's thesis). Kocaeli University, Kocaeli.

Hardiman, P. T., Well, A. D., & Pollatsek, A. (1984). Usefulness of a balance model in understanding the mean. Journal of Educational Psychology, 76(5), 792–801.

Hammerman, J. K., Rubin, A. (2003). Reasoning in the presence of variability. In C. Lee (Ed.), *Proceedings of the Third International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-3 CD ROM).* Mt. Pleasant: Central Michigan University.

Hobden, S. (2014). When statistical literacy really matters: Understanding published Information about the HIV/AIDS epidemic in South Africa. *Statistics* Education Research Journal, 13(2), 72–82.

İş, Ç. (2003). A cross-cultural comparison of factors affecting mathematical literacy *of students in programme for internatıonal student assessment (PISA)* (Unpublished master's thesis). Middle East Technical University, Ankara.

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.

Jansen, H. (2010, May 11). *The logic of qualitative survey research and its position in the field of social research methods*. Retrieved from http://www.qualitative-research.net/index.php/fqs/article/view/1450/2946#g2

Kazak, S. (2016). İstatistiksel akıl yürütme gelişimi üzerine teorik çerçeveler. In E. Bingölbali, S. Arslan, İ.Ö. Zembat, (Eds.), *Matematik Eğitiminde Teoriler*, (201-213). Ankara: Pegem.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). Adding it up: Helping children lea rn mathematics. Washington, DC: National Academy Press.

Kirsch, I. & Mosenthal, P. (1990) Understanding the news. *Reading research* quarterly, 22(2), 83-99.

Konold, C., Pollatsek, A., Well, A., & Gagnon, A. (1997). Students analyzing data: Research of crit- ical barriers. In J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics: Proceedings of the 1996 IASE Round Table Conference* (pp. 151-167). Voorburg, The Netherlands: International Statistical Institute.

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.

Koparan, T. (2012). *Proje tabanlı öğrenme yaklaşımının öğrencilerin istatistiksel okuryazarlık seviyelerine ve istatistiğe yönelik tutumlarına etkisi.* (Unpublished dissertation), Karadeniz Technical University, Trabzon.

Koparan, T. & Güven, B. (2015). The effect of project-based learning on students' statistical literacy levels for data representation. *International Journal of Mathematical Education in Science and Technology,* 46(5), 658-686.

Leavy, A., & O'Loughlin, N. (2006). Preservice teachers' understanding of the mean: moving beyond the arithmetic average. Journal of Mathematics Teacher Education, 9, 53–90.

Lewis, F. W. (2016). *Middle School Students' Understanding of Variability,* (Doctoral dissertation). ProQuest LLC.

Luke, A., and Freebody, P., 1997, Shaping the social practices of reading. In Constructing Critical Literacies: Teaching and Learning Textual Practice, edited by S. Musprati, A. Luke and P. Freebody (St. Leonards, Australia: Allen and Unwin), pp. 185–225.

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

McKean, E. (Ed.). (2006). Concise oxford American dictionary. New York: Oxford University Press.

Makar, K. (2014). Young children's explorations of average through informal inferential reasoning. *Educational Studies in Mathematics, 86*(1), 61-78.

Makar, K., & Confrey, J. (2005). "Variation Talk": Articulating meaning in statistics. *Statistical Education Research Journal, 4* (1), 27-54.

Memnun, D. (2013). Ortaokul yedinci sınıf öğrencilerinin çizgi grafik okuma ve çizme becerilerinin incelenmesi. *International Periodical For The Languages, Literature and History of Turkish or Turkic,* 8*(12),* 1153-1167.

Merriam, S. B. (2009). Qualitative research: A guide to design and implementation. San Francisco, CA: Jossey-Bass.

Metcalfe, J., Simpson, D., Todd, I. & Toyn, M. (2013). *Thinking Through New Literacies for Primary and Early Years.* London: Sage.

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

Miller, D. L. (1993). Making the connection with language. *Arithmetic Teacher*, *40*(6), 311-317.

Ministry of National Education (MNE). (2018). Ortaokul Matematik Dersi Öğretim Programı Ankara: Milli Eğitim Basımevi.

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

Monteiro, C., Ainley, J. (2015). Investigating the interpretation of media graphs among student teachers. *International Electronic Journal of Mathematics Education, 2*(3), p. 187-207.

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

Nisbet, S. (2001). Representing categorical and numerical data. In J.Bobis M.Mitchelmore, B.Perry (Eds.), *Proceedings of the 24th Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 378-385). Sydney: MERGA.

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.

OECD (2001), Knowledge and Skills for Life-First Results from PISA 2000, Paris.

Özen, M. (2013). *Investigatıon of pre-service mathematics teachers' critical thinking processes through statistical and probabilistic knowledge in the context of popular media texts.* (Unpublished masters' thesis), Middle East Technical University, Ankara.

Özgün-Koca, S. A. (2001). *The Graphing Skills of Students in Mathematics and* Science Education. (ERIC Document Reproduction Service No. ED 464 804).

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

Reading, C. (2004). Student description of variation while working with weather data. *Statistics Education Research Journal, 3*(2), 84-105.

Reading, C., & Shaughnessy, J. M. (2000). Student perceptions of variation in a sampling situation. In T. Nakahar & M. Koyama (Eds.), *Proceedings of the 24th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 89–96). Hiroshima: Hiroshima University.

Reading, C., & Shaughnessy, M. (2004). Reasoning about variation. In D. Ben-Zvi & J. Garfield (Eds.) *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 201–226). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Rumsey, D. J. (2002) Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*, *10*(3). Retrieved from www.amstat.org/publications/jse/v10n3/rumsey2.html

Sharma, S. (2017). Definitions and models of statistical literacy: a literature review. *Open Review of Educational Research*, *4*(1), 118-133.

Sharma, S., Doyle, P., Shandil, V., & Talakia'atu, S. (2011). Developing statistical literacy with year 9students. Set: Research Information for Educational Research, 1, 43–60.

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

Shaugnessy, J.M. (2003b). *The development of secondary students' conceptions of variability.* (Annual Report year 1. NSF Grant No. REC0207842. Portland, OR: Portland State University.

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957–1010). Reston, VA: NCTM.

Shaugnessy, J. M. & Ciancetta, J. (2002). Students' understanding of variability in a probability environment. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.215.5833&rep=rep1&type=pdf

Shaugnessy, J. M. & Pfannkuch, M. (2002). How Faithful is Old Faithful? Statistica Thinking: A story of variation and prediction. *The Mathematics Teacher,* 95, 252-259.

Shaugnessy, J. M., Watson, J. M., Moritz, J. B., Reading, C. (1999, April.) *School mathematics students' acknowledgement of statistical variation: There is more life than centers.* Paper presented at the Research Pre-session of the 77th Annual meeting of the National Council of Teachers of Mathematics, San Francisco, CA.

Shah, P., Freedman, E., & Vekiri, I. (2005). The comprehension of quantitative information in graphical displays. In P. Shah & A. Miyake (Eds.), The Cambridge handbook of visuospatial thinking (pp. 426–476). New York: Cambridge University Press.

Shaughnessy, J. M., & Zawojewski, J. S. (1999). Secondary students' performance on data and chance in the 1996 NAEP. The Mathematics Teacher, 92, 713–718.

Scheaffer, R. L. (2006). Statistics and mathematics: On making a happy marriage. In G. F. Burrill & P. C. Elliott (Eds.), *Thinking and reasoning about data and* chance: Sixty-eighth yearbook (pp. 309–322). Reston, VA: NCTM.

Schield, M. (1999). Statistical literacy: Thinking critically about statistics. *Of Significance Journal. The Association of Public Data Users*, *1*(1). Retrieved from, www.statlit.org/pdf/1999SchieldAPDU.pdf

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

Tortop, T. (2011). *7th-grade students' typical errors and possible misconceptions in graphs concept before and after the regular mathematics instruction* (Unpublished master's thesis). Middle East Technical University, Ankara.

Townsend, M. (2006). Developing statistical literacy in youth: statistics Canada's education outreach program. Retrieved from www.statlit.org/PDF/2006TownsendStatLitInYouth.pdf

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., Karp, K. S., & Bay-Williams J. M. (2013). *Elementary and middle school mathematics: Teaching developmentally* (8th Ed.). Boston, MA: Pearson Education, Inc.

Wallman, K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, *88*(421), 1–8.

Watson, J. (1997). Assessing statistical thinking using the media. In I. Gal and J. Garfield (Eds.), *The Assessment Challenge in Statistics Education*, (pp. 107-122). Amsterdam: IOS Press and the International Statistical Institute.

Watson, J. M., & Kelly, B. A. (2005). The winds are variable: student intuitions about variation. School Science and Mathematics, 105, 252–269.

Watson, J. M., & Kelly, B. A. (2008). Sample, random and variation: The vocabulary of statistical literacy. *International Journal of Science and Mathematics Education*, *6*(4), 741-767.

Watson, J. M., & Moritz, J. B. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics, 37*, 145–168.

Watson, J.M., & Moritz, J. B. (2000). The longitudinal development of understanding of average. *Mathematical Thinking and Learning, 2*(1-2), 11-50.

Watson, J. M., & Moritz, J. B. (2000b). Development of understanding of sampling for statistical literacy. Journal of Mathematical Behavior, 19(1), 109-136.

Watson, J. M., Kelly, B. A., Callingham, R. A. & Shaugnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology, 34*(1), 1-29.

Watson, J. M. (2006). *Statistical literacy at school: Growth and goals.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Watson, J. M. & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal, 2*(2), 3-46.

Watson, J. M., Skalicky, J., Fitzallen, N. & Wright, S. (2009). Production and manufacturing: All-sorts of practical applications for statistics. *Australian Primary Mathematics Classroom, 3*(14), 4-13.

Watson, J. M., Chick, H. & Callingham, R. (2014). Average: the juxtaposition of procedure and context. *Mathematics Education Research Journal, 26*(3), 477–502. doi: 10.1007/s13394-013-0113-4

Watson, J. M., & English, L.D. (2015). Statistical Literacy in the Elementary School: Opportunities for Problem Posing**.** In F. M. Singer, N.F. Ellerton, J. Cai (Eds.), *Mathematical Problem Posing*, pp. 241-256.

Whitaker, D., & Jacobbe, T. (2017). Students' understanding of bar graphs and histograms: Results from the LOCUS assessments. Journal of Statistics Education, 15(2), 90-102. doi: 10.1080/10691898.2017.1321974

Yeşildere, S. (2010). İlköğretim matematik öğretmen adaylarının matematiksel alan dilini kullanma yeterlikleri. *Boğaziçi Üniversitesi Eğitim Dergisi, 24*(2), 61-70.

Yolcu, A. (2012). An investigation of eighth grade students' statistical literacy, attitudes towards statistics and their relationship. Unpublished master's thesis, Middle East Technical University, Ankara.

Yolcu, A. (2014). Middle School Students' Statistical Literacy: Role of Grade Level and Gender. *Statistics Education Research Journal, 13*(2), 118-131.

APPENDICES

A. METU HUMAN SUBJECTS ETHICS COMMITTEE APPROVAL

UYGULAMALI ETİK ARAŞTIRMA MERKEZİ
APPLIED ETHICS RESEARCH CENTER

ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

DUMLUPINAR BULVARI 06800
ÇANKAYA ANKARA/TURKEY
T: +90 312 210 22 91
F: +90 312 210 79 59
ueam@metu.edu.tr
www.ueam.metu.edu.tr

Sayı: 28620816 / 298

11 MAYIS 2018

Konu: Değerlendirme Sonucu

Gönderen: ODTÜ İnsan Araştırmaları Etik Kurulu (İAEK)

İlgi: İnsan Araştırmaları Etik Kurulu Başvurusu

Sayın Prof. Dr. Mine Işıksal BOSTAN

Danışmanlığını yaptığınız Emine Çatman AKSOY' un **"Yedinci Sınıf Öğrencilerinin Sütun ve Çizgi Grafiğinde Verilen Ortalama ve Yayılım Kavramları Hakkında İstatiksel Okuryazarlık Seviyelerinin İncelenmesi"** başlıklı araştırması İnsan Araştırmaları Etik Kurulu tarafından uygun görülerek gerekli onay **2018-EGT-078** protokol numarası ile **11.05.2018 - 30.12.2018** tarihleri arasında geçerli olmak üzere verilmiştir.

Bilgilerinize saygılarımla sunarım.

Prof. Dr. Ş. Halil TURAN
Başkan V

Prof. Dr. Ayhan SOL
Üye

Prof. Dr. Ayhan Gürbüz DEMİR
Üye

Doç. Dr. Yaşar KONDAKÇI
Üye

Doç. Dr. Zana ÇITAK
Üye

Doç. Dr. Emre SELÇUK
Üye

Dr. Öğr. Üyesi Pınar KAYGAN
Üye

B. PERMISSION OBTAINED FROM MINISTRY OF EDUCATION

T.C.
ANKARA VALİLİĞİ
Milli Eğitim Müdürlüğü

Sayı : 14588481-605.99-E.13419350
Konu : Araştırma İzni

17.07.2018

ORTA DOĞU TEKNİK ÜNİVERSİTESİNE
(Öğrenci İşleri Daire Başkanlığı)

İlgi: a) MEB Yenilik ve Eğitim Teknolojileri Genel Müdürlüğünün 2017/25 nolu Genelgesi.
b) 21/06/2018 Tarihli ve E.16 sayılı yazınız.

Üniversiteniz Matematik ve Fen Bilimleri Eğitimi bölümü yüksek lisans öğrencisi Arş. Gör. Emine ÇATMAN AKSOY'un **"Yedinci Sınıf Öğrencilerinin Sütun ve Çizgi Grafiğinde Verilen Ortalama ve Yayılım Kavramları Hakkındaki İstatistiksel Okuryazarlık Seviyelerinin İncelenmesi"** konulu uygulama talebi Müdürlüğümüzce uygun görülmüş ve uygulamanın yapılacağı İlçe Milli Eğitim Müdürlüğüne bilgi verilmiştir.

Görüşme formunun (13 sayfa) araştırmacı tarafından uygulama yapılacak sayıda çoğaltılması ve çalışmanın bitiminde bir örneğinin (cd ortamında) Müdürlüğümüz Strateji Geliştirme (1) Şubesine gönderilmesini rica ederim.

Vefa BARDAKCI
Vali a.
Milli Eğitim Müdürü

Güvenli Elektronik İmzalı
Aslı İle Aynıdır.
/21/08/2018

Adres: Emniyet Mah. Alparslan Türkeş Cad. 4/A Yenimahalle/ANKARA
Elektronik Ağ: www.meb.gov.tr
e-posta: istatistik06@meb.gov.tr

Bilgi için: D. KARAGÜZEL
Tel: 0 (312) 221 02 17
Faks: 0 (312) 221 02 16

Bu evrak güvenli elektronik imza ile imzalanmıştır. https://evraksorgu.meb.gov.tr adresinden 010a-cb6e-34ca-af3f-0d73 kodu ile teyit edilebilir.

C. STATISTICAL LITERACY TEST

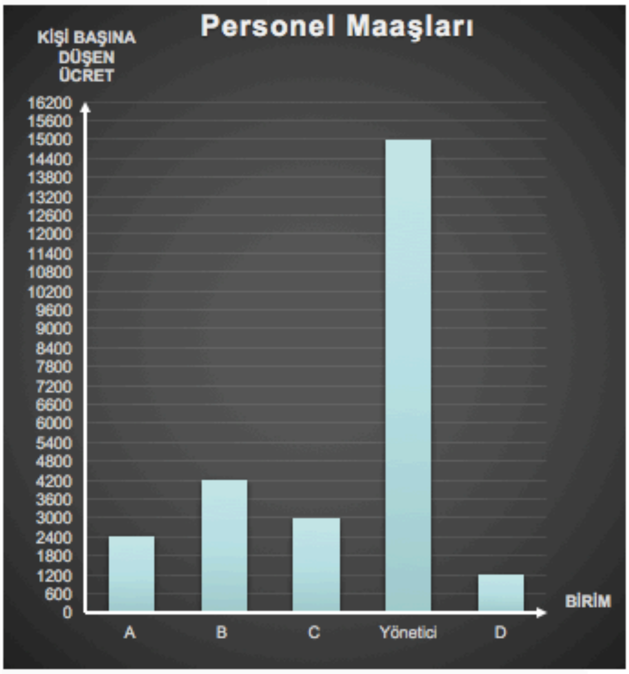
İSTATİSTİKSEL OKURYAZARLIK TESTİ

Soru 1:

İŞ DÜNYASI

www.ısdünyası.com **İŞ DÜNYASINDAN SON HABERLER** 18 Aralık 2017

Ortalama Maaş 4300 Türk Lirası!



İnşaat sektöründe hizmet veren İnşaSAN şirketi çalışanlarına yandaki veriler doğrultusunda **ortalama** 4300 TL maaş verdiklerini belirtti.

Birim	A	B	C	D	Yönetici
Birimde Çalışan Personel Sayısı	25	9	10	8	8

a) Yukarıdaki haberde geçen **ortalama** kelimesinden ne anlıyorsunuz?

b) Yukarıdaki haberde geçen **ortalama maaş** nasıl hesaplanmıştır? Cevabınızı açık bir şekilde belirtiniz.

c) Şu an 2900 TL maaş aldığınız bir işte çalıştığınızı varsayınız. Haberdeki iş yerinde herhangi bir birime yerleştirilecek olsanız verilen <u>grafiği ve ortalama maaşı göz önüne alarak</u> şu anki işinizi bırakıp bu iş yerinde çalışmak ister misiniz?

Evet / Hayır

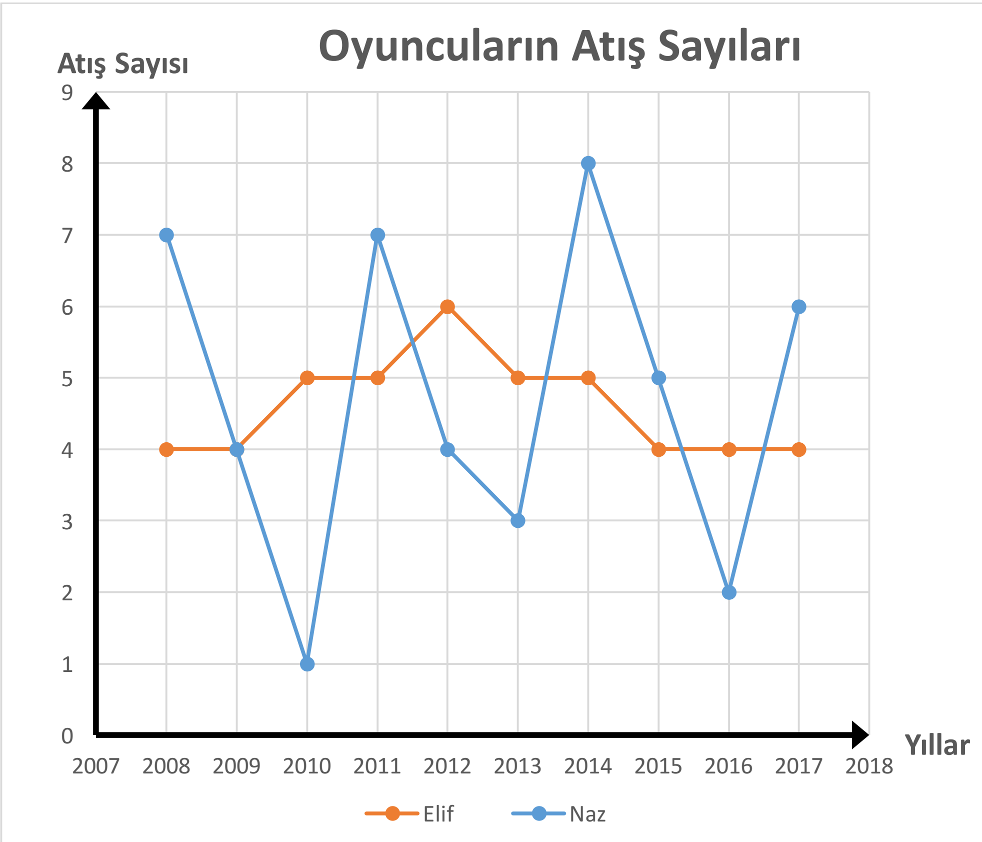
- Bu karara nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

  ………
  ………
  ……………………………………………………………………………

d) Matematikte kullandığınız "Ortalama" kelimesini kendi cümleleriniz ile tanımlayınız. Ortalama kavramı ile ilgili bir örnek veriniz.

Soru 2:

Yıldız Milli Kadın Basketbol takımı antrenörü final maçında oynayacak oyuncuları seçmek için her bir oyuncunun son 10 yıldaki final maçlarında takımları için attıkları sayıları incelemektedir. Aşağıdaki grafik antrenörün seçmek de kararsız kaldığı iki oyuncunun attıkları sayıları göstermektedir.



a) Her iki oyuncunun son 10 yılda final maçlarındaki ortalama atış sayılarını bulunuz. Her iki oyuncu için nasıl bulduğunuzu gösteriniz.

Elif'in ortalama atış sayısı :

Naz'ın ortalama atış sayısı:

b) Antrenör final maçı için Naz'ı seçme kararı almıştır. Antrenörün bu kararına katılıyor musunuz?

Evet / Hayır

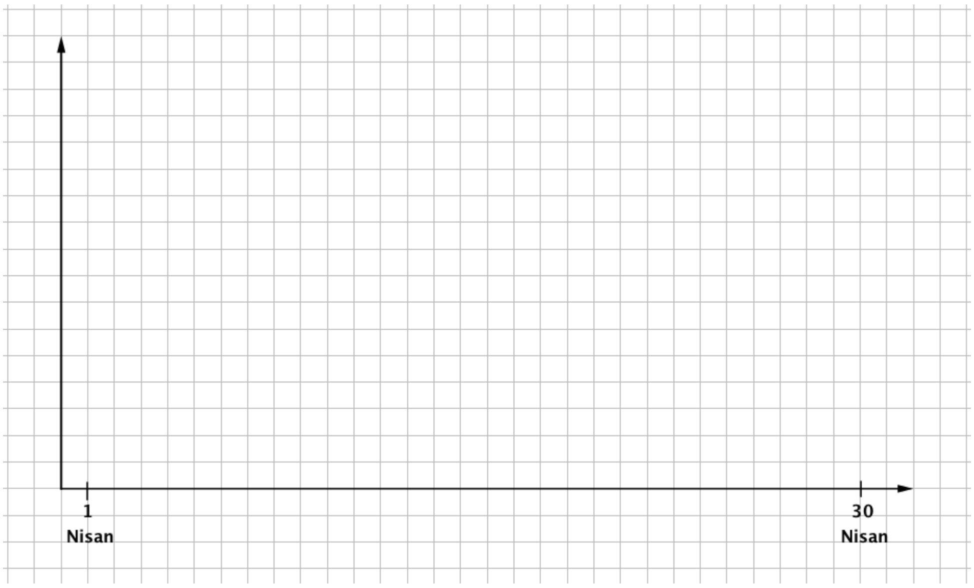
- Bu sonuca nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

  ………
  ………
  ………………………………………………………………………………

Soru 3:

"Meteorolojiden alınan verilere göre Nisan ayı ortalama sıcaklığı yaklaşık 17 derecedir." Bu haberi manşetinde kullanacak olan bir gazeteci haberini bir grafikle desteklemek istiyor.

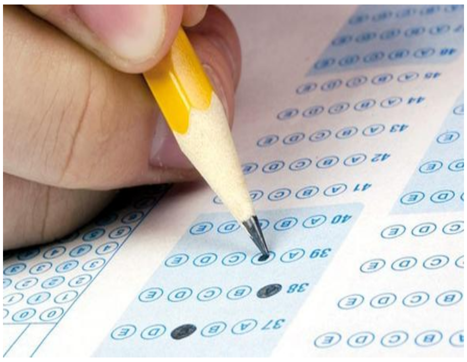
a) Gazetecinin haberinde kullanabileceği bir çizgi grafiği çiziniz.



b) Yukarıda çizdiğiniz grafiği haberde verilen ortalama sıcaklığı göz önünde bulundurarak yorumlayınız.

c) Matematikte kullandığınız "açıklık" kelimesini tanımlayınız.

Soru 4:



102 öğrencimiz Eylül ayındaki yabancı dil sınavında en yüksek puanı aldı. Son 5 yıldaki ortalamamız 42 öğrenci! Siz de gelin başarın!

a) Broşürde geçen **ortalama** kelimesinden ne anlıyorsunuz?

b) Sizce broşürde belirtilen **ortalama öğrenci sayısı** nasıl hesaplanmıştır? Cevabınızı açık bir şekilde ifade ediniz.

c) Şu anki kursunuzda yılda ortalama 30 öğrencinin en yüksek puanı aldığını biliyorsunuz. Her iki kurstan da belirtilen yıllarda eşit sayıda öğrencinin sınava katıldığını biliyorsanız, broşürü göz önüne alarak şimdiki kursunuzdan ayrılıp bu kursa kayıt olmayı düşünür müsünüz?

Evet / Hayır

- Bu karara nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

  ……
  ……
  …………………………………………………………………………………

Soru 5:

Aşağıda çeşitli tatil bölgelerinde otellere sahip olan Metin Mutlu ile yapılan röportajdan bir kesit sunulmuştur.

Yaza Yaklaşırken…

….

S: Otellerinizdeki paket seçenekleri nelerdir? Bu paketlerin içeriğini neye göre belirliyorsunuz?

M: Müşterilerimizin kalacakları gün sayısı paketleri belirlemedeki en büyük etken.

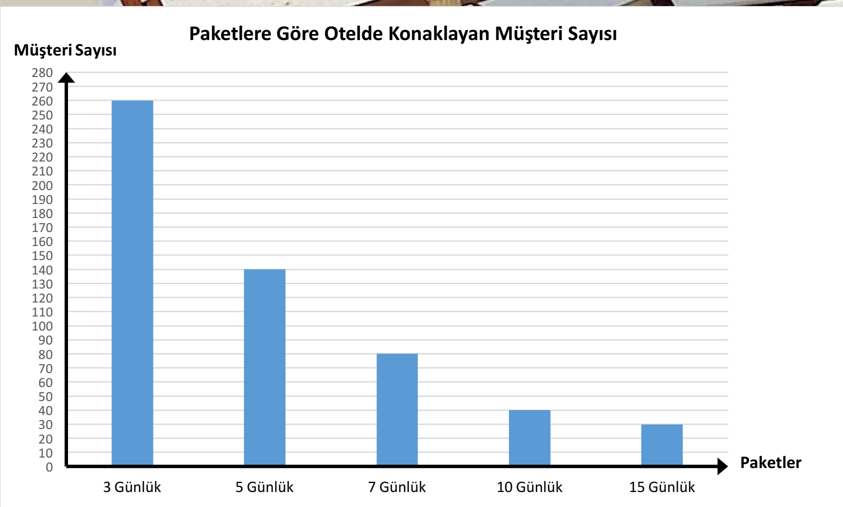
S: Siz kaç günlük paketler sunuyorsunuz?

M: 3, 5, 7 günlük gibi paketlerimiz olmakla birlikte her yıl paketlerimizi gözden geçiriyoruz.

S: Paketlerdeki gün sayısını belirlemedeki kriteriniz nedir? Bu sizin için epey bir zor olmalı.

M: Her yılın sonunda müşterilerimizin kaç günlük paketleri tercih ettiğini gösteren bir araştırma yapıyoruz. Bu araştırmanın sonuçlarına göre paketlerimizin kaç gün olacağına ve fiyatlarına karar veriyoruz. Örneğin, bir otelimizin geçen yılın verilerini gösteren bir grafik şu an elimde. Bu grafiğe göre müşterilerimiz ortalama 5 gün tatil yapmış. Bu sebepten dolayı bu yıl 5 günlük paketleri en fazla sayıda yapma kararı aldık.

…



a) Sizce Metin Bey röportajda geçen **5 günlük ortalamayı** grafikteki verileri göz önüne alarak nasıl hesaplamıştır? Cevabınızı açık bir şekilde ifade ediniz.

b) <u>Verilen grafiği göz önüne alarak</u> Metin Bey'in 5 günlük paketleri en fazla yapma kararına katılıyor musunuz?

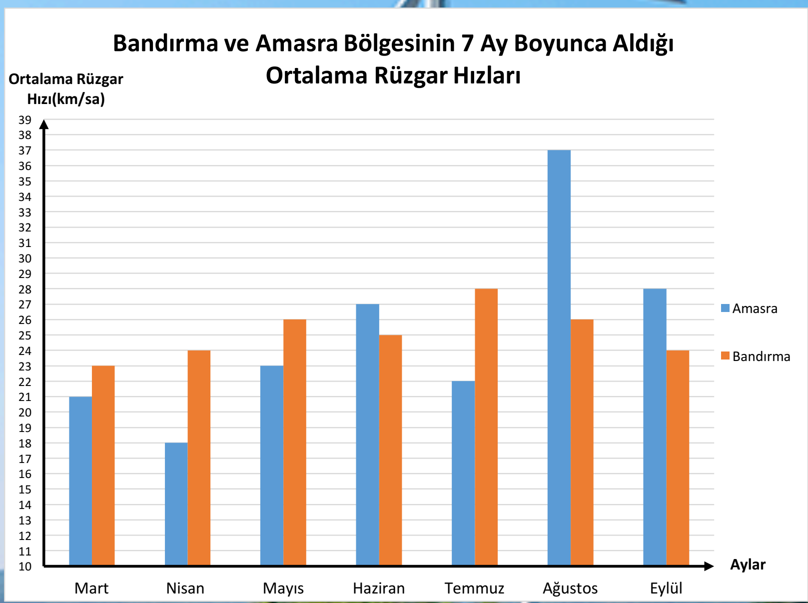
Evet / Hayır

- Bu sonuca nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

  ……
  ……
  ………………………………………………………………………………

Soru 6:

Fabrikaların enerji ihtiyacını karşılamak üzere rüzgar santralleri kuran bir firma kendileri için en karlı yatırımı yapmak üzere yaptığı araştırma sonucuna göre Bandırma ve Amasra bölgelerinin kendileri için karlı olacağını düşünmektedir. Aşağıdaki grafik iki bölgenin belirli aylar boyunca aldıkları ortalama rüzgar hızlarını göstermektedir.



a) Her iki bölgenin belirtilen aylarda aldığı rüzgar hızlarının ortalamasını bulunuz. Her iki bölge için nasıl bulduğunuzu gösteriniz.

Amasra Bölgesinin ortalama rüzgar hızı:

Bandırma Bölgesinin ortalama rüzgar hızı:

b) Şirket yöneticileri Amasra bölgesine santral kurma kararı almıştır. Eğer rüzgar hızının dalgalı olmasının bakım maliyetlerini artırdığı bilgisi göz önüne alınırsa sizce şirketin aldığı karar uygun mudur?

Evet / Hayır

- Bu sonuca nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

  ………
  ………
  ………………………………………………………………………………

Soru 7:

Ankara Şoförler Odası fabrika servis hızlarının saatte ortalama 70 km olması gerektiğini belirlemiş olup hız limitine uymayan şoförlere cezai işlem uygulanacağını bildirmiştir.

Aşağıdaki grafik bir servisin günün belirli saatlerinde ölçülen hızını göstermektedir.

Şoförler odası bu grafiğe dayanarak servisin hızı konusunda bir sorun olmadığını düşünerek şoföre herhangi bir cezai işlem uygulamamıştır. Derneğin aldığı karara katılıyor musunuz?

Evet / Hayır

- Bu sonuca nasıl ulaştınız? Cevabınızı açık bir şekilde belirtiniz.

  ……
  ……
  ………………………………………………………………………………

D. TURKISH SUMMARY/TÜRKÇE ÖZET

YEDİNCİ SINIF ÖĞRENCİLERİNİN ORTALAMA VE DEĞİŞİM KAVRAMLARI İLE İLGİLİ İSTATİSTİKSEL OKURYAZARLIKLARININ SÜTUN VE ÇİZGİ GRAFİĞİNDE İNCELENMESİ

Giriş

İstatistik eğitimi ile ilgili ilkeler ve ölçme raporu (GAISE,2005), ulaşılacak ana hedefin istatistiksel okuryazarlık olduğunu belirtmiştir. Bu rapora göre amaç; öğrencilerin liseden vatandaşlık, istihdam, aile gereksinimleri ile başa çıkmak ve sağlıklı, mutlu ve üretken bir hayata sahip olmak için yeterli istatistiksel akıl yürütme becerisine sahip olmalarını sağlamaktır (GAISE, 2005). Diğer bir deyişle, GAISE (2005) okulların bir amacının da istatistiksel okuryazar olan bireyler yetiştirmek olması gerektiğini dile getirmiştir.

Günlük yaşamda bireylerin sağlık, eğitim, politika gibi birçok alanda karar vermeleri gerekmektedir (Halpern, 1997). İnsanlar bu alanlarda yüzdeler, çeşitli grafik türleri, ortalamalar ve olasılıklar ile karşılaşmaktadır (Wallman, 1993). Bir başka deyişle, bireyler günlük yaşamlarında istatistiksel bilgi ile karşılaşırlar ve bilgili bir vatandaşın bu bilgileri anlayıp sonrasında bu bilgiyi kullanarak çeşitli kararlar alması gerekmektedir (Towsend, 2006; Wallman, 1993). Fakat günlük yaşamda veya medyada karşılaşılan istatistiksel mesajlar ön yargı veya öznellik içerebilir (Gal, 2004). Bazı istatistiksel bilgiler reklamlara, argümanlara veya önerilere sadece onları biraz daha güvenilir kılmak için eklenebilir (Ben-Zvi & Garfield, 2004) ve bu istatistiksel bilgiler genellikle sunulan reklamların veya argümanların sahipleri tarafından hazırlanır (Shield, 1999). Bu sebeple bireylerin sunulan istatistiksel bilgileri anlamasının yanında ön yargı içeren veya yanlış yönlendiren istatistiksel mesajları eleştirel bir şekilde değerlendirebilmeleri gerekmektedir (Gal, 2004; Watson, 2006).

Rumsey (2002) ise istatiksel okuryazarlığın bireylere sunulan bilgileri eleştirel bir şekilde değerlendirebilmek için gerekli olduğunu belirtmiştir. Buna paralel olarak, Gal (2002) istatistiksel okuryazarlığın bilgi toplumunda yaşayan bireyler için kilit bir beceri olduğunu dile getirmiştir. Dahası, istatistiksel okuryazarlık çocukların 21. yüzyılda karar verebilmesi için gereklidir çünkü çocuklar çeşitli veri temsilleri ile kolayca yanlış şekilde bilgilendirilebilir (Watson & English, 2015).
Önceden belirtildiği üzere, istatistiksel okuryazarlık okulda elde edinilmesi beklenen bir beceridir (GAISE, 2005; Gal, 2002; Ben-Zvi & Garfield, 2004). Dolayısıyla farklı ülkelerdeki bir çok öğretim programı istatistiksel okuryazarlığın önemini vurgulamış, istatistik ve olasılık kavramlarını öğretim programlarının içerisine dâhil etmiştir. Örneğin, NCTM (National Council of Teachers of Mathematics, 2000) öğrencilerin karar vermede önem arz eden istatistiksel yöntemleri anlamaları gerektiğini belirterek istatistiksel okuryazarlığın önemine değinmiştir. Avustralya Eğitim Konseyi (1991) istatistiksel okuryazar olan bireyler yetiştirmek için şansın sosyal hayattaki kullanımını anlama, açıklama ve istatistiğin günlük yaşamımıza olan etkilerini anlama kazanımlarını öğretim programlarına eklemiştir. Benzer şekilde ülkemizdeki matematik öğretim programında yer alan beş öğrenme alanından biri veri işlemedir. Bu öğrenme alanı araştırma sorusu oluşturmak, verilerin temsili, merkezi eğilim ve yayılım ölçüleri gibi temel konuları kapsamaktadır. Diğer ülkelerdeki öğretim programları gibi Türkiye'deki matematik öğretim programı da hesaplanan istatistiksel değerleri gerçek hayat durumlarında yorumlamaya ve var olan istatistiksel bilgiye göre karar verebilmeye önem vermiştir (MONE, 2018).

Öğretim programlarının yanında istatistiksel okuryazarlık yaklaşık 20 yıldır önemli bir araştırma konusu hâline gelmiştir. Alanda öncü olan araştırmacılardan biri olan

Watson (1997) istatistiksel okuryazarlığı üç aşamalı hiyerarşik bir yapı olarak tanımlamıştır. Bu aşamaların ilki istatistik ve olasılık alanındaki temel terminolojiyi anlamaktır. İkinci aşama ise istatistik ve olasılık ile ilgili olan kavramları anlayıp farklı bağlamlarda yorumlayabilmeyi içerir. Son aşamada ise istatistiksel okuryazar olan bir birey gerçek yaşamda karşılaştığı iddiaları eleştirel olarak değerlendirebilmelidir. Yani, Watson (1997) için istatistiksel okuryazarlık bireylerin gerçek yaşamlarında karşılaştıkları istatistiksel mesajları anlayabilme, yorumlayabilme ve eleştirel bir şekilde değerlendirebilme becerisidir. Sonraki yıllarda Watson ve arkadaşları tarafından bu üç aşamalı yapı kullanılarak birçok araştırma yapılmıştır (Watson & Moritz, 1999, 2000, 2000b; Watson & Callingham, 2003). Yapılan çalışmaların sonuçlarına göre, Watson 2006 yılında istatistiksel okuryazarlığın ana bileşenlerini gösteren bir model geliştirmiştir. Yedi bölümden meydana gelen bu modeldeki bölümlerden biri veri ve şans konularını içeren bölümdür. Bu bölüm veri toplama, şans, ortalama ve çıkarım gibi alt bileşenlerden oluşur. Bu bileşenin yanında farklı araştırmacılar tarafından da önemi sıklıkla belirtilen bir bileşen bulunmaktadır (Cobb & Moore, 1997; Ben Zvi, 2004). Bu, değişim bileşenidir. Cobb and Moore (1997) istatistiğe ihtiyaç duyulmasının sebebinin bu bileşenin yani değişim bileşeninin var olması olduğuna inanmaktadır. Ayrıca Konold ve Higgins (2003) ortalamayı gürültülü bir ortamda bulunan sinyal olarak nitelendirmektedir. Gürültü kelimesi ile ortamda bulunan değişimden bahsedilmektedir (Konold & Higgins, 2003). Watson (2006)'ın modelinde bulunan ortalama ve değişim kavramları Türkiye'de ortaokul matematik öğretim programında yer alan iki önemli konudur. Bu sebeple, bu çalışmada ortalama ve değişim kavramlarına odaklanılacaktır.

Literatürde öğrencilerin ortalama ve yayılım kavramlarını anlayışları üzerine yapılan çeşitli çalışmalar bulunmaktadır (Mokros & Russel, 1995; Watson & Moritz, 2000, Ucar & Akdoğan, 2009; Shaugnessy & Ciancetta, 2002; Watson, Kelly, Callingham & Shaugnessy, 2003). Ortalama kavramı ile ilgili olan çalışmalar, öğrencilerin ortalama kavramını tanımlamada zorluk yaşadıklarını

(Watson & Moritz, 2000), ortalamayı veriyi temsil eden bir değer olarak göremediklerini (Mokros &Russell, 1995), iki veri grubunu karşılaştırmada kullanamadıklarını (Gal, Rotschild & Wagner, 1989; Watson &Moritz,1999) göstermiştir. Değişim kavramı ile ilgili yapılan çalışmaların sonuçları da herhangi bir farklılık göstermemiştir. Çoğu öğrencinin değişim kavramını tanımlama ve yorumlamada zorluk çektiği (Watson & Kelly, 2008), ortalama kavramı gibi değişim kavramını da iki grubu karşılaştırırken kullanmadığı görülmüştür (Shaugnessy, 2003b). Öte yandan, Shaugnessy ve Pfannkuch (2002) öğrencilerin tablo halinde sunulan bir verinin grafiğini çizdiklerinde, verideki değişimi fark edebildiklerini gözlemlemiştir. Benzer şekilde, Bright ve Friel (1998) grafiklerin öğrencilerin sunulan verileri yorumlayabilmelerinde etkili olduğunu belirtmiştir. Ayrıca, grafiklerin ortalama kavramının yorumlanması üzerinde etkisi olabilir (Enisoglu, 2014). Farklı bir deyişle, grafikler ortalama ve değişim kavramlarının yorumlanması üzerinde etkili bir role sahip olabilir.

Aynı zamanda, Watson, Chick ve Callingham (2014) her gerçek yaşam bağlamının öğrencilerin ortalama kavramını anlayışları üzerinde aynı etkiyi doğurmadığını gözlemlemiştir. Çalışmalarında, hava durumu ile ilgili olan bir bağlamda öğrencilerin ortalamayı havadaki değişimi de göz önüne alarak kolayca yorumlayabilmelerine rağmen ev fiyatlarının ortalamasını yorumlarken zorluk yaşadıklarını gözlemlemişlerdir. Benzer şekilde, McGatha, Cobb ve Mc Clain (2002) çoğu öğrencinin hava durumu ile ilgili bir bağlamda verideki değişimi fark ettiğini gözlemlerken basketbol ile ilgili bir bağlamda öğrencilerden sadece bir kaçının değişimi fark edebildiğini gözlemlemiştir. Diğer bir deyişle, öğrencilerin ortalama ve değişim kavramları ile ilgili anlayışlarını farklı gerçek yaşam durumlarında incelemek önemlidir çünkü yukarıdaki yazarların bahsettiği gibi farklı gerçek yaşam durumları farklı sonuçların bulunmasına olanak sağlayabilir.

Yukarıda açıklanan sebepler dolayısıyla bu çalışmanın amacı ortaokul yedinci sınıf öğrencilerinin ortalama ve değişim kavramları ile ilgili istatistiksel okuryazarlıklarını grafik temsilleri ile farklı gerçek yaşam durumlarında incelemektir. Ortaokul matematik öğretim programında yer alan iki grubu karşılaştırmada aritmetik ortalama ve açıklığı kullanır kazanımı doğrultusunda öğrencilerin ortalama ve değişim kavramı ile ilgili istatistiksel okuryazarlıkları, karşılaştırma gereken durumlarda da incelenecektir.

Öğrencilerin yedinci sınıfa kadar öğrendikleri grafik temsilleri sütun, çizgi ve daire grafikleridir. Sütun ve çizgi grafiklerinin iki veri grubunu karşılaştırmayı gerektiren durumlar için uygun grafik temsilleri oldukları düşünülmesi sebebiyle bu çalışmanın amacı yedinci sınıf öğrencilerinin ortalama ve değişim kavramları ile ilgili istatistiksel okuryazarlıklarını sütün ve çizgi grafikleri üzerinde incelemektir.

Çalışmanın Amacı

Bu çalışmanın amacı ortaokul yedinci sınıf öğrencilerinin ortalama ve değişim kavramları ile ilgili istatistiksel okuryazarlıklarını sütun ve çizgi grafiğinde sosyal veya bilimsel bağlamlarda sunulan veriler kullanarak incelemektir. Bu bağlamda, bu çalışma yedinci sınıf öğrencilerinin ortalama ve değişim kavramları ile ilgili istatistiksel okuryazarlık seviyelerini sütun ve çizgi grafiğinde sosyal veya bilimsel bağlamlarda sunulan veriler kullanarak belirlemeyi ve farklı istatistiksel okuryazarlık seviyelerine sahip öğrencilerin ortalama ve değişim kavramları ile ilgili tanımlarını, yorumlarını ve değerlendirmelerini araştırmayı amaçlamaktadır.

Önemli Terimlerin Tanımları

Ortalama: Ortalama birçok sayıyı temsil eden bir sayı veya ölçü olarak tanımlanmıştır (Van de Wall, 2013).

Bu çalışmada ortalama kavramı sütun ve çizgi grafiğinde sosyal veya bilimsel bağlamlarda sunulan verileri temsil eden bir ölçüyü ifade etmektedir.

Değişim: Değişim kavramı değişimin tasviri veya ölçüsü olarak tanımlanmıştır. (Reading & Shaugnessy, 2004, p.202).

Bu çalışmada, değişim kavramı sütun ve çizgi grafiğinde sosyal veya bilimsel bağlamlarda sunulan verilerdeki değişimi temsil eden ölçüyü ifade etmektedir.

Sütun Grafiği: Değişkenlerin sayısal değerlerinin yüksekliğinin veya uzunluğunun eşit genişliğe sahip dikdörtgenler ile temsil edildiği diyagramdır (Oxford American Dictionary, 2006, p. 66).

Çizgi Grafiği: İki sürekli veri arasındaki ilişkinin iki noktayı birleştiren çizgilerle gösterildiği temsil türüdür.

Sosyal veya Bilimsel Bağlam: Sosyal veya bilimsel bağlam gazete, araştırma raporları, broşür, dergi gibi günlük hayatta özellikle medyada karşılaşılan yazılı kaynaklardır. Bu çalışma sosyal bağlam olarak genellikle araştırmacı tarafından hazırlanan gazeteler, broşürler ve dergiler içerirken bilimsel bağlam olarak yine araştırmacı tarafından hazırlanan bazı araştırma sonuçlarını içermektedir.

Yöntem

Çalışma Deseni

Fraenken ve Wallen (2006) bir popülasyonun beceri veya bilgi gibi belirli yönlerinin veya özelliklerinin betimlenmesinde tarama tipi araştırma çalışmalarının oldukça faydalı olduğunu belirtmiştir. Bu doğrultuda, yedinci sınıf öğrencilerinin istatistiksel okuryazarlıklarını incelemek amacıyla tarama tipi araştırma deseni kullanılmıştır. Ayrıntılı olarak bu çalışma seçilen örneklemden gerekli veriyi tek seferde toplamayı gerektirdiğinden araştırmanın deseni kesitsel tarama olarak belirlenmiştir. Öğrencilerin testteki sorulara sundukları cevaplar öğrencilerin ortalama ve değişim kavramları ile ilgili tanımları, yorumları ve değerlendirmelerini belirlemek amacıyla derinlemesine incelenmiştir.

Katılımcılar

Bu çalışmada örneklem uygun örnekleme yöntemi kullanılarak belirlenmiştir. Bu bağlamda çalışmanın örneklemini 2017-2018 öğretim yılında Ankara'nın Akyurt ve Çankaya ilçesinde iki devlet okulunda öğrenim gören 164 (82 kız ve 82 erkek) 7. sınıf öğrencisi oluşturmaktadır.

Veri Toplama Aracı

Çalışmanın verileri araştırmacı tarafından hazırlanan İstatistiksel Okuryazarlık Testi (İOT) ile toplanmıştır.

İstatistiksel Okuryazarlık Testi

Yedinci sınıf öğrencilerinin ortalama ve değişim kavramaları ile ilgili istatistiksel okuryazarlıklarını incelemek amacıyla araştırmacı tarafından İstatistiksel Okuryazarlık Testi (İOT) hazırlanmıştır. Ölçeğin maddeleri hazırlanırken Milli Eğitim Bakanlığı (MEB) Ortaokul Matematik Dersi Öğretim Programı'nda yer alan ortalama ve değişim kavramları ve sütun ve çizgi grafiği ile ilgili kazanımlar göz önünde bulundurulmuştur. Ayrıca Watson'ın 3 hiyerarşik adımdan oluşan istatistiksel okuryazarlık teorik çerçevesi soruların hazırlanmasında göz önüne alınmıştır. Bu bağlamda, ortalama ve değişim kavramlarının tanımlanması, yorumlanması ve değerlendirilmesi hakkında sorular hazırlanmıştır. Hazırlanan ölçek 46 8. sınıf öğrencisiyle yapılan pilot çalışma sonucunda tekrar düzenlenmiştir. Son durumda, İOT 7 açık uçlu sorudan oluşmaktadır. Bu sorulardan ikisi alan yazınından adapte edilirken, diğerleri araştırmacı tarafından hazırlanmıştır. Ayrıca asıl çalışmanın katılımcılarından rastgele seçilen 82 öğrencinin cevapları alanında uzman iki kişi tarafından değerlendirilerek sonuçların güvenilirliği test edilmiştir. Cronbach alpha güvenilirlik katsayısı .92 olarak bulunmuştur.

Verilerin Analizi

Bu çalışmanın amacı ortaokul yedinci sınıf öğrencilerinin ortalama ve değişim kavramları ile ilgili istatistiksel okuryazarlıklarını sütun ve çizgi grafiğinde sosyal veya bilimsel bağlamlarda sunulan veriler kullanarak incelemektir. Bu bağlamda, bu çalışma yedinci sınıf öğrencilerinin ortalama ve değişim kavramları ile ilgili istatistiksel okuryazarlık seviyelerini sütun ve çizgi grafiğinde sosyal veya bilimsel bağlamlarda sunulan veriler kullanarak belirlemeyi amaçlamaktadır. Öğrencilerin istatistiksel okuryazarlık seviyelerini belirleyebilmek için Watson ve Callingham'ın 2003 yılında Watson (1997)'ın üç aşamalı hiyerarşik yapısını genişleterek oluşturdukları 6 seviyeden oluşan istatistiksel okuryazarlık teorik çerçevesi kullanılmıştır. Bu çerçeve Watson'ın (2006) modelinde bahsettiği tüm bileşenler ile ilgilidir. Diğer bir deyişle, çerçevedeki her bir seviye, öğrencilerin veri toplama, veri temsili ve ortalama kavramları ile ilgili sahip oldukları istatistiksel okuryazarlık özelliklerinden bahsetmektedir. Bu çalışma sadece ortalama ve değişim kavramları ile ilgili olduğundan öncelikle öğrencilerin ortalama ve değişim kavramları ile ilgili sahip olduğu özellikler teorik çerçeve incelenerek belirlenmiştir. Belirlenen özellikler öğrencilerin cevaplarını incelemek için faydalı olmasına rağmen toplanan verilerin analizi için daha detaylı özelliklere ihtiyaç duyulmuştur. Bu sebeple Watson (2006)'ın kitabındaki ortalama, değişim ve çıkarım ile ilgili bölümleri, Watson'ın ilgili makaleleri, Shaugnessy' nin 2003 yılında yayımladığı makalesi ve çalışmadaki öğrenci cevapları incelenerek Watson ve Callingham (2003)'ın çerçevesinde değişikliklere gidilmiştir. Teorik çerçeveye son hâlinin verilmesinin ardından öğrencilerin istatistiksel okuryazarlık seviyeleri her bir soru için ayrı olarak belirlenmiştir. Bunun sebebi farklı bağlamların öğrencilerin istatistiksel okuryazarlık seviyeleri üzerindeki rolünü incelemektir. Ayrıca, öğrencilerin istatistiksel okuryazarlık seviyeleri teorik çerçevedeki her iki seviye birleştirilerek, Seviye 1-2 veya Seviye 3-4 gibi sunulmuştur. Bunun sebebi İOT'nin Watson (1997)'nın üç aşamalı hiyerarşik yapısına göre hazırlanmış olmasıdır. Diğer bir deyişle, ortalama ve değişim kavramlarının tanımlanması, yorumlanması ve değerlendirilmesi hakkında sorular hazırlanmıştır. Daha önceden bahsedildiği gibi analiz için kullanılan çerçeve bu hiyerarşik yapının detaylandırılmış hâlidir. Kavramların yorumlanması 3. Seviyede başlarken 4. Seviyede devam eder. Öte yandan kavramların eleştirel bir şekilde değerlendirilmesi 5 ve 6. Seviyede beklenen bir beceridir. Örneğin, ortalama kavramını sosyal veya bilimsel bir bağlamda eleştirel bir şekilde değerlendiren bir öğrencinin istatistiksel okuryazarlık seviyesi Seviye 5 veya Seviye 6'dır. Dolayısıyla bu çalışmada o öğrencinin istatistiksel okuryazarlık seviyesi Seviye 5-6 olarak belirlenmiştir. Farklı bir örnekte, değişim kavramının eleştirel bir şekilde değerlendirilmesini içeren bir soruda öğrencilerin istatistiksel okuryazarlık seviyesinin teorik çerçeveye göre Seviye 5-6 olması beklenir. Fakat, analizler öğrenci cevaplarına göre yapıldığı için öğrencilerin istatistiksel okuryazarlık seviyeleri Seviye 1-2 veya Seviye 3-4 olabilir.

Bulgular ve Tartışma

Öğrencilerin Ortalama Kavramı ile ilgili İstatistiksel Okuryazarlık Seviyeleri

Bu çalışmanın odaklandığı kavramlardan biri ortalama kavramıdır. Bu çalışma öğrencilerin ortalama kavramı ile ilgili istatistiksel okuryazarlık seviyelerini belirlemeyi ve farklı istatistiksel okuryazarlık seviyelerinde bulunan öğrencilerin ortalama kavramı ile ilgili tanımlarını, yorumlarını ve değerlendirmelerini araştırmayı amaçlamıştır.

Toplanan veriler incelendiğinde grafik türü fark etmeksizin çalışmaya katılan yedinci sınıf öğrencilerinin çoğunluğunun istatistiksel okuryazarlık seviyeleri Seviye 3-4 olarak belirlenmiştir. Okulların amacı kullanılan teorik çerçevedeki en yüksek istatistiksel okuryazarlık seviyesine sahip bireyler yetiştirmek olmasına rağmen (Chick & Pierce, 2011) bu çalışmada ortalama ile ilgili sütun ve çizgi grafiğini içeren tüm sorularda bu seviyeye ulaşan fazla öğrenci ile karşılaşılmamıştır. Diğer bir deyişle, çalışmaya katılan öğrencilerin ortalama kavramını sütun veya çizgi grafiği üzerinde sosyal veya bilimsel bağlamlarda yorumlayabilirken, çoğunun sunulan ortalama kavramını eleştirel bir şekilde değerlendiremediği söylenebilir. Aynı zamanda, çok az sayıda öğrenci ortalama kavramını tanımlarken ortalamanın verileri temsil etme doğasından bahsetmiştir. Literatürde bulunan bazı çalışmalar da öğrencilerin istatistiksel kavramların tanımlanması ve değerlendirilmesinde yorumlanmasına nazaran daha fazla zorluk yaşadıkları sonucuna ulaşmıştır (Watson, 2006; Watson & Callingham, 2003; Yolcu, 2012). Aşağıda belirtilen sebepler böyle bir bulgu edinilmesine yol açmış olabilir.

Bu sebeplerden ilki Türkiye'deki matematik öğretim programı olabilir. Türkiye'de bulunan matematik öğretim programının istatistiksel kavramların eleştirel bir şekilde değerlendirmesine, kavramların yorumlanması kadar önem vermediği gözlemlenmiştir. Diğer bir deyişle, öğrenciler ortalama kavramının farklı bağlamlarda yorumlanmasına alışkın fakat değerlendirmesine alışkın olmadıklarından dolayı ortalama kavramının sunulan sosyal veya bilimsel bağlamlarda eleştirel bir şekilde değerlendirilmesinde zorlanmış olabilirler.

Böyle bir bulguya, çalışmaya katılan öğrencilerin öğretmenleri de sebebiyet vermiş olabilir. Özen (2013) öğretmenlerin sunulan bazı istatistiksel iddaları eleştirel bir şekilde değerlendiremediği sonucuna ulaşmıştır. Aynı zamanda, eğer öğretmenler sunulan veri setine karşı eleştirel bir tutuma sahip değilse, bu öğretmenlerin öğrencilerinin de böylesi bir tutuma sahip olması beklenemez (Chick &Pierce, 2011; Watson, 2006). Dolayısıyla öğretmenlerin gerçek yaşam durumlarındaki istatistiksel mesajlara karşı sahip oldukları tutumlar, bu çalışmadaki öğrencilerin ortalama kavramını eleştirel değerlendirmede yaşadıkları zorlukların altında yatan bir neden olabilir.

Öte yandan öğretmenlerin ortalama kavramını tanımlarken matematiksel dili doğru bir şekilde kullanmaması (Yolcu, 2012), bu çalışmadaki öğrencilerin ortalama kavramını tanımlamada zorluk yaşamalarına sebebiyet vermiş olabilir. Literatürde öğretmen adaylarının matematiksel dili doğru bir şekilde kullanmadığı görülmüş olup (Yesildere, 2010) matematik öğretmenlerinin genellikle kavramların öğretiminde uygun terminolojiyi kullanmak yerine günlük dili kullandıkları sonucuna ulaşılmıştır (Miller, 1993). Bu çalışmaya katılan öğrencilerin öğretmenleri ortalama kavramının öğretiminde uygun matematiksel terminolojiyi kullanmıyorsa, elde edilen bulgular şaşırtıcı değildir.

Bu çalışmada öğrencilerin ortalama kavramı ile ilgili istatistiksel okuryazarlıklarının belirlenmesinin ardından farklı istatistik okuryazarlık seviyelerinde bulunan öğrencilerin ortalama kavramını nasıl tanımladıkları, yorumladıkları ve değerlendirdikleri detaylı bir şekilde incelenmiştir. İlk olarak öğrencilerin ortalama kavramını tanımlamaları ve yorumlamaları incelendiğinde, çoğu öğrencinin ortalamayı tanımlamak ve yorumlamak için merkezi eğilim ölçülerini kullandıkları fakat en çok kullanılan ölçünün aritmetik ortalama kavramı olduğu görülmüştür. Ortanca ve tepe değer kavramları az sayıda öğrenci tarafından kullanılmıştır. Literatürde bulunan bir çok çalışma da aynı sonuçlara ulaşmıştır (Brown & Silver, 1989; Enısoglu, 2014; Ucar & Akdogan, 2009; Watson, 2006; Yolcu, 2012). Bu durum çalışmaya katılan öğrencilerin öğretmenlerinin sadece aritmetik ortalama kavramına odaklanmasından kaynaklanmış olabilir çünkü Leavy ve O'Loughlin (2006) öğretmenler arasında ortalama kavramı ile ilgili en yaygın olan fikrin aritmetik ortalama kavramı olduğunu ifade etmiştir. Diğer bir neden ise aritmetik ortalama kelimesi ile ortalama kelimesi arasındaki benzerlik olabilir (Enisoglu, 2014). Alternatif bir bakış açısı olarak, bu bulgu Türkiye'deki matematik öğretim programından da kaynaklı olabilir çünkü yedinci sınıf matematik öğretim programında bulunan bir kazanım şu şekildedir: "Bir veri grubuna ait ortalama, ortanca ve tepe değeri bulur ve yorumlar." (MoNE, 2018, p. 70). Diğer bir deyişle, öğretim programı aritmetik ortalama kavramı yerine ortalama kavramını kullandığından dolayı bu çalışmadaki öğrencilerin ortalama kavramını duyduğunda aklına aritmetik ortalama kavramı gelmiş olabilir.

Aritmetik ortalama kavramı öğrenciler tarafından ortalama kavramını tanımlama ve yorumlamada çok fazla tercih edilmesine rağmen, bazı bulgular öğrencilerin bu kavramla ilgili kavramsal bir anlamaya değil işlemsel bir anlamaya sahip olduklarını göstermiştir. Bu bulgulardan bir tanesi aritmetik ortalama algoritmasının yanlış bir şekilde uygulanmasıdır. Bu bulgu literatürdeki bir çok çalışma ile tutarlılık göstermektedir (Cai, 2000; Maverach, 1983; Pollatsek et al. 1981; Watson & Moritz, 1999). Örneğin, bazı öğrenciler aritmetik ortalama algoritmasının ters çevrilmiş hâlini kullanmıştır; yani, veri kümesindeki veri sayısının toplamını veri kümesindeki verilerin toplamına bölmüştür. Çoğu öğrenci ise ağırlıklı ortalama hesaplamakta zorluk çekmiştir. Bu bulgular çalışmaya katılan öğrencilerin aritmetik ortalama kavramı ile ilgili kavramsal anlamalarının yeterli olmadığını açık bir şekilde yansıtmıştır çünkü eğer öğrenciler herhangi bir kavram hakkında kavramsal bir anlamaya sahip olsalardı, o kavram ile ilgili herhangi bir algoritmayı karşılaştıkları her durumda doğru bir şekilde uygulayabilirlerdi (Kilpatrick, Swafford & Findell, 2001).

Öğrencilerin aritmetik ortalama kavramı ile ilgili işlemsel bir anlamaya sahip olduğunu gösteren diğer bir bulgu da öğrencilerin ortalama tanımları ve yorumları arasındaki tutarsızlıktır. Örnek olarak, çalışmadaki bazı öğrenciler ortalama kavramını tanımlamak veya kavramın anlamını verilen sosyal veya bilimsel bağlamlarda yorumlamak için yaklaşık veya tahmini kelimelerini kullanırken, ortalamanın hesaplanması istenildiğinde ya verilerin aritmetik ortalamasını hesaplamış ya da tüm verileri toplamıştır. Fakat bu öğrenciler hesaplamalarının tanımları veya yorumları ile tutarlı olup olmadığına dikkat etmemiştir. Bazı öğrenciler ise ortalama kavramını artimetik ortalama olarak tanımlarken bu tanımlarını ortalama kavramını yorumlamak için kullanmamıştır. Bu bulguların altında yatan neden öğrencilerin aritmetik ortalama kavramı ile ilgili işlemsel bir anlamaya sahip olmaları olabilir. Başka bir deyişle, bu öğrenciler aritmetik ortalamanın sadece toplama ve bölme işlemi ile ilgili olduğunu fakat elde edilen sayının neyi temsil ettiğini bilmiyor olabilir (Konold & Higgins, 2003; Mokros &

Russell, 1995). Dolayısıyla öğrenciler ortalama kavramını verilen bağlamlarda aritmetik ortalama olarak yorumlayamamış veya ortalama kavramını yorumlayabilmek için kendi anlayışlarını kullanmış olabilirler.

Öğrencilerin ortalama kavramını tanımlamaları ve yorumlamalarının yanında verilen sosyal veya bilimsel bağlamlarda değerlendirmeleri incelendiğinde ise çoğu öğrencinin sunulan ortalamayı eleştirel bir şekilde değerlendiremediği görülmüştür. Aynı zamanda çoğu öğrenci verilen ortalama hakkında ya hiç şüphe duymamış ve verilen ortalamayı kabul etmiş ya da kendine özgü cevaplar sunmuştur. Bu bulgu Özen (2013)'in çalışması ile paralellik gösterir. Özen (2013) öğretmen adaylarının popüler medya metinlerini eleştirel bir şekilde değerlendiremediğini, ya verilen bilgiyi doğrudan kabul ettiklerini ya da kendi fikirlerini ortaya attıklarını gözlemlemiştir. Bu çalışmada bazı öğrencilerin verilen ortalamayı değerlendirmeye çalıştığı gözlemlenmiştir, fakat bu öğrenciler değerlendirmelerini veri kümesindeki sadece birkaç değere odaklanarak yapmıştır. Başka bir deyişle, bu öğrenciler veri kümesindeki tüm verilerin verilen ortalamanın üzerinde etkisi olacağını düşünmemiştir. Bu durum, öğrencilerin ortalama kavramını veri kümesindeki tüm verileri temsil ettiğini bilmemelerinden kaynaklı olabilir. Eğer bu çalışmadaki öğrenciler ortalamanın verilen bağlamdaki bütün veriyi temsil eden bir ölçü olduğunu bilselerdi (Mokros & Russell,1995) değerlendirmelerini yapmak için verilen tüm veriye odaklanabilirlerdi.

Öğrencilerin değerlendirme sorularında göstermiş olduğu bu genel bulguların yanı sıra bazı değerlendirme sorularının bağlamları farklı bulgulara yol açmıştır. İlk olarak, İOT'de ortanca kullanılmasını gerektiren iki soruya yer verilmiştir. Bir soruda veriler sütun grafiği üzerinde sunulurken, diğer soruda çizgi grafiği üzerinde sunulmuştur. Her iki soruda sunulan grafiğin öğrencilerin verideki uç değerleri fark etmesine yardımcı olacağı düşünülmüştür fakat elde edilen bulgular bazı araştırmacıların bulgularıyla tutarlılık göstermiştir (Watson & Callingham, 2003; Watson & Moritz, 1999b). Diğer bir deyişle, veriler çizgi veya sütun grafiği üzerinde sunulmasına rağmen öğrenciler verilen uç değerleri fark edememiş dolayısıyla sunulan ortalamayı eleştirel bir şekilde değerlendirememiştir. Bu bulgu öğrencilerin hangi merkezi eğilim ölçüsünün hangi durumlarda kullanılması konusunda herhangi bir tartışma içerisinde bulunmamalarından kaynaklanıyor olabilir. Çünkü bazı araştırmacılar üç merkezi eğilim ölçüsünün farklı bağlamlardaki etkisinin öğrenciler ile açık bir şekilde tartışılmadan fark edilemeyeceğini dolayısıyla öğrencilerin yüksek istatistiksel okuryazarlık seviyelerine ulaşamayacağını dile getirmiştir (Hobden, 2014; Watson, 2006; Watson & Callingham, 2003; Sharma, 2017).

Bazı öğrencilerin ise çizgi grafiği içeren bağlamda tek bir değere odaklanarak verilen ortalamayı değerlendirdiği fark edilmiş ve bu değerin veri kümesindeki uç değer olduğu görülmüştür. Diğer bir deyişle, aslında bu öğrencilerin veri kümesindeki uç değeri fark ettiği fakat bu uç değerin verilen ortalama üzerindeki etkisini düşünmedikleri söylenebilir. Bu öğrenciler sunulan ortalamaya karşı herhangi bir şüphe duymamalarına rağmen bu öğrencilerin fikirlerinin uç değerlerin ortalama üzerindeki etkisinin tartışılmasında başlangıç olarak kullanılabileceği düşünülmüştür.

İOT'de bir soruda ortalama olarak tepe değer mekezi eğilim ölçüsünü kullanmayı gerektirmektedir çünkü bu soruda öğrencilere kategorik bir veri sunulmuştur. Şaşırtıcı bir şekilde, ortanca kullanmayı gerektiren sorular ile karşılaştırıldığında daha fazla öğrencinin sunulan ortalamayı eleştirel bir şekilde değerlendirebildiği görülmüştür. Öğrenciler doğrudan tepe değerin, oteldeki 3 günlük paketlerin, ortalama olacağını ifade etmeseler de bu öğrenciler sunulan ortalama değerin yanlış olduğunu ifade etmiş ve müşteriler tarafından en fazla sayıda tercih edilen paketin kullanılması gerektiğini açıklamışlardır. Bu bulguya dayanarak, bu sorunun bağlamının tepe değer merkezi eğilim ölçüsünün kritik bir şekilde değerlendirilmesinde faydalı olduğu söylenebilir.

Öğrencilerin Değişim Kavramı ile ilgili İstatistiksel Okuryazarlık Seviyeleri

Bu çalışmanın odaklandığı kavramlardan biri değişim kavramıdır. Bu çalışma öğrencilerin ortalama kavramı ile ilgili istatistiksel okuryazarlık seviyelerini belirlemeyi ve farklı istatistiksel okuryazarlık seviyelerinde bulunan öğrencilerin değişim kavramı ile ilgili tanımlarını, yorumlarını ve değerlendirmelerini araştırmayı amaçlamıştır.

Toplanan veriler incelendiğinde grafik türü fark etmeksizin çalışmaya katılan yedinci sınıf öğrencilerinin çoğunluğunun istatistiksel okuryazarlık seviyeleri Seviye 1-2 olarak belirlenmiştir. Sadece çizgi grafiğini içeren bir soruda çoğu öğrenci Seviye 3-4'e ulaşabilmiştir. Sütun ve çizgi grafiğini içeren bütün sorularda en yüksek istatistiksel okuryazarlık seviyesine ulaşan az sayıda öğrenci olduğu görülmüştür. Başka bir deyişle, bu çalışmaya katılan öğrencilerin değişim kavramını hem yorumlamada hem de eleştirel bir şekilde değerlendirmede zorluk yaşadıkları söylenebilir. Aynı zamanda çoğu öğrencinin değişim kavramı ile ilgili bir ölçü olan açıklık kavramını tanımlayamadığı görülmüştür. Literatürde bulunan bazı çalışmalar da öğrencilerin istatistiksel kavramların tanımlanması ve değerlendirilmesinde yorumlanmasına nazaran daha fazla zorluk yaşadıkları sonucuna ulaşmıştır (Watson, 2006; Watson & Callingham, 2003; Yolcu, 2012). Farklı olarak bu çalışmada öğrencilerin değişim kavramının sütun veya çizgi grafiği üzerinde sosyal veya bilimsel bağlamlarda yorumlanmasında da zorlanma yaşadıkları bulgusu elde edilmiştir. Çeşitli sebepler böyle bir bulgu edinilmesine yol açmış olabilir.

Bunlardan ilki Türkiye'deki matematik öğretim programı olabilir. Daha önceden belirtildiği gibi Türkiye'de bulunan matematik öğretim programının istatistiksel kavramların eleştirel bir şekilde değerlendirmesine kavramların yorumlanması kadar önem vermediği gözlemlenmiştir. Bu sebeple çalışmaya katılan öğrenciler değişim kavramının eleştirel bir şekilde değerlendirilmesinde zorluk yaşamış olabilir. Bu çalışmadaki öğrencilerin değişim kavramının yorumlanmasında zorluk yaşamalarının sebebi ise öğretim programının değişim kavramının farklı bağlamlarda yorumlanmasına ortalama kavramının yorumlanması kadar önem vermemesi olabilir. Öğretim programı incelendiğinde değişim kavramının yorumlanması ile ilgili tek kazanımın açıklık kavramının yorumlanması olduğu görülmüştür. Dolayısıyla bu çalışmada değişim kavramının yorumlanması ile ilgili edinilen bu bulgu şaşırtıcı değildir.

Bu çalışmada öğrencilerin değişim kavramı ile ilgili istatistiksel okuryazarlıklarının belirlenmesinin ardından farklı istatistik okuryazarlık seviyelerinde bulunan öğrencilerin değişim kavramını nasıl tanımladıkları, yorumladıkları ve değerlendirdikleri detaylı bir şekilde incelenmiştir. İlk olarak öğrencilerin açıklık kavramını tanımlamaları incelendiğinde, çoğu öğrencinin, belirlenen istatistik okuryazarlık seviyesi fark etmeksizin açıklık kavramını tanımlayamadığı görülmüştür. Bu bulgu Watson ve Kelly (2008)'nin çalışması ile paralellik gösterirken Yolcu (2012)'nun çalışmasında elde ettiği bulguya tezat bir durum oluşturur. Yolcu (2012) çalışmasında sekizinci sınıf öğrencilerinin değişim kavramını açıklık kavramı ile doğru bir şekilde tanımladığını gözlemlemiştir. Bu bulguya dayanarak Yolcu (2012) öğrencilerin değişim kavramı ile ilgili işlemsel bir anlayışa sahip oldukları sonucuna ulaşmıştır çünkü öğrenciler değişim ile ilgili herhangi bir fikirden bahsetmeden sadece bir işlemden bahsetmiştir. O hâlde, Yolcu (2012)'nun görüşüne dayanarak bu çalışmaya katılan öğrencilerin açıklık kavramını tanımlayamamaları değişim kavramı ile ilgili işlemsel anlamalarının yeterli seviyede olmadığını gösterebilir. Türkiye'deki matematik öğretim programının 6. sınıf kazanımlarından birinin öğrencilerin açıklık kavramını tanımlaması olmasına rağmen böylesi bir bulguya ulaşılmasının nedeni çalışmaya katılan öğrencilerin öğretmenlerinin bu kazanıma yeterli derecede önem vermemesi olabilir. Eğer 6. sınıfta öğretmenler sadece açıklığın tanımını öğretip veri kümesindeki en yüksek değerden en küçük değer çıkarıldığında elde edilen değerin neyi temsil ettiğini veya kavramın ne zaman kullanılacağının faydalı olduğunu öğrenciler ile tartışmamış ise bu çalışmaya katılan yedinci sınıf öğrencilerinin açıklık kavramını tanımlamakta zorlanması şaşırtıcı bir sonuç değildir.

Öğrencilerin değişim kavramını yorumlamaları incelendiğinde ise tanımlamalarından çok farklı bir bulgu elde edilmemiştir. Diğer bir deyişle, çoğu öğrenci hava durumu ile ilgili bir bağlamda değişim kavramını yorumlayamamıştır. İOT'de öğrencilerden Ankara'nın Nisan ayında verilen ortalama sıcaklığı ile ilgili bir çizgi grafiği çizmeleri istenerek öğrencilerin değişim kavramını yorumlamaları incelenmiştir. Fakat çoğu öğrencinin herhangi bir grafik çizmediğinden dolayı değişim kavramını yorumlayamadığı sonucuna ulaşılmıştır. Bu sonuç Watson ve Kelly (2005)'nin çalışması ile çelişkili bir sonuçtur çünkü araştırmacılar öğrencilerin değişim kavramının grafik ile gösterilmesindeki performanslarının inceledikleri diğer değişkenlerden yüksek olduğunu gözlemlemiştir. Dolayısıyla çalışmada elde edilen bu bulgu öğrencilerin çizgi grafiği çizmede yaşadıkları zorlanmadan kaynaklı olabilir. Literatürdeki bazı çalışmalar da öğrencilerin çizgi grafiği çizmede sütun grafiği ve daire grafiğine nazaran zorluk yaşadıklarını göstermiştir (Capraro, Kulm & Hammer, 2005). Aynı zamanda az sayıda öğrencinin eksen isimleri, grafik başlığı gibi grafik öğelerine dikkat ettiği görülmüştür. Elde edilen bu bulgulardan, öğrencilerin grafik çizme etkinlikleri ile çok fazla haşır neşir olmadıkları, bunun yerine genellikle hazır grafikler verilerek verilerin okunması veya yorumlanması istenildiği sonucuna ulaşılabilir.

Öte yandan öğrencilerin açıklık tanımları ve değişim kavramını yorumlamaları karşılaştırmalı olarak incelendiğinde, bazı öğrencilerin açıklık kavramını tanımlayabildikleri ya da en azından açıklığın fark ile ilgili olduğunu bildikleri görülmüştür. Fakat bu öğrenciler aynı performansı değişim kavramının yorumlanmasında gösterememiştir. Elde edilen bu bulgu öğrencilerin açıklık kavramı ile ilgili işlemsel bir anlayışa sahip olduğunu gösterebilir çünkü öğrencilerin bir veri kümesinin açıklığını nasıl hesaplayacağını bilmelerine rağmen elde edilen farkın ne için kullanıldığı ile ilgili herhangi bir fikre sahip olmadıklarını göstermiştir. Bu öğrencilerin açıklığın değişim kavramının yorumlanmasında kullanılabilecek bir ölçü olduğunun farkında olmadıklarını göstermiştir. Bu bulgulardan, açıklık kavramının yorumlanmasının matematik öğretim programında yer almasına rağmen, çalışmaya katılan öğrencilerin sadece açıklık kavramının hesaplanmasına odaklandıkları sonucuna ulaşılabilir.

Öğrencilerin değişim kavramını değerlendirilmeleri incelendiğinde ise iki veri grubunun karşılaştırılmasını içeren sorularda ortalama kavramına nazaran daha fazla sayıda öğrencinin değişim kavramını eleştirel bir şekilde değerlendirebildiği görülmüştür. Diğer bir deyişle, Shaugnessy ve Pffannkuch (2002)'un çalışmasına paralel olarak çizgi ve sütun grafiklerinin iki veri grubunu karşılaştırmayı gerektiren durumlarda değişim kavramının değerlendirilmesine yardımcı olduğu görülürken bir tane veri kümesi olduğunda aynı bulguya ulaşılmamıştır. Bu durum öğrencilerin iki veri grubunu karşılaştırırken dikkatlerinin veri grupları içerisindeki değişime yönelmesi ile açıklanabilir. Başka bir şekilde ifade etmek gerekir ise öğrenciler veri grupları arasındaki değişimi incelerken veri grupları içerisindeki değişimi fark etmiş olabilir. Bu duruma tek veri kümesi içeren sorunun bağlamı da sebebiyet vermiş olabilir. Gal (2004) ve Watson (2006)'ın da belirttiği gibi eğer öğrenciler verilen bir sorunun bağlamını anlamadıysa, öğrencilerden değişim kavramını eleştirel bir şekilde değerlendirebilmesi beklenmemelidir.

Değişim kavramının değerlendirilmesinden elde edilen diğer bir bulgu ise ortalama kavramının değerlendirilmesinde gözlenen bir bulgu ile benzerlik göstermektedir. Değişim kavramının değerlendirilmesini içeren tüm sorularda çoğu öğrenci değerlendirmesini yapabilmek için sunulan veri kümesinden ya yalnızca bir değere ya da bir kaç değere odaklanmıştır. Fakat Watson (2006)'a göre bir veya birkaç değere odaklanma değişim kavramının değerlendirilmesinde bir başlangıç niteliğindedir. Dolayısıyla bu çalışmaya katılan öğrencilerin değişim kavramını verilen bağlamlarda eleştirel bir şekilde değerlendirmeye başladıkları söylenebilir.

Öte yandan, değişim kavramının değerlendirilmesinin iki veri grubunun karşılaştırılmasını gerektiren sorulardan elde edinilen bir diğer bulgu ise, veri grupları sütun grafiği üzerinde verildiğinde daha fazla sayıda öğrencinin değişim kavramını eleştirel bir şekilde değerlendirebilmesidir. Bu durumun bir sebebi sütun grafiğini içeren soruda değişim kavramının önemini belirten ifade olabilir. Soruda rüzgâr hızının dalgalı olmasının bakım masraflarını arttırdığı bilgisi öğrencilere sunulmuştur. Bu bilgi öğrencilerin rüzgar hızındaki değişimleri fark etmesine olanak sağlamış olabilir. Bir diğer sebep ise Capraro ve arkadaşlarının da (2005) belirttiği gibi öğrencilerin sütun grafiğine olan aşinalığı olabilir. Matematik öğretim programında 6. sınıfta yer alan bir kazanım da bu duruma sebebiyet vermiş olabilir. Bu kazanım öğrencilerin aritmetik ortalama ve açıklık kavramını iki veri grubunun karşılaştırılmasında kullanması gerektiğini belirtirken, karşılaştırmalarda kullanılan grafik türü sütun grafiğidir. Öğrencilerin bu tarz etkinliklere olan alışkanlığından dolayı çizgi grafiğine nazaran daha fazla sayıda öğrenci veriler sütun grafiği üzerinde sunulduğunda değişim kavramını eleştirel bir şekilde değerlendirmiş olabilir. Fakat sadece bir öğrenci sütun grafiğini içeren karşılaştırma sorusunda rüzgâr hızlarının açıklığından bahsetmiştir. Diğer öğrenciler veri gruplarındaki değişimden nicel olarak değil nitel olarak bahsetmiştir. Başka bir deyişle, öğrencilerin açıklık kavramını değişim kavramının yorumlanmasında kullanmadıkları gibi değerlendirilmesinde de kullanmadıkları görülmüştür. Bu bulgu, öğrencilerin sadece bir veri kümesinin açıklığını nasıl hesaplayacağını bilip elde edilen farkın ne için kullanıldığı ile ilgili herhangi bir fikre sahip olmadıklarından kaynaklı olabilir.

Son olarak, değişim kavramının iki veri grubunun karşılaştırılmasını gerektiren durumların incelendiği sorulardan ortalama kavramı ile ilgili de bir bulgu elde edilmiştir. Literatürdeki birçok çalışma öğrencilerin ortalama kavramını iki veri grubunu karşılaştırırken kullanmadığı sonucuna ulaşmıştır (Gal et al., 1989; Shaugnessy, 2003b; Watson & Moritz, 1999). Dolayısıyla bu çalışmada öğrencilerden öncelikle ortalamaları aynı fakat değişimleri farklı olan iki veri grubunun ortalamalarını bulmalarını istenmiş olup öğrencilerin ortalamanın aynı olmasını fark etmesinden sonra veri grupları içerisindeki değişimi değerlendirip değerlendiremedikleri incelenmiştir. Fakat çalışmaya katılan öğrencilerin çoğunun veri gruplarının ortalamalarını hesaplamalarına rağmen karşılaştırma yaparken kullanamadıkları, veri grupları içerisindeki birkaç değere göre karşılaştırmalarını yaptıkları görülmüştür. Literatürdeki bir çok çalışma da benzer bulguya ulaşmıştır (Gal et al., 1989; Konold & Pollatsek, 2002; Shaugnessy, 2003b; Watson & Moritz, 1999). Bu durum Gal (2005)'ın da belirttiği gibi öğrencilerin iki veri kümesini karşılaştırırken ortalama kavramını kullanmayı düşünmemelerinden kaynaklı olabilir. Ayrıca, Konold ve Pollatsek (2002)'in düşüncesiyle paralel olarak, bu durum öğrencilerin ortalama kavramını veri kümesindeki tüm verileri temsil ettiğini bilmemelerinden kaynaklı olabilir. Eğer bu çalışmadaki öğrenciler ortalamanın verilen bağlamdaki bütün veriyi temsil eden bir ölçü olduğunu bilselerdi (Mokros & Russell,1995), belki de iki veri grubunu karşılaştırırken kullanabilirlerdi.

Öneriler

Bu çalışma yedinci sınıf öğrencilerinin ortalama ve değişim kavramları ile ilgili istatistiksel okuryazarlıklarının sütun ve çizgi grafiğinde incelenmesine odaklanmıştır. Çalışmanın bulgularına dayanarak gelecek çalışmalar için bazı önerilerde bulunulabilir:

İlk olarak bu çalışmada uygun örnekleme yöntemi kullanılmıştır. Bulguların diğer yedinci sınıf öğrencilerine genellenebilmesi için seçkisiz örnekleme yöntemi kullanılarak elde edilen katılımcılarla benzer çalışmalar yapılabilir. Aynı zamanda ortaokul öğrencileri ile yapılacak olan boylamsal çalışmalar öğrencilerin ortalama ve değişim kavramı ile ilgili istatistiksel okuryazarlıklarının zaman içerisinde gelişip gelişmediğini göstermesi açısından faydalı olabilir. Ayrıca bu çalışma devlet okulu öğrencileri ile yapıldığından dolayı özel okul öğrencileri ile yapılacak çalışmalar okul türünün öğrencilerin istatistiksel okuryazarlıkları üzerindeki rolünü gösterebilir.

Öte yandan bu çalışmanın bulguları İOT'deki sorular ile sınırlıdır. Benzer çalışmalar farklı testler geliştirilerek veya İOT'nin farklı bağlamlar veya farklı grafik türleri ile geliştirilmesi ile tekrarlanabilir.

Bu önerilerin yanı sıra öğrenciler ile yapılacak klinik görüşmeler öğrencilerin düşünme süreçlerinin derinlemesine incelenmesine olanak sağlayabilir. Klinik görüşmeler vasıtası ile ortalama ve değişim kavramlarının yorumlanması veya değerlendirilmesi daha detaylı bir şekilde incelenebilir ve bu görüşmeler öğrencilerin ortalama ve değişim kavramlarını yorumlama ve değerlendirmede zorluk yaşamalarının muhtemel sebeplerini ortaya çıkarabilir.

Son olarak, bu çalışmanın analizinde kullanılan teorik çerçeve için bazı önerilerde bulunulabilir: Çalışmanın bulguları öğrencilerin yorumlarını veya değerlendirmelerini ortalama ve değişim kavramlarının tanımları ile ilgili kavramsal anlamalarını kullanarak yapamadıklarını göstermiştir. Başka bir deyişle, çalışmanın bulguları çoğu öğrencinin sunulan veri kümesinin sadece aritmetik ortalamasını veya açıklığını hesapladığını fakat neden bu değerleri hesapladıkları hakkında fazla bir bilgiye sahip olmadıklarını göstermiştir. Dolayısıyla öğrenciler yorumlarını veya değerlendirmelerini yaparken aritmetik ortalama veya açıklık kavramlarını kullanamamış, sorulardaki bağlamların veya grafiklerin yardımı ile ortalama veya değişim kavramlarını verilen bağlamlar içerisinde yorumlamış veya değerlendirebilmiştir. Fakat kullanılan çerçevede verilen tanımların kavramsal olması ile ilgili herhangi bir özellik bulunmamaktadır. Başka bir ifadeyle çerçeve, tanımın işlemsel veya kavramsal olmasına bakmaksızın her türlü tanımı kabul etmektedir. Bu sebeple teorik çerçevede yorumların ve değerlendirmelerin bağlamdaki bilgilerin yanı sıra ortalama ve değişim kavramı ile ilgili kavramsal anlamların da kullanılarak yapılması gerektiğini gösteren değişiklikler yapılabilir.