ASSESSMENT OF PRESERVICE MATHEMATICS TEACHERS' KNOWLEDGE FOR TEACHING STATISTICS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

OKTAY MERCİMEK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
SECONDARY SCIENCE AND MATHEMATICS EDUCATION

DECEMBER 2013

Approval of the thesis:

ASSESSMENT OF PRESERVICE MATHEMATICS TEACHERS' KNOWLEDGE FOR TEACHING STATISTICS

submitted by **OKTAY MERCİMEK** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Secondary Science and Mathematics Education Department, Middle East Technical University** by,

Prof. Dr. Canan ÖZGEN ____________________
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Ömer GEBAN ____________________
Head of Department, **Secondary Sci. and Math. Edu.**

Assoc. Prof. Dr. Ayhan Kürşat ERBAŞ ____________________
Supervisor, **Secondary Sci. and Math. Edu. Dept., METU**

Examining Committee Members:

Assoc. Prof. Dr. Erdinç ÇAKIROĞLU ____________________
Elementary Education Dept., METU

Assoc. Prof. Dr. Ayhan Kürşat ERBAŞ ____________________
Secondary Science and Mathematics Education Dept., METU

Assoc. Prof. Dr. Nihat BOZ ____________________
Secondary Sci. and Math. Edu. Dept., Gazi University

Assoc. Prof. Dr. Bülent ÇETİNKAYA ____________________
Secondary Science and Mathematics Education Dept., METU

Assist. Prof. Dr. Ömer Faruk ÖZDEMİR ____________________
Secondary Science and Mathematics Education Dept., METU

Date: 11.12.2013

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 : Oktay MERCİMEK

Signature :

ABSTRACT

ASSESSMENT OF PRESERVICE MATHEMATICS TEACHERS' KNOWLEDGE FOR TEACHING STATISTICS

Mercimek, Oktay

Ph.D., Department of Secondary Science and Mathematics Education

Supervisor: Assoc. Prof. Dr. Ayhan Kürşat ERBAŞ

December 2013, 187 pages

The purpose of this study is to assess preservice teachers' mathematical knowledge for teaching statistics (MKT-S). For this purpose, MKT-S instrument consisting of two dimensions, 'content knowledge' (CK) and 'pedagogical content knowledge' (PCK) was developed, and applied to 659 preservice middle school mathematics teachers (PTs).

Confirmatory factor analysis showed that CK and PCK are two different dimensions of mathematical knowledge of teaching statistics. It was found that CK factor scores were highly correlated with PCK factor scores. The reliability levels were 0.65 for CK factor scores and 0.76 for PCK factor scores.

Analysis of CK items revealed that only some PTs were able to (a) evaluate center of data has extreme cases; and (b) construct a histogram from data that has decimal numbers.

Analysis of PCK items also revealed that nearly less than a quarter of PTs were able to (a) see the connections between different types of graphics; (b) recognize an alternative correct approach for handling a statistics problem; (c) offer correct examples for both arithmetical mean and median; (d) identify both logical and illogical parts of a student's answer; (e) diagnose why a students' error occurred; and (f) provide feedback that targeted to solve students' misunderstanding.

MKT-S instrument developed in this study has several implications for teacher education. MKT-S instrument can be used to evaluate efficiency of PTs' mathematical knowledge for teaching statistics. Instrument can be adapted for in-service teachers. Deficiencies revealed in this study can be used develop effective theoretical statistics courses and teaching statistics courses.

Keywords: Preservice mathematics teachers, knowledge for teaching mathematics, content knowledge, pedagogical content knowledge, statistics, averages, graphs

ÖZ

MATEMATİK ÖĞRETMENİ ADAYLARININ İSTATİSTİK ÖĞRETİMİNE YÖNELİK BİLGİLERİNİN DEĞERLENDİRİLMESİ

Mercimek, Oktay

Doktora, Orta Öğretim Fen ve Matematik Alanları Eğitimi Bölümü

Tez Yöneticisi: Doç. Dr. Ayhan Kürşat Erbaş

Aralık 2013, 187 sayfa

Bu çalışmanın amacı matematik öğretmen adaylarının istatistik öğretimine yönelik matematiksel bilgilerinin (MKT-S) ölçülmesidir. Bu amaçla, ‘alan bilgisi’ (CK) ve ‘pedagojik alan bilgisi’ (PCK) olmak üzere iki bölümden oluşan MKT-S ölçeği geliştirilmiş ve 659 ilköğretim matematik öğretmeni adayı (ÖA) üzerinde uygulanmıştır.

Doğrulayıcı faktör analizi, CK ve PCK faktörlerinin, istatistik öğretimine yönelik matematiksel bilginin iki ayrı boyutu olduğunu göstermiştir. CK puanlarının PCK puanlarıyla yüksek derecede ilişkisinin olduğu saptanmıştır. Ayrıca CK faktör puanlarının güvenirliği 0,65 düzeyinde iken PCK faktör puanlarının güvenirliği 0,76 düzeyindedir.

CK maddelerinin analizi sonucunda sadece bazı ÖA'larının (a) veride olağan dışı ölçümler olduğu durumlarda verinin merkezini doğru değerlendirebildikleri ve (b) ondalık sayılar içeren bir verinin histogram grafiğini oluşturabildikleri bulunmuştur.

PCK maddelerinin analizi sonucunda ise ÖA'larının dörtte birinden azının (a) farklı grafik tipleri arasındaki bağlantıyı görebildikleri; (b) bir istatistik probleminin alternatif doğru cevabını fark edebildikleri;(c) hem aritmetik ortalama hem de ortanca konuları için doğru örnekler önerebildikleri; (d) bir öğrencinin cevabındaki hem doğru hem de yanlış yönleri saptayabildikleri; (e) bir öğrenci hatasının neden kaynaklandığını açıklayabildikleri; ve (f) öğrencideki yanlış anlamaları giderebilecek şekilde geri dönüt sağlayabildiklerini bulunmuştur.

Bu çalışmada geliştirilen MKT-S ölçeğinin sonuçlarının öğretmen eğitimi için çeşitli çıkarımları bulunmaktadır. Geliştirilen ölçek ÖA'larının istatistik öğretimine yönelik matematiksel bilgilerinin etkinliğini değerlendirmede kullanılabilir ya da çalışan öğretmenlere yönelik adapte edilebilir. Çalışmada ortaya çıkarılan eksiklikler, daha etkili teorik istatistik dersleri ve istatistik öğretimi derslerinin geliştirilmesinde kullanılabilir.

Anahtar Kelimeler: Matematik öğretmen adayları, alan bilgisi, pedagojik alan bilgisi, istatistik, ortalamalar, grafikler

To My Wife
Deniz ÇAĞLAR MERCİMEK

And

To My Son
Kerem Türkay MERCİMEK

ACKNOWLEDGEMENTS

I would like to thank my supervisor Assoc. Prof. Dr. Ayhan Kürşat ERBAŞ for his encouraging support and guidance throughout the study. He always supported me, in spite of his intensive work program.

I also would like to thank Assoc. Prof. Dr. Erdinç ÇAKIROĞLU and Assoc. Prof. Dr. Bülent ÇETİNKAYA for their comments, suggestions and contributions throughout the study.

I also thank Assoc. Prof. Dr. Nihat BOZ and Asst. Prof. Dr. Ömer Faruk ÖZDEMİR for valuable comments for this study.

Special thanks to Prof. Dr. Ahmet KAÇAR, İbrahim KEPÇEOĞLU and Abdülkadir ÖNER for their great support during data collection process, and to Tevfik KÜÇÜKALİOĞLU for his contributions during instrument development process.

I also wish to express my deep gratitude to the preservice middle school mathematics teachers who voluntarily participated in my study only for the sake of science because without them this dissertation would not exist.

I also wish to thank Lütfi İNCİKABI and Eralp BAHÇIVAN for reviewing early drafts of this manuscript and sharing their valuable opinions with me.

Last but not the least; my deepest thanks go to my wife and my son. I always felt their all sweetness and light throughout my study.

TABLE OF CONTENTS

ABSTRACT ... v
ÖZ ... vii
ACKNOWLEDGEMENTS ... x
TABLE OF CONTENTS ... xi
LIST OF TABLES ... xv
LIST OF FIGURES ... xvii
LIST OF ABBREVIATIONS ... xix
CHAPTERS
1 INTRODUCTION ... 1
1.1 Problem Statement ... 4
1.2 Significance of the Study ... 5
1.3 Definitions of Important Terms ... 7
2 LITERATURE REVIEW ... 9
2.1 Defining Teacher Knowledge ... 9
2.2 Assessing Teacher Knowledge ... 15
2.3 Turkish Preservice Middle School Mathematics Teachers' Opportunities for Learning and Teaching Statistics ... 20
2.4 Statistics Knowledge of Students and Teachers: Misconceptions and Difficulties ... 22
2.5 Summary of Literature Findings ... 27
3 METHODOLOGY ... 29
3.1 Research Design ... 29
3.2 Instrument Development ... 29
3.2.1 Framework of the Instrument ... 30
3.2.2 Content Selection Process of the Instrument ... 32
3.2.3 Limitations on Content Coverage ... 35
3.2.4 Preparing Test Blueprint and Developing Items ... 36

3.2.5 Sources for Item Development 37
3.2.6 Item Trial 39
3.2.7 Pilot Testing of Instrument 40
3.3 Data Collection and Analysis 43
3.3.1 MKT-S Instrument 43
3.3.2 Population and Sample 45
3.3.3 Data Recording and Scoring Method for MKT-S Instrument 46
3.3.3.1 Qualitative Data Recording 46
3.3.3.2 Developing Scoring Rubric for Final Implementation. 47
3.3.3.3 Analysis of Missing Data 48
3.3.4 Validity and Reliability Evidences for MKT-S instrument 48
3.3.5 Factor Analysis of MKT-S Instrument 50
4 RESULTS 53
4.1 Validation of the Mathematical Knowledge for Teaching Statistics Instrument 53
4.1.1 Confirmatory (Item) Factor Analysis Results of MKT-S 53
4.1.2 Model Improvement 55
4.1.3 Concurrent Validity Evidences for MKT-S instrument 58
4.1.4 Psychometric Properties of MKT-S Instrument 60
4.1.4.1 Reliability 60
4.1.4.1.1 Reliability of CK Factor 61
4.1.4.1.2 Reliability of PCK Factor 62
4.1.4.2 IRT Parameters for MKT-S Instrument 63
4.1.4.2.1 Item Parameters for CK Items 63
4.1.4.2.2 Item Parameters for PCK Items 63
4.1.5 Relationship Between Content Knowledge Scores and Pedagogical Content Knowledge Scores 65

4.2 Preservice Middle School Mathematics Teachers' Content Knowledge in Statistics ... 66
4.2.1 Preservice Teachers' Content Knowledge of Central Tendency ... 66
4.2.2 Preservice Teachers' Content Knowledge of Graphics ... 70
4.3 Prospective Middle School Mathematics Teachers' Pedagogical Content Knowledge in Statistics ... 78
4.3.1 Preservice Teachers' Ability to See Connections (PCK-1) ... 79
4.3.2 Preservice Teachers' Ability to Evaluate a Student's Correct Work (PCK-2) ... 80
4.3.3 Preservice Teachers' Ability to Select Appropriate Examples (PCK-3) ... 84
4.3.4 Preservice Teachers' Ability to Evaluate Students' Arguments (PCK-4) ... 87
4.3.5 Preservice Teachers' Ability to Diagnose Misconceptions (PCK-5) ... 90
4.3.6 Preservice Teachers' Ability to Provide Feedback (PCK-6) ... 93
4.4 Summary of Results ... 96
5 DISCUSSION, CONCLUSION AND IMPLICATIONS ... 99
5.1 Discussion and Conclusion ... 99
5.2 Limitations of the Study ... 105
5.3 Implications ... 106
5.4 Recommendations for Future Research ... 108
REFERENCES ... 109
APPENDICES
A. BLUEPRINT AND COMPLETE ITEMS FOR ITEM BANK ... 117
B. PERMISSION LETTER FROM JUAN ANTONIO GARCÍA CRUZ ... 125
C. COMPLETE MKT-S INSTRUMENT FOR PILOT IMPLEMENTATION ... 127
D. BLUEPRINT AND COMPLETE MKT-S INSTRUMENT FOR FINAL IMPLEMENTATION ... 135
E. MODEL FIT INFORMATION FOR MODEL I AND MODEL II USING MLR ... 143
F. COMPLETE LIST OF RESULTS FOR ITEM F.1C ... 149

G. ORIGINAL SAMPLE ANSWERS FOR ITEM F.1C 151
H. ORIGINAL SAMPLE ANSWERS FOR PCK-2 159
I. ORIGINAL SAMPLE ANSWERS FOR PCK-3 167
J. ORIGINAL SAMPLE ANSWERS FOR PCK-4 175
K. ORIGINAL SAMPLE ANSWERS FOR PCK-5 179
L. ORIGINAL SAMPLE ANSWERS FOR PCK-6 183
CURRICULUM VITAE 187

LIST OF TABLES

TABLES

Table 2.1. Sub-domains and Objectives of PCK of TEDS-M 13

Table 2.2. Summary Table on Definitions for Pedagogical Content Knowledge...... 14

Table 2.3. Primary Items by Sub-domains and Blocks of TEDS-M.......................... 18

Table 2.4. Lower Secondary Items by Sub-domains and Blocks of TEDS-M 18

Table 2.5. Elementary Mathematics Teacher Education Program in Turkey 21

Table 3.1. Cognitive Knowledge Domains for Content Knowledge 31

Table 3.2. Sub-domains and Objectives of PCK used in TEDS-M 32

Table 3.3. Statistics Content Standards from 2005 Turkish National Curriculum 33

Table 3.4. Comparison of Statistics Content Standards for Middle School Grades .. 34

Table 3.5. Test Blueprint and Item Distribution 37

Table 3.6 . Distribution of Preservice teachers' Answer for Item P.2. 41

Table 3.7 . Cross-tabulation of Results of P.13A and P.13B 42

Table 3.8. Distribution of Number of Not Reached Items for Two Booklets............ 43

Table 3.9. Distribution of All Items in Final Implementation. 44

Table 3.10. Distribution of the Items to Four Blocks 44

Table 3.11. Summary of Three Booklets in Final Implementation............................ 45

Table 3.12. Distribution of Sample 46

Table 4.1. Fit Indices for Model I and Model II for MLR Estimator. 54

Table 4.2. Fit Indices for Model I and Model II for WLSMV Estimator. 55

Table 4.3. Modification Indices for Model II. 55

Table 4.4. Fit Indices for Model IIA and Model IIB for WLSMV Estimator. 56

Table 4.5. Results for Comparing Chi-square Values.. 56

Table 4.6. Comparison of Results with TEDS-M Study.. 57

Table 4.7.Correlation Between Factor Scores and ISP-I Grades. 58

Table 4.8. Tests of Normality 59

Table 4.9. Descriptive Statistics and t-test Results 60

Table 4.10. IRT Parameters for CK Items 63
Table 4.11. IRT Parameters for PCK Items 64
Table 4.12. Preservice Teachers Estimates for Item F.1C 67
Table 4.13. Some Arguments of Preservice Teachers who Defend Arithmetical Mean 68
Table 4.14. Some Arguments of Preservice Teachers who Mistakenly Consider "Value of 13" 68
Table 4.15. Preservice Teachers who Explicitly Discarded the "Value of 13" 70
Table 4.16. Preservice Teachers' Score Distribution for Item F.10 71
Table 4.17 Distribution of Preservice Teacher's Correct Responses for Item F.2..... 79
Table 4.18. Data Used in Item F.11 81
Table 4.19. Preservice Teachers' Score Distribution for Item F.11 81
Table 4.20. Some Answer Examples of Preservice Teachers for Item F.11. 82
Table 4.21. Preservice Teachers' Score Distribution for Item F.3 85
Table 4.22. Graphics Selected by Preservice Teachers and Their Reasons 85
Table 4.23. Preservice Teachers' Score Distribution for Item F.4 88
Table 4.24. Preservice Teachers' Score Distribution for Item F.5A. 91
Table 4.25. Preservice Teachers' Score Distribution for Item F.5B. 94

LIST OF FIGURES

FIGURES

Figure 2.1. The Network of Knowledge of Effective Teaching 10

Figure 2.2. Domain Map for Mathematical Knowledge for Teaching 12

Figure 2.3. Framework and an Example for Observing Pedagogical Knowledge..... 19

Figure 2.4. A Graphical Representation of SOLO Taxonomy.................................. 23

Figure 2.5. Items from Garcia, Cruz and Garret 24

Figure 2.6. Bivarite Graph from Taylor’s (1993) Study 27

Figure 3.1. Item P.2 from Pilot Instrument. 41

Figure 3.2. Item Response Function of P.2 42

Figure 3.3. Qualitative Data Recording Cycle for Each Open-ended Item 47

Figure 3.4. Proposed Models for MKT-S Instrument 51

Figure 4.1 Standardized Loadings for Model I and Model II Using MLR Estimator. 54

Figure 4.2. Standardized Factor Loadings for Model IIB. 57

Figure 4.3. Histograms for CK and PCK. 59

Figure 4.4. IRT Score Distribution and Information Function for CK Scores. 61

Figure 4.5. IRT Score Distribution and Information Function for PCK Scores. 62

Figure 4.6. Scatterplot of CK and PCK Scores 65

Figure 4.7. Translated Version of Item F.1C 66

Figure 4.8. Translated Version of the Item F.10 71

Figure 4.9. Interpreting Data as Bivariate 72

Figure 4.10. Frequency Polygon without Intervals 72

Figure 4.11. Separate Bars without Intervals 72

Figure 4.12. Adjacent Bars without Intervals 73

Figure 4.13. A Graph that is Similar to Greatest Integer Function 73

Figure 4.14. Frequency Polygon and Separate Bars (interval length=1) 73

Figure 4.15. Increasing Bars with Equal Measurement in Each Interval (Error 1) ... 74

Figure 4.16. Increasing Bars with Equal Measurement in Each Interval (Error 2).... 74
Figure 4.17. Increasing Bars with Equal Measurement in Each Interval (Error 3).... 74
Figure 4.18. Equal Separate Bars with Equal Measurement in Each Interval 75
Figure 4.19. Equal Adjacent Bars with Equal Measurement in Each Interval........... 75
Figure 4.20. Histogram with Four Intervals 76
Figure 4.21. Histogram with Six Intervals 76
Figure 4.22. Histogram Forced to Have Five Intervals.. 77
Figure 4.23. Separated Bars with Five Intervals 77
Figure 4.24. Sample for Correct Histogram... 78
Figure 4.25. Teacher and Students' Way of Thinking from F.11 Item Stem............. 81
Figure 4.26. Translated Version of the Item F.3. 84
Figure 4.27. Translated Form of Item F.4. 87
Figure 4.28. Translation of Stem for Item F.5A and F.5B Including Item F.5A 90

LIST OF ABBREVIATIONS

TIMSS	Trends in International Mathematics and Science Study
PCK	Pedagogical Content Knowledge
CK	Content Knowledge
MNE	Ministry of National Education, Turkey
NCTM	National Council of Teachers of Mathematics, USA
MKT-S	Mathematical Knowledge for Teaching Statistics
IRT	Item Response Theory
MIRT	Multidimensional Item Response Theory
CCSSI	Common Core State Standards Initiative
CCSS	Common Core State Standards
MoE	Ministry of Education, Singapore
CHE	Council of Higher Education, Turkey
ISP-I	Introduction to Statistics and Probability-I
TEDS-M	Teacher Education and Development Study in Mathematics
NAEP	National Assessment of Educational Progress
MT21	Mathematics Teaching in the 21st Century
MLR	Maximum Likelihood Estimation With Robust Standard Error
WLSMV	Robust Weighted Least Square Estimator
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
RMSAE	Root Mean Square Error of Approximation
TLI	Tucker-Lewis Index
CFI	Comparative Fit Index
WRMR	Weighted Root-Mean-Square Residual

CHAPTER 1

INTRODUCTION

> Statistics ... "the most important science in the whole world: for upon it depends the practical application of every other science and of every art; the one science essential to all political and social administration, all education, all organization based upon experience, for it only gives the results of our experience."
>
> -Florence Nightingale

As Moore (1998) point out “Statistics is a general intellectual method that applies wherever data, variation, and chance appear. It is a fundamental method because data, variation, and chance are omnipresent in modern life” (p. 1254). Since the statistics is in everywhere from newspapers to television to inform people (Boslaugh & Watters, 2008), it is also can be used to mislead people’ decisions (Huff, 1954). People are easily fascinated by the amount of money they may win from a lottery but they are not informed about the chance of winning (Utts, 2003). Public opinion poll results are another example of the statistics that can manipulate the decisions of less educated citizens. Different companies can publish different results on the same issue depending on the choice of sample and statistical method (Balcı & Ayhan, 2004). Even advertisements can have misleading statistical results (Ha, 2012; Huff, 1954). Therefore, statistics is required for every person to take advantage of the full rights of citizenship (Franklin et al., 2005), and “a working knowledge of statistics is the best check against the proliferation of misleading or outright false claims” (Boslaugh & Watters, 2008, p. xiii)

Even though importance of statistics is much clear in this century, statistics has considered as a subject that supposed to be taught at college level before 1970s. During that period, several organizations spread the word about the importance of statistical reasoning in elementary and middle school level (Cooper, 2002). Putting more emphasis on statistical reasoning on every level of education started globally in the early 1990s (Watson, 2006). This global movement also affected Turkish mathematics curriculum, and Ministry of National Education revised mathematics curriculum from scratch in 2005 (Babadoğan & Olkun, 2006). New concepts that never taught before added to middle school mathematics curriculum, and these concepts were especially apparent on the 'Statistics and Probability' section of the new curriculum (MNE, 2005a). However, this global implementation also increased the global awareness about issues related to teaching and learning statistics. This implementation did not cause a complete working knowledge of statistics, and deficiencies in the conceptual understanding of statistics detected by researchers globally for students (Toluk Uçar & Akdoğan, 2008; Garcia, Cruz & Garret, 2008), preservice teachers (Bruno & Espinel, 2009; Moneiro & Ainley, 2006) and in-service teachers (Russell & Mokros, 1990). Therefore, one can say that there is a never-ending deficiency cycle for statistics knowledge in education systems. Deficiencies of students continue when they become pre-service teachers (Taylor, 1993; Cooper, 2002). College education do not solve the issue and the deficiencies, even for simple statistics concepts such as arithmetical mean, also apparent for in-service teachers (Russel and Mokros, 1990).

It could be difficult to break this cycle. However, one way to weaken this cycle is to understand the faulty ingredients of the preservice teachers' knowledge to teach statistics. This understanding requires an appropriate test to assess the preservice teachers' knowledge to teach statistics, and building this kind of test requires defining teacher knowledge.

Some researchers assumed that the content knowledge is equivalent to teaching knowledge; it can be acquired through traditional undergraduate mathematics courses; and this knowledge is the only knowledge to teach mathematics (Ball & Wilson, 1990). However, Cobb (1992) warns us that the undergraduate statistics courses may not be successful to support the statistics knowledge of preservice teachers:

> “Basic [*statistics*] concepts are hard, misconceptions persistent. As [*university*] teachers, we consistently overestimate the amount of conceptual learning that goes on in our courses, and consistently under-estimate the extent to which misconceptions persist after the course is over.” (p. 10, italics added by researcher for clarification)

Shulman (1986) attempted to clarify knowledge that is required to teach a specific subject, subject matter knowledge, and he described this knowledge as having three categories: subject matter content knowledge (or content knowledge), pedagogical content knowledge and curriculum knowledge. His subject matter content knowledge refers to facts, concepts, and theorems in a domain. It also includes why a fact or concept is true, and how knowledge is generated in the domain. He described the pedagogical content knowledge (PCK) as “goes beyond knowledge of subject matter per se to the dimension of subject matter knowledge for teaching” (p. 9). Shulman’s third category, curriculum knowledge, includes understanding why a topic is included in curriculum, why we teach it in a certain level and other alternative curriculum materials to teach.

Even though Shulman’s definition (1986, 1987) for teacher knowledge, especially pedagogical content knowledge (PCK) has become very popular among mathematics education researchers, most of the research targeted inservice teachers. Standards also established for describing the qualities of in-service teachers (MNE, 2007; NCTM, 2000). However, recent attempt for theorizing the teacher knowledge of preservice teachers came as international comparative studies, MT21 (Schmidt et al., 2007) and TEDS-M (Tatto et al., 2008). These studies influenced mainly by theory of Shulman (1986, 1987) and adapted the work of Fan and Cheong (2002). These research projects hypothesized the preservice teacher knowledge (mathematical knowledge for teaching) consisting of two dimensions: (mathematical) content knowledge and (mathematics) pedagogical content knowledge. They also hypothesized that pedagogical content knowledge has at least three components: Curricular knowledge; knowledge of planning for mathematics teaching and learning (pre-active); and enacted mathematics knowledge for teaching and learning (interactive).

These international studies (MT21 and TEDS-M) significantly contributed to PCK definition for preservice teachers. Researcher not just identified three components of PCK but also defined each component operationally using expected objectives. Even though some form of these objectives independently studied by researchers, researchers managed to refine objectives for preservice teachers, and these objectives have potential to be implemented in teacher education programs as standards.

Assessment method for the teacher knowledge is also as important as the theorizing the teacher knowledge. In general, teacher knowledge (CK, PCK or both) assessment studies use qualitative research designs (Baxter & Lederman, 1999). In these studies, researchers generally use classroom observations, interviews and lesson plans as data sources regarding assessment. Because of the design of these studies, it was possible to examine teacher knowledge structures only for a few teachers. Even though these studies supplied hints about the structures of particular teachers' knowledge, their samples were shallow to paint the general picture of teacher knowledge. The nature of the qualitative designs, which used classroom observations or teacher interviews, is also not suitable for large-scale assessment because these studies require long-term observations and need high-quantity investment. On the other side, large scale studies (Hill et al., 2008, Krauss et al., 2008; Tatto et al., 2008) concentrated on quantitative results of the instruments. They contributed mainly to methodological aspects of assessing teacher knowledge in large-scale settings. However, these studies did not make any efforts to discuss the qualitative findings related to deficiencies in teacher knowledge.

1.1 Problem Statement

Kleickmann et al. (2013) stated that "despite the importance attributed to teachers' knowledge of subject matter, the understanding of how the learning opportunities available during teacher education and professional development affect the development of subject-specific knowledge is still limited" (p. 92). In order to understand development of subject-specific knowledge, researchers need to assess the current state of the teacher knowledge. However, assessing the teacher knowledge also requires developing qualitative and/or quantitative assessment tools to describe teacher knowledge.

Therefore, primary purpose of this study was to assess preservice middle school mathematics teachers' mathematical knowledge of teaching statistics, understanding the relationship between its components, and investigating the adequacy of this knowledge in order to paint the general picture for preservice teachers' mathematical knowledge for teaching statistics.

To fulfill the primary purpose, secondary purpose of the study is developing an instrument, which is reliable, valid for preservice teachers and suitable for large-scale assessment, to assess mathematics knowledge of teaching statistics. The following questions guided the study:

1. Will the instrument developed in this study be valid and reliable for measuring preservice middle school mathematics teachers' mathematical knowledge for teaching statistics concepts, specifically averages and graphs?
2. What kinds of deficiencies do preservice teachers have in their content knowledge regarding middle school statistics concepts, specifically averages and graphs?
3. What kinds of deficiencies do preservice teachers have in their pedagogical content knowledge regarding middle school statistics concepts, specifically averages and graphs?

1.2 Significance of the Study

Hill, Ball, Sleep and Lewis (2007) stated that there are three important contemporary reasons to call out for a system of teacher assessment that is both professionally relevant and broadly credible. These are political demand that students should be taught by qualified teachers, establishing evidence on the effects of the teacher education programs, and distinguishing what makes teachers professional that is the professional knowledge and skills not possessed by any educated person.

However, this study will not just add valuable information for instrument development efforts about assessing teacher knowledge but also will contribute to mathematics education literature as a large-scale study on teacher knowledge (Adler et al., 2005). Methodology and results of this research will guide future studies about developing and validating instruments for teacher knowledge. It will also make it possible to compare the structures of teacher knowledge for different cultures by comparing both quantitative and qualitative findings of MKT-S instrument.

Researchers have stated a need for assessing mathematics teachers' pedagogical content knowledge and the need for an assessment tool (Ball & McDiarmid, 1988; Grouws & Schultz, 1996). The instrument constructed by researcher will help to researchers who will conduct experimental research on developing pedagogical content knowledge. Without a proper measure for teacher knowledge, many research and development efforts on teacher education will have limited aspects. For example, exploring the effects of any professional development program, i.e. for increasing the quality of instruction, needs longitudinal studies to see effects of the professional development program on teachers' knowledge and student gaining (Hill et al., 2004).

Another benefit of measuring *mathematical knowledge for teaching* is to assess whether teacher education programs are successful for providing knowledge required to teach statistics (Kleickmann, 2013). In the current mathematics teacher education program in Turkey, on average, there are three or four courses that may help preservice teachers to gain knowledge for teaching statistic. Even though it is naturally assumed that preservice will gain enough knowledge to teach statistics when passed those courses, researchers state that even highly educated adults have problems with statistical concepts (Utts, 2003). Therefore, this assumption needs further testing, and this assumption cannot be tested without a proper measurement tool.

Large-scale teacher knowledge assessment studies (Krauss et al., 2008; Tatto et al., 2008) also bring certain issues that need be resolved. First of all, these large-scale studies covered general mathematics content domain, and each sub-domain (e.g.

geometry, algebra, statistics) measured with a few items. Therefore, researchers were not able make conclusion for a specific sub domain. Second issue is also related to general coverage of mathematics domain. Statistics educators state the differences between mathematical reasoning and statistical differences, and delMas (2004) argue that "Statistics may be viewed as similar to disciplines such as physics that utilize mathematics, yet have developed methods and concepts that set it apart from mathematical inquiry" (p. 84). Therefore, conclusions about teacher knowledge that are drawn on general mathematics domain may not reflect the situation for knowledge for teaching statistics. These large-scale studies also focus mostly on psychometric properties of the teacher knowledge instruments and their reports lack the information about teacher knowledge itself. So, this study will try to resolve these issues by conducting a large-scale study that specifically targets knowledge for teaching statistics, and reporting not just quantitative but also qualitative findings for knowledge for teaching statistics of preservice teachers.

1.3 Definitions of Important Terms

Mathematical Knowledge for Teaching (MKT): Defined as "comprising two main subsets of knowledge: mathematical content knowledge and pedagogical content knowledge" (Tatto et al., 2008, p. 20)

Content Knowledge (CK): set of fundamental assumptions, definitions, concepts, and procedures. It also corresponds to 'subject matter content knowledge' dimension of Shulman (1986).

Pedagogical Content Knowledge (PCK): "It represents the blending of content and pedagogy into an understanding of how particular topics, problems, or issues are organized, represented, and adapted to diverse interests and abilities of learners, and presented for instruction" (Shulman, 1987, p. 8). It also includes "… representations most useful for teaching an idea and learners' typical errors and misconceptions" (Hill, Schiling, & Ball, 2004, p.12).

Preservice Middle School Mathematics Teachers: Preservice teachers who are enrolled in a primary school mathematics education department of an education faculty and will be eligible to teach in Grade 5 through Grade 8 in Turkey. Middle School mathematics teachers also called lower secondary mathematics teachers in some cultures.

CHAPTER 2

LITERATURE REVIEW

This section starts with a report for discussions on teacher knowledge and methods to assess teacher knowledge. Then, opportunities for Turkish preservice teachers to master their skills for middle school statistics content and teaching statistics was reported. After that, current research on statistics knowledge of students and teachers were presented. Finally, a summary of literature findings was supplied.

2.1 Defining Teacher Knowledge

Researchers and teacher educators need to clarify what knowledge teachers require to teach effectively. Lee Shulman was one of the pioneers that started theorizing teacher knowledge (Shulman, 1986; Shulman, 1987). Shulman attempted to describe the knowledge that is required to teach a specific subject and he described the teaching knowledge as having three categories: Subject matter content knowledge, pedagogical content knowledge (PCK) and curriculum knowledge. His subject matter content knowledge refers to facts, concepts, and theorems in a domain. It also includes why a fact or concept is true, and how knowledge is generated in the domain. He describes the PCK as

> … goes beyond knowledge of subject matter per se to the dimension of subject matter knowledge for teaching
> …the most useful forms of representations of ideas, the most powerful analogies, illustrations, examples and demonstrations, …an understanding of what makes the learning of specific topics easy or difficult: the conceptions and preconceptions that students of different ages and backgrounds bring with them to the learning (Shulman, 1986, p.9).

Shulman's third category, curriculum knowledge, includes understanding why a topic is included in curriculum, why we teach it in a certain level, and other alternative curriculum materials to teach.

An, Kulm and Wu (2004) defined teacher knowledge as the knowledge of effective teaching. According to them, knowledge of effective teaching consisted of three sub-dimensions; namely knowledge of content, knowledge of curriculum, and knowledge of teaching. They also placed 'teaching' (similar to PCK in Shulman's definition) as a core component of knowledge of effective teaching. The teaching dimension also had five sub-scales: knowing students' thinking, building on students' math ideas, promoting students' thinking, addressing students' misconception and engaging students in mathematics learning (see Figure 2.1). They also assert that teachers' beliefs impacts network of teachers' knowledge.

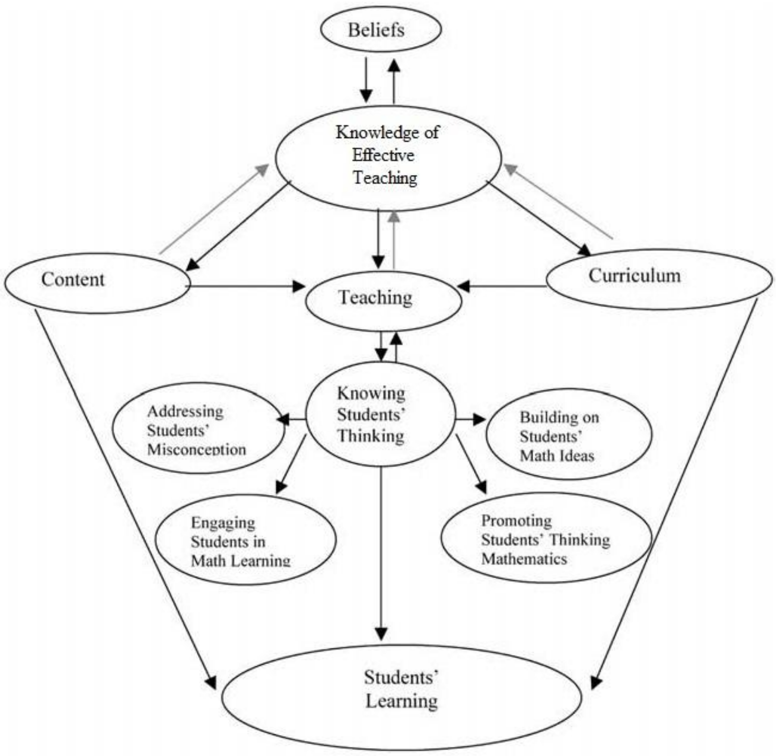


Figure 2.1. The Network of Knowledge of Effective Teaching (Adapted from An, Kulm & Wu, 2004)

Another attempt to distinguish between different components of teacher knowledge, namely PCK and content knowledge (subject matter content knowledge in Shulman's definition), came from German researchers (Krauss, Brunner, Kunter, Baumert, Blum, Neubrand, & Jordan, 2008). They conducted a study on 198 secondary mathematics teachers to explore the relationship between the PCK and content knowledge (CK). Their study also compared teachers with respect to their teacher training program which qualifies them whether to teach in Gymnasium (GY), an academic track, or non-Gymnasium, e.g., Realschule, Sekundarschule.

Their PCK test consisted of three subscales:

i. Task, knowledge of mathematical tasks
ii. Student, knowledge of student misconceptions and difficulties
iii. Instruction, knowledge of mathematics specific instructional strategies.

The study mainly resulted that GY and NGY teachers differed in their both PCK and CK level. Moreover, they found that cognitive connectedness, latent correlation between CK and PCK, is dependent on the level of mathematical expertise. Even though loadings for indicators were not significantly different, the latent correlation between PCK and CK was 0.61 in the NYG group and 0.96 for the GY group. Very strong relationship between PCK and CK in the GY group raised the question whether PCK and CK is separable constructs for these highly knowledgeable teachers. Another result was that PCK and CK form one body of connected knowledge that almost indistinguishable in the group of GY teachers. However, for the NYG group PCK and CK categories were separate constructs. Their results may imply that it is very difficult to construct CK or PCK items for highly knowledgeable teachers. For example, a highly knowledgeable teacher may offer more than one approaches to handle a mathematical task using his/her deeply connected content knowledge without thinking pedagogical aspects of the task.

An additional attempt for conceptualizing the mathematics teacher knowledge (see Figure 2.2) comes from Hill, Ball & Schilling (2008) as a product of their progress on measuring mathematical knowledge for teaching (Ball, 2002; Hill, Schilling & Ball, 2004; Hill, Rowan, Ball, 2005). In their 2008 article, Hill, Ball, & Schilling defined teacher knowledge (mathematical knowledge of teaching) having two major dimensions: Subject matter knowledge and pedagogical content knowledge. Their subject matter knowledge included 'common content knowledge', knowledge at the mathematical horizon' and 'specialized content knowledge'. Their pedagogical content knowledge also included three subdimensions: 'Knowledge of content and students (KCS)', 'knowledge of content and teaching (KCT)', and knowledge of curriculum'. They define KCS as "content knowledge intertwined with knowledge of how students think about, know or learn this particular content" (Hill at al., 2008, p. 375).

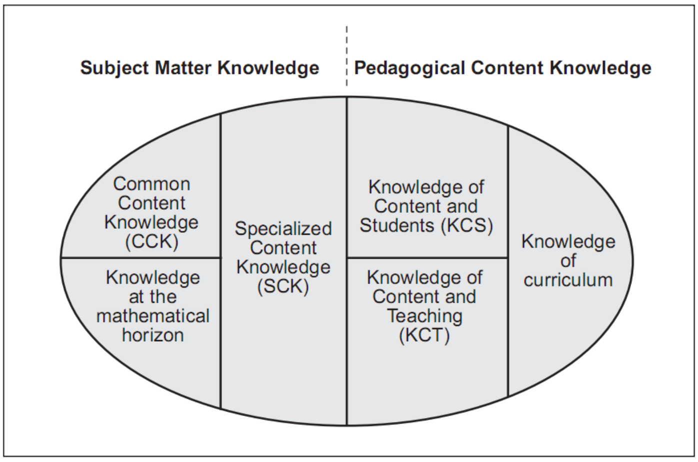


Figure 2.2. Domain Map for Mathematical Knowledge for Teaching (Hill, et al., 2008; p. 337)

Hill et al.'s (2008) domain map for mathematical knowledge for teaching implies that subject matter knowledge and pedagogical content knowledge are separate constructs, and more importantly 'Knowledge of Content and Teaching' and 'Knowledge of Content and Students' can be independently observable from each other.

Even though teacher knowledge studies generally focused on inservice teachers, recent attempt for theorizing the teacher knowledge of preservice teachers came as international comparative studies, MT21 (Schmidt et al., 2007) and TEDS-M (Tatto et al., 2008). These studies influenced mainly by theory of Shulman (1986, 1987) and adapted the work of Fan and Cheong (2002). These research projects hypothesized the mathematical knowledge for teaching having two dimensions: CK and PCK. They also hypothesized that PCK has at least three components: curricular knowledge; knowledge of planning for mathematics teaching and learning (pre-active); and enacted mathematics knowledge for teaching and learning (interactive). Table 2.1 shows the objectives for components of PCK.

Table 2.1. Sub-domains and Objectives of PCK of TEDS-M. (Tatto et al., 2008, p. 39)

Mathematical curricular knowledge	Establishing appropriate learning goals Selecting possible pathways and seeing connections within the curriculum Identifying the key ideas in learning programs Knowledge of mathematics curriculum
Knowledge of planning for mathematics teaching and learning (pre-active)	Planning or selecting appropriate activities Choosing assessment formats Predicting typical students' responses, including misconceptions Planning appropriate methods for representing mathematical ideas Linking didactical methods and instructional designs Identifying different approaches for solving mathematical problems Planning mathematical lessons
Enacting mathematics for teaching and learning (interactive)	Analyzing or evaluating students' mathematical solutions or arguments Analyzing the content of students' questions Diagnosing typical students' responses, including misconceptions Explaining or representing mathematical concepts or procedures Generating fruitful questions Responding to unexpected mathematical issues Providing appropriate feedback

Tatto et al. (2008), claim that choice of verbs is helpful for distinguishing between pre-active and interactive dimensions of the categories. As seen from Table 2.1, pre-active levels usually related to planning phase of the mathematics lessons. However, some of this objectives can also be helpful for interactive phases of mathematical lessons. For example, 'Identifying different approaches for solving mathematical problems' objective can be necessary during interactive phase when students did not understand the usual approach of the teacher. This objective can also be useful during teaching session when students come up with a different but correct approach.

At the end of the literature review for defining pedagogical knowledge, the tabular summary of PCK ideas (see Table 2.2), which is originally constructed by Park and Oliver (2008), extended to include teacher knowledge studies reviewed in this study.

Table 2.2. Summary Table on Definitions for Pedagogical Content Knowledge

Scholars	Knowledge of									
	Purposes for teaching a subject matter	Student understanding	Curriculum	Instructional strategies and representations	Media	Assessment	Subject matter	Context	Pedagogy	Task
Shulman(1987)	D	O	D	O			D	D	D	
Tamir(1988)		O	O	O		O	D		D	
Grossman (1990)	O	O	O	O			D			
Marks (1990)		O		O	O		O			
Smith and Neale (1989)	O	O		O			D			
Cochran et al. (1993)		O		N			O	O	O	
Geddis et al. (1993)		O	O	O						
Fernandez-Balboa and Stiehl (1995)	O	O		O			O	O		
Magnusson et al. (1999)	O	O	O	O		O				
Hasweh (2005)	O	O	O	O		O	O	O	O	
Loughran et al. (2006)	O	O		O			O	O	O	
Krauss et al. (2008)		O		O			D			O
An, Kulm and Wu (2004)		O	O	O			O			
Hill et al. (2008)		O	O	O			D			
Fan and Cheong (2002)		O	O	O			D			

Note. Extended from Park and Oliver (2008), p.265; D Author placed this subcategory outside of PCK as a distinct knowledge base for teaching; N author did not discuss this subcategory explicitly (Equivalent to blank but used for emphasis); O author included this subcategory as a component of PCK

2.2 Assessing Teacher Knowledge

This section presents several assessment tools that are found on the education literature to assess the knowledge of teachers. Some assessment tools focused on both content knowledge and pedagogical content knowledge of teachers, whilst some focused on only a component of teacher knowledge, e.g. subject matter content knowledge (or content knowledge) and PCK. Additionally, some of these tools used for measurement purposes while some used for only comparison purposes.

Hill, Shilling and Ball (2004) were one of the frontier researches on teacher knowledge (mathematical knowledge for teaching) that includes much technical details on constructing and testing an instrument. As they reported, total of 138 multiple choice mathematics item constructed by the researchers at the Study of Instructional Improvement. Their guiding idea during item construction period was "What mathematical knowledge is needed to help students learn mathematics?" (p.15) and their interest was "what and how subject-matter knowledge is required for teaching" (p. 15). The items they constructed tapped into one of two domains: 'knowledge of content' and 'knowledge of students and content'. After pilot testing of the items, they constructed three testlets. Each testlet consisted 11 to 15 items and 3 items were constant across test forms for testlet equating. Their participants, for final administration of testlets, were total of 1552 in service teachers. As a result of the study, they found that two of the forms were in three-factor structure while one form was in two-factor structure. They conducted IRT analysis to assess the reliability of the scales. Reliability was changed from one form to another form and lowest reliability was 0.71 while highest reliability 0.78. However, in another study (Hill et al., 2008) where same testlets used, reliabilities of these forms dropped and changed between 0.58 and 0.69 when there are fewer participants than initial study.

In their late article (Hill et al., 2008), they tried to clarify a component of PCK: Knowledge of Content and Students (KCS). They found that KCS is a multidimensional construct; however, cause of multidimensionality was not the specification of the domain. They explain that "different amounts on mathematical reasoning, knowledge of students, and perhaps even on a special kind of reasoning about students' mathematical thinking” (p. 395) caused multidimensionality. Even though they tried to construct KCS items that teachers would use knowledge of students, their follow up interviews showed that about forty percent of teachers used mathematical reasoning and twenty percent of teachers used test-taking skills to find correct answer in a multiple-choice KCS item

An et al. (2004) compared the U.S and Chinese teachers’ knowledge of effective teaching structures, and participants were 28 mathematics teachers who teach in fifth to eighth-grade levels from 12 schools in Texas, U.S. and 33 mathematics teachers who teach in fifth and sixth-grade levels from 22 schools in Jiangsu, China. They used Mathematics Teaching Questionnaire that has four open-ended items with each item having two or three parts. They also conducted interviews and class observations with selected teachers to validate their findings. Their primary focus was to compare the knowledge differences between United States and Chinese teachers. Thus, they did not score the teachers’ responses but used qualitative analysis to understand the nature of mathematics teacher knowledge possessed by teachers of different cultures. Their results indicated that when a student cannot solve a problem, most Chinese teachers think students forgot the prior knowledge while U.S. teachers think students did not understand the prior knowledge. Accordingly, 93% of U.S. teacher used various approaches for teaching fraction addition by focusing on the connection with concrete or pictorial models, whilst only 42% of the Chinese teachers used concrete models to develop this students’ knowledge. Most Chinese teachers focus on procedures and rules while only a quarter of the U.S. teachers think that procedures and rules are effective. They also found Chinese teachers put more emphasis on conceptual understanding. This research shows that structure of knowledge of effective teaching is quite different for different cultures. Therefore, teacher knowledge itself may have different dimensions for different cultures.

Chick, Baker, Pham and Cheng (2008) also prepared a questionnaire for assessing the mathematics teachers’ knowledge for decimals, and questionnaire consisted of 17 open-ended items. They also analyzed the questionnaire only qualitatively and described the teacher knowledge according to their framework.

Manizade (2006) tried to develop a questionnaire on geometry topic that has only PCK component. She used Delphi methodology to construct 10 open-ended geometry related PCK items. Using Delphi methodology, 20 participants with different expertise evaluated the items in three rounds. In each round, she modified the items according to expert opinions. Even though she constructed a ten open-ended item PCK questionnaire at the end of third round, instrument was not tested on teachers; therefore, reliability of the questionnaire could not be reported.

Krauss et al. (2008) used open-ended items to develop their PCK and CK tests. They used 21 item for PCK test and 13 item for CK test. They used classical test theory to analyze the reliability of these two tests. Their instrument's reliability level for CK scores was 0.77 and reliability level of PCK scores was 0.83.

TEDS-M project (Tatto et al., 2008) was one of the largest projects for comparing teaching knowledge levels of preservice teachers cross nationally. They were also one of the first researchers who used both multiple choice and open-ended items simultaneously for assessing preservice teachers' knowledge. They developed 45 items for CK dimension and 25 items for PCK dimension for primary teachers; 37 items for CK dimension and 12 items for PCK for lower secondary teachers. Their CK included algebra, geometry and 'number and data'. Additionally, their PCK dimension included three parts, namely 'mathematical curricular knowledge', 'knowledge of planning' and 'enacting students.' As mentioned at the previous section of this chapter, they defined PCK dimension having 18 objectives, which helped them to construct items. They also had three content knowledge subdomains. Thus, measuring each objective for each content knowledge subdomain requires at least 54 items. Since there are 25 PCK items for primary future teachers and 12 PCK items for lower secondary future teachers in their PCK dimension, it is clear that each objective was not measured for each sub-domain (algebra, geometry, etc.) in the TEDS-M project. In fact, they admit that they have to combine 'mathematical curricular knowledge' and 'knowledge of planning' parts into 'curriculum and planning' for primary future teacher and all sub-domains of the MPCK into a single part for lower secondary teachers. According to authors, this was necessary due to limited number of items in each sub-domain and reporting reliable scores. They also used balanced incomplete block design during the test administration in such a way that each booklet included two blocks. They constructed five blocks for primary items and three blocks for lower secondary items. Table 2.3 shows the distribution of primary items and Table 2.4 shows the distribution of lower secondary items for TEDS-M project.

Table 2.3. Primary Items by Sub-domains and Blocks of TEDS-M (Tatto et al., 2008, p. 66)

Subdomain	Number of Items in Assessment Blocks					Total Items
	B1	B2	B3	B4	B5	
Algebra	2	3	4	4	2	15
Geometry	2	2	3	3	2	12
Number and data	5	4	3	2	4	18
MathPed1 (Curriculum and Planning)	2	2	3	2	4	13
MathPed2 (Enacting)	2	3	1	4	2	12
Grand Total	13	14	14	15	14	70

Table 2.4. Lower Secondary Items by Sub-domains and Blocks of TEDS-M (Tatto et al., 2008, p. 67)

Subdomain	Number of Items in Assessment Blocks			Total Items
	B1	B2	B3	
Algebra	6	3	3	12
Geometry	4	5	3	12
Number	2	2	5	9
Data	1	1	2	4
MathPed	3	5	4	12
Grand Total	16	16	17	49

Their primary purpose was to compare the preservice teacher across nations, not to define preservice teachers' pedagogical knowledge in detail. Therefore, it may be acceptable that not each objective of PCK matched by all content knowledge subdomains but matched the content knowledge subdomains generally. It may also show that it was not easy or possible to write items for a particular PCK objective for every content knowledge subdomains. TEDS-M project also did not publish the technical details about instrument analysis at the time of this review. Thus, the reliability levels for their instrument cannot be reported.

There was limited number of studies in literature that was related to assessing the pedagogical knowledge of teachers in statistics related topics. In one of them, Burges (2007) examined the four inservice teachers' knowledge for teaching statistics (content knowledge and pedagogical content knowledge). The study was based on classroom observations and teacher interviews. He actually did not measure the teacher knowledge; instead, he observed whether content knowledge and pedagogical content knowledge 'used' and 'not used' during teaching phase. His framework together with an evaluation of a teacher is presented in Figure 2.3.

		Statistical knowledge for teaching			
		Content knowledge		Pedagogical content knowledge	
		Common knowledge of content (CKC)	Specialised knowledge of content (SKC)	Knowledge of content and students (KCS)	Knowledge of content and teaching (KCT)
Thinking	Need for data				
	Transnumeration				
	Variation				M
	Reasoning with models			M	
	Integration of statistical and contextual				M
Investigative cycle					M
Interrogative cycle					
Dispositions					

Key: = direct evidence of that knowledge used; = indirect evidence of that knowledge; M = missed opportunity related to that knowledge.

Figure 2.3. Framework and an Example for Observing Pedagogical Knowledge (Burges, 2007, p.87)

As seen from the Figure 2.1, he recorded the moments when a teacher used specific knowledge directly or indirectly or missed an opportunity to use pedagogical knowledge needed for clarifying the topics to students. Since the framework constructed for observation purposes only, it is not useful for large-scale measurement purposes.

A second study identified as assessing PCK for statistics was conducted by Pinto Sosa (2010). He also used mostly qualitative techniques to assess the pedagogical knowledge. The two participating teachers were specialized in Psychology and Education. Since the main purpose was to understand the nature of pedagogical knowledge possessed by these two teachers, this study cannot be interpreted as a measurement study.

Another issue regarding to assessment of teacher knowledge is constructing items that do not favor any specific learning theory (e.g., constructivism and behaviorism). Manizade (2006) explained that items, which were all open-ended, were constructed to be ideologically free, such as “What instructional strategies and/or tasks would you use during the next instructional period?” (p. 145). Hill et al. (2008) also explained that they constructed multiple-choice items in a way that correct answers do not based on any particular learning theory because they strictly relied on empirical evidence, and theories are propositional and arguable.

2.3 Turkish Preservice Middle School Mathematics Teachers’ Opportunities for Learning and Teaching Statistics

In order to be hired as a middle school mathematics teacher in Turkey, a high school graduate has to complete four-year elementary mathematics teacher education program. Mathematics education departments has to follow a program that is regulated by Council of Higher Education (CHE, 2006a) and departments can only change 25 % of the program (CHE, 2006b). This program is presented in Table 2.5.

To become certified in teaching, a math teacher candidate has to register for 56 credits (38%) of mathematics courses, 14 credits (10%) of mathematics education courses, 26 credits (18%) of education courses, 34 credits (23%) of general culture courses and 16 credits (11%) of elective courses. According to regulations (CHE, 2006b), departments can define elective courses, and modify up to 25% of the program by replacing the mandatory courses or changing the sequence of the courses.

Table 2.5. Elementary Mathematics Teacher Education Program in Turkey

Year	First Semester Courses	Cat.	Second Semester Courses	Cat.
1	Int. Mathematical Structures	M	Abstract Mathematics	M
	History of the Turkish Rep. I	C	Geometry	M
	Turkish I	C	History of the Turkish Rep. II	C
	Foreign Language I	C	Turkish II	C
	Technology I	C	Foreign Language II	C
	Introduction to Education	ES	Technology II	C
			Educational Psychology	ES
2	Calculus I	M	Calculus II	M
	Linear Algebra I	M	Linear Algebra I	M
	Physics I	C	Physics II	C
	Educational Research Methods	ES	Instructional Tech. and Material Dev.	ES
	Principles and Methods of Instruction	ES	Elective Course	EL
	Elective Course	EL		
3	Calculus III	M	Differential Equations	M
	Analytical Geometry I	M	Analytical Geometry II	M
	Statistics and Probability I	M	Statistics and Probability II	M
	Introduction to Algebra	M	Teaching Mathematics II	ME
	History of Science	C	Community Service	C
	Teaching Mathematics I	ME	History of Turkish Education	ES
	Elective Course	EL	Assessment and Evaluation	ES
4	Elementary Number Theory	M	Mathematical Philosophy	M
	Fundamentals of Guidance and Counseling	ES	Practice Teaching in Mathematics	ME
	History of Mathematics	M	Turkish Education System	ES
	Classroom Management	ES	Elective Course	EL
	Special Education	ES	Elective Course	EL
	School Experience	ME		
	Elective Course	EL		

Note. Cat. denotes Category of Course; M denotes Mathematics; C denotes Culture; ES denotes Educational Sciences; ME denotes Mathematics Education and EL denotes Elective.

Council of Higher Education (CHE, 2006a) also mandated the objectives of the courses offered from elementary mathematics education program. ‘Statistics and Probability I’ course must covers these concepts: (a) Basic concepts, frequency distributions, histogram and frequency polygon, graphical representation of categorical data and applications; (b) parametric and nonparametric central tendency measures and applications; (c) parametric and nonparametric dispersion measures and applications; (d) skewness and kurtosis; (e) basic concepts of probability theory, addition and multiplication rules, Bayes' theorem, probability distribution table, expected value and applications; and (f) basic concepts of discrete probability distributions, binomial distribution, Poisson distribution, hypergeometric distribution and applications

Thus, 'Statistics and Probability I' covers almost all statistics topics that taught in middle grades in Turkey. Therefore, this course may provide opportunities for preservice teachers to master their statistics knowledge for the topics they supposed to teach. However, this course does not include any teaching essence. As one can see from Table 2.5, other courses that may help preservice middle school mathematic teachers to learn teaching statistics include 'Teaching Mathematics I' and 'Teaching Mathematics II'. Preservice teachers may also benefit from courses that might have statistics content such as 'Statistics and Probability II', 'Assessment and Evaluation', 'School Experience' and 'Practice Teaching in Mathematics'.

2.4 Statistics Knowledge of Students and Teachers: Misconceptions and Difficulties

Misconceptions (about statistics concepts) or difficulties (experienced during statistical problem solving) are related to this study in two ways. Firstly, preservice teachers should be aware of the misconceptions or difficulties in order to have a sound pedagogical content knowledge for teaching statistics. Second reason is that preservice teachers could also have the same misconceptions or difficulties as student teachers.

Most of the research about misconceptions and difficulties in statistics education is related to central tendency topics such as mean (Cooper, 2002). Researchers generally used SOLO (structure for observed learning outcome) taxonomy to evaluate statistics understanding of students or teachers. Figure 2.4 illustrates the five levels of the SOLO taxonomy (Biggs & Tang, 2007).

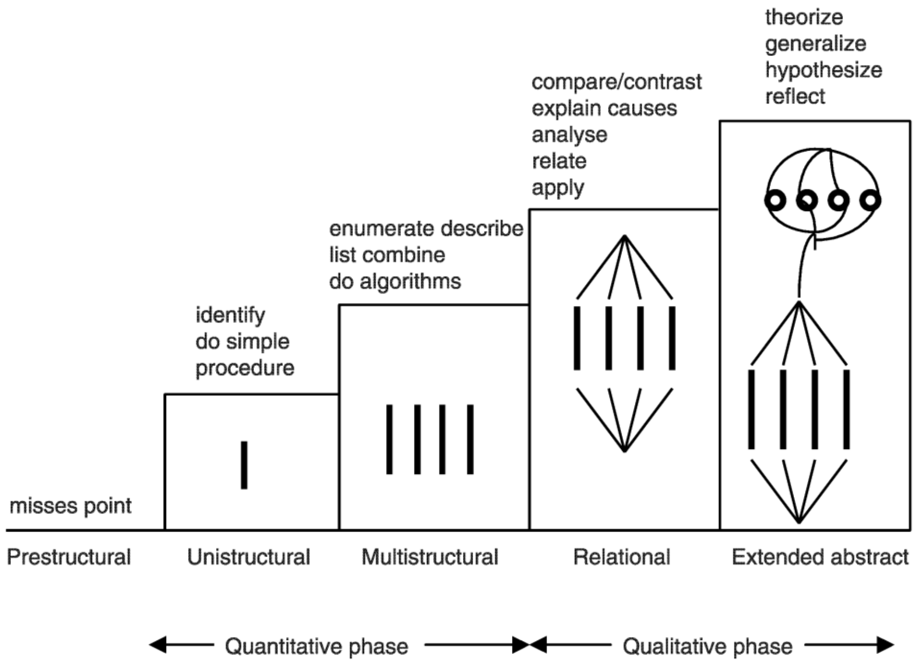


Figure 2.4. A Graphical Representation of SOLO Taxonomy [Biggs & Tang, 2007, p. 79]

Five levels of SOLO taxonomy can be summarized as follows (Biggs & Tang, 2007; Groth & Bergner, 2006):

- Pre structural: Response is not relevant to task
- Unistructural: Responses include one aspect of relevance, far from complete
- Multistructural: Responses include more than one aspect of relevance, however, lists these aspects without making connections
- Relational: Goes more than multistructural level and discusses relations and makes connections
- Extended Abstract: In addition to relational level, it goes beyond the requirements of the task

Groth and Bergner (2006) investigated preservice elementary and middle school teachers' procedural and conceptual knowledge of mean, median and mode; and used a modified version of SOLO taxonomy to analyze responses of teachers. Their participants were 46 preservice elementary and middle school teacher, and they classified preservice teachers' answers into four groups, namely 'unistructural/

concrete symbolic', 'multistructural/concrete symbolic', 'relational/concrete symbolic' and 'extended abstract/formal-1'. They found that more than half of the preservice teachers were in either unistructural or multistructural level of SOLO taxonomy. Moreover, preservice teachers in these levels described the processes for finding mean, median and mode. According to their study, only a few preservice teachers were in the extended abstract level that requires discussing situations when one of the three measures was better than other (Groth & Bergner, 2006).

Garcia, Cruz and Garret (2008) conducted a study with 227 students about the concept of arithmetic mean. One hundred and thirty of these students were high school students and ninety-seven of them were education faculty students who were majoring in Mathematics or Education. They collected data using an open-ended item and a multiple-choice item. Both of these items were shown at Figure 2.5. Garcia et al. (2008) also analyzed result using SOLO taxonomy.

Problem "Time taken over 100 metres": *When asked by their PE teacher, 10 students independently and simultaneously recorded the time taken by another student to run 100 m. The times recorded (in seconds) were the following:*

15.05; 14.95; 15.05; 15; 10; 15; 14.90; 15; 14.95; 15

What time should the teacher consider as the estimation of the real time taken by the student, and why?

Problem "In a science class": *Nine students weighed a small object with the same instrument. The weight recorded by each student (in grams) is as follows:*

6'2	*6'0*	*6'0*	*15'3*	*6'1*	*6'3*	*6'2*	*6'15*	*6'2*

The students want to find out as accurately as possible the real weight of the object. Which of the following methods would you recommend?

(Mark only one of the following answers)

a. Use the number repeated most, which is 6.2.
b. Use 6.15 as this is the recording with most decimal places.
c. Add up the 9 numbers and divide by 9.
d. Discard the number 15.3 and add up the other 8 numbers and divide by 8.

Figure 2.5. Items from Garcia, Cruz and Garret (2008; p. 54)

For the 'Time taken over 100 meters' problem in their study, some *prestuructural responses* were "10 seconds should be considered as this is the time of the runner who came first; The real time run by the student should be 10 seconds because that would be the ideal time for 100 meters" and "It's 15 seconds, because it depends on a student's fitness and ability to do this type of sport or activity over a distance of 100 meters, and the student has to be well trained" (p. 55). Some *unistructural responses* were "10 seconds should be considered as it's the shortest time", "10 seconds because in a 100 meter race the minimum time is 9.60 seconds, and for me that would be almost impossible and if the teacher wants to consider a time, it has to be 10 seconds. Also 10 seconds is the closest time", and "15.05 seconds should be considered because it's repeated twice" and "The longest time should be considered as the seconds pass quickly" (p. 55). Some *multistructural responses* were "The time the teacher should consider as an estimate of the real time run by the student is 15 seconds because this time is the most repeated in the data" and "The median should be used, as this is the value between the fifth and sixth positions" (p. 56); some *transitional responses* were "The sum of the times recorded is 144.9. Dividing this we get 14.49. So the time is 14.9 seconds" and "The teacher should consider the time of 14.90 as the estimation because this is the number closest to the average of 14.49" (p. 56); and none of the responses were classified as *relational response*.

For the 'In a science class' problem, options "use the most repeated number, which is 6.2" and "use 6.15, as this is the value with most decimal points" classified as multistructural responses; option "add up all the values and divide the total by the total number of data" classified as transitional response; and option "discard the number 15.3 and add up the other 8 numbers and divide by 8" classified as relational response.

Their results indicated that none of the students was able to give an answer, which is classified as relational response when the item is presented in an open-ended item format. However about 13% of the students were able to give an answer which is classified as relational response when the item is presented in a multiple-choice item format. Contrasting open-ended item answers with multiple-choice item answer for university students revealed that students who gave a relational response to multiple-choice item were almost equally distributed among prestructural, unistructural, multistructural and transitional response categories. Thus it can be concluded that some students were able to pick most logical option among multiple choices even though they cannot reach the same conclusion in their words in a free response environment.

As in most statistics topics, many students also misinterpret histograms. Bruno and Espinel (2009) analyzed the construction and evaluation errors made by preservice teachers. According to the results, teacher candidates constructed the histograms making several mistakes. In addition, teacher candidates constructed the histograms where the rectangles are separated, labeled the axes incorrectly or omitted the zero frequency intervals.

Another study investigated the conceptual errors of undergraduate students related to histograms (Lee & Meletiou-Mavrotheris, 2003). Students were asked "What goes on the vertical axis and horizontal axis when constructing a histogram for describing the distribution of salaries for individuals that are 40 or older and have not yet retired?" Most common misconception students held was the interpretation that histograms are two variable scatterplots. Thirty one percent of the students had a mistake by stating that age should be on X-axis and salary should be on the Y-axis. Other similar mistakes were salary on X and Age on Y; age on X and frequency of salary on Y. Some students also have misconception that histograms are displays of raw data, and individuals should be on X-axis while salary on Y-axis.

Burgess (2002) explored the data sense of 30 preservice teachers and found that preservice teachers mostly draw graphs at the end of investigation without backing up any idea. Graphs produced by preservice teachers were also mostly incomplete or inappropriate for the considered variables.

Taylor's (1993) doctoral thesis examined statistical content knowledge of 35 elementary and middle school preservice teachers from a large university in USA. Subjects were assessed on ten questions that developed by author according to recommendations of Mathematical Association, American Statistical Association and NCTM. Author analyzed whether students met minimal level of statistical competency. Author found that preservice teachers generally met minimal competency levels for items that require simple numeric calculations or little explanation. Author also interviewed 8 preservice teachers to understand conceptual errors of these preservice teachers. She found that preservice teacher could not interpret a box plot when the numbers are not given. Preservice teachers also interpreted quadratic trend of a scatter plot (Figure 2.6) and draw a curve that best fits to data.

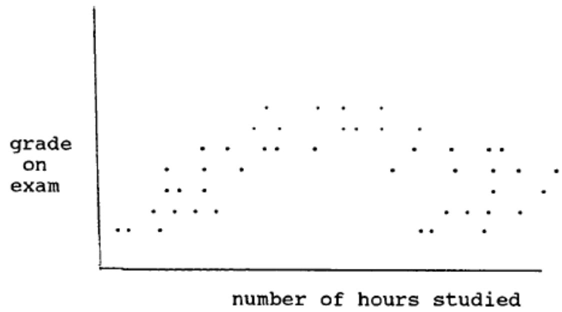


Figure 2.6. Bivarite Graph from Taylor's (1993) Study

Even though most preservice teachers were able to draw a curve that best fits to data, they described this curve as bell (normal distribution) curve. These preservice teachers were assigning one variable curve to bivariate data.

2.5 Summary of Literature Findings

Examination of opportunities to acquire mathematical knowledge for teaching statistics showed that Turkish preservice teacher had limited opportunities to acquire both content knowledge and pedagogical content knowledge for teaching statistics.

Literature review indicates that most researchers focused on how central tendency measures understood by students or how histograms are drawn by students. Unfortunately, there is a lack of articles that focused on pie or line graphs.

Up to this point, several definitions for PCK and how PCK assessed by several researchers have been reported. Then some statistics education manuscripts that related to focus of this study were reported. As it is seen from the Table 2.2, there was consensus among most scholars that 'student understanding' and 'instructional strategies and representations' are both classified as a piece of pedagogical content knowledge. 'Subject matter content knowledge' (or content knowledge) divides scholars into two groups: One group of researchers places this knowledge outside of the PCK while other group of researchers places it in the PCK. However, Krauss et al. (2008) and Blömeke, Houang and Suhl (2011) showed that content knowledge and pedagogical content knowledge are statistically separable constructs.

Several different ways presented in literature for assessing the pedagogical knowledge for inservice and preservice teachers. It seemed that most of the researchers relied on classroom observations or teacher interviews to catch glimpses of PCK for inservice teachers. However, it was not an appropriate option for preservice teachers. Another threat for measuring PCK during classroom observations is that a teacher may use a small proportion of her/his pedagogical knowledge while teaching a particular topic in a particular classroom. Paper and pencil test, however, may simulate various teaching and learning scenarios in a limited time.

Creating paper and pencil tests for measuring PCK construct requires researchers to choose an appropriate item format for developing PCK items. Multiple choice or other dichotomously scored items take less time to score and preparing rubric for these types of items is also easier. However, measuring a complex knowledge such as PCK may not be appropriate with multiple-choice items for every PCK dimension. On the other hand, scoring open-ended items requires much more time, and creating rubrics are much more difficult. But main advantage of using open-ended problems for measuring PCK concept is that answers of this type of items help researchers to understand how a teacher thinks and uses his/her available knowledge for a given situation. Open-ended items also help researchers to evaluate the possible options for developing professional education settings for inservice teachers or developing undergraduate courses for preservice teachers.

CHAPTER 3

METHODOLOGY

The methodology chapter explains the research design, description of the population and sample, description of MKT-S instrument that used for data collection, procedure by which the study was conducted, rubric preparation, description of the data analysis methods for validity and reliability analysis of the MKT-S instrument.

3.1 Research Design

Primary purpose of this study is assessing preservice middle school mathematics teachers' mathematical knowledge for teaching statistics, understanding the relationship between CK scores and PCK scores, and investigating the ingredients of CK and PCK. In this regard, an instrument will be developed and distributed to the participants, and therefore; the methodology of the study is cross-sectional survey since researcher would collect the data 'at just one point in time' (Fraenkel & Wallen, 2009). However, this study can also be interpreted as correlational study since researcher will also examine the relationship between components of the MKT-S instrument.

3.2 Instrument Development

Researchers (Hill, Ball & Schilling, 2008) argue that two sets of criteria are important to building measures for teacher knowledge. Their first set of criteria is the conceptualizing the domain. They advise beginning by proposing construct, elaborate the theoretical or empirical basis for the construct, delineate the boundaries of the construct, and specify how it is related to other constructs. Their second criterion is based on the analysis of pilots of the tests items to assess whether the conceptualization is correct and adequate, and whether the instruments meet several measurement related criteria.

Downing (2006) recommends defining purpose of the test as a first step in test development process. The purpose of the instrument was to assess preservice middle school mathematics teachers' mathematical knowledge of teaching statistics, and investigating the adequacy of this knowledge in order to paint the general picture for preservice teachers' mathematical knowledge for teaching statistics.

Second step in test development process is defining content of the test, and this step requires both identifying the cognitive process and delineating the content coverage (Linn, 2006). Identifying the cognitive processes also requires deciding the framework for instrument.

3.2.1 Framework of the Instrument

This study adopted the theoretical framework of Knowledge for Teaching Mathematics instrument of TEDS-M [(Tatto et al., 2008) see also MT21 (Schmidt et al., 2007)] for measuring mathematical knowledge of teaching statistics of preservice teachers. Four main reasons led to choose this framework:

- This framework defines components clearly in a fashion of expected objectives
- These objectives can be set as a teacher education standards, and can be measurable with paper and pencil tests
- Framework is specially developed and appropriate for preservice mathematics teachers' knowledge structure.
- Assessment type is suitable for large-scale assessment.

This framework was influenced mainly by theory of Shulman (1986, 1987), and adapted the work of Fan and Cheong (2002). In these research projects, the mathematical knowledge for teaching were hypothesized to have two dimensions: CK and PCK.

CK dimension of TEDS-M framework (Tatto et al., 2008) has three main cognitive domains that are parallel to TIMSS 2007 framework (Mullis et al., 2005). These main cognitive domains are knowing, applying and reasoning. Knowing domain includes recall, recognize, compute, measure and classify/order; applying domain includes select, represent, model, implement and solve routine problems; and reasoning analyze, generalize, synthesize/integrate, justify and solve non-routine problems. These domains are further explained in Table 3.1.

*Table 3.1. Cognitive Knowledge Domains for Content Knowledge**

Domain /subdomain	Definition
Knowing	
Recall	Recall definitions; terminology; number properties; geometric properties; notation
Recognize	Recognize mathematical objects, shapes, numbers and expressions; recognize mathematical entities that are mathematically equivalent.
Compute	Carry out algorithmic procedures for addition, multiplication, division, subtraction with whole numbers, fractions, decimals, and integers; approximate numbers to estimate computations; carry out routine algebraic procedures.
Retrieve	Retrieve information from graphs, tables, or other sources; read simple scales.
Measure	Use measuring instruments; use units of measurement appropriately; estimate measures.
Classify/ Order	Classify/group objects, shapes, numbers, and expressions according to common properties; make correct decisions about class membership; order numbers and objects by attributes.
Applying	
Select	Select an efficient/appropriate operation, method, or strategy for solving problems where there is a known algorithm or method of solution.
Represent	Display mathematical information and data in diagrams, tables, charts, or graphs; generate equivalent representations for a given mathematical entity or relationship.
Model	Generate an appropriate model, such as an equation or diagram, for solving a routine problem.
Implement	Follow and execute a set of mathematical instructions; draw figures and shapes according to given specifications.
Solve Routine Problems	Solve routine or familiar types of problems (e.g., use geometric properties to solve problems); compare and match different representations of data; use data from charts, tables, graphs, and maps to solve routine problems.
Reasoning	
Analyze	Determine and describe or use relationships between variables or objects in mathematical situations; use proportional reasoning; decompose geometric figures to simplify solving a problem; draw the net of a given unfamiliar solid; visualize transformations of three-dimensional figures; compare and match different representations of the same data; make valid inferences from given information.
Generalize	Extend the domain to which the result of mathematical thinking and problem-solving is applicable by restating results in more general and more widely applicable terms.
Synthesize/ Integrate	Combine (various) mathematical procedures to establish results, and combine results to produce a further result; make connections between different elements of knowledge and related representations, and make linkages between related mathematical ideas.
Justify	Provide a justification for the truth or falsity of a statement by reference to mathematical results or properties.
Solve Non-routine Problems	Solve problems set in mathematical or real-life contexts where future teachers are unlikely to have encountered closely similar items, and apply mathematical procedures in unfamiliar or complex contexts; use geometric properties to solve non-routine problems.

* Adapted from TIMSS 2007 cognitive domain assessment framework (Mullis et al., 2005).

It was also hypothesized that PCK has at least three components: curricular knowledge; knowledge of planning for mathematics teaching and learning (pre-active); and enacted mathematics knowledge for teaching and learning (interactive). These components are further explained in Table 3.2.

Table 3.2. Sub-domains and Objectives of PCK used in TEDS-M. (Tatto et al., 2008, p. 39)

Mathematical curricular knowledge	Establishing appropriate learning goals Selecting possible pathways and seeing connections within the curriculum Identifying the key ideas in learning programs Knowledge of mathematics curriculum
Knowledge of planning for mathematics teaching and learning	Planning or selecting appropriate activities Choosing assessment formats Predicting typical students' responses, including misconceptions Planning appropriate methods for representing mathematical ideas Linking didactical methods and instructional designs Identifying different approaches for solving mathematical problems Planning mathematical lessons
Enacting mathematics for teaching and learning	Analyzing or evaluating students' mathematical solutions or arguments Analyzing the content of students' questions Diagnosing typical students' responses, including misconceptions Explaining or representing mathematical concepts or procedures Generating fruitful questions Responding to unexpected mathematical issues Providing appropriate feedback

3.2.2 Content Selection Process of the Instrument

Since Turkey uses a national curriculum provided by Ministry of National Education, statistics content is naturally defined by this national curriculum. After several failures of Turkish elementary students on international assessment programs, MNE decided to change the national curriculum on 2005, from 'behaviorist approach' to 'constructivist approach' (Babadoğan et al., 2006; Koç et al., 2007). This change also introduced new statistics topics to the Turkish mathematics curriculum, namely median, mode, range, quartiles, standard deviation, histograms (MNE, 2005a; MNE, 2005b). Table 3.3 summarizes the statistics content in 2005 curriculum.

Even though identifying statistics content was an easy process, It should be taken to account that MNE had changed the content of mathematics teaching program in past without a prior notice so MNE can change this content again without a prior notice.

Table 3.3. Statistics Content Standards from 2005 Turkish National Curriculum [Adapted from MNE (2005a, 2005b)]

Grade	Standards *[Students will be able to…]*
5	• Construct and interpret line graphs • Construct and interpret two-way tables • Organize data using schemes • Explain and calculate arithmetical mean
6	• Construct research problems regarding to an issue, choose an appropriate sample and collect data • Show and interpret data using appropriate statistical representations • Explain the situations where bar graphs can cause misinterpretations • Calculate and interpret the arithmetical mean and range • Make assumptions based on the data
7	• Construct and interpret bar and line graphs that are based on more than one property • Construct and interpret pie graphs • Make assumptions on real life situations based on statistical representations • Make assumptions based on the data • Explain the situations where line and picture graphs can cause misinterpretations • Calculate and interpret median, mode and quartile range
8	• Construct and interpret histograms • Calculate standard deviation • Make assumptions on real life situations based on statistical representations, central tendency measures and standard deviation

Another issue for using a time specific curriculum is the evolving nature of the curriculum. An instrument, prepared according to specific objectives of a curriculum, can be invalid when objectives changed slightly for a newer version of the curriculum. Therefore, several countries' statistics content, such as United States of America and Singapore were also analyzed to find common concepts that taught in middle grades.

Even though there is no national curriculum in United States of America, there are institutions that lead to set content and cognitive standards for teaching mathematics such as NCTM and Common Core State Standards Initiative. According to NCTM standards (NCTM, 2000), expectations for statistics content are divided into three categories: formulate questions, use appropriate statistical methods and develop inferences. Another attempt to define standards for school mathematics in USA came recently as a collaborative work of National Governors Association Center for Best Practices (NGA) and Council of Chief State School Officers (CCSSO). These two agencies collaboratively founded Common Core State Standards Initiative and defined Common Core State Standards (CCSS) for English language and mathematics. Resulting standards (CCSSI, 2010) were become very popular and were accepted by almost all of the states (Porter, McMaken, Hwang & Yang, 2011;

CCSSI, 2013). CCSS defines the goals for students to be reached in each grade level. Even though these standards do not specify content that should be applied for each grade, they may help one to understand the contents that should be taught at a specific grade. Even though cognitive alignment of CCSS and NCTM standards is different (Porter etal.,2011) content standards are very similar at the end of eight grade.

Singaporean curriculum is another curriculum that examined for middle school statistics content. This country is of special interest because of its ranking in international studies (Mullis et al., 2012). After reviewing statistics content standards for several countries, Table 3.4 was prepared to understand similarities and differences among counties

Table 3.4. Comparison of Statistics Content Standards for Middle School Grades

Content	Turkey	CCSS	NCTM	Singapore
Graphical representations				
Line graphs	C		C	C
Bar graphs	C	L	C	C
Pie charts	C		H	C
Dot plots		C	C	
Histograms	C	C		C
Box plots		C	C	
Scatterplots and fit line		C	C	
Association between two quantities		C	c	
Two-way tables	C	C	c	c
Comparing different representations			C	
Sampling				
Research problem posing	C	c	c	
Sampling	C	C	C	
Random sampling		C	c	
Data collection	C	C	C	C
Numerical vs. categorical data distinction			C	c
Comparing data sets/ samples		C	C	
Variability				
Range	C	c	C	H
Interquartile range	C	C	C	H
Mean absolute deviation		C		
Standard deviation	C	H	H	H
Center				
Arithmetical Mean	C	C	C	C
Median	C	C	C	H
Mode	C		C	H
Comparisons between averages	C	C	C	

Note. C denotes explicitly discussed topic; c denotes inexplicitly discussed topic; L denotes topics discussed only for lower grades (primary school) compared to Turkish curriculum; H denotes topics discussed only for higher grades (high school) compared to Turkish curriculum.

Comparison of several curriculums showed that Singaporean curriculum covered only basic concepts while CCSS and NCTM standards covered a wide range of concepts. Turkish curriculum covered more concepts than Singaporean curriculum and fewer concepts than CCSS and NCTM. Differences between Turkish and American standards were especially apparent for graphical representations such as dot plots, box plots and scatterplots; and comparing the characteristics of two different samples.

3.2.3 Limitations on Content Coverage

Since teacher knowledge and statistics content are too broad to cover in this study, both teacher knowledge coverage and statistics content coverage was limited to construct a fifty-minute length instrument. Statistics contents that were included in the middle school curriculum was decided to be limited to the central tendency related topics such as mode, median and arithmetical mean; and graphic related topics such as histograms, data clustering, bar graphs and pie graphs because these topics constituted the majority of middle school statistics curriculum

All PCK objectives in MT21 framework were too broad to cover in this study. Thus, objectives were examined, and objectives that are appropriate to assess statistics knowledge were included in the study. For example, ‘choosing assessment formats’ objective was not selected for this study because this objective was related to general mathematics education. Some objectives were also too broad and could be defined as a combination of other objectives. For example ‘planning mathematical lessons’ can be measurable as combination of ‘planning or selecting appropriate activities’, ‘selecting possible pathways and seeing connections within the curriculum’, ‘Identifying different approaches for solving mathematical problems’ and so on.

Then PCK objectives that will be included in the test were limited to six main objectives and defined to guide the item development process. Definitions for the selected PCK objectives are as follows:

- PCK-1 is "*Selecting possible pathways and seeing connections within the statistics curriculum*", and defined as "Future teacher should be able see connections between statistics topics and know how a statistics topic can be related with another topic. This is also includes seeing connections between topics that taught in different grades".

- PCK-2 is "*Identifying different approaches for solving statistical problems*", and defined as "Future teacher should see and value that some statistical questions can be handled using different approaches that are all correct.

- PCK-3 is "*Planning or selecting appropriate methods and activities for representing statistical ideas*", and defined as "Future teacher should be able to plan a lesson by selecting appropriate methods and identifying key ideas. Activities involved in methods should match the key statistical ideas and learning goals in the curriculum. This objective also includes selecting appropriate examples".

- PCK-4 is "*Analyzing or evaluating students' statistical solutions or arguments*", and defined as "Future teachers should experiment with different teaching approaches and activities, and monitor the results, using conventional tests, and by carefully listening to students and evaluating information" (Garfield, 1995).

- PCK-5 is "Predicting *or diagnosing typical students' responses, including misconceptions*", and defined as "Future teacher should be able to (a) know how regular student will respond to statistical question, (b) predict a misconception and (c) identify a previously constructed misconception".

- PCK-6 is "*Providing appropriate feedback*", and defined as "Future teacher should be able to assess and question the student learning aligned with learning goals (Pfannkuch & Dani Ben-Zvi, 2011) and able provide appropriate feedback after diagnosing students' responses in a way that given feedback improves students learning" (Chickering, Gamson & Poulsen, 1987).

3.2.4 Preparing Test Blueprint and Developing Items

After defining PCK objectives, a test blueprint prepared to guide item construction to assure content coverage. Mathematics education literature related to both middle school students' statistics knowledge and (preservice) teacher's statistics knowledge were searched while also considering definitions of both CK and PCK objectives in mind. Then 19 items was constructed by the researcher benefiting from four sources for the item development: (a) directly from literature, (b) indirectly from literature, (c) from teaching experiences of teachers and (d) from the teaching experiences of researcher (these sources further explained under the title 'Sources for Item Development'). Items were placed in the test blueprint until a sufficient coverage reached (see Table 3.5 for distribution of items to test blueprint).

Scoring rubric was prepared by researcher during item development process. A five point analytic scoring found appropriate for open-ended items where appropriate answer gets 4 point and wrong or irrelevant responses get 0 point. For example, Item B.5B (corresponds to item F.5B on final version of the instrument and detailed results can be found on section 4.4) was constructed for PCK-6 objective, which is "providing appropriate feedback that improves students' learning". Scoring rubric for this item was in this way:

4 Points: Answer includes two aspects of efficient feedback: (i) clear explanation of what was the error of the student and (ii) additional detailed feedback that mentions how student recover this error.

3 Points: Answer includes two aspects of efficient feedback: (i) clear explanation of what was the error of the student and (ii) additional detailed feedback that mentions how student recover this error.

2 Points: Feedback provided to students only mentions what was the error of student or (ii) feedback only explains how to find median of the data.

1 Point: (a) Answer does not explain students' error or misconception clearly and no feedback provided or (b) Answer does not include explanation for student's error and only offers a general feedback that is not related to median concept.

0 Point: (a) Answer does not explain students' error or misconception correctly and no feedback provided or (b) Answer does not include explanation for student's error and only offers a feedback that is not meaningful or appropriate.

Then three mathematics education experts (Panel A) in a panel reviewed the appropriateness of the items and rubric, and possible revisions were discussed. After panel, some items were partly changed, and item bank was ready (see Appendix A for complete item bank) to be tested on preservice teachers.

Table 3.5. Test Blueprint and Item Distribution

Concept	CK			PCK					
	K	A	R	PCK_1	PCK_2	PCK_3	PCK_4	PCK_5	PCK_6
Mean	B.7A B.7B		B.7C B.7D			B.10		B.5A	B.5B
Graphs		B.1	B.8A B14	B.3	B.12 B.6	B.4 B.9	B.11 B.13	B.2	B.8B

Note: Items labeled differently in all revisions of MKT-S instrument in order to prevent confusion. B, in front of item numbers, denotes item sequence in Item Bank (Appendix A)

3.2.5 Sources for Item Development

Mathematics education literature has limited information about pedagogical knowledge on statistics concepts. Therefore, only item B.7C on the item bank constructed using the information *directly from literature*. This item was adapted from the work of Garcia et al. (2008) with the permission (see Appendix B).

Items B.2, B.4, B.5A, B.5B, B.7A, B.7B, B.7D and B.10 were constructed using the information presented in statistics textbooks or articles indirectly. For example, items B.5A and B.5B were constructed from the information reported in Barr (1980). In this article, author asked students to find the median of a data that was presented in tabular form, and article reported that most students had a misconception or found the median of the given data inaccurately. Therefore, the question itself and answers given by these students were created the opportunity to write a PCK item about predicting students' misconception and/or providing an appropriate feedback for a student who had a misconception.

During the item development process, researcher asked a group of experienced inservice teachers' advice for the situations where students could provide unexpected answers for a statistics question that could presented as scenario for a PCK item. One teacher, who is also a graduate student in mathematics education department, commented and gave two specific examples while he encountered during teaching or assessment of statistics topics. Items B.11 and B.13 were constructed from the information presented by this experienced mathematics teacher.

Another source for the items came from the teaching experiences of the researcher. Items B.1, B.3, B.6, B.9, B.12 and B.14 were constructed by the researcher. who works as research/teaching assistant in elementary mathematics education department in a mid-sized university on the Northern area of Turkey since 2007. Before the development of the instrument, he worked two years as helping teaching assistant, and two years as main teaching assistant for the courses 'Introduction to Statistics and Probability I', 'Introduction to Statistics and Probability II', 'Methods for Teaching Mathematics I' and 'Methods for Teaching Mathematics II'. Therefore, researcher gained valuable experience about preservice middle school mathematics teachers' both overall statistics knowledge and overall pedagogical knowledge.

3.2.6 Item Trial

Linn (2006) stated that “Even skilled and experienced item writers sometimes produce flawed items that are ambiguous, have no correct answer, or are unintentionally offensive to some groups of test takers” (p. 32). Item bank was tested on 69 preservice middle school mathematics teachers from Institution A to understand preservice teachers’ reactions to particular items. The main purpose of testing was to see whether items were understood in a way that is parallel to the intended purpose of the items, and scoring rubric was appropriate.

Each test form of 69 preservice teachers clearly examined, and some problematic answer categories were observed for some items, which was possibly due to the wording of the items. For example, in Item B.5A, preservice teachers were asked “What is the situation for students who gave 5 as answer?”. While most preservice teachers understood the item correctly and supplied an appropriate or inappropriate misconception for why students found 5, some preservice teachers gave unexpected answers. Most frequent unexpected answer to this particular item was “because students make calculation error”. Even though this answer was correct, it was not parallel to expected answer. Therefore, item was adjusted and final form of the item was “What is the situation for students who gave 5 as answer if we assume no calculation error?

All items except item B.11, which had unexpected preservice teachers answers, were edited in respect to observed misunderstandings. Item B.11 was the most problematic item and it seemed that there was no clear pattern for unexpected answers. Therefore, this item was removed from the item bank.

To justify the edited item bank, six volunteered preservice teachers, who participated in item bank trial study, were selected for interview. Then these preservice teachers interviewed in a semi-structured interview format where the main question was “What do you understand from this question? Can you please think aloud?” During the interview, necessary adjustments were made until all preservice teachers had the intended understanding behind the item.

After adjustments, another panel of three mathematics education experts (Panel B) reviewed second versions of the items and revisions found appropriate. Then it was decided that the second version of the instrument would be tested on a slightly larger sample from different institutions in order to evaluate items both qualitatively and quantitatively. This process resulted in second version of the instrument (see Appendix C for detailed test blueprint and for complete item bank).

3.2.7 Pilot Testing of Instrument

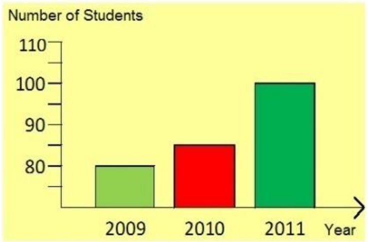
Pilot test form of the instrument printed in two-booklet form. These forms, Form A and From B, included all items in different order to prevent possible preservice teacher interaction. Then these forms tested by researcher in fifth-minutes sessions on 164 preservice teachers from three different public universities (50 preservice teachers from University B, 61 preservice teachers from University C and 53 preservice teachers from University D). Two different forms were distributed to preservice teachers evenly in every classroom where two adjacent preservice teachers get different forms. As a result, 83 form of Form A and 81 form of Form B were collected from preservice teachers. Then these forms analyzed by researcher according to previously constructed rubric

First analysis for pilot study was data screening. During data screening five forms of Form A and seven forms of Forms B removed from the analysis due to large number of missing values. This resulted in total of 78 form of Form A and total of 74 form of Form B.

All 152 forms analyzed using IATA program. This program was found appropriate for analyzing pilot test for two reasons. First of all this program is suitable for analyzing mixed type test, which includes multiple choice and constructed response items. Second reason to use IATA was having two forms of booklet, and IATA was used to analyze both forms simultaneously (Cartwright, 2013).

After analyzing the items with IATA, two problematic items identified and these items were the Item P.2 and Item P.13A. Item P.2 is seen on Figure 3.1.

Ayşe teacher gave a graphic (on right) to students at the beginning of the lesson and wanted students to interpret the graphic. The graphic was explaining the distribution of students who registered in last three years.



This graph is a typical example of
A) Sampling method is important for decisions on population
B) Some graphics can be mistakenly interpreted in some cases
C) Which deviation measure (e.g. standard deviation) can better represent data
D) Which central tendency measure (e.g. mean) can better represent data

Figure 3.1. Item P.2 from Pilot Instrument.

This item was adapted from a graph that was included in Guidebook for Middle School Mathematics Teachers (MNE, 2005a). In the Guidebook, a similar graph was given for an example to explain situations where some graphics may mistakenly be interpreted by middle school students. Therefore, the answer for the item was option B. Table 3.6 shows the distribution of the preservice teachers' answer to Item P.2.

Table 3.6 . Distribution of Preservice teachers' Answer for Item P.2.

Option	Frequency	Percent
Omitted	9	5.9
A	26	17.1
B	20	13.2
C	25	16.4
D	72	47.4
Total	152	100.0

Results showed that the correct answer was the least selected option (B) and further analysis of item response function of P.2 (see Figure 3.2) showed that this item did not discriminate low proficiency preservice teachers from high proficiency preservice teachers. Therefore, the probability of getting this item correct by a high proficiency preservice teacher was almost equal to the probability of getting this item correct by a low proficiency preservice teacher.

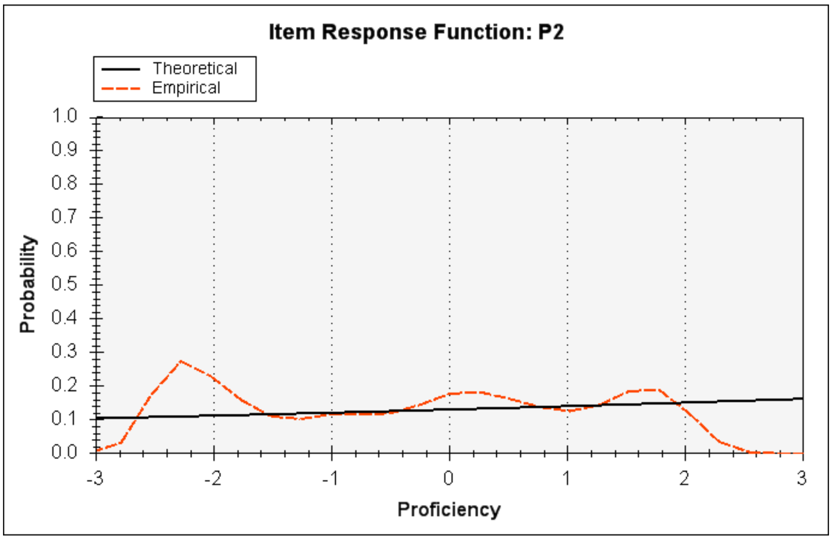


Figure 3.2. Item Response Function of P.2

Another problematic item from pilot study was Item P.13A. IATA program was not able to provide statistics for this item. The reasons behind this result were checked and it was found that difficulty of Item P.13A was 0.94, which means 94% of the preservice teachers solved this item correctly. Further examination also revealed that the score gained from Item P.13B was almost completely depending on Item P.13A, which can be seen on Table 3.7 . Therefore, it has been decided to remove Item P.2 and Item P.13A from instrument.

Table 3.7 . Cross-tabulation of Results of P.13A and P.13B

Booklet	P.13A Options	P.13B Score					Total
		0	1	2	3	4	
Form-A	A	1	-	-	-	-	1
	B*	-	-	9	6	3	18
	C	1	-	-	-	-	1
Form-B	A	2	-	-	-	-	2
	B*	9	2	41	8	3	63
	C	3	-	-	-	-	3

*correct answer; "-" denotes zero frequencies

Another problem observed from the analysis of the pilot study was the rate of 'not reached' responses, and this situation was consistent with the preservice teachers' feedback after participating in pilot study. Some preservice teacher gave feedback after pilot test administration, and they stated that fifty minutes was not enough for them to go through all items. Average number of items that preservice teachers had not a chance to see was 1.7 items. Distribution of *number of not reached items* for two booklets can be seen on Table 3.8.

Table 3.8. Distribution of Number of Not Reached Items for Two Booklets

Number of Not Reached Items	Booklet		Total
	Form A	Form B	
0	20	18	38
1	21	23	44
2	18	15	33
3	10	8	18
4	5	5	10
5	3	0	3
6	1	3	4
7	0	1	1
9	0	1	1

High number of not reached items led to consider Balanced Incomplete Blocks design (BIB) in final implementation.

3.3 Data Collection and Analysis

Final version of the MKT-S instrument submitted to the Middle East Technical University of Human Researches Ethic Committee for the approval of the study. After approval, permissions from participating universities to conduct survey on preservice middle school mathematics teachers also gained through Ethics Committee. For confidentiality issues, participants were not required to write their names or any other information. Name of the university, gender, class level and grade received from the 'Introduction to Probability & Statistics-I' course were asked to participants. Therefore, confidentiality was not issue for the study.

3.3.1 MKT-S Instrument

After making adjustments on pilot instrument, there were six items in CK category and ten items in PCK category on final version of MKT-S instrument. Distribution of these items is in the Table 3.9, and items can be seen on Appendix D.

Table 3.9. Distribution of All Items in Final Implementation.

Concept	CK			PCK					
	K	A	R	PCK_1	PCK_2	PCK_3	PCK_4	PCK_5	PCK_6
Mean	F.1A F.1B		F.1C F.1D			F.3		F.5A	F.5B
Graphs		F.8	F.10	F.2	F.7 F.11	F.6	F.4	F.9	F.12

Note: Items labeled differently in all revisions of MKT-S instrument in order to prevent confusion. F denotes item sequence in Final version of MKT-S instrument (Appendix D)

Because of high number of not reached items in pilot administration, 3 different booklets which are result of BIB design were decided to be used in final implementation. Even though BIB design has some disadvantages if the test scores will be used as a decision criterion for individuals, it has advantages when the purpose of the test is diagnosing the current situations of subgroups (Gonzales & Rutkowski, 2010). Since the purpose of this study is to gather maximum information from preservice teachers, BIB design allows testing more items and collecting more information in a limited time. For the same purpose, BIB was used in many national and international exams such as NAEP, TIMSS and TEDS-M (Johnson, 1992; Rutkowski et al.,2010; Tatto et al, 2008).

Because of the low number of items in CK category, all CK items constituted one block and retained in all three booklets. Ten PCK items were divided into three blocks considering pilot test results, time required to solve each item and distribution of item on blueprint. This process resulted in four blocks, which are summarized on Table 3.10.

Table 3.10. Distribution of the Items to Four Blocks

Block	Dimension	Items	Number of Items
Block I	CK	F.1A, F.1B, F.1C, F.1D, F8, F10	6
Block II	PCK	F.2, F.3, F.4	3
Block III	PCK	F.5A, F.5B, F.6, F.7	4
Block IV	PCK	F.9, F.11, F.12	3

Then each booklet contained Block I and two blocks from Block II, Block III or Block IV. Final forms of booklets are summarized in Table 3.11.

Table 3.11. Summary of Three Booklets in Final Implementation

Booklet	Blocks			Number of CK items	Number of PCK Items	Total Number of Items
A	Block I	Block II	Block III	6	7	13
B	Block I	Block II	Block IV	6	6	12
C	Block I	Block III	Block IV	6	7	13

3.3.2 Population and Sample

As explained in second chapter, all preservice middle school mathematics teachers in Turkey has to take 'Introduction to Statistics & Probability-I' course approximately at the fall semester of their third year and this course covers the all topics that is required to solve items in this study (CHE, 2006a). Other important courses that have to be taken in order to solve items in this study are 'Methods of Teaching Mathematics-I' and 'Methods of Teaching Mathematics-II'. These courses are also offered in their third year. Therefore, target population of the study is all third and fourth year preservice middle school mathematics teachers in Turkey. Reports from OSYM, a national institution responsible from centrally placing high school graduate students to universities according to students' preferences, has been used to calculate the number of target population. In 2008, 2092 preservice teachers were accepted to middle school mathematics teacher license program and these preservice teachers were in their fourth year in the program during the final implementation (OSYM, 2008). In 2009, 3156 preservice teachers were accepted to middle school mathematics teacher license program and these preservice teachers were in their third year in the program during the final implementation (OSYM, 2009). Thus, the total number of the preservice teachers for target population size was at most 5248 preservice teachers.

In final administration, MKT-S was applied to 659 preservice middle school mathematics teachers (approximately 13 % of population) from eight public universities. Sample was not selected randomly and two factors taken into account during sampling: (1) university capacity for preservice middle school mathematics teachers and (2) convenience to travel between universities. Table 3.12 presents the distribution of preservice teachers to universities. The final implementation sample consisted from 421 (65.7%) third year and 220 (34.3) fourth year preservice teachers.

Furthermore, 435 (68.5%) of the participants were female and 200 (31.5%) of them were male preservice teachers.

Table 3.12. Distribution of Sample

Institution	Region	N
University A	Northern Anatolia	46*
University B	Northern Anatolia	56*
University C	Northern Anatolia	126*
University D	North-west Anatolia	64*
University E	Middle Anatolia	40
University F	Middle Anatolia	79
University G	Eastern Anatolia	148
University H	Western Anatolia	100
TOTAL		659

* Different preservice teachers participated in Final implementation

Final instrument was applied to preservice teachers from eight institutions across Turkey by researcher in fifty minutes sessions. Final instrument was especially applied at the end of spring semester (May 2012) to assure that each participant received required knowledge to solve items. Each booklet distributed evenly in every classroom to make sure each group had approximately equal number from each booklet.

3.3.3 Data Recording and Scoring Method for MKT-S Instrument

All forms scored by the researcher during pilot study analysis. For the analysis of forms gathered from final implementation, a different approach than pilot study preferred, and data recorded both qualitatively and quantitatively.

3.3.3.1 Qualitative Data Recording.

In this approach, each open-ended item analyzed separately, and two different spreadsheet files constructed for data recording. The purpose of the first spreadsheet file was recording qualitative data while the other one used for preservice teacher data. Data recording process for each item summarized in Figure 3.3.

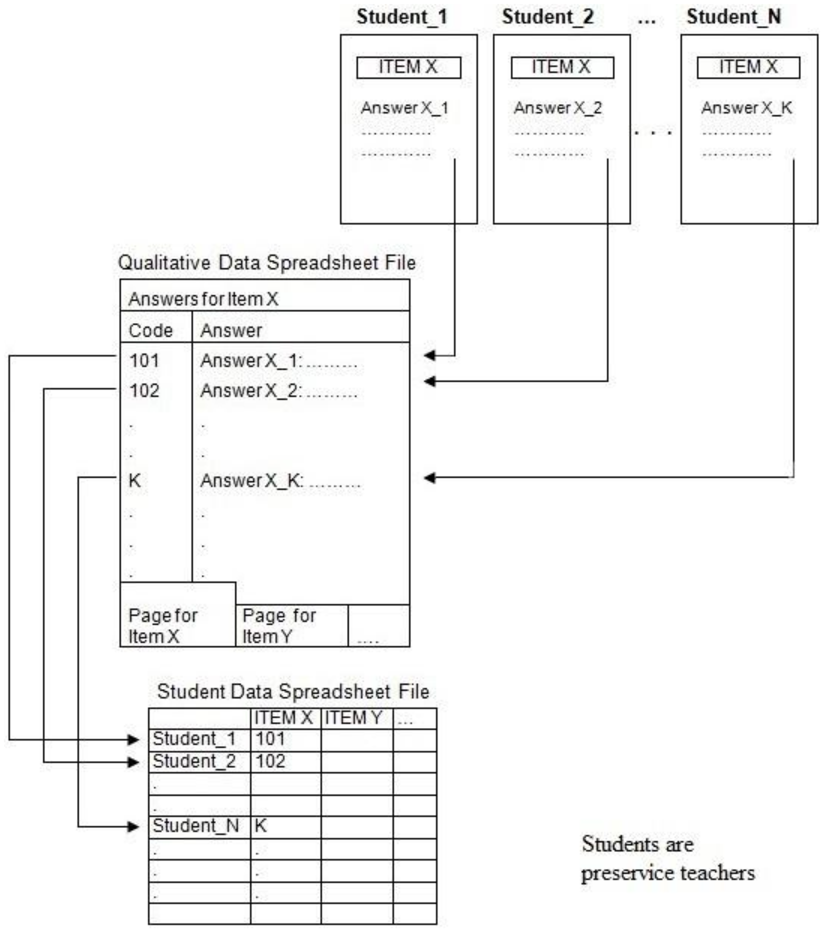


Figure 3.3. Qualitative Data Recording Cycle for Each Open-ended Item

Each different answer for a particular item recorded to qualitative spreadsheet file, and a unique code given for each unique answer. Then the same unique code also recorded to preservice teacher data spreadsheet file for corresponding preservice teacher. This process resulted in two complete spreadsheet files. To assess the coding reliability, randomly selected 50 forms also reviewed by an experienced mathematics teacher who teaches mathematics for nine years and has Master's in Arts Degree in Statistics. Then the qualitative data spreadsheet files for researcher and mathematics teacher compared for fifty forms and coding files matched 82 % in average for fifty forms.

3.3.3.2 Developing Scoring Rubric for Final Implementation.

To rate each unique answer for an open-ended item, each unique answer was placed to rubric under the appropriate title by researcher. The titles in the rubric describe properties of each score point from 4 to 0. In some cases, there was more than one title for a particular score point because some items had more than one possible correct (or wrong) answer categories. After each answer placed under the corresponding score title, rubric file sent to experts (Panel B) for review, and a meeting in two weeks planned to discuss the results. In this file, there was also an empty space for each unique answer for experts to state their ideas. Mathematics teacher educators were expected to rate the each unique answer' place on the rubric. If the answer was not appropriate for corresponding score point, they expected to provide a possible score point for that particular answer. After giving 15 days for review time, a meeting held at METU to discuss ratings of the answers. After reaching consensus on score point allocations for the preservice teachers' answers, extended rubric was constructed.

After that, preservice teacher data spreadsheet file imported to SPSS 17 software and converted to preservice teacher score data file using 'Recode' option of SPSS 17 software.

3.3.3.3 Analysis of Missing Data

The treatment of missing values also poses an important problem in item analysis. There are two main types of missing values in the missing data. One type occurs naturally because of the design of the three booklets. Therefore, there are designed missing values for each of the booklet. The other type of the missing values comes from the omitted or not reached responses. In this study, omitted responses treated as 'wrong answer' and 'not reached items' are treated as 'missing value'. Therefore, 'not reached' items did not affect the preservice teachers' scores. If a preservice teacher omitted more than half of the items for both CK dimension and PCK dimension, s/he discarded from the analysis of the study.

3.3.4 Validity and Reliability Evidences for MKT-S instrument

Messick (1992) defined validity as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores or other modes of assessment" (p. 1). In general, three types of validity evidence is collected to make sure that the inferences drawn from assessment are accurate and these are content-related evidence of validity, criterion-related evidence of validity and construct-related evidence of validity (Frankel & Wallen, 2009).

Content-related evidence was collected through expert opinion. First, a test blueprint, or table of specifications, was prepared. Items were constructed until number of the items was sufficient, and contents of the items distributed on the test blueprint adequately for a standard one-class-time test. Then items and the test blueprint were reviewed by mathematics education experts, and found appropriate to measure the intended content.

Criterion related evidence was collected in two ways. First form of evidence was collected in the form of concurrent validity. Since it is known that the content knowledge of the preservice teachers on statistics is taught as a part of 'Introduction to Probability and Statistics-I' course, it is expected that preservice teachers' scores on the CK dimension of the test is positively correlated with preservice teachers' course grades of Introduction to Probability and Statistics-I. Second evidence was collected by assessing whether test scores discriminate two different groups, namely third year and fourth year preservice teachers.

Even though third year preservice middle school mathematics teachers acquire knowledge to solve items presented in the instrument, fourth year preservice middle school mathematics teachers acquire additional statistics content knowledge and pedagogical knowledge from the courses taught in fourth year or from student teaching activities, which is also offered in fourth year. Therefore, it is expected that there will be little variation between third and fourth year preservice middle school mathematics teachers' scores.

Construct-related evidence of validity was collected through confirmatory factor analysis to validate hypothesized factor structure of the MKT-S instrument. Since the MKT-S instrument is in mixed-item format where 3 items are dichotomous (binary), and 13 items are polychotomous (ordinal or ordered categorical), all variables in the study is classified as categorical variables where normality assumption was not applicable.

Frankel and Wallen (2009) define reliability as “the consistency of the scores obtained- how consistent they are for each individual from one administration of an instrument to another and from one set of items to another” (p. 154). Since the instrument is in mixed type item format, Cronbach’s Alpha coefficient is not suitable for the reliability analysis. In a recent article, Cronbach’s and Shavelson (2004) stated that “A much more significant report on the measuring instrument is given by the residual (error) variance and its square root, the standard error of measurement” (p. 410). Therefore, standard error of measurement (SEM) will be used to evaluate the reliability of the factor scores.

3.3.5 Factor Analysis of MKT-S Instrument

The aim of factor analysis was to confirm two-factor structure of the instrument instead of generating a new model for explaining teacher knowledge. Therefore, confirmatory factor analysis was the main tool for exploring data (Stevens, 2002; see also Hurley et al., 1997). As explained in second section, generally two different views exist among researchers for relation between content knowledge and pedagogical content knowledge (Park and Oliver, 2008). Some researchers accept content knowledge as a part of pedagogical content knowledge while others argue that content knowledge and pedagogical content knowledge are two different forms of teacher knowledge. Researchers still have problems for defining pedagogical content knowledge (Graeber & Tirosh, 2008) Thus, two models were constructed based on literature review. Model I represents the teacher knowledge model that content knowledge and pedagogical content knowledge are both constitute a single form teacher knowledge, and serves as a null model for teacher knowledge. Model II was the proposed model for the teacher knowledge that content knowledge and pedagogical content knowledge are two different categories of teacher knowledge. These models are presented in Figure 3.4.

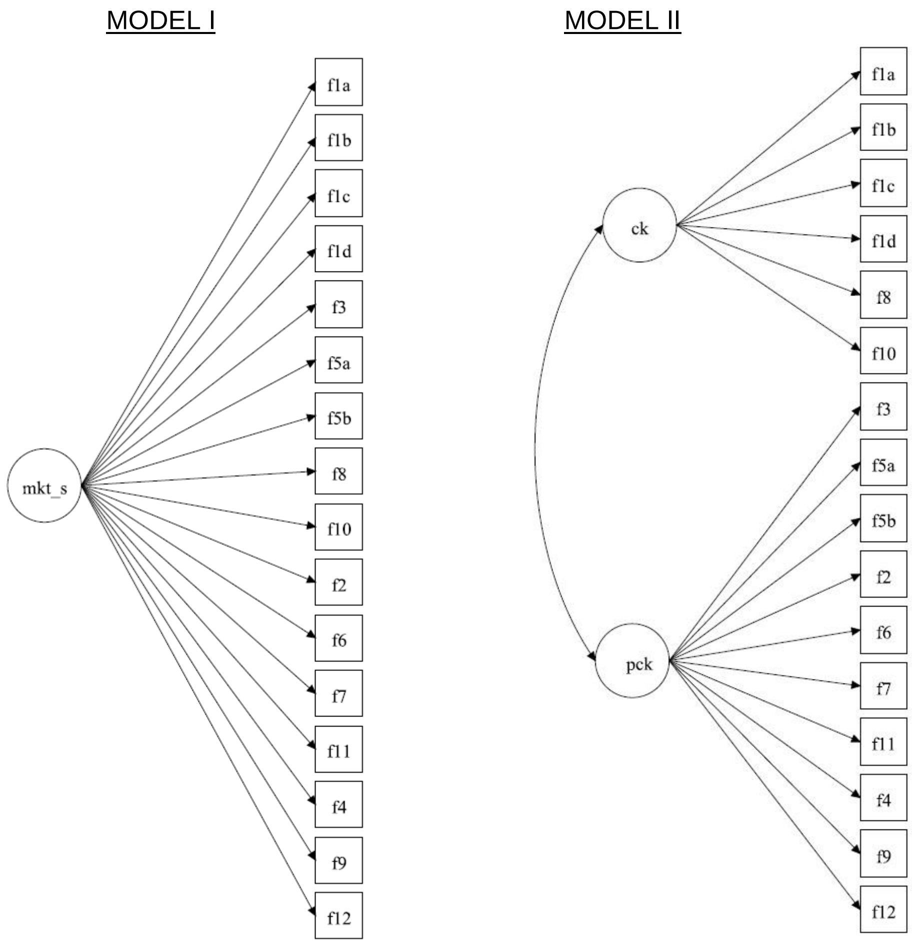


Figure 3.4. Proposed Models for MKT-S Instrument

To conduct confirmatory factor analysis, several software options searched and Mplus 7.1 software (Muthen & Muthen, 1998-2012) was found to be most comprehensive software for handling confirmatory factor analysis of categorical variables especially in the case of multi-factor solutions (Brown, 2006). Mplus has also capacity to handle multiple form structure of the instrument (Blömeke, Houang & Suhl, 2011)

Default estimator for categorical variables is robust weighted least squares (WLSMV) in Mplus 7.1 (Muthen & Muthen, 1998-2012). WLSMV estimator uses full diagonal weight matrix, however this matrix is not inverted during estimation process. Therefore diagonal weight matrix does not have to be positive definite, and this brings additional advantages when analyzing categorical variables.

Another confirmatory factor analysis option is using multidimensional item response theory (MIRT). It is also available in Mplus using maximum likelihood estimation with robust standard error (MLR). Unlike other item response theory software, Mplus does not directly provide difficulty (*b*) and discrimination (*a*) values for each item. However, the software provides factor loadings and threshold values, and these values can be convertible to conventional item response theory (IRT) *a* and *b* values (Muthen & Muthen, 2006).

There is not much theoretical difference between conventional confirmatory factor analysis and multidimensional item response theory analysis (Muthen et all, 1991), and Brown (2006) states the difference as "IRT … relates characteristics of items and characteristics of individuals to the probability of endorsing a particular response category…Whereas CFA aims to explain the correlations among test items" (p. 396).

In practice, MLR estimator reports only Loglikelihood, AIC, BIC and Adjusted BIC when there are thirteen variables in the analysis, and can be quite time consuming when there are more than three factors. Whereas WLSMV reports fit indices (χ^2, RMSAE, TLI, CFI and WRMR) and modification indices independently from the number of variables studied. WLSMV provided χ^2 statistics cannot be used directly for comparison of non-nested models (Muthen & Muthen, 1998-2012), however, MLR provided Loglikelihood, AIC, BIC and Adjusted BIC can be used for model comparison (Raftery, 1995; Blömeke et al., 2011)

Even though there are several fit indices that can be used to evaluate model fit when χ^2 statistics is significant, Mplus software supply limited number of fit indices when dealing with ordered categorical variables. Therefore, it is possible to evaluate model fit using χ^2, RMSAE, TLI, CFI and WRMR fit indices. Another fit index which is not reported directly by Mplus is χ^2/df (Byrne, 2010), and it can be calculated from a Mplus output easily. Yu (2002) reviewed fit indices for outcomes that have severe non-normality and for binary outcomes. In summary, she found that cut-off values for fit indices that indicates good model fit are χ^2 p-value $\geq$ 0.05, CFI $\geq$ 0.95, RMSAE $\geq$ 0.05 and WRMR $\leq$ 1.0 when the sample size is larger than 500 for severely non-normal or binary outcomes.

CHAPTER 4

RESULTS

This chapter consisted of four sections. Validation process of the MKT-S instrument explained in the first section. Results regarding to content knowledge dimension of the MKT-S instrument presented in second section. Results regarding to pedagogical content knowledge dimension of the MKT-S instrument presented in third section. A summary of results presented in the last section of this chapter

4.1 Validation of the Mathematical Knowledge for Teaching Statistics Instrument

This part of the chapter aimed to answer first research problem, and provided information about validation processes of the MKT-S instrument including results for confirmatory factor analysis, evidences for concurrent validity and reliability analysis of scores obtained from the instrument.

4.1.1 Confirmatory (Item) Factor Analysis Results of MKT-S

To validate the factor structure of the MKT-S instrument, it was needed to test that proposed two-factor model (Model II) better fits to data than one-factor model (Model I). To achieve this goal, results were acquired using MLR estimator. One-factor solution (Model I) contained 68 parameters while two factor solution (Model II) contained 69 parameters. A chi-squared difference test conducted for assessing the fit of these two models. Results are summarized in Table 4.1. Chi-Squared difference test results showed that two-factor model (Model II) significantly ($\Delta\chi^2(1)$=7.95, p< 0.01) better fitted to data than one-factor model (Model I). Standardized factor loading are shown in Figure 4.1, and a detailed Mplus output could be seen at Appendix E.

Table 4.1. Fit Indices for Model I and Model II for MLR Estimator.

Model	Log Likelihood	Scaling Correction Factor	Number of Parameters	$BIC_{adj.}$*	χ^2 Difference	*df* $(\Delta\chi^2)$	*p*-value $(\Delta\chi^2)$
Model I	-8449.10	1.0279	68	17123.68	7.95	1	0.0048
Model II	-8441.15	1.0274	69	17111.09			

* Adjusted Bayesian Information Criterion

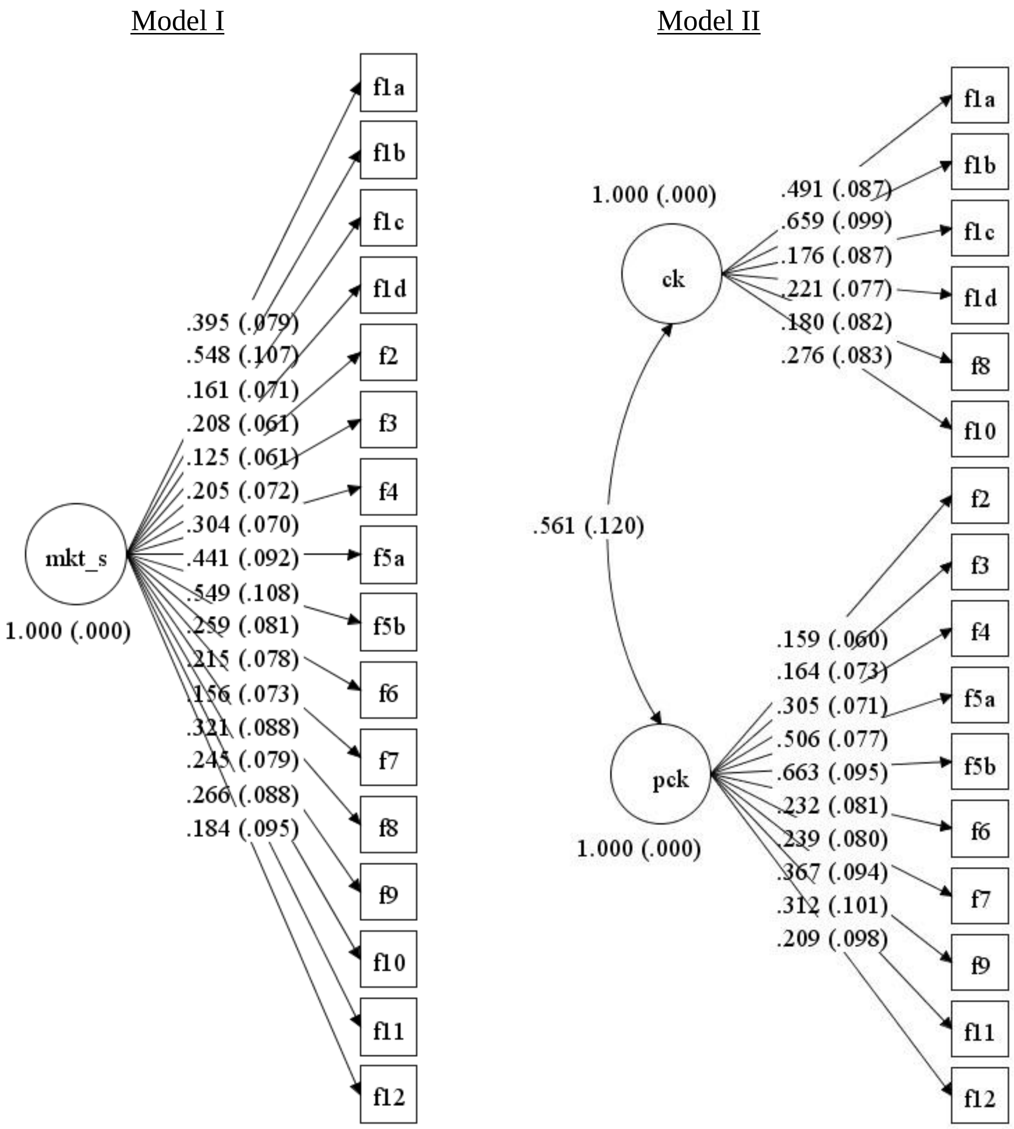


Figure 4.1 Standardized Loadings for Model I and Model II Using MLR Estimator.

Since MLR estimator is only useful for model comparison, fit of these two models also tested using WLSMV estimator and results are summarized in Table 4.2.

Table 4.2. Fit Indices for Model I and Model II for WLSMV Estimator.

Model	χ^2	df	p-value	χ^2/df	RMSAE	CFI	TLI	WRMR
Model I	166.05	104	0.0001	1.596	0.030	0.781	0.748	0.961
Model II	151.33	103	0.0014	1.469	0.027	0.830	0.802	0.915

Even though Model II seemed to fit to data better than Model I, Chi-squared differences cannot be computed directly from WLSMV output because Chi-square differences for WLSMV is not distributed as chi-square. However, there is a DIFFTEST option in Mplus, which utilizes chi-square testing for nested models. When these two models tested using DIFFTEST command, Model II showed significantly better fit than Model I, $\Delta\chi^2(1)$=11.549, p< 0.001.

Results of both estimators significantly favored Model II. These results led to conclude that items contained in MKT-S instrument do not uniformly measure a single construct (teacher knowledge). Instead, pedagogical content knowledge (PCK) and content knowledge (CK) are two different constructs that both of them had their own characteristics.

4.1.2 Model Improvement

After confirming two-factor structure of the instrument, modification indices reported by Mplus were examined. Mplus reported two modification indices that may improve the fit of the Model II, and these indices can be seen in Table 4.3. Recommended modifications were correlating item F.1D with F.1C, and correlating F5B with F5A. Recommended correlations clearly made sense because both indices were related to items which share same stem even though they seek different information.

Table 4.3. Modification Indices for Model II.

Pair	Modification Index	Expected Parameter Change
F1D with F1C	13.211	0.195
F5B with F5A	12.676	0.336

First modification conducted was correlating item F1D with F1C resulting in Model IIA and second modification conducted was correlating item F1D with F1C resulting in Model IIB. Confirmatory factor analysis results, using WLSMV estimator, are summarized in Table 4.4.

Table 4.4. Fit Indices for Model IIA and Model IIB for WLSMV Estimator.

Model	χ^2	*df*	*p*-value	χ^2/df	RMSAE	CFI	TLI	WRMR
Model IIA (F1D with F1C)	137.945	102	0.0103	1.352	0.023	0.873	0.851	0.870
Model IIB (F5B with F5A added to Model IIA)	124.999	101	0.0530*	1.237	0.019	0.915	0.900	0.821

*Model significantly fitted to data at 0.05 level

Even though it seemed each modification improved model fit, DIFFTEST command of Mplus had to be applied to test chi-square differences. The test results are summarized in Table 4.5.

Table 4.5. Results for Comparing Chi-square Values

Compared Models	$\Delta\chi^2$	*df*	*p*-value
Model II and Model IIA	14.189*	1	0.0002
Model IIA and Model IIB	10.708*	1	0.0011

*Differences tested using DIFFTEST option of Mplus

DIFFTEST results showed first modification significantly improved fit of Model II , so Model IIA fitted to data better than Model II ($\Delta\chi^2(1)$=14.189, p< 0.001), and second modification significantly increased fit of Model IIA so Model IIB fitted to data better than Model IIA ($\Delta\chi^2(1)$=10.708, p< 0.01).

Final model, Model IIB, significantly fitted to data, $\chi^2(101)$=124.999, p>0.05 (Barret,2007). Most of the other fit indices also showed good fit of model. For example, χ^2/df was 1.137 and it was lower than most conservative cut-off value of 2. RMSAE was 0.019, and it was lower than 0.05. WRMR was 0.821 and it was lower than 1.0. On the other hand, CFI and TLI indices showed poor fit.

Since the χ^2 statistic was not significant, it was concluded that proposed model (Model II) is currently best model that fitted to data and standardized loadings of Model IIB are shown at Figure 4.2

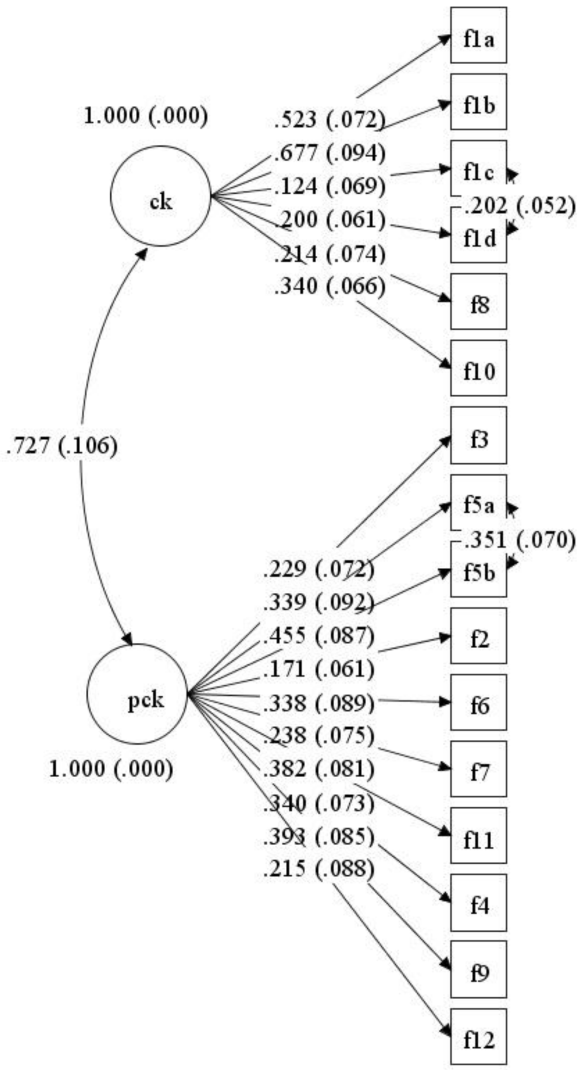


Figure 4.2. Standardized Factor Loadings for Model IIB.

Next step in validation process was comparing results with previously conducted TEDS-M study. Since this study and TEDS-M study used similar framework, it was possible to compare results by following the method that explained by Blömeke, Houang and Suhl (2011). To achieve this goal, researcher also constrained the factor loadings to be same within each factor. Table 4.6 shows the comparison of this study with TEDS-M study.

Table 4.6. Comparison of Results with TEDS-M Study.

Model	Factor Loading for CK items	Factor loadings for PCK items	R^2	
			CK	PCK
TEDS-M	0.34 (0.00)***	0.30 (0.01)***	0.12 (0.00)	0.09 (0.00)
Current Study	0.306 (0.028)***	0.299 (0.023)***	0.094 (0.017)	0.089 (0.014)

***p< 0.001. Parenthesis represent standard errors.

A comparison of the results showed that this study had similar finding for both CK and PCK factors. Only clear differences were observed for CK factor were average loading was 0.34 and R^2 was 0.12 for TEDS-M study while average loading was 0.306 and R^2 was 0.094 for this study. Since TEDS-M study covered more items than this study, it was concluded that differences were small and arbitrary. Even though TEDS-M study covers a broad range of topics and this study covers only some of statistics topics, results of MKT-S instrument was consistent with MKT-S study.

4.1.3 Concurrent Validity Evidences for MKT-S instrument

After validating factor structure of the MKT-S instrument, factor scores for CK and PCK were calculated using Maximum Likelihood (MLR) estimator of Mplus. Then tests were conducted for factor scores against predetermined variables.

First, the correlation between factor scores and preservice teachers' Introduction to Statistics and Probability-I (ISP-I) grades were checked. The results are summarized in Table 4.7.

Table 4.7.Correlation Between Factor Scores and ISP-I Grades.

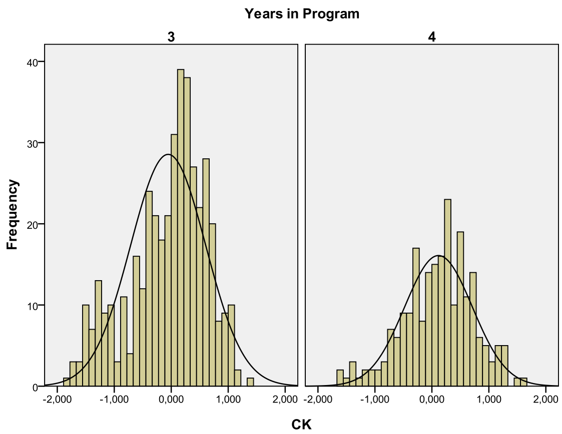
Pairs	Pearson *r*	*n*
ISP-I grade - CK	0.305***	555+
ISP-I grade - PCK	0.273***	555+

+ 104 preservice teachers did not state their ISP-I grades
*** Significant at 0.001 level

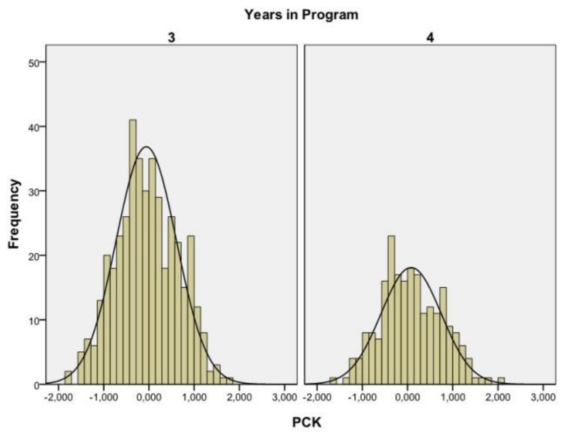
Correlation between CK score and ISP-I grade was 0.305 ($p<0.001$), and correlation between Mathematics PCK score and ISP-I grade was 0.273 ($p<0.001$). Small but significant positive correlation was found between factor scores and preservice teachers' ISP-I grades. Since the ISP-I course covers much broader content than this study, it was concluded that scores obtained from MKT-S was instrument consistent with preservice teachers' ISP-I course grades.

As a second step, factor score differences were tested for third and fourth year preservice middle school mathematics teachers. First the assumptions to conduct a MANOVA test were checked. Box's test, regarding to equality of covariance matrices, results showed that covariance matrices were same for third year and fourth year preservice teachers, Box's M=5.116, F(3, 5337424)=1.699, p=0.165.

Then the normality of CK and PCK scores were inspected using both histograms and normality tests. Histograms are shown at Figure 4.3 and normality tests are presented in Table 4.8.



(a)



(b)

Figure 4.3. Histograms for CK and PCK.

Table 4.8. Tests of Normality

Variable	Years in Program	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	*df*	*p*	Statistic	*df*	*p*
CK	3	.098	421	.000***	.960	421	.000***
	4	.053	220	.200	.989	220	.082
PCK	3	.039	421	.140	.994	421	.107
	4	.052	220	.200	.993	220	.383

***Significant at 0.001 level

Normality tests showed that all scores were normally distributed except CK scores for third year preservice teachers. However, as seen on the left hand side of Figure 4.3 (a), histogram of CK scores for third year preservice teachers can be accepted as normally distributed. After checking assumptions, MANOVA test was conducted.

Field (2005) recommends using Pillai's Trace statistics when groups differ along more than one variable and MANOVA test results revealed that differences existed among third year and fourth year preservice teachers, F(2, 638)=5.076, p=0.007, partial eta squared=0.016. Then independent samples-t-tests were conducted for follow up analysis and findings are reported on Table 4.9.

Table 4.9. Descriptive Statistics and t-test Results

Factor	n*		$\bar{x}$		t	df	p-value	Cohen's d
	3	4	3	4				
CK	421	220	-0.055	0.117	-3.177	639	.002	0.264
PCK			-0.046	0.084	-2.317	639	.021	0.193

* 18 preservice teachers did not state their years in program

Results showed that fourth year preservice teachers CK factor score was significantly ($p<0.01$) higher than third year preservice teachers, and fourth year preservice teachers PCK factor score was significantly ($p<0.05$) higher than third year preservice teachers. Even though differences were significant, effect size for CK factor was small and effect size for CK factor was barely small.

4.1.4 Psychometric Properties of MKT-S Instrument

4.1.4.1 Reliability

There are several ways of getting a reliability coefficient for scores obtained using item response theory. IRT uses the test information to describe the accuracy of the test at each level of proficiency. In IRT approach, standard error of measurement of a proficiency level is inversely related to value of test information function of that proficiency, and defined as

$$SEM(\theta) = \frac{1}{\sqrt{I(\theta)}}$$ (Embretson and Reise, 2000).

Additionally IATA software was used to assess the reliability of scores obtained from the MKT-S instrument because IATA displays test information function along with the IRT scores, and provides a holistic reliability coefficient which is defined as the proportion of variability in observed scores that can be explained by variation in true scores, and computed using following formula,

$$\text{reliability} = \sqrt{1 - \frac{average\ SEM^2}{variance\ of\ test\ scores}} \qquad \text{(Cartwright, 2013).}$$

4.1.4.1.1 Reliability of CK Factor

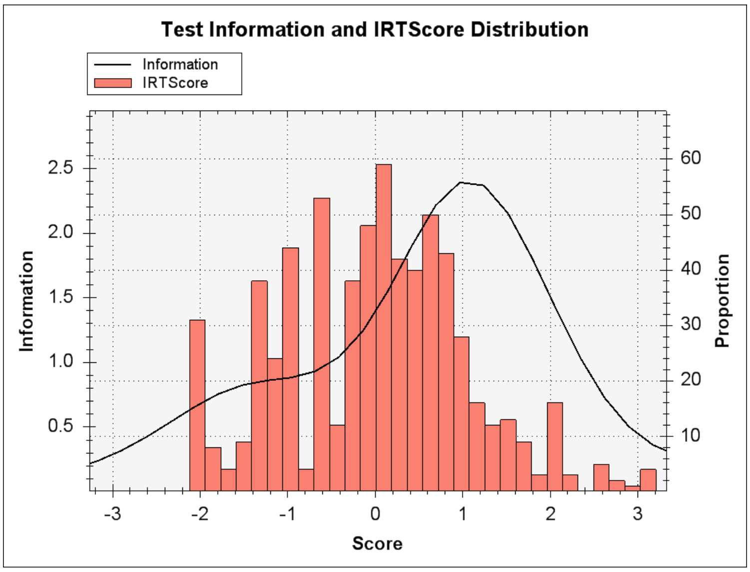


Figure 4.4. IRT Score Distribution and Information Function for CK Scores.

IRT score distribution and information function, $I(\theta)$, for CK scores are shown at Figure 4.4. For CK scores, maximum information occurred at $\theta = 1$, and information value was about 2.4 for this point. Standard error of measurement, which corresponds to information value of 2.4, was 0.64 for $\theta = 1$. Reliability coefficient provided by IATA, which is based on average SEM^2 of IRT scores, was 0.65. Information was generally high between $\theta = 0$ and $\theta = 2$, and smallest standard error of measurement also occurred between these points. These results showed that mostly difficult items influenced the reliability of the CK scores. Therefore, CK scores were more reliable for high ability preservice teachers.

4.1.4.1.2 Reliability of PCK Factor

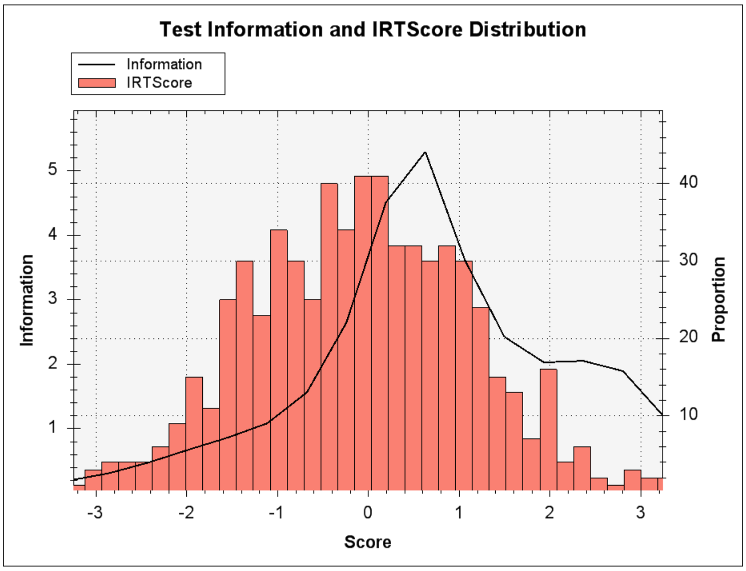


Figure 4.5. IRT Score Distribution and Information Function for PCK Scores.

IRT score distribution and information function, $I(\theta)$, for PCK scores are shown at Figure 4.5. For PCK scores, maximum information occurred at $\theta = 0.6$, and information value was about 5.3 for this point. Standard error of measurement, which corresponds to information value of 5.3, was 0.43 for $\theta = 0.6$. Reliability coefficient provided by IATA, which is based on average SEM^2 of IRT scores, was 0.76. Information was generally high between $\theta = -0.2$ and $\theta = 3$, and smallest standard error of measurement also occurred between these points. These results showed that mostly difficult items influenced the reliability of the PCK scores. Therefore, PCK scores were also more reliable for high ability preservice teachers.

4.1.4.2 IRT Parameters for MKT-S Instrument

4.1.4.2.1 Item Parameters for CK Items

Table 4.10 summarizes the item parameters (IRT) for CK items. Difficulties were high for 'reasoning' type items (F1C, F1D and F10), and difficulties were especially very high for the last levels of Item F1C and F1D. Difficulty level for 'applying' type item (F8) was moderate, and difficult level for 'knowing' type items (F1A and F1B) were low. Discrimination levels were usually low for 'reasoning' type items while discrimination levels were slightly higher for 'knowing' type items.

Table 4.10. IRT Parameters for CK Items

Item	Concept	Cognitive Type	*a* (Discrimination)	*b* (Difficulty)	
				Value	Level
F1A	Mean	Knowing	1.021	-0.190	1
F1B	Mean	Knowing	1.589	-0.619	1
F1C	Mean	Reasoning	0.324	1.012	1
				1.253	2
				1.398	3
				6.623	4
F1D	Mean	Reasoning	0.411	-0.182	1
				1.056	2
				2.002	3
				6.482	4
F8	Graph	Applying	0.331	0.719	1
F10	Graph	Reasoning	0.520	0.687	1
				1.285	2
				1.613	3
				1.823	4

4.1.4.2.2 Item Parameters for PCK Items

Table 4.11 summarizes item parameters (IRT) for PCK items. Difficulties for PCK items ranged from 0.670 to 5.173 for the last levels. Taking partial credit for an answer was especially easy for items F2, F7, F11 and F12. However, taking partial credit for Item F3 was very difficult.

Difficulties of first and second level were parallel for the items F3, F4, F5A, F5B, F7, F9, and F12. Difficulties of third and fourth level were also parallel for the items

F5A, F5B and F7. The gap between second and third level was most apparent for the items F7 and F12.

Discrimination levels for PCK items ranged from 0.292 to 1.606, and discrimination levels were for mean related PCK items were usually higher then graph related items.

Table 4.11. IRT Parameters for PCK Items

Item	Concept	PCK Objective	*a* (Discrimination)	*b* (Difficulty) Value	Level
F2	Graph	PCK_1	0.292	-3.045	1
				0.045	2
				2.853	3
F3	Mean	PCK_3	0.301	1.492	1
				1.585	2
				3.973	3
				5.173	4
F4	Graph	PCK_4	0.581	0.028	1
				0.074	2
				0.115	3
				1.024	4
F5A	Mean	PCK_5	1.065	-0.063	1
				-0.040	2
				0.557	3
				0.670	4
F5B	Mean	PCK_6	1.606	0.210	1
				0.238	2
				0.910	3
				1.235	4
F6	Graph	PCK_3	0.433	0.469	1
				1.453	2
F7	Graph	PCK_2	0.447	-2.667	1
				-2.369	2
				2.980	3
				3.186	4
F9	Graph	PCK_5	0.715	0.283	1
				0.324	2
				0.499	3
				2.512	4
F11	Graph	PCK_2	0.596	-1.383	1
				-0.621	2
				-0.169	3
				0.831	4
F12	Graph	PCK_6	0.388	-1.219	1
				-1.075	2
				2.235	3
				5.034	4

4.1.5 Relationship Between Content Knowledge Scores and Pedagogical Content Knowledge Scores

In this part, the result regarding the relationship between CK factor scores and PCK factor scores was presented. To address this problem, the scatterplot of CK and PCK, which is shown at Figure 4.6, was examined.

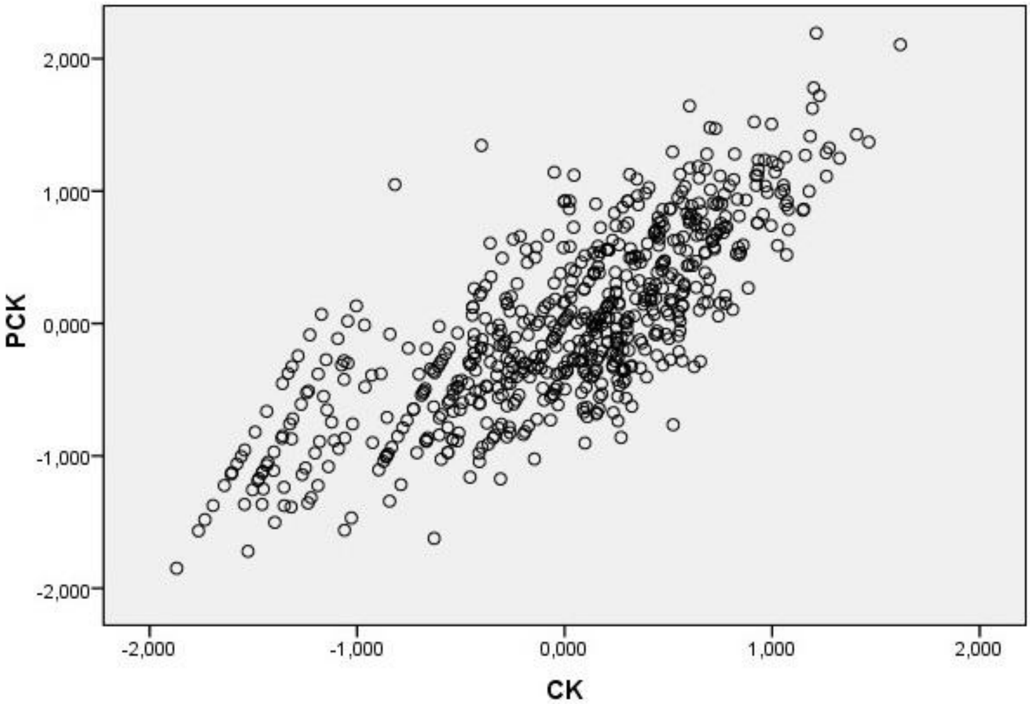


Figure 4.6. Scatterplot of CK and PCK Scores

After the linear relationship between CK and PCK scores was seen, the Pearson correlation coefficient was checked for these scores. It was found that correlation between CK scores and PCK scores was very high ($r=0.78$, $p<0.001$). This high correlation coefficient implied that content knowledge and pedagogical content knowledge dimensions were closely related to each other for preservice teachers. Therefore, a high content knowledge score was generally corresponding to a high pedagogical content knowledge score.

4.2 Preservice Middle School Mathematics Teachers' Content Knowledge in Statistics

In this section, survey results regarding the second research problem were presented. To address this problem, one mean related item and one graphics related item selected for detailed analysis. Item F.1C was selected as a mean related item and Item F.10 for the graphics related item because there were comparable research findings related to these items in literature. Thus, it was possible to compare and contrast content knowledge of preservice teachers with other cultures.

4.2.1 Preservice Teachers' Content Knowledge of Central Tendency

Item F.1C was selected as an average related item because this item requires preservice teachers to think simultaneously about arithmetical mean, median, mode and distribution of the data.

The Item F.1C requires preservice teachers to think the scenario presented in Figure 4.7.

> When asked by their teacher, 11 students independently and simultaneously recorded the time taken by another student to run 100m. The times recorded (in seconds) were the following:
>
> 15,05 14,97 13 15 14,98 15 14,93 15,06 14,96 15 14,96
>
> Arithmetical mean of this data is 14,81 seconds. What would be a good estimate for running 100 m. for this student considering arithmetical mean, mod, median and whole data? Explain how you reached this conclusion?

Figure 4.7. Translated Version of Item F.1C. [Adapted from Garcia Cruz & Garrett (2008) with Permission]

Preservice teachers were expected to give a single estimation for this item, and their estimates for the running time of the student were summarized in Table 4.12.

Table 4.12. Preservice Teachers Estimates for Item F.1C

Estimates of preservice teachers*	N	%
Omitted	148	22.46
lower than 10	6	0.91
13-13.92	5	0.76
14	6	0.91
14.03	3	0.46
14.5	12	1.82
14.6-14.62	3	0.46
14.7	8	1.21
14.75	5	0.76
14.8	9	1.37
14.81	108	16.39
14.82-14.85	7	1.06
14.86-14.89	5	0.76
14.9	52	7.89
14.905-14.925	8	1.21
14.93-14.935	20	3.03
14.95	16	2.43
14.96-14.965	22	3.34
14.97	10	1.52
14.975	3	0.46
14.98	53	8.04
14.99	13	1.97
15	133	20.18
15.01	1	0.15
higher than 15.06	3	0.46
Total	659	100

*Some low frequency answers were collapsed for summary reasons. However complete list of results for Item F.1C can be found at Appendix F

Majority of the data points, presented in Item F1.C, were between 14.93 and 15.06, and a value of 13 was a clearly a measurement error since all students measure the running time of same student simultaneously. However, results indicated that more than one third of the preservice teachers reached to an estimation which was lower than 14.93. Results also indicated that most common estimate for running time lower than 14.93 were arithmetical mean, and preservice teachers defended their estimates with several arguments. Some of these arguments were summarized in Table 4.13.

Table 4.13. Some Arguments of Preservice Teachers who Defend Arithmetical Mean

No.	Argument of preservice teacher*
1	Arithmetical mean is more realistic
2	Because it is arithmetical mean
3	Arithmetical mean is always trustable
4	Because arithmetical mean considers all values
5	Because arithmetical mean is best estimator
6	Because arithmetical mean is middle point which represents all values
7	Arithmetical mean is a generalization of all numbers
8	I preferred arithmetical mean because there is no outliers
9	It is arithmetical mean and also close to mode value
10	Because arithmetical mean is equally distant to all values
11	We can use arithmetical mean because there in not much difference between two ends.
12	Mod, median and arithmetical mean are close to each other. Therefore I use arithmetical mean because it is affected from whole data
13	Arithmetical mean is close to other values
14	Since all values are close to each other, I used arithmetical mean
15	If we consider the error rates for each measurement, arithmetical mean is most suitable one
16	Because arithmetical mean is more meaningful than mode and median

* Original arguments are presented in Appendix G, Part 1.

Preservice teachers, who defended arithmetical mean as estimator, based their arguments on the nature of arithmetical mean, and most of them could be acceptable answers if data is normally distributed, free of errors or free of outliers. However, these preservice teachers could not read the data thoroughly and did not consider value of 13 as an erroneous measurement. Some preservice teachers gave mistakenly special attention to the value of 13 and some these ideas summarized in Table 4.14.

Table 4.14. Some Arguments of Preservice Teachers who Mistakenly Consider "Value of 13"

No.	Argument of preservice teacher*
1	My estimate is 14,03 because it is the average of the smallest and the highest number
2	I picked 14 because it is values are between 13 and 15
3	My estimate is 14.90 because 13 lowers the average of the data
4	Values are generally higher than 14.90 but I also considered value of 13 [*estimate stated as 14.90*]
5	Values are piling around 15. However it should be lower than 15 because 13 will lower the average [*estimate stated as 14.50*]
6	Values are generally between 14.95 and 15.05. However 13 could lower this average [*estimate stated as 14.80*]
7	It should be between 13 and 15 but more close to 15 [*estimate stated as 14.50*]
8	Between 13 and 15, since the mode is 15, it should be more close to 15 [*estimate stated as 14.80*]

* Italics added by researcher for clarification, and original arguments are presented in Appendix G, Part 2.

These preservice teachers explicitly stated that they considered value of 13 while estimating the running time of the student. Even though some of them were able to identify that values are generally between 14.93 and 15.06, they insisted on considering value of 13 in the estimation process.

Another group was the preservice teachers who estimated the running time around 14.90. Some of them estimated 14.93 because their argument was that the average of arithmetical mean (14.81), median (14.98) and mode (15) is 14.93. Some of them stated that the average should be somewhere between arithmetical mean, median and mode. Some of these preservice teachers sensed that arithmetical mean was too low estimate for running time. Therefore, they made up a solution that involved three different averages for estimating the center of data.

In other cases, estimates were usually between 14.98 and 15. Preservice teachers, who estimated running time as 14.98, usually defended their estimates using the fact that median was 14.98; and preservice teachers, who estimated running time as 15, usually defended their estimates using the fact that mode was 15. Some of these arguments based on the distribution of the data where value of 13 considered as an outlier instead of an erroneous measurement. Since they considered the value of 13 as an outlier, they reached to a conclusion that data were skewed to right so arithmetical mean was not an appropriate measure for the center. In some cases, they supported their idea that median (or mode) was also stronger estimate for the center because mode and median values were very near to their estimates.

In very rare cases, preservice teachers explicitly stated that value of 13 caused problems during estimation process. Some of these answers summarized in Table 4.15.

Table 4.15. Preservice Teachers who Explicitly Discarded the "Value of 13"

No.	Argument of preservice teacher*
1	Arithmetical mean gives wrong information because of the value of 13 so I chose median as estimate. It is also more trustable than mode
2	I took the average of all values except value of 13 [*estimate stated as 14.90*]
3	I took the average of mode and median. I did not considered the arithmetical mean because of the value of 13
4	My estimate is 14.99. The reality of the value of 13 is open to discussion so I averaged the all other values [*averaging process was not stated explicitly*]

* Italics added by researcher for clarification, and original arguments are presented in Appendix G, Part 3.

Even though limited number of preservice teachers explicitly discussed the trustworthiness of the data point of 13, according to classification of Garcia Cruz and Garrett (2008) [or extended abstract level according to Groth and Bergner (2006)] only two preservice teachers gave the relational response as "take the average of all values except value of 13". One of these preservice teachers' estimates was 14.90, and this value was not accurate for the method he or she described for modification process.

In summary, the depth of the preservice middle school mathematics teachers' content knowledge related to average concept was limited to the fact that arithmetical mean is not trustable when the distribution of the data skewed. Two preservice teachers had much deeper knowledge than this level, and it was considered that these cases were extreme and not generalizable to all preservice middle school mathematics teachers.

4.2.2 Preservice Teachers' Content Knowledge of Graphics

Item F.10 was constructed to measure the preservice middle school mathematics teachers' content knowledge related to graphics concept. This item requires preservice teachers to construct a histogram from an extra ordinary data. Data presented in item F.10 was consisted of values that had one decimal point. Our experience with statistics textbooks indicated that histogram construction examples generally use data that are consisted of integers. Since the procedures in these examples described for integers, this item requires preservice teachers to think thoroughly the logic behind histogram construction, and to extend their histogram knowledge for the data that includes decimal points. The translated version of the Item F.10 presented in Figure 4.8.

A teacher wanted students to collect data to construct histogram. A student wanted to use the weights of students in her classroom as data so she brought a digital scale from her house. Then she measured weight of 30 students using this digital scale. The table shows the measurements of these 30 students. (unit is kilogram)

16,1	16,5	16,6	16,7	16,8	17,6	17,6	17,8	18,0	18,1	19,5	19,8	20,0	20,1	21,8
21,9	22,0	22,1	22,1	22,2	22,4	22,7	22,7	23,5	23,5	23,6	24,5	24,5	24,7	26,8

Using this data, construct a histogram that has 5 intervals.

Figure 4.8. Translated Version of the Item F.10

Preservice teachers score distribution for this item summarized in Table 4.16. Results indicated that one quarter of preservice teachers omitted this item while another quarter of preservice teachers' answer was completely wrong. Some preservice teachers, who omitted this item, honestly admitted that they had no idea about what a histogram is.

Table 4.16. Preservice Teachers' Score Distribution for Item F.10

Score	Frequency	Percent
Omitted	170	25,80
0	176	26,71
1	103	15,63
2	46	6,98
3	25	3,79
4*	139	21,09
Total	659	100,00

*Preservice teachers who draw a correct histogram with 4, 5 or 6 intervals got a score of 4.

Qualitative analysis of zero scoring preservice teachers' answer revealed several important findings. Some of these preservice teachers constructed a graph that is completely different from histogram. For example, some preservice teachers draw a scatterplot that shown at Figure 4.9. These preservice teachers interpreted data as bivariate and used one axis for data presented in Item. They used another axis either for the data point itself or for the order of corresponding data point. This interpretation resulted in a data that looked similar to $\{(16.1, 16.1), (16.5, 16.5), \dots , (26.8, 26.8)\}$ or $\{(16.1, 1), (16.5, 2), \dots , (26.8, 30)\}$

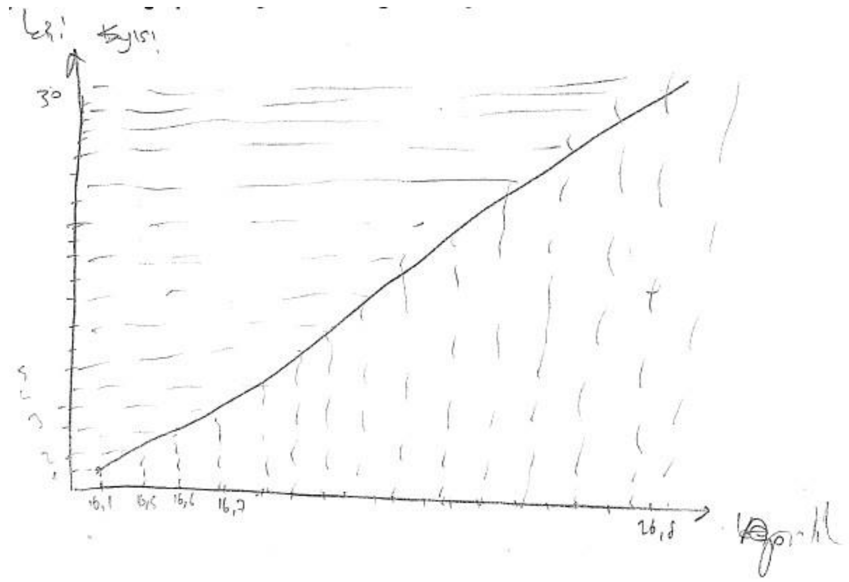


Figure 4.9. Interpreting Data as Bivariate

Another group of zero scoring preservice teachers used regular frequency table for their graphs. Some of them draw line graphs that presented in Figure 4.10. Some of them draw bar graphs that presented in Figure 4.11, and some of them draw adjacent rectangles that was similar to histogram, which presented in Figure 4.12.

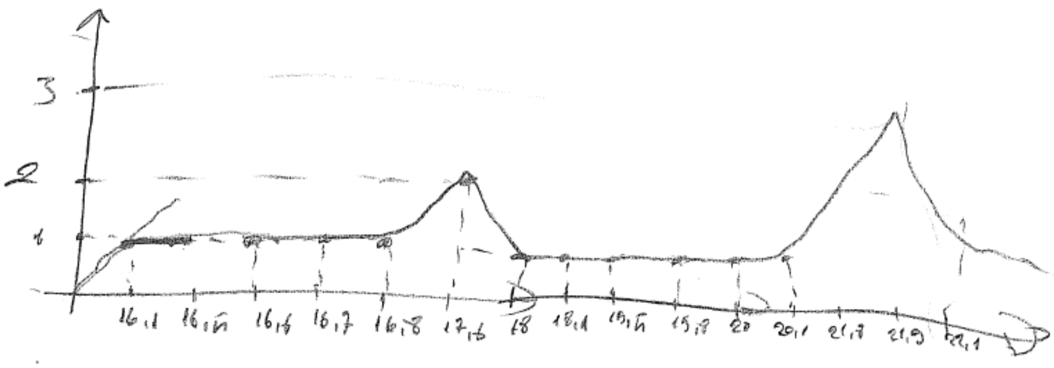


Figure 4.10. Frequency Polygon without Intervals

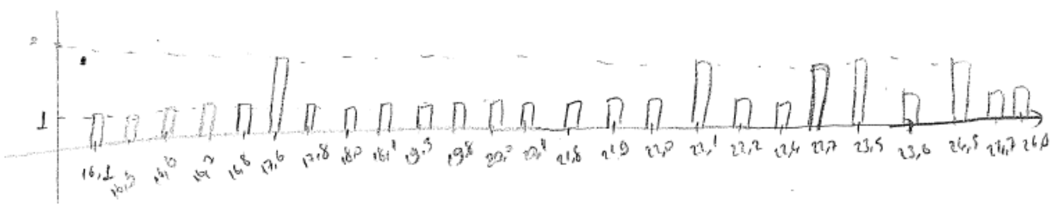


Figure 4.11. Separate Bars without Intervals

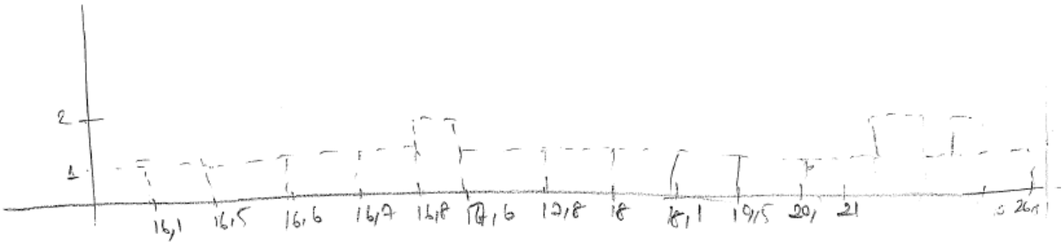


Figure 4.12. Adjacent Bars without Intervals

Another group of zero scoring preservice teachers grouped measurements that had same integer parts. This resulted in a frequency table with 11 intervals where the interval length was "1". Graphs of these preservice teachers also had different patterns. One of them drew a graph that was similar to greatest integer function graph, which presented in Figure 4.13. Another two preservice teachers presented this information using a graph that was similar to histogram that presented in Figure 4.14

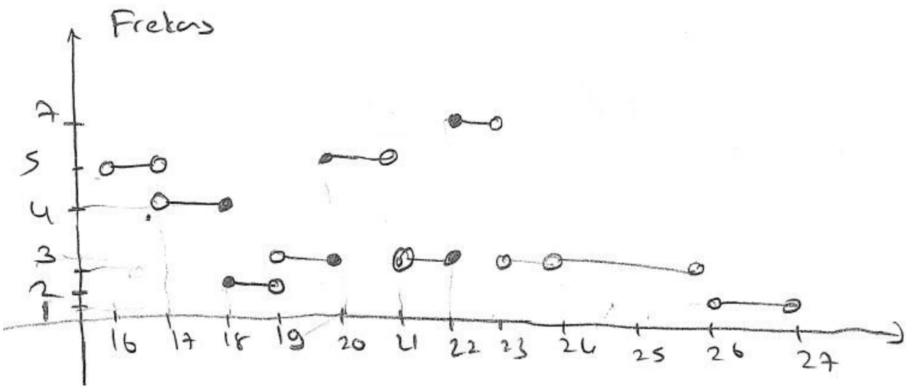


Figure 4.13. A Graph that is Similar to Greatest Integer Function

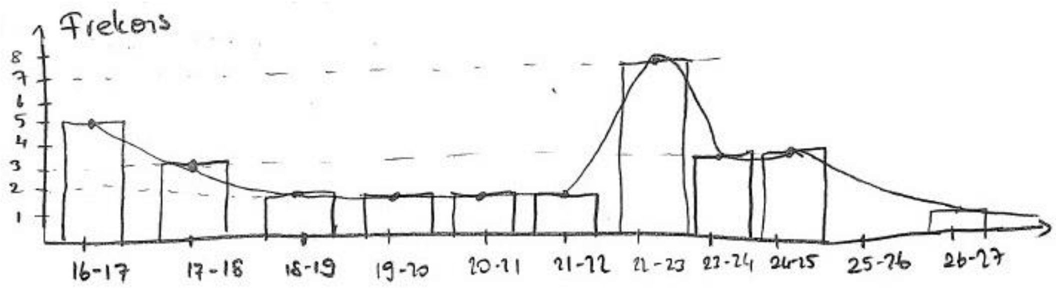


Figure 4.14. Frequency Polygon and Separate Bars (interval length=1)

Majority of zero scoring preservice teachers had an idea that constructing a histogram with five intervals requires dividing data (30 measurements) into five equal parts (so every part should have had 6 measurements). After dividing data into five equal groups, they draw graphs in several ways. Some preservice teachers draw rectangles (adjacent or separate) that height of each rectangles increased. Three different graphs, which presented in Figure 4.15, Figure 4.16 and Figure 4.17, represent this situation.

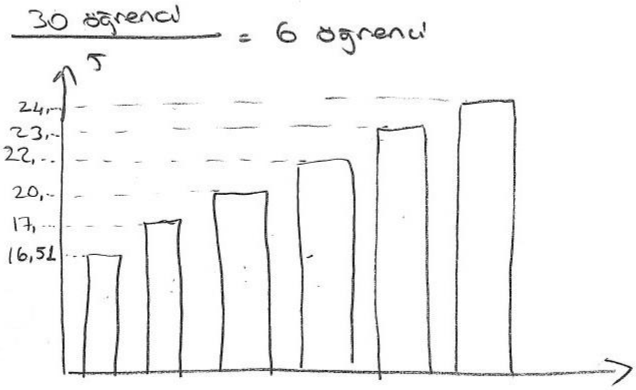


Figure 4.15. Increasing Bars with Equal Measurement in Each Interval (Error 1)

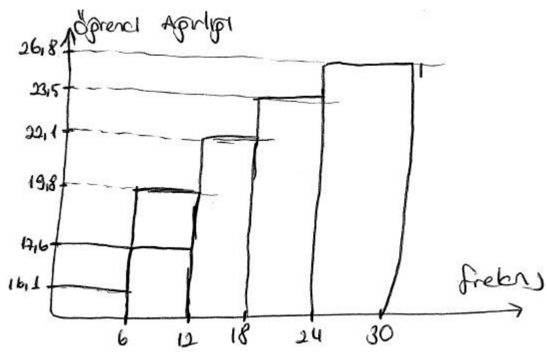


Figure 4.16. Increasing Bars with Equal Measurement in Each Interval (Error 2)

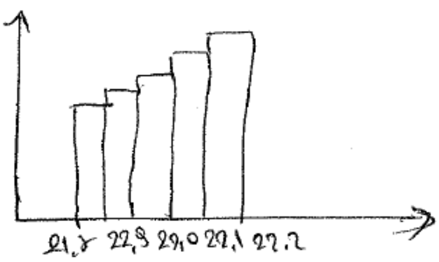


Figure 4.17. Increasing Bars with Equal Measurement in Each Interval (Error 3)

Most preservice teachers, who divided data into five equal groups, draw graphs looked similar to histograms. Their graphs consisted of same level of rectangles that were separate as in Figure 4.18 or adjacent as in Figure 4.19.

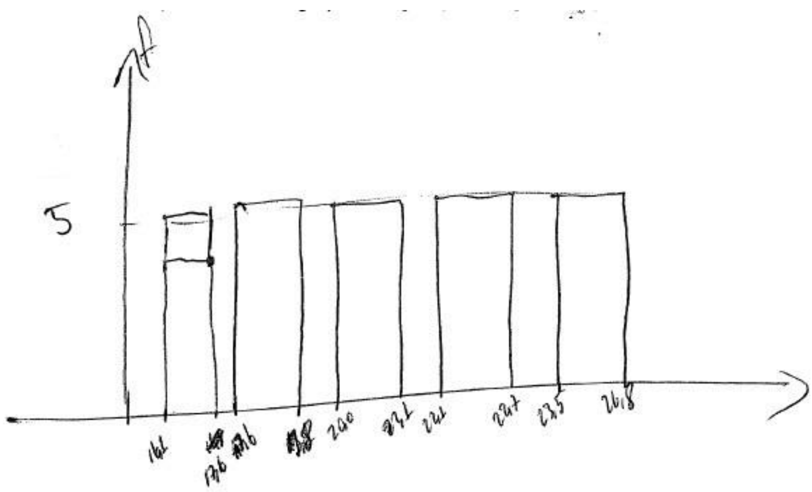


Figure 4.18. Equal Separate Bars with Equal Measurement in Each Interval

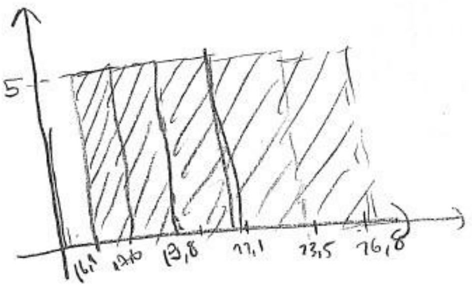


Figure 4.19. Equal Adjacent Bars with Equal Measurement in Each Interval

During item development, the data of this item was designed in a way that would reveal the depth of preservice teachers' histogram knowledge. This data required preservice teachers to know that interval length can be decimal number in some special cases. Range of the data in Item F.10 was 10.7. Therefore each interval should be little larger than 2,14 in order to construct a histogram with five intervals. Using interval length as 2 or 3 does not result in a histogram with five intervals. Using interval length as 2 results in a histogram with six intervals, and using interval length as 3 results in a histogram with four six intervals.

A small number of preservice teachers accepted interval length as 3 (or larger in rare cases). Therefore, they constructed a histogram with less than 5 intervals such as in Figure 4.20.

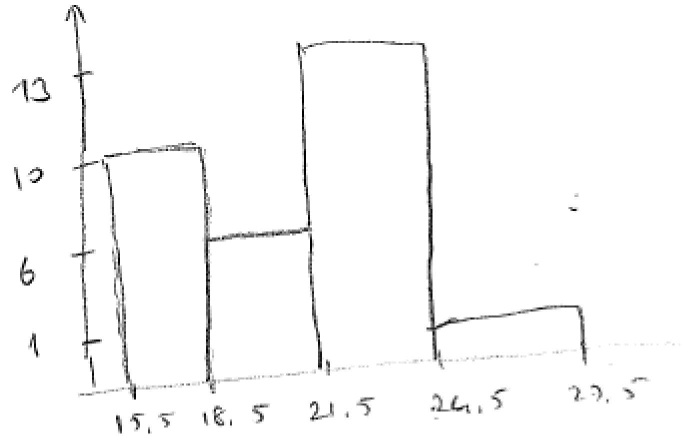


Figure 4.20. Histogram with Four Intervals

It was also seen that a large number of preservice teachers accepted interval length as 2. This process normally results in a histogram with six intervals and some preservice teachers' drawings, such as Figure 4.21, were appropriate for this situation

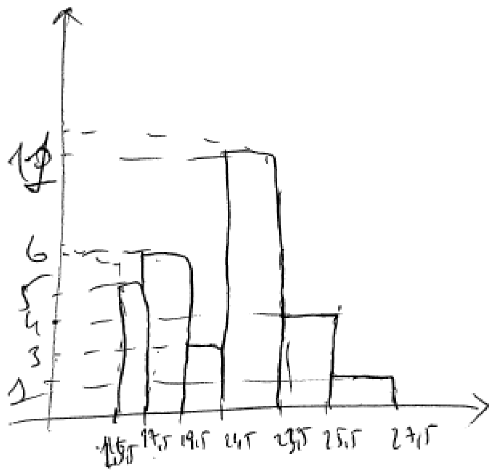


Figure 4.21. Histogram with Six Intervals

However, a very large number of preservice teachers, who accepted interval length as 2 (or 2.1), draw a histogram with five intervals. Upon the examination of their graphs, it was seen that these preservice teachers mostly combined fifth and sixth interval. In other words, they added last data point (26.8) to the fifth interval in order to draw a histogram with five intervals. An example for this situation is presented in Figure 4.22.

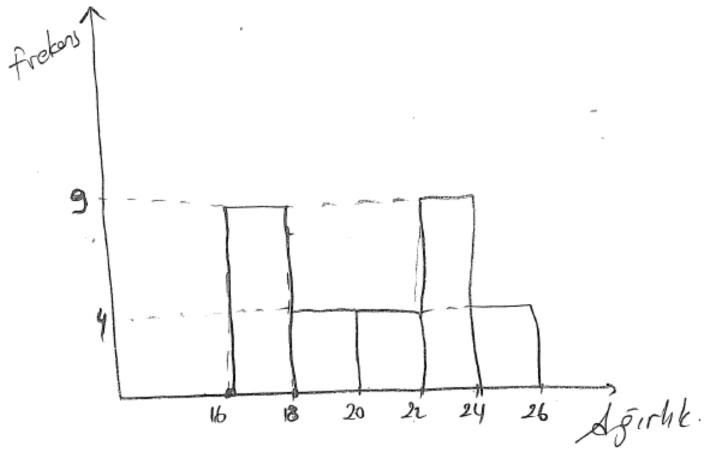


Figure 4.22. Histogram Forced to Have Five Intervals

In other cases, preservice teachers accepted the interval length between 2.2 and 2.5. However using an appropriate interval length did not always resulted in a correct histogram. Some of these preservice teachers constructed a histogram where rectangles were not adjacent such as in Figure 4.23.

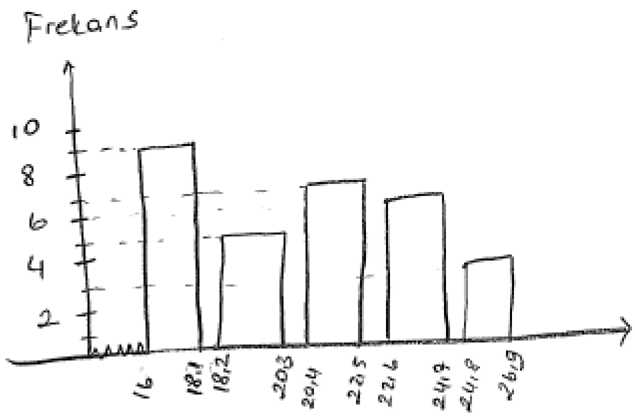


Figure 4.23. Separated Bars with Five Intervals

Last group of preservice teachers, who were wanted to discuss, were those preservice teacher who had deep knowledge about constructing histograms. These preservice teachers constructed a histogram with five intervals using an appropriate interval length such as in Figure 4.24.

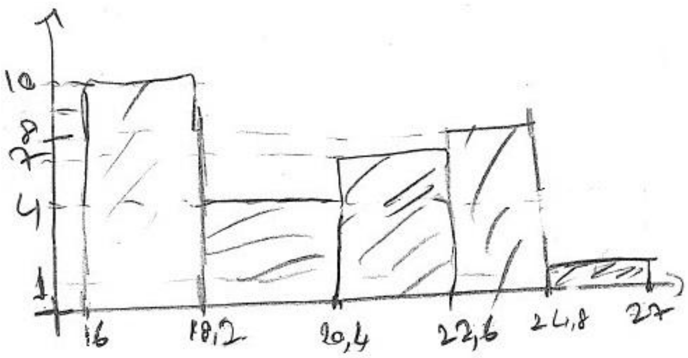


Figure 4.24. Sample for Correct Histogram

However, preservice teachers, who had the deep knowledge of constructing histogram, consisted nearly ten percent of sample. Therefore, it was assumed that only one tenth of preservice teachers, who had a deep knowledge of histogram, will graduate from the education faculties.

4.3 Prospective Middle School Mathematics Teachers' Pedagogical Content Knowledge in Statistics

In this section, survey results regarding the third research problem were presented. To address this problem, one item for each PCK objectives selected for detailed analysis. The content of the items also considered, and all three mean concept related items of PCK (F.3, F.5A and F.5B) selected, and these items were represented PCK-3, PCK-5 and PCK-6 objectives. The other three items were selected from graphics concept related items of PCK (F.2, F.11 and F4) which were represented PCK-1, PCK-2 and PCK-4 objectives.

4.3.1 *Preservice Teachers' Ability to See Connections (PCK-1)*

PCK-1 is "Selecting possible pathways and seeing connections within the statistics curriculum", and defined as "Future teacher should be able see connections between statistics topics and know how a statistics topic can be related with another topic. This is also includes seeing connections between topics that taught in different grades".

To address this objective, a graphic related complex multiple-choice item (Item F.2) was constructed, and this item requires preservice teacher to think about the following scenario,

> "consider a specific data Mrs. Fatma used last year to create a picture graph, and help her to decide whether that the same data can also meaningfully be used to create (I) a pie graph, (II) a bar graph, (III) a line graph, and (IV) a histogram."

The picture graph that presented to preservice teachers was summarizing the number of three different flavors of candies. Preservice teacher expected to know that data represents frequencies for a categorical variable. Therefore, this data can be used to create a pie graph or a bar graph, but cannot be used to create a meaningful a line graph or a histogram. Preservice teachers got 1 point for every graph evaluated correctly.

Table 4.17 Distribution of Preservice Teacher's Correct Responses for Item F.2.

Score	N	%*
Omitted	11	2.5
0	8	1.8
1	54	12.3
2	149	33.9
3	137	31.2
4	80	18.2
Not applied	220	

*According to 439 preservice teachers who had a chance to see the item

Table 4.17 shows the distribution of preservice teacher's correct responses. Results showed that only 18% of the preservice teachers, who answered this item, were able to evaluate the appropriateness of the data for all graph types. The ratio of the preservice teachers who evaluated three graphs correctly was 31%, and the ratio of the preservice teachers who evaluated two graphs correctly was 34%. Therefore, it was concluded that majority of the preservice teachers were not able evaluate the graphics where categorical data can be used.

4.3.2 Preservice Teachers' Ability to Evaluate a Student's Correct Work (PCK-2)

PCK-2 is "Identifying different approaches for solving statistical problems", and defined as "Future teacher should see and value that some statistical questions can be handled using different approaches that are all correct."

To address this objective, a graphic related open-ended item (Item F.11) were constructed, and this item requires preservice teacher to think about data grouping activity that is needed before constructing a histogram using the following scenario,

> "A couple of students are working together and trying to group data into five intervals. The students' method looks different than Mr. Mehmet's rubric. If you were their teacher to evaluate and score the answer of these students according the criteria provided below, which score do you assign and explain why?"
>
> a) Completely correct (4 points)
> b) Mostly correct (3 points)
> c) Half correct (2 points)
> d) Mostly wrong (1 point)
> e) Completely wrong (0 point)

The students' answer, which was given in the stem of the item, was actually an alternative correct answer. This answer did not group the data in traditional fashion, and groping started from the largest number on the data.

In this item, Mr. Mehmet's rubric was prepared in traditional way and summarized in Figure 4.25(a), and students' way of working was different and summarized in Figure 4.25(b). The actual data was also given in Table 4.18.

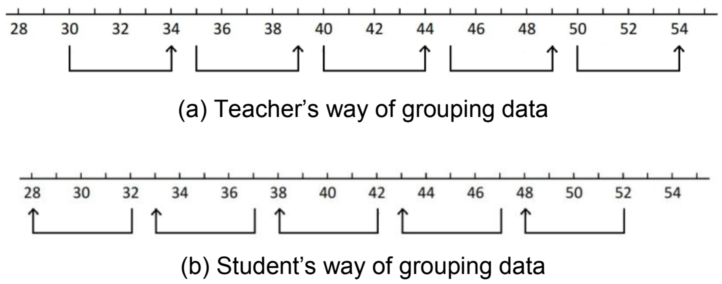


Figure 4.25. Teacher and Students' Way of Thinking from F.11 Item Stem.

Table 4.18. Data Used in Item F.11

Data Points															
30	33	34	34	35	35	35	36	37	38	38	39	39	40	40	40
41	41	41	42	45	45	45	45	46	46	46	47	48	50	51	52

Table 4.19 summarizes the score distribution for this item. Nearly one third of preservice teachers (33%) thought that 'the student answer' given in the item stem was wrong in some ways. Only 27% of the preservice teachers thought that the answer given in the stem was also an acceptable answer for the teacher's question.

Table 4.19. Preservice Teachers' Score Distribution for Item F.11

Score	N	%*
Omitted	69	15.8
0	36	8.2
1	63	14.4
2	48	11.3
3	102	23.4
4	117	26.9
Not applied	224	

*According to 435 preservice teachers who had a chance to see the item

Table 4.20 gives some examples for preservice teachers' assigned scores and their explanation for corresponding score. Some answer was worth noting such as "we did not learn grouping in this way". Therefore, this preservice teacher thought that the answer should have to be strictly must in line with the way they learn. Some answers were conflicting with itself such as "There is no number such as 28 and 29, so students way of grouping is wrong". Therefore, saying that students way of thinking was wrong but teachers way of thinking was correct because of the stated reason were conflicting with itself because the teacher's way of grouping ended on 54 and there was no number such as 53 and 54 in the data.

Table 4.20. Some Answer Examples of Preservice Teachers for Item F.11.

Score Given by Preservice Teacher	Preservice Teachers' Reasons for Their Scores
0	Frequencies of the teacher and students does not match
0	There could be some numbers between 52 and 54 and teachers rubric covers these numbers
0	They have to start from 54
0	They have to start from the smallest value
0	There is no number such as 28 and 29, so students way of grouping is wrong
0	We always sort the values from smallest to largest
0	Students have serious misconceptions about grouping the data
0	Because cumulative frequencies will be much different
0	Because we did not learn grouping in this way
0	Students' grouping covers the data but it is a coincidence, not always true
1	This way of grouping may result in data loss
1	Since the table is wrong, graphic will be wrong too.
1	Frequencies are close to teacher's rubric
1	Since the table is wrong, anything that will be computed from the table will be wrong too.
1	Only interval width is correct
1	Students started from 28, however the smallest data is 30
1	Even though procedure is correct, the answer is wrong because they used different intervals
1	It is correct that they cover whole data but their starting points is wrong
2	Students mistake is a result of lack of attention, because their method is correct if they started from 30.
2	At least they know frequency concept correctly. Starting from the largest number as a result of misconception does not mean their method is completely wrong.
2	They got the interval width correct, starting point wrong. They also know frequency concept.
2	They got the logic behind grouping concept but they do not know how to start grouping
2	All intervals are shifted half of interval width
2	The answer is correct. They just started from the largest number
3	Median will be affected from cumulative frequencies that are computed using this answer
3	Some numbers will be included in different interval
3	Both teacher's rubric and students' answer disregard some numbers but we usually start from the smallest number
3	Teacher's rubric and students' answer are not much different graphic will not be same
3	Even though interval width does not change, frequencies are different because they use different intervals. However, method is mostly correct
3	The only thing different is starting values but frequencies will be affected
4	Students' answer is correct but useless
4	The important thing is whether there is at least a student in all intervals [*it is true if the frequencies of the each intervals different than zero**]
4	There nothing 100% correct in statistics. Since they cover the whole data, their answer is correct as well
4	It is correct since teacher did not state to start from 30.

*italics is not an actual preservice teacher explanation, it was paraphrased by researcher, and * original arguments are presented in Appendix H.

Other preservice teachers, who considered the students' work was completely wrong, usually defended their arguments using the fact that students did not start grouping from the smallest data point.

Preservice teachers, who stated that students' answer is mostly wrong, usually based their arguments on the interval width, which was same for teacher's rubric and students' answer. Some preservice teachers also considered the consequences of grouping such as graphics. These preservice teachers thought that if there is a flaw in the grouping process, graphics constructed or the statistics computed from table will be completely wrong.

Preservice teachers, who stated that students' answer was half correct, usually valued the students' knowledge about frequency concept. These preservice teachers taught that students actually know computing frequencies for corresponding intervals. Some of the preservice teachers also valued the students' partial knowledge about grouping. For example, they stated that students actually knew how to group data but students did not pay attention to start from the smallest data point.

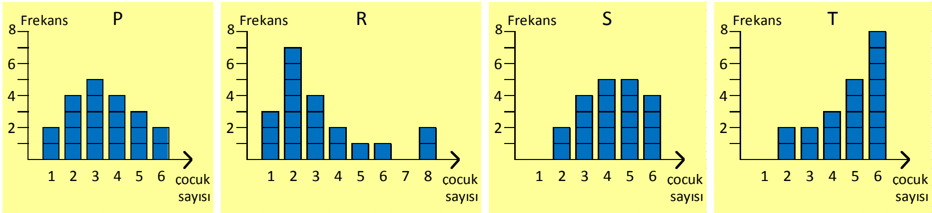
There were also some exceptions for some preservice teachers, who stated that students' answer is mostly correct or completely correct. For example, some preservice teacher thought that students' answer is correct but the answer did not deserve full credit because it did not follow the traditional method. In another exception, a preservice service teacher assigned full credit to students' answer but thought that the students' approach was useless. Another exception also revealed a possible misconception about grouping data. One preservice teacher stated that students' answer was correct because there were observations for each interval. So this preservice thought that grouping must be done in a way that all frequencies for corresponding intervals must be different than zero.

4.3.3 *Preservice Teachers' Ability to Select Appropriate Examples (PCK-3)*

PCK-3 is "Planning or selecting appropriate methods and activities for representing statistical ideas", and defined as "Future teacher should be able to plan a lesson by selecting appropriate methods and identifying key ideas. Activities involved in methods should match the key statistical ideas and learning goals in the curriculum. This objective includes selecting appropriate examples".

Item F.3 were constructed to address this objective. This item requires preservice teachers to consider four different types of data, which presented in graphic form, and think about the scenario presented in Figure 4.26,

A teacher wants to explain the situations where arithmetical mean and median are appropriate central tendency measures. He wants to give an example for mean and an example for median. For this purpose he examines the following graphics



Which graphic or graphics would you choose for this purpose if you were the teacher? And explain why?

Figure 4.26. Translated Version of the Item F.3.

Table 4.21 shows the score distribution of preservice teachers for item F.3. Results show that a large proportion (45%) of preservice teachers has no idea or definitely wrong ideas for appropriateness of mean or median for a specific data. Twenty nine percent of the preservice teachers can only have ideas about mean or median. Most of them gave clear explanation for the mean concept and a fuzzy explanation for median while only a few of them gave clear explanation for median but a fuzzy explanation for mean. Results also showed that only eleven percent of preservice teachers gave clear and appropriate explanations for both mean and median concepts.

Table 4.21. Preservice Teachers' Score Distribution for Item F.3

Score	N	Valid %*
Omitted	75	17
0	121	27.5
1	27	6.1
2	128	29.1
3	41	9.3
4	47	10.7
Total	439	
Not Applied	220	

* According to 439 preservice teachers who had a chance to see the item F.3

Table 4.22 shows some example for preservice teachers' work of this item.

Table 4.22. Graphics Selected by Preservice Teachers and Their Reasons

Example for		Reasons to choose these examples*
Mean	Median	
P	P	It should be approximately normal data for both examples.
P	P	Data should be symmetric for mean example; P is also appropriate for median example because there is a frequency for each number.
P	R	P is for mean example because values are close to each other; R is for median because it is easy to sort from smallest to largest.
P	S	P is for mean example because it is a normal distribution; S is for median because I can show that median is between 4 and 5.
P	S	P is for mean example because it is smooth [or balanced] distribution; S is for median because there are odd number of observations.
P	S	P is for mean example because values are close to each other; S is for median because there are five data points.
P	S	P is for mean example because it is a normal distribution; S is for median because values in the middle are very close to each other.
P	T	P is for mean because most of the values are in the middle. T is for mean because it is sorted according to number of persons.
P	T	P is for mean example because values are close to each other; T is for median because T has the highest frequency number.
S	P	S is for mean because mean will be at the point where the data is piling up; P is for median because it is easy to find median from P.
S	P	S is for mean because mean will be at the point where the data is piling up; P is for median because there is no outliers.
S	P	S is for mean example because values are close to each other; P is for median because increase and decrease is regular.
P	S	Mean can be easily estimated from P. S is for median because it would be interesting to have two medians.
R	T	R is for mean because mean is affected from outliers; T is for median because median is appropriate for skewed distributions.
P	P	P is appropriate for both examples because all others have two medians which may cause problems.
P	P	It is easier compute both from P.
P	P	P is appropriate for both examples because the most frequent data point is also median.
P	R	P is for mean because frequencies are not much equal to each other; R is for median because there are odd number of observations.
R	P	R is for mean because it is skewed to the right; P is for median because it is a normal distribution.

Table 4.22. Continued

Example for		Reasons to choose these examples*
Mean	Median	
R	P	R is for mean because there are outliers; P is for median because frequencies are different.
R	R	There is a disconnection on the data, and this may result in lasting [permanent] learning.
R	S	R is for mean because this data has unbalanced distribution; S is for median because this data has balanced distribution.
R	S	R is for mean because it is heterogeneous; S is for median because S has fewer groups than others.
R	S	R is for mean because standard deviation is larger; S is for median because standard deviation is smaller.
S	S	Since it is symmetrical, it is appropriate for both examples.
S	S	Since the values are close to each other, it is appropriate for both examples.
S	S	S is appropriate for both examples because the values are more close in this data
T	P	T is for mean because it is regularly increasing; P is for median because it is a normal distribution.
T	R	T is for mean because there are outliers. R is appropriate for median [*no reason stated*]
T	S	T is for mean because it has largest mean; S is for median because it has largest median
T	S	T is for mean because differences are larger between observations; S is for median because differences are smaller between observations.
T	T	T is appropriate for both of them because we need to sort data to find median
T	T	T is appropriate for both of them because values are different, and can be sorted easily.

* Original arguments are presented in Appendix I.

As seen from the Table 4.22, it was not a rare situation that preservice teacher chose the same data to be used for both mean and median concepts. In some cases, preservice teacher defined the data they chose inappropriately. For example, they defined Data S having odd number of observations. In fact, all data examples given for this item, including Data S, had twenty observations. Since the purpose of this item was to make sure that preservice teachers would focus on the distributions of the data, data examples were different in shape but not in the observation number. Actually, some preservice teachers stated that a graphic has odd number observations, when the graphics had odd number of different observations. They were actually referring to the number of columns a graphic had. This was also the case when a preservice teacher stated that ‘Data S has five data points’, ‘Data R has odd number of observations’ or ‘Data S has fewer groups than others’

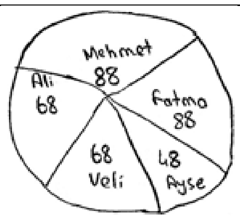
It was also worth to note that some preservice teachers thought that ‘Data T is an already ordered data’ or ‘data T is ordered from smallest to largest’. These preservice teachers disregarded the idea that a data, which presented in graphical form, is an ordered data independently from the shape or distribution of the data. Quantitative and qualitative analysis of Item F.3 showed that more than half of the preservice teachers had problems to identify an appropriate example for mean concept, and more than three quarter of preservice teachers had problems for providing an appropriate example for median concept.

4.3.4 Preservice Teachers' Ability to Evaluate Students' Arguments (PCK-4)

PCK-4 is "Analyzing or evaluating students' statistical solutions or arguments", and defined as "Future teachers should experiment with different teaching approaches and activities, and monitor the results, using conventional tests, and by carefully listening to students and evaluating information (Garfield, 1995)".

To address this objective, a graphics related open-ended item (Item F.4) constructed, and this item requires preservice teachers to think about a hypothetical student's answer. The answer represented a pie chart construction activity using angles. Hypothetical student's answer was starting in a way that keeps the proportional aspects of pie graphs. After a point, student disregarded proportions to find solution. In this item, it was expected from preservice teachers to evaluate correct or wrong aspects of the student's answer. A translated form of Item F.4 is given at Figure 4.27

Student	Weekly Allowance
Ayse	10 TL
Ali	15 TL
Veli	15 TL
Mehmet	20 TL
Fatma	20 TL
Total	80 TL



Student's explanation for the graphic he constructed for given data:

"Total was 80, and 320 was the multiple of 80 that was nearest to 360. Therefore, I multiplied every number by 4 to get 320 in total. Then 360-320=40 left. Since there are five numbers, I divided 40 by 5, and got 8. Then I added 8 to all numbers I got before."

Analyze the student's response, and evaluate the correct or wrong aspects?

Figure 4.27. Translated Form of Item F.4.

As seen from Table 4.23, this item was one of the least omitted items by preservice teachers. Almost 92 % of the preservice teachers stated their ideas. However, most frequent score for this item was zero. This meant, 156 (36%) preservice teachers answered this item either “Student’s answer is completely wrong” or “Student’s answer is completely correct”.

Table 4.23. Preservice Teachers’ Score Distribution for Item F.4

Score	N	Valid %*
Omitted	37	8.4
0	156	35.5
1	15	3.4
2	25	5.6
3	93	21.1
4	113	25.7
Total	439	
Not Applied	220	

* According to 439 preservice teachers who had a chance to see the item F.3

Qualitative analysis of the preservice teacher revealed several deficiencies regarding evaluation of students’ arguments. Some preservice teachers evaluated the students’ work as correct using following arguments:

- Correct but an unnecessarily long solution approach.
- It is correct since student divided in equal proportions, multiplied and added. Operations affected all number equally at the same time
- Logic is correct. Since he distributed 360 degrees in proportional way, pie graph is also correct.
- Student has a very logical solution approach.
- It is correct since he found equal angles for equal allowances.
- Students’ geometrical thinking is very nice. He solved the question correctly while thinking step by step.
- It is very logical but I could not understand whether this method is generalisable
- Result is correct but it does not mean that the method is correct.

These preservice teachers mostly thought that students work should be correct since the answer treated allowances “equally” (multiplying all number with 5 then adding 8 to all). Even though some preservice teachers questioned the correctness or generalizability of the method, they stated that the result is correct. Another teacher stated the method should be correct since equal allowances corresponded to equal angels. Even though this statement is correct for any pie chart, it is not a sufficient condition to hold proportional properties of the pie charts.

Some of the preservice teachers honestly stated that they knew the student’s results were not correct but they could not find where student made mistake. Some preservice that the students answer was an approximation.

- Approximate answers can be found using this method but the correct answer cannot be found.
- Student can find approximate answers using this method but it cannot be applied to other examples

The qualitative analysis also revealed that preservice teachers, who think that solution is wrong, mostly focused on final product of the student. These preservice teachers mostly valued the results, and disregarded the solution process of the student.

- There is nothing correct for this response.
- He should have multiplied with 4.5 instead of 4. Therefore, the solution is wrong.
- The student is not aware that he needs to use direct proportion to 360.
- It is wrong to find 320 as a nearest multiple to 360. The rest is nonsense.
- It is good that he found 320 as a nearest multiple to 360. However, the solution is not valid since the results are wrong.

These preservice teachers did not usually considered the appropriateness of the each part of the student’s answer but focused on the final product. It was also seen that preservice teachers, who stated that students’ solution is completely wrong, usually solved the question using their own methods and then compared their results with students' answer.

Some preservice teachers only evaluated the correctness of a part of the students' answer, and stated "It is good that he found 320 as a nearest multiple to 360. However, it would have been better if he used regular direct proportion rules." Original arguments of the preservice teachers are presented in Appendix J.

4.3.5 Preservice Teachers' Ability to Diagnose Misconceptions (PCK-5)

PCK-5 is "Predicting or diagnosing typical students' responses, including misconceptions", and defined as "Future teacher should be able to (a) know how regular student will respond to statistical question, (b) predict a misconception and (c) identify a previously constructed misconception".

To address this objective, a mean related item (Item F.5A) constructed, and preservice teachers were expected to analyze the results of a homework assignment using the scenario presented in Figure 4.28

Mrs. Ayşe asked her students to find the median of the following data as a homework assignment.

Homework Question

Table shows the number of broken biscuit in package for 21 packages. What is the median of this data?

Number of broken biscuits in each package	Frequency
2	5
4	6
6	7
9	2
11	1
Total	21

A day later, she collected the students' homework assignments and took notes for the answers given by students. At that time she has seen that many students gave 5 or 6 as an answer even though the correct answer is 4. Then she wrote her comments on the students homework papers about why the students made mistake

F.5A. If we assume that students did not make any computing errors, what would be the reason that many students found 5 as the median of this data?

Figure 4.28. Translation of Stem for Item F.5A and F.5B Including Item F.5A

There were two possible good explanations for this item. First one was an obvious pattern that students could have been found 5 as a median because they disregarded observations and instead find median of the frequencies. The second pattern was not that much obvious. The second pattern is related to weighted mean, and the weighted mean of this data is also 5. Before final administration, it was expected to see that most preservice teachers, who were able to solve this item, would identify the first pattern. However, analysis results showed that frequency of second pattern was close to frequency of first pattern, and a few of the preservice teachers identified both patterns. Table 4.24 shows the score distribution of this item.

Table 4.24. Preservice Teachers' Score Distribution for Item F.5A.

Score	N	Valid %*
Omitted	65	14.6
0	83	18.6
1	32	7.2
2	121	27.2
3	47	10.5
4	96	21.6
Total	444	
Not Applied	215	

* According to 444 preservice teachers who had a chance to see the item F.5A

As it can be seen from the table, 15% of the preservice teacher omitted this item. This could be sign that this item is difficult and/or different from what the preservice teacher used to in their regular classrooms. Results also showed that only 22% of the preservice teachers clearly identified a possible misconception.

Qualitative analysis of the preservice teachers' answer showed that many preservice teachers gave inacceptable answers for this item. Since the item stated "many students", preservice teachers expected to give general reasons for students' mistakes for computing median of the data. Some of the preservice teacher teachers' answers were related to a rare combination of the numbers from the table presented in the item. Examples to this kind of answer were as follows:

- They may find 5 because there are 5 numbers

- They may divided most often number, which is 11, by 2
- They may divide total number of packages by real median, 21/4, and find 5.
- Students know median as the frequency of the smallest observation

In these examples, a preservice teacher even claimed that dividing total number of packages by real median could be considered as general situation for these students.

Some preservice teacher did not give specific reasons for students' errors, and gave answers such as "students do know what the median is", "Student does not know what median and frequency are" and "There is no 5 as a broken biscuit number. I need to see their homework to understand what they did."

There were also other explanations that worth attention such as "students are mixing median with mode" or "They may have a misconception and they found mode instead of median". In fact, the mode of the presented data was 6. Therefore, this situation showed that these preservice teachers had in fact problems with both median and mode concepts. Other examples of finding inappropriate student error were as follows:

- They may think that frequencies do not affect median
- Because students saw 5 as median
- They may have problems while computing the median formula which is "Number of Biscuits x frequency" [weighted mean formula]
- They may use the formula (n-1)/2

In a rare situation, preservice teacher's answer was not interpretable and was not related to median concept and s/he stated, "Students thought that data come from sample instead of population". This preservice teacher's answer was probably related to standard deviation concept and it was not clear why this preservice teacher included sample and population concepts for discussing median concept.

In a rare situation, even though the median of the data is stated as 4 in the item, a preservice teacher claimed that the median is 5, and therefore s/he concluded that students’ answer was correct. Original arguments of the preservice teacher, discussed for this PCK objective, are presented in Appendix K.

4.3.6 Preservice Teachers’ Ability to Provide Feedback (PCK-6)

PCK-6 is “Providing appropriate feedback”, and defined as “Future teacher should be able to assess and question the student learning aligned with learning goals (Pfannkuch & Dani Ben-Zvi, 2011) and able provide appropriate feedback after diagnosing students' responses in a way that given feedback improves students learning (Chickering, Gamson, & Poulsen, 1987)”.

To address this objective, Item F.5B, which has the same base information with item F5.A, has been used. This item required preservice teacher think about the reasons that caused students to made a mistake on a particular assignment and provide appropriate and specific feedback for the supplied condition. Since the purpose of the item was focusing on feedback strategies, a slightly different stimulus for this item was used. Preservice teachers asked to think about following scenario,

> “If we assume that students did not make any computing errors, how would you comment on students, who gave 6 as an answer, in such a way that permits students to fix their errors? (also considering the reason that caused them to make mistake)”

Since the item was especially focused on feedback capabilities of the preservice teachers, students who gave 6 as an answer was pointed. It was expected that preservice teacher would easily identify these students’ error, which are finding mode instead of median or finding median without considering frequencies. However, preservice teachers were not limited with these specific errors, and they were free to supply feedback on any mistake they found.

Table 4.25 shows the preservice teachers’ score distribution for this item. As it can be seen from the table 89 (20%) preservice teachers omitted this item, and a success rate was lower than item F5.A. Results showed that only 7% of the preservice teachers explained an appropriate feedback clearly while 9% of the preservice teachers explained an appropriate but not sufficient feedback.

Table 4.25. Preservice Teachers' Score Distribution for Item F.5B.

Score	N	Valid %*
Omitted	89	20
0	104	23.4
1	47	10.5
2	135	30.4
3	39	8.7
4	30	6.7
Total	444	
Not Applied	215	

* According to 444 preservice teachers who had a chance to see the item F.5A

In fact this PCK objective is closely related to PCK-5, which is diagnosing typical students' responses. In order to provide appropriate feedback, preservice teacher has to consider how students' error occurred or what students missed to solve problem correctly.

Qualitative analysis also revealed that preservice teachers most commonly used the description of median -or how compute median from data- as a feedback strategy. This pattern was also consistent among preservice teacher who clearly identified the reason that caused students to make mistake.

Most of the preservice teachers were on the score category of 2. Therefore, feedback understanding of the most of the preservice teachers was telling students directly what was the error they made on the task without guiding students about how can recover their error. Examples to this kind answer were as follows:

- I would say to students that 'you find mode by selecting the most often number'
- I would write to students' homework that "Being most often number does not mean it is the median of the data"
- They may find 6 because the largest frequency is 7

Even though some preservice teachers tried to supply a feedback that is more than mentioning student's error, additional information was not clear. Examples to this kind answer were as follows:

- Students are mixing median with mode. So I would explain these topics again
- Students are mixing median with mode. So I would explain the differences between mode and median
- Students are mixing median with mode. So I would explain most often number is not always the median of the data

For example, some preservice teachers claimed that they would explain the differences between mode and median but they did not explain how they would do that.

Some preservice teachers supplied feedback that is nor specifically relevant with the median concept, and can be used in any teaching situation. Examples to this kind answer were as follows:

- I would say "think again"
- I would say "be more careful"
- I would ask why they think so, then I explain what they are missing
- I would ask "what the [median] is". Then I ask how they found this answer.

Generality of the last sentence is not trivial directly. However, 'median' word between brackets can be replaced by any concepts to make it related with another concept.

In some cases preservice teachers could not identified the student answer correctly. Therefore, their feedback to recover student's error was not appropriate. Examples to this kind answer were as follows:

- Students found the median of the frequencies

- We may correct them by asking how many packages we have. Then we may remind the formula, “Number of Biscuits x frequency” [weighted mean formula]
- I would say to student “do not mix it up with arithmetical mean”.
- I would explain using examples that median value will be close to the arithmetical mean value

In rare cases, preservice teachers supplied a feedback that could be a key part of an appropriate feedback, such as:

- I would say, “There are as many packages as frequencies”

This feedback most probably targeted students who computed the median of the data without considering the frequency column of the table. However, feedback does not inform students about their errors. Student even may know the meaning of the frequency, and teacher’s statement about his/her homework may not make any sense at all. This and other examples for preservice teachers’ answers are presented in Appendix L in original format.

4.4 Summary of Results

In this section of the chapter, the results of the three research problems were summarized. Before answering research questions, MKT-S instrument that consisted of content knowledge (CK) and pedagogical content knowledge (PCK) was developed. Final version of the instrument had sixteen items, six items for CK and ten items for PCK.

In answering the first research problem “Will the instrument developed in this study be valid and reliable for measuring preservice middle school mathematics teachers’ mathematical knowledge for teaching statistics concepts, specifically averages and graphs?”, two models were tested for assessing the structure of the MKT-S instrument based on the literature. It was found that two-factor solution (CK and

PCK as separate constructs) better explained the structure of the MKT-S than one-factor solution (CK and PCK are inseparable constructs). The validity of MKT-S instrument was assessed using preservice teachers' ISP-I course grades and preservice teachers' year in the program. It was found that ISP-I course grades positively correlated with both CK and PCK scores. It was also found that preservice teachers' year in the program had small impact on both CK and PCK scores. Reliability of CK items was 0.65 while reliability of PCK items was 0.76. Both dimensions of the instrument were more reliable for high achieving preservice teachers.

It was also found that CK and PCK were highly correlating dimensions of MKT-S instrument ($r=0.78$, $p<0.001$). Results also showed that having a high CK score did not always resulted in high PCK score.

In answering to the second research problem "What kinds of deficiencies do preservice teachers have in their content knowledge regarding middle school statistics concepts, specifically averages and graphs?", content knowledge of preservice teachers was examined using answers of an average related item and a graphics related item. For average related content knowledge, an item where preservice teachers required estimating the average of a data was examined. It was found that a high number of preservice teachers relied on arithmetical mean as an efficient estimator even for the case that arithmetical mean was not trustable. About thirty percent of the preservice teachers' estimate was meaningful for estimating the average of a data that has questionable measurements. Most of this group was used median or mode in the estimation process and only two preservice teachers offered cleaning data before estimating the average. For graphics related content knowledge, an item where preservice teachers required constructing a histogram for a data that is unusual for them was examined. It was found that only twenty-one percent of the preservice teachers were able construct a histogram that has no mistake regarding to properties of histogram while only ten percent of the preservice teachers' histogram were accurate for the given data.

In answering to the third research problem “What kinds of deficiencies do preservice teachers have in their pedagogical content knowledge regarding middle school statistics concepts, specifically averages and graphs?”, one item for each of six PCK objectives was examined. Therefore, the results of the six items, which are F.2, F.11, F.3, F.4, F.5A and F.5B, were reported. PCK-1 objective is about seeing connections between statistics topics. It was found that only eighteen percent of the preservice teachers successfully interpreted whether a picture graph can be related to a pie graph, a bar graph, a line graph, and a histogram. PCK-2 objective is identifying different approaches for solving statistical problems. It was found that only 27 percent of the preservice teachers were able interpret that (a) given student answer was an alternative approach to handle the question, and (b) it was possible to get a correct answer that was different from teacher’s rubric. PCK-3 objective is planning or selecting appropriate methods and activities for representing statistical ideas. It was found that only eleven percent of the preservice teachers were able to provide and explain an appropriate example for both arithmetical mean concept and median concept. It was also found that twenty-nine percent of preservice teachers were able to provide only arithmetical mean (or in some rare cases, only for median concept). PCK-4 objective is analyzing or evaluating students' statistical solutions or arguments. It was found that only twenty-six percent of the preservice teachers were able clearly evaluate both logical parts and illogical parts of a student’s answer related to construction process of a pie graph. PCK-5 objective is predicting or diagnosing typical students' responses, including misconceptions. It was found that only twenty-two percent of the preservice teachers were able to explain what would be reason that a student made a mistake while computing the median of a data in a tabular form. PCK-6 objective is providing appropriate feedback. It was found that only seven percent of the preservice teachers were able to provide appropriate feedback for a student who made a mistake while computing the median of a data in a tabular form.

CHAPTER 5

DISCUSSION, CONCLUSION AND IMPLICATIONS

This chapter of the study presents discussion and conclusion of the results, limitations, implications, and finally, recommendations for future research studies.

5.1 Discussion and Conclusion

Primary purpose of this study was to assess preservice middle school mathematics teachers' mathematical knowledge for teaching statistics, understand the relationship between its components, and investigate the adequacy of this knowledge. For this aim, MKT-S instrument was developed, and its validity and reliability was investigated.

MKT-S instrument included two dimensions namely content knowledge and mathematic pedagogical content knowledge, for measuring preservice teachers' knowledge that is required to teach statistics topics from Grade 5 to Grade 8. Other researchers also included these dimensions for assessing preservice teachers' knowledge for teaching (Tatto et al., 2008) or very similar dimensions for assessing inservice teachers' knowledge for teaching (Krauss et al., 2008).

It was found that MKT-S instrument has two dimensions, and content knowledge and pedagogical content knowledge are two different knowledge forms of teaching knowledge. The structure of MKT-S instrument was also compared with other researchers' results (Blömeke, Houang, & Suhl, 2011), and found that results of MKT-S instrument was in line with these researchers.

To provide concurrent validity evidences, the relationship between preservice teachers' 'Introduction to Statistics and Probability (ISP-I)' grades and MKT-S scores (CK and PCK scores) was examined. It was found that there was a small relationship between ISP-I grades and CK scores (r=305, p<0.001), and a small relationship between ISP-I grades and PCK scores (r=273, p<0.001). This result can be explained by the nature of the items in the MKT-S instrument. Even though items were part of the ISP-I course content; items were required preservice teachers to think on more abstract level such as the type of the outlier: Is it a possible outlier or a measurement error.

The third year and fourth year preservice teachers' CK and PCK scores were also compared for validating MKT-S instrument, and it was anticipated that fourth year preservice teacher should have better CK and PCK score than third year preservice teachers since they acquire an additional year of training in the mathematics education program. Even though fourth year preservice teachers significantly got better CK and PCK scores than third year preservice teachers, effect sizes for differences were very small. This result suggested that preservice teachers gain a small amount of information for teaching statistics topics during fourth year in the program.

It was found that the reliability of the CK scores was 0.65. Even though reliability is lower than industry standard of 0.7, the low number of items that consists CK scores could explain this situation, and larger number of items may result in a more reliable CK instrument. Reliability was also higher for high CK scores and this could be due to absence of items that has medium difficulty. Since the one purpose of the instrument was to assess the adequacy of content knowledge of the preservice teachers, the items generally aimed this purpose, and items were challenging to seek deep information of preservice teachers.

Reliability of PCK scores was 0.76, and it was little higher than industry standard of 0.70. This level of reliability can be considered enough for the purposes of this study. In this study, the general pedagogical levels of preservice middle school mathematics teachers regarding to statistics topics were tried to be pictured, and aim was not defining cut-off values that important decisions (such as hiring for a job or passing from a course) will be made upon these values. Low level of the reliability is also can be explained by content of the PCK items. Even though both average related PCK items and graphics related PCK items aim to measure pedagogical knowledge for statistics topics, confirmatory factor analysis results show that mean loading for average related items were higher than mean loadings of graphics related items. Even though PCK scores aimed to picture the general pedagogical levels of preservice middle school mathematics teachers, it makes sense that an average related PCK items will seek different information from a graphics related PCK item.

Low reliability levels of MKT-S instrument can also be connected to content coverage. As Shulman (1987) stated "Pedagogical content knowledge … presents the blending of content and pedagogy into an understanding of how particular topics, problems, or issues are organized, presented, and adapted to diverse interest and abilities of learners, and presented for instruction" (p. 8). Therefore, it may be asserted that content is a key factor for pedagogical content knowledge (or teacher knowledge general), and teachers' knowledge is differently organized for different concepts of statistics curriculum. However, because of the insufficient number of items, it was not possible test whether knowledge structures differs for concepts of statistics, and it was assumed that both content knowledge and pedagogical content knowledge of the preservice teachers would be parallel for both central tendency and graph related topics.

It was found that PCK scores were highly correlated with CK scores ($r=0.78$, $p<0.001$). The result is similar to finding reported by other researchers. For example, Krauss et al. (2009) found similar latent correlation ($r=0.79$) between content knowledge and pedagogical content knowledge, and correlation was even higher ($r=0.96$) for teachers who possess high level of CK and PCK. Even though it was explicitly tested and found that two-factor structure better fitted the data, a high correlation among factors of an instrument brings the question that whether factors could be collapsed to construct a single factor. It was found that this result was pretty much in line with the nature of pedagogical knowledge because it is a trivial fact that teaching a mathematics topic properly for any person requires an understanding about topic but knowing mathematical content does not always result in good teaching (Borko et al., 1992). Therefore, it was an expected result for this study. However, fully testing the first research problem requires including samples other than preservice mathematics teachers such as mathematics majors who are not interested in teaching.

Some items were analyzed to understand the adequacy of content knowledge and pedagogical content knowledge of preservice teachers. Analysis of average related content knowledge item revealed that a high numbers of preservice teachers estimated the average of the data using arithmetical mean even for the situations where data has extreme cases. Preservice teachers' dependency on arithmetical mean can be explained by usage of arithmetical mean in inferential statistic course. In Inferential statistics course, preservice teachers greatly use arithmetical mean to estimate confidence intervals or to compare the means of two different groups. Preservice teachers' answers such as "because arithmetical mean is best estimator", "arithmetical mean is a generalization of all numbers" or "arithmetical mean is always trustable" support this claim. Another reason could be the fact that these topics are relatively new topics in mathematics curriculum and some mathematics educators still do not understand differences between these three types of averages. Explanations related to average topics on the national teacher guide reflects this situation. For example, usage of the averages (MNE, 2009, p.275.) explained by the sentences;

> "… Aritmetik ortalama, ortanca ve tepe değeri istatistikte yer alan ortalama çeşitleridir. Aritmetik ortalama duyarlı ortalama iken diğerleri duyarlı olmayan ortalamalardır. Amaca uygun ortalama çeşidinin kullanılması gerektiği vurgulanır… Veri grubunda çok büyük ve çok düşük değerlerin olması durumunda ortanca, aritmetik ortalamadan daha sağlıklı bilgi verir. Bunun nedeni sözü edilen değerlerin ortancayı etkilemesidir."
>
> [… *Arithmetical mean, median and mode are three types of averages in statistics. Arithmetical mean is a sensitive average while others not sensitive averages. Teachers should stress using a type of averages that suits the purpose…When there are very large and very small values in data, median gives more healthier information than arithmetical mean because these values affect median*]

Using the word “purpose” in this explanation falsely implies that those three types of averages have different purposes, and the second sentence falsely implies that outliers affect the median value.

Analysis of the histogram related content knowledge item revealed that the most of Turkish preservice mathematics teachers, who subject to teach histogram to middle school students, could not construct a histogram from extra ordinary data. This result can be explained by the fact that textbooks generally include histogram construction examples that are based on ordinary data. It was also observed that histogram drawings were not accurate for some preservice teachers who can handle extra ordinary data. However, this result does not conflict with results reported for other nations (Bruno & Espinel, 2009; Lee & Meletiou-Mavrotheris, 2003).

Analysis of pedagogical knowledge items also revealed several deficiencies. Some of these deficiencies are directly related to preservice teachers’ content knowledge since pedagogical content knowledge of a specific topic requires a good understanding about content of the specific topic. For example, a preservice teacher cannot give an appropriate example for teaching median concept if he/she does not know how median differs from arithmetical mean for summarizing center of the data.

Preservice teachers’ procedural knowledge and conceptual knowledge about a topic also plays an important role for shaping pedagogical content knowledge (Eisenhart et al., 1993). Some preservice teachers evaluated a student’s correct answer as completely incorrect because student’s solution did not follow usual procedural knowledge for grouping data into intervals. However, student’s solution was correct and analyzing this solution required conceptual understanding about grouping data.

Giving feedback is an important step in the teaching process (Chickering, Gamson, & Poulsen, 1987; Garfield, 1995), and most of the preservice teachers’ understanding for feedback in this study was explaining how to solve the question (re-teaching) without targeting the students’ misconception. Even though re-teaching can be less time consuming than targeting each student’s misconception (Shute, 2008), preservice teachers should be able provide task specific feedback (Maverech, 1983).

Some of these deficiencies related to way that preservice teacher learn statistics concepts in their education from first grade to end of teaching preparation program. For example, most preservice teachers do not possess knowledge about connection between different types of graphics. This result should be considered very normal since Turkish curriculum does not explicitly discuss how a type of graphic is related to another type of graphic. Teaching program usually focuses how a graphics constructed for *given* data; and how the resulting graphics interpreted. Since students (elementary students or preservice teachers) master their graphics skill on already given data, which is appropriate for the graphics under consideration, they rarely judge why the given data is appropriate or what kind of data could be inappropriate. However, the latest revision of the Turkish middle school curriculum could change the situation. In the latest teaching program, authors added a new objective (MNE, 2013, p. 41) to the eight grade level as follows:

> “8.4.1.2. Araştırma sorularına ilişkin verileri uygunluğuna göre daire grafiği, sıklık tablosu, sütun grafiği çizgi grafiği veya histogramla gösterir ve bu gösterimler arasında dönüşümler yapar”
>
> [*8.4.1.2. Students represent data, which is related to the research question, in appropriate form, and use pie charts, frequency table, bar graph, line graph or histogram. Students also make transformations among these representations*]

This objective could provide opportunities for inservice and preservice teachers to think about different data types. It could also make these teachers more aware about the idea that some representations could not be meaningful for a particular type of data.

Other deficiencies were related to two different kinds of knowledge source that preservice teachers acquire pedagogical knowledge. First source is mathematics teaching method courses where teachers acquire pedagogical knowledge actively, and second source is all learning environments where preservice teachers acquire pedagogical knowledge passively from their learning experiences (Kennedy, 1998;

Llinares & Krainer, 2006). It is also known that teachers disregard knowledge that is acquired from the methods courses, and instead tend to teach topics similar to a way that they learnt (Llinares & Krainer, 2006; Lortie, 1975; Zeichner & Tabachnick, 1981).

5.2 Limitations of the Study

In this study, MKT-S instrument, which consists of 16 items, was developed regarding to statistics topic that are taught in middle schools in Turkey. The first limitation of the study was statistics topics that were covered in this study. Because preservice teachers participated in this study voluntarily, testing time had to be appropriate for them to focus on whole instrument. Therefore, testing time limited to a single lesson length. Because of the 50 minutes of test length, it was not possible cover all statistics topics. However, study covered most of the statistics topics that are taught in middle school level such as mean related topics and graphics related topics.

Another limitation of the study was covering six pedagogical objectives among eighteen objectives that were expected from preservice teacher. This study was limited to these six objectives because of several reasons. Most important reason was that these selected objectives had structures that can be clearly formulated for statistics education to construct items. Second, it was wanted to construct pedagogical knowledge items for both mean and graphics topic for each objective if possible. Third, fifty minutes of testing time was an important issue.

Number of participants was also limited by limiting number of institutions to eight. Since the researcher was solely responsible for data collection to make sure all preservice teachers participated in the study had equal conditions during test implementation, it was not feasible to travel all institutions around Turkey. Therefore, eight public institutions were selected as diverse as possible.

5.3 Implications

First and the most important implication of this study is based on the result that most preservice teachers does not have enough knowledge to teach statistics topics. This implication suggests that it is needed to reevaluate the adequacy of courses in the middle school mathematics teacher education program. The number of courses related to teaching of mathematics may be increased and separate courses, such as teaching statistics, teaching geometry and teaching statistics, can be designed for each component of mathematics. This study also showed that there is little statistics teaching knowledge differences between third and fourth year preservice teachers. This finding implied that an additional year of study, in the mathematics teacher education program, had little effect on the preservice teachers' mathematical knowledge for teaching statistics levels. This result led the recommendation that a special care also must be given to fourth year courses in reevaluation process.

Largest employer of mathematics teachers is Ministry of National Education (MNE), and results of this study may have implication on the selection process of mathematics teachers. Currently, MNE hires middle school mathematics teachers using only the results of a national standardized exam, which does cover pedagogical knowledge. This exam consisted of four sections, namely general culture knowledge (15%), general ability (15%), general pedagogical knowledge (20%), and mathematics content knowledge (50%). Mathematics content knowledge section consisted of five sections, namely calculus (28%), algebra (18%), geometry (18%), applied mathematics (16%), and pedagogical content knowledge (20%). Therefore, teacher-hiring process is mostly based on the content knowledge, and pedagogical content knowledge affects 10% of a preservice mathematics teachers score. However, results showed that a high score on content knowledge does not always imply a high score on pedagogical content knowledge. Since the pedagogical content knowledge is most essential knowledge for teaching profession, it is recommended to add a larger pedagogical content knowledge section to national exam.

The instrument designed in this study allows evaluation of professional development efforts for preservice teacher in teaching statistics. Researchers can use the MKT-S instrument in two ways. First way to use MKT-S instrument is comparing another sample's both content knowledge and pedagogical content knowledge to sample of this study using the confirmatory factor analysis results of the MKT-S instrument. Second way to use MKT-S instrument is comparing a sample's content knowledge and pedagogical knowledge in pre- and post-treatment settings. Pre- and post-treatment scores may also be computed using confirmatory factor analysis results of this study to understand the sample's position before and after treatment compared to this study. Even though classical test theory can be used score preservice teachers factor scores of MKT-S instrument, this study used IRT scoring of the factor scores, which took account of both difficulty of an item compared to other items and difficulty of each score level of item. However, it should be also noted that reliability levels also estimated using IRT because other methods are not possible when instrument administered balanced incomplete booklet design. Therefore, researchers may implement complete MKT-S instrument in order to analyze psychometric properties of the instrument under classical test theory.

Another implication of the study is related with instrument development efforts. As discussed in method section, some items, multiple-choice or free response, do not work parallel to the intended purpose of the item. In some occasions, distracters may better work than correct answer in multiple choice items. In other cases, teachers may respond to open-ended questions in a way that makes impossible to implement rubric. Therefore, explicit item trials, maybe more than once, required to understand the nature of the each item. It is also observed that scoring and recording open-ended items for large number of teachers take great amount of time. Therefore, it is advised to split open-ended items, which requires long complex answers, into manageable pieces that each piece requires shorter free-response answers. For example, Item F.3 required preservice teacher to supply appropriate examples for both arithmetical mean and median concepts. This structure of the item made it very difficult record and score answers of preservice teacher. However, this item could be split into two items where one item deal with arithmetical mean concept and other item deals with median concept. In rare cases, it was also observed that some free-response items, which requires choosing an option from the list and explanation on why they chose the option, understood as a solely multiple choice items because options numbered similar to multiple-choice items (a, b, c, d, and e). Therefore, it would be better to present options without numbers for open-ended items to prevent confusion.

5.4 Recommendations for Future Research

Recommendations for future research studies were presented as below:

1. MKT-S instrument is focused on statistics topics such as averages and graphs. However, there are still other topics in mathematics curriculum left out in this study. It is suggested that other topics also included in PCK studies. As it was seen in this study, using open-ended items to investigate pedagogical knowledge of preservice teachers for topics, which was not investigated before, provide valuable information. Therefore, it is also suggested using open-ended items for topics that have limited literature support.
2. MKT-S instrument that was developed in this study only included six objective of the pedagogical knowledge that found appropriate for preservice teachers. Therefore, it is suggested that future studies may identify and focus other objectives of the pedagogical content knowledge that is appropriate for preservice teachers.
3. MKT-S instrument that was developed in this study included mostly open-ended items in the study. It may be suggested developing a test that consists solely of selection type items, particularly multiple-choice items. Answers of preservice teaches to open-ended items may provide a valuable base for constructing multiple-choice items. This type of items may contribute positively to the reliability estimation of scores, reaching to high number of participants in the studies and completing the evaluation procedures in short time duration.
4. Because of the design of this study, the predictive validity of MKT-S instrument could not be checked. Therefore, it is suggested to other researchers to design longitudinal studies that monitor preservice teachers after graduating education faculty, and observe the effect of their pedagogical knowledge to their teachings.

REFERENCES

Adler J., Ball, D., Krainer, K., Lin, F. L., & Novotna, J. (2005). Reflections on an emerging field: Researching mathematics teacher education. *Educational Studies in Mathematics, 60*(3), 359-381.

An, S., Kulm, G., & Wu, Z. (2004). The pedagogical content knowledge of middle school mathematics teachers in China and the U.S. *Journal of Mathematics Teacher Education, 7*(2), 145–172.

Babadoğan, C., & Olkun, S. (2006). Program development models and reform in Turkish primary school mathematics curriculum. *International Journal for Mathematics Teaching and Learning*. Retrieved May 12, 2013, from http://www.cimt.plymouth.ac.uk/journal/babadogan_olkun.pdf

Balcı, Ş. & Ayhan, B. (2004). Seçmen tercihlerinin belirlenmesine yönelik yapılan kamuoyu araştırmalarında güvenilirlik ve geçerlilik problemleri: "28 Mart 2004 yerel seçimleri örneği". *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *11*,135-168

Ball, D. L. (2002). Knowing mathematics for teaching: Relations between research and practice. *Mathematics and Education Reform Newsletter*, *14*(3), 1-5.

Ball, D. L., & McDiarmid, B. (1988). Research on teacher learning: Studying how teachers knowledge changes. *Action in Teacher Education*, *10*(2), 17-23.

Ball, D. L., & Wilson, S. M. (1990). *Knowing the subject and learning to teach it: Examining assumptions about becoming a mathematics teacher*. (Research Report No.90-7). East Lansing, MI: NCRTL, Michigan State University.

Barr, G. V., (1980). Some student ideas on the median and the mode. *Teaching Statistics*, *2*(2), 38-41.

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, *42*, 815-824.

Baxter, J. A., & Lederman, N. G. (1999). Assessment and measurement of pedagogical content knowledge. In J. Gess-Newsome & N. G. Lederman (eds.), *Examining pedagogical content knowledge* (pp.147-161). Netherlands: Kluwer Academic Publishers.

Biggs, J., & Tang, C. (2007). Teaching for Quality Learning at University. Berkshire, England : Open University Press.

Blömeke, S., Houang, R. T., & Suhl, U. (2011). TEDS-M: Diagnosing teacher knowledge by applying multidimensional item response theory and multiple-group models. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, *4*, 109-129. Retrieved May 10, 2013, from http://www.ierinstitute.org/dissemination-area.html

Borko, H., Eisenhart, M., Brown, C. A., Underhill, R. G., Jones, D., & Agard, P. C. (1992). Learning to teach hard mathematics: Do novice teachers and their instructors give up too easily?. *Journal for Research in Mathematics Education*, 194-222.

Boslaugh, S. & Watters, P. A. (2008). Statistics in a Nutshell. Sebastopol, CA: O'Reilly Media.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Byrne, B. M. (2010). *Structural equation modeling with AMOS. Basic concepts, applications and programming* (2nd ed.). New York, NY: Taylor & Francis Group.

Bruno, A., & Espinel, M. C. (2009). Construction and evaluation of histograms in teacher training. *International Journal of Mathematical Education in Science and Technology*, *40*(4), 473-493.

Burgess, T. (2002). Investigating the 'data sense'of preservice teachers. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics*, Cape Town. Voorburg, The Netherlands: International Statistical Institute. Retrieved August 18, 2011, from http://iase-web.org/documents/papers/icots6/6e4_burg.pdf

Burgess, T. A. (2007). *Investigating the nature of teacher knowledge needed and used in teaching statistics* (Doctoral dissertation, Massey University, Palmerston North, New Zealand). Retrieved March 25, 2012, from www.stat.auckland.ac.nz/~iase/publications

Cartwright. F., (2013). IATA Manual. Retrieved June 25, 2013, from www.polymetrika.org

CCSSI. (2010). *Common Core State Standards for mathematics*. Retrieved March 3, 2011, from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf

CCSSI (2013). *In the states*. Retrieved March 25, 2013, from http://www.corestandards.org/in-the-states

CHE, 2006a. Elementary Mathematics Teacher License Program. Retrieved March 25, 2008, from http://www.yok.gov.tr/documents/10279/49665/ ilkogretim_matematik /cca48fad-63d7-4b70-898c-dd2eb7afbaf5

CHE, 2006b. Announcement About New Programs for Education Faculties. Retrieved from March 25, 2008, http://www.yok.gov.tr/documents/10279/49665/aciklama_programlar/aa7bd091-9328-4df7-aafa-2b99edb6872f

Chick, H. L., Baker, M., Pham, T., & Cheng, H. (2006). Aspects of teachers' pedagogical content knowledge for decimals. In *30th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 297-304).

Chickering, A. W., Gamson, Z. F., & Poulsen, S. J. (1987). Seven principles for good practice in undergraduate education. *AAHE Bulletin*, *39*(7), 3-7.

Cobb, G.W. (1992). Teaching statistics. In L.A. Steen (Ed) *Heeding the Call for Change: Suggestions for Curricular Action* (pp. 3-43). Washington, D.C.: Mathematical Association of America

Cooper, L. L. (2002). An assessment of prospective secondary mathematics teachers' preparedness to teach statistics (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3078386)

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, *64*(3), 391-418.

delMas, R. (2004). A comparison of mathematical and statistical reasoning. In D. Ben-Zvi & J. Garfield, (Eds.). *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 79-95). The Netherlands: Kluwer Academic Publishers.

Downing, S. M. (2006). Twelve steps for effective test development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 3-25). Mahwah, NJ: Lawrence Erlbaum Associates

Eisenhart, M., Borko, H., Underhill, R., Brown, C., Jones, D., & Agard, P. (1993). Conceptual knowledge falls through the cracks: Complexities of learning to teach mathematics for understanding. *Journal for Research in Mathematics Education*, *24*(1), 8-40.

Embretson, S. E. , & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

Fan, L., & Cheong, N. P. (2002). Investigating the sources of Singaporean mathematics teachers' pedagogical knowledge. In D. Edge & B. H. Yeap (Eds.), *Mathematics education for a knowledge-based era : Vol. 2* (pp. 224-23 1). Singapore: AME.

Field, A. P. (2005). *Discovering statistics using SPSS* (2nd edition). London, England: Sage

Fraenkel, J. R., & Wallen, N. E. (2009). *How to design and evaluate research in education* (7th ed.). Boston, MA: McGraw Hill Higher Education.

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.* Alexandria, VA: American Statistical Association.

García Cruz, J. A., & Garrett, A. J. (2008). Understanding the Arithmetic Mean: A Study with Secondary and University Students. *Journal of the Korea Society of Mathematical Education Series D: Research in Mathematical Education,12*(1), 49-66.

Garfield, J. (1995). How students learn statistics. *International Statistical Review*, *63*, 25-34.

Gonzalez, E., & Rutkowski, L. (2010). Principles of multiple matrix booklet designs and parameter recovery in large-scale assessments. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, *3*, 125-156. Retrieved July 25, 2012, from http://www.ierinstitute.org/dissemination-area.html

Graeber, A., & Tirosh, D. (2008). Pedagogical content knowledge: Useful concept or elusive notion? In P. Sullivan & T. Woods (Eds.), *International handbook of mathematics teacher education: Vol. 1. Knowledge and beliefs in mathematics teaching and teaching development* (pp. 117–132). Rotterdam, the Netherlands: Sense Publishers.

Grouws, D. A., & Schults, K.A. (1996). Mathematics teacher education. In J. Sikula, T. J. Buttery & E. Guyton (Eds.), Handbook of research on teacher education. New York, NY: Mcmillan.

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

Ha, L. (2012). Sampling in international advertising research. In S. Okazaki (Ed.) *Handbook of Research on International Advertising*. Glos, England: Edward Elgar Publishing.

Hill, H. C., Ball, D. L., Sleep, L., & Lewis, J. M. (2007). Assessing Teachers' Mathematical Knowledge: What Knowledge Matters and What Evidence Counts? In F. Lester (Ed.), *Handbook for Research on Mathematics Education* (pp. 111-155). Charlotte, NC: Information Age Publishing.

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education, 39*(4), 372–400.

Hill, H.C., Schilling, S.G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal*, *105*, 11-30.

Hill, H.C., Rowan, B., & Ball, D.L. (2005) Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*, 371-406.

Huff, D. (1954). How to Lie with Statistics. London, England: Penguin Books.

Hurley, A. E., Scandura, T. A., Schriesheim, C. A., Brannick, M. T., Seers, A., Vandenberg, R. J., & Williams, L. J. (1997). Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. *Journal of Organizational Behavior*, *18*(6), 667-683.

Johnson, E. G. (1992). The design of the national assessment of educational progress. *Journal of Educational Measurement*, *29*(2), 95-110.

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology, 100*(3), 716-725.

Kennedy, M. M. (1998). *Learning to teach writing: Does teacher education make a difference?*. New York, NY: Teachers College Press.

Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S.,& Baumert, J., (2013). Teachers' Content Knowledge and Pedagogical Content Knowledge: The Role of Structural Differences in Teacher Education. *Journal of Teacher Education*, *64*(1), 90-106. doi: 10.1177/0022487112460398

Koc, Y., Isiksal, M. & Bulut S. (2007). Elementary school curriculum reform in Turkey. *International Education Journal*, *8*(1), 30-39.

Lee, C., and Meletiou-Mavrotheris, M., (2003), Some difficulties of learning histograms in introductory statistics. In *2003 Proceedings of the American Statistical Association*, Statistics Education Section [CD-ROM], pp. 2326 - 2333. Alexandria, VA: American Statistical Association.

Linn, R. L. (2006). The standards for educational and psychological Testing: Guidance in test development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 27-38). Mahwah, NJ: Lawrence Erlbaum Associates

Llinares, S., & Krainer, K. (2006). Mathematics (student) teachers and teacher educators as learners. In A. Gutierrez & P. Boero (Eds.), *Handbook of Research on the Psychology of Mathematics Education: Past, Present and Future* (pp. 429–459). Rotterdam: Sense Publishers.

Lortie, D. C. (1975). *Schoolteacher: A sociological study*. Chicago, IL: University of Chicago Press.

Mevarech, Z. R. (1983). A deep structure model of students' statistical misconceptions. *Educational studies in mathematics, 14*(4), 415-429.

Manizade, A. (2006). Designing measures for assessing teachers' pedagogical content knowledge of geometry and measurement at the middle school level (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3235090)

Messick, S. (1992). Validity of test interpretation and use. In M. C. Alkin (ed.) *Encyclopedia of Educational Research* (6th ed., pp. 1487-1495). New York, NY: Macmillan

MNE (Ministry of National Education) (2005a) *Ilkogretim matematik dersi 6-8 sınıflar öğretim programı ve klavuzu*. Ankara, Turkey: MNE.

MNE (Ministry of National Education) (2005b) *Ilkogretim Matematik Dersi 1-5 siniflar Ogretim Programi ve Klavuzu*. Ankara, Turkey: MNE.

MNE (2007). *Matematik öğretmeni özel alan yeterlilikleri*. Retrieved October 25, 2010, from http://otmg.meb.gov.tr/belgeler/ogretmen_yeterlikleri_kitabi/%C3%96%C4%9Fretmen_Yeterlikleri_Kitab%C4%B1_matematik_%C3%B6%C4%9Fretmeni_%C3%B6zel_alan_yeterlikleri_ilk%C3%B6%C4%9Fretim_par%C3%A7a_10.pdf

MNE (Ministry of National Education) (2009) *Ilkogretim matematik dersi 6-8 sınıflar öğretim programı ve klavuzu*. Ankara, Turkey: MNE.

MNE (Ministry of National Education) (2013) *Ortaokul Matematik Dersi (5, 6, 7 ve 8. Siniflar) öğretim programi*. Retrieved August 25, 2013, from http://ttkb.meb.gov.tr/www/ogretim-programlari/icerik/72

Moore, D. (1998). Statistics among the liberal arts. *Journal of the American Statistical Association, 93*, 1253–1259.

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College

Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan C. Y., Arora, A. & Erberber, E.. (2005). *TIMSS 2007 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College

Muthen, B., Kao, Chih-Fen & Burstein, L. (1991). Instructional sensitivity in mathematics achievement test items: Applications of a new IRT-based detection technique. *Journal of Educational Measurement*, 28, 1-22.

Muthén, L.K. and Muthén, B.O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén

Muthen, L. K., & Muthen, B. O. (2006).IRT in Mplus. Retrieved March 25, 2013, from http://www.statmodel.com/download/MplusIRT1.pdf

NCTM (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM

OSYM (2008). Table 4: University Student Capacities for Bachelor Degree Level. Retrieved September 23, 2010, from http://dokuman.osym.gov.tr/pdfdokuman/arsiv/2008/2008_OSYS_TERCIH_KILAVUZU/6_tablo4.pdf

OSYM (2009). Table 4: University Student Capacities for Bachelor Degree Level. Retrieved September 23, 2010, from http://dokuman.osym.gov.tr/pdfdokuman/arsiv/2009/2009_OSYS_TERCIH_KILAVUZU/tablo4.pdf

Park, S., & Oliver, J. S. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education*, *38*(3), 261-284.

Pfannkuch, M., & Ben-Zvi, D. (2011). Developing teachers' statistical thinking. In *Teaching statistics in school mathematics-challenges for teaching and teacher education* (pp. 323-333). Netherlands: Springer.

Pinto Sosa, J. E. (2010). Conocimiento didáctico del contenido sobre la representación de datos estadísticos: Estudios de casos con profesores de estadística en carreras de psicología y educación. (Doctoral dissertation, Universidad de Salamanca, Salamanca, Spain). Retrieved March 25, 2012, from www.stat.auckland.ac.nz/~iase/publications

Porter, A., McMaken, J., Hwang, J., & Yang, R. (2011). Common Core standards: The new U.S. intended curriculum. *Educational Researcher*, *40*(3), 103–116.

Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology* (pp. 111-163). Cambridge, England: Basil Blackwell.

Russell, S. J., & Mokros, J. R. (1990). What's typical: Teachers' descriptions of data. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International Large-Scale Assessment Data Issues in Secondary Analysis and Reporting. *Educational Researcher*, *39*(2), 142-151.

Schmidt, W., Tatto, M.T., Bankov, K., Blomeke, S., Cedillo, T., Cogan, L., Han, S. I., Houang, R., Hsieh, F. J., Paine, L., Santillan, M., & Schwille, J. (2007). *The preparation gap: Teacher education for middle school mathematics in six countries* (MT21 report). East Lansing, MI: Michigan State University. Retrieved February 25, 2010, from http://usteds.msu.edu/MT21Report.pdf

Shute, V. J. (2008). Focus on formative feedback. *Review of educational research, 78*(1), 153-189.

Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, *15*(2), 4-14

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard educational review*, *57*(1), 1-23.

Stevens, J. (2002). *Applied Multivariate Statistics for the Social Sciences (4th ed.).* Mahwah, NJ: Lawrence Erlbaum Associates.

Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2008). *Teacher Education and Development Study in Mathematics (TEDS-M): Conceptual framework.* East Lansing, MI: Teacher Education and Development International Study Center, College of Education, Michigan State University.

Taylor, S. E. (1993). An exploratory study of the statistical competencies of certain future elementary and middle school teachers (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9328809)

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

Utts, J. (2003). What Educated Citizens Should Know about Statistics and Probability. *The American Statistician*, *57*(2), 74-79.

Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, N.J.: L. Erlbaum Associates.

Yu, C.-Y. (2002). Evaluation of model fit indices for latent variable models with categorical and continuous outcomes. Unpublished dissertation. Retrieved March 25, 2010, from Mplus website http://www.statmodel.com/download/Yudissertation.pdf

Zeichner, K. M., & Tabachnick, B. R. (1981). Are the effects of university teacher education "washed out" by school experiences? *Journal of Teacher Education*, *32*(3), 7–11.

APPENDIX A

BLUEPRINT AND COMPLETE ITEMS FOR ITEM BANK

ITEM BANK BLUE PRINT

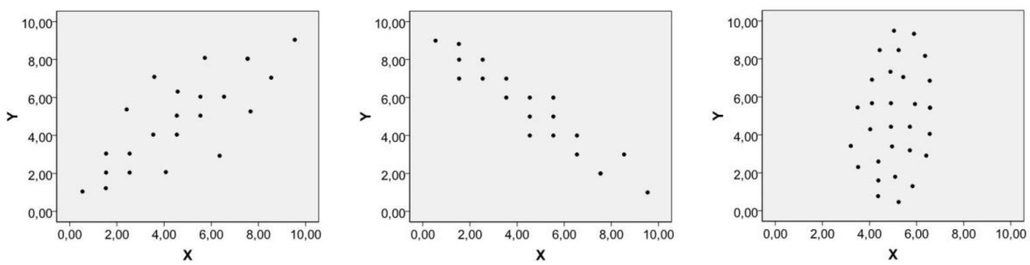
Table A.1.

CONCEPT	CONTENT KNOWLEDGE			PEDAGOGICAL CONTENT KNOWLEDGE					
	K	A	R	PCK_1	PCK_2	PCK_3	PCK_4	PCK_5	PCK_6
MEAN	B.7A B.7B		B.7C B.7D			B.10		B.5A	B.5B
GRAPHICS		B.1	B.8A B14	B.3	B.12 B.6	B.4 B.9	B.11 B.13	B.2	B.8B

ITEM BANK

ITEM B.1. Aşağıda üç farklı durum için serpilme (saçılma) diyagramları verilmiştir?

(I) (II) (III)



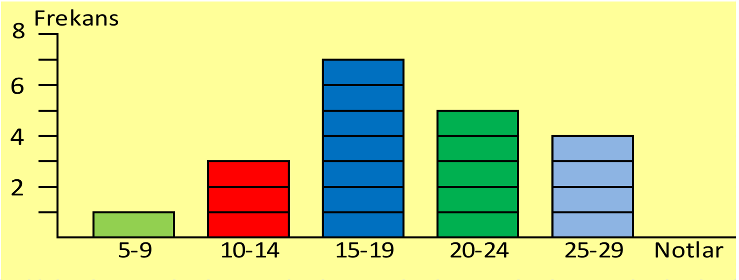
Birinci grafikteki korelasyon katsayısı r_1, ikinci grafikteki korelasyon katsayısı r_2, ve üçüncü grafikteki korelasyon katsayısı r_3 olmak üzere aşağıdaki karşılaştırmalardan hangisi doğrudur?

A) $|r_1| < |r_2| < |r_3|$ B) $|r_1| < |r_3| < |r_2|$ C) $|r_2| < |r_1| < |r_3|$

D) $|r_2| < |r_3| < |r_1|$ E) $|r_3| < |r_1| < |r_2|$ F) $|r_3| < |r_2| < |r_1|$

ITEM B.2.

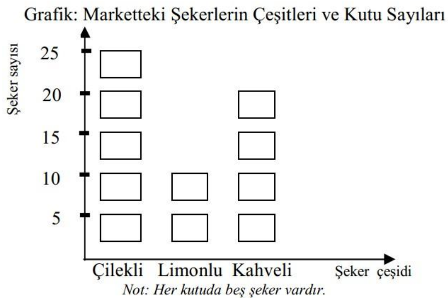
Aralık	Frekans
5-9	1
10-14	3
15-19	7
20-24	5
25-29	4



Bir sınıfın 30 soruluk bir testteki doğru cevaplarının sayısı için bulunan veri yukarıda soldaki tabloda gösterilmiştir. Sağda ise bir öğrencinin bu gruplandırılmış veri için çizdiği grafik görülmektedir. Öğrencinin **bu grafikte** yaptığı hata veya hatalar nelerdir? (açıklayınız)

ITEM B.3. Fatma öğretmenin öğrencileri geçen sene sınıfta yandaki nesne grafiğini oluşturmuşlardır.

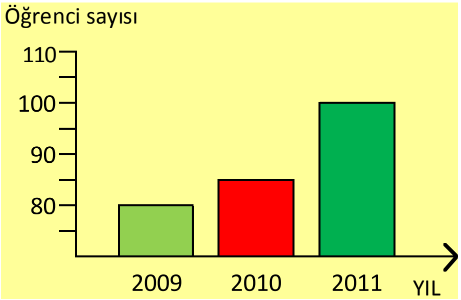
Fatma öğretmen bu yıl öğreteceği grafik konularının hangisinde veya hangilerinde aynı verinin kullanılabileceğini düşünmektedir.



Bu durumda Fatma öğretmene yardım etmek için aşağıdaki **ifadeleri değerlendiriniz?**

	Doğru	Yanlış
I. Bu nesne grafiğindeki verinin Daire Grafiği olarak da ifade edilmesi anlamlıdır.	(D)	(Y)
II. Bu nesne grafiğindeki verinin Sütun grafiği olarak da ifade edilmesi anlamlıdır.	(D)	(Y)
III. Bu nesne grafiğindeki verinin Çizgi Grafiği olarak da ifade edilmesi anlamlıdır.	(D)	(Y)
IV. Bu nesne grafiğindeki verinin Histogram olarak da ifade edilmesi anlamlıdır.	(D)	(Y)

ITEM B.4. Ayşe öğretmen derse başlarken öğrencilerine, okullarına yeni başlayan öğrencilerin son üç yıla göre dağılımı gösteren yandaki grafiği vermiş ve öğrencilerin bu grafiği yorumlamalarını istemiştir.



Ayşe öğretmen bu örnekle aşağıdaki konulardan hangisine giriş yapmış olabilir?

A) Örneklemin seçilme yönteminin popülasyon hakkında karar verirken önemli olduğu konusuna
B) Grafiklerinin bazı durumlarda yanlış anlamalara yol açabileceği konusuna
C) İstatistiklerle gerçek yaşam durumları için görüş oluşturulabileceği konusuna
D) Hangi yayılma ölçüsünün veriyi daha iyi temsil edebileceği konusuna
E) Hangi eğilim ölçüsünün veriyi daha iyi temsil edebileceği konusuna

ITEM B.5: Ayşe öğretmen öğrencilerine ödev olarak ortanca (medyan) ile ilgili aşağıdaki soruyu sormuştur. Bir gün sonra ödev kağıtlarını toplayarak öğrencilerin bulduğu cevapları not almış ayrıca her öğrencinin ödev kağıdına buldukları sonuçlarla ilgili yorum yapmıştır.

ÖDEV SORUSU

Aşağıdaki veri için ortancayı hesaplayınız.

Bisküvi paketlerindeki kırık bisküvi sayısı	Frekans
2	5
4	6
6	7
9	2
11	1
Toplam	21

Öğrencilerin verdiği cevaplar ve oranları

Cevap	Oran
4	%20
5	%43
6	%21
7	%14
8	%2

B.5A. Bu soruya 5 cevabını veren öğrenciler için nasıl bir durumun söz konusu olduğunu düşünüyorsunuz?

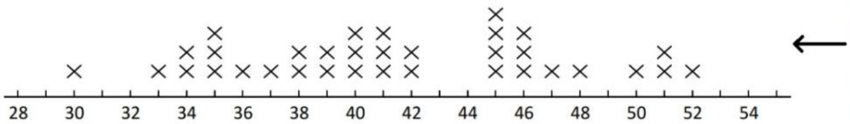
B.5B. Bu öğretmenin soruya 6 cevabını veren öğrencilerin ödev kağıdına yaptığı yorum ne olmalıdır?

ITEM B.6. Mehmet öğretmen öğrencilerine sayıların gruplandırılması konusunu anlatırken aşağıdaki çalışma kağıdını hazırlamış ve ders başlangıcında öğrencilerine gruplar halinde çalışmalarını söyleyerek dağıtmıştır. Mehmet öğretmen öğrencilerinden veriyi 5 gruba ayırmasını istemiştir. Sayıların dağılımını öğrencilerin daha iyi anlaması için bir de grafik çizmiştir.

ÇALIŞMA KAĞIDI

Aşağıda geçen yıl 32 öğrenciden oluşan sekizinci sınıf öğrencilerinin ağırlıkları görülmektedir. Bu veriyi 5 gruba ayırınız.

30	33	34	34	35	35	35	36	37	38	38	39	39	40	40	40
41	41	41	42	45	45	45	45	46	46	46	47	48	50	51	52

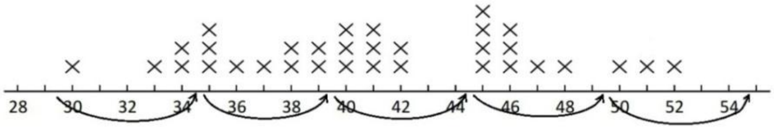


Öğretmenin gruplandırmaya yardımcı olması amacıyla çalışma kağıdına çizdiği grafik

Mehmet öğretmen kendisine de aşağıdaki cevap anahtarını hazırlamıştır.

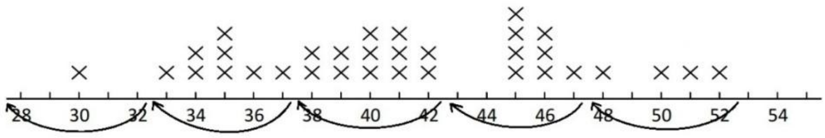
Mehmet öğretmenin cevap anahtarı:

Aralık	Frekans
30-34	4
35-39	9
40-44	8
45-49	9
50-54	3



Mehmet öğretmen daha sonra sınıfı gezerek grupların çalışmalarını izlemiş ve her gruba yorumlar yapmıştır. Bu sırada gruplardan birinin farklı bir yöntemle çalıştığını görmüştür. Bu gurubun yöntemi aşağıdaki gibidir.

Aralık	Frekans
28-32	4
33-37	9
38-42	8
43-47	9
48-52	3



Bu durumda Mehmet öğretmenin yerinde olsanız bu grubun yöntemi için nasıl bir yorumda bulunursunuz?

ITEM B.7. Bir sporcunun 100 metrelik bir mesafeyi koşma süresi, on bir öğrenci tarafından birbirinden bağımsız olarak kaydedilmektedir ve her bir öğrenci kendi yöntemini kullanmaktadır. Bu on öğrencinin bulduğu süreler saniye türünden aşağıdaki gibidir.

15,05 14,95 10 15 14,96 15 14,90 15 14,95 15,05 14,91

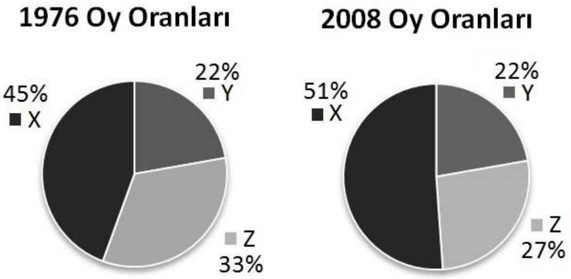
B.7A. Bu verinin ortancası nedir? …….

B.7B. Bu verinin modu nedir? …….

B.7C. Bu verinin ortalaması 14,52 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod ve ortancayı göz önüne alarak tahmin ediniz?

B.7D. Sizce 10 saniye değeri bu veriye ne kadar uygundur? Bu değer hakkında ne düşünüyorsunuz?

ITEM B.8. Fatma öğretmen öğrencilerini dört gruba ayırarak, bir Avrupa ülkesinde 1976 ve 2008 yıllarında yapılan seçimlerde üç partinin oy dağılımlarını gösteren aşağıdaki grafikleri incelemesini istemiştir. Ayrıca bu ülkede seçmen sayısının her geçen yıl artığı bilgisini vererek 2008 yılındaki durumu, 1976 yılı ile nasıl karşılaştırabileceklerini sormuştur.



Fatma öğretmen sınıfı gezerek gruplardan aşağıdaki yorumları almıştır ve **kesinlikle yanlış yorum yapan** guruba müdahele ederek neden yanlış yaptıklarını açıklamıştır.
Grupların Yorumları:
I) 2008 yılında X partisinin oy sayısı 1976 yılına göre %6'dan fazla artmıştır.
II) 2008 yılı ile 1976 yılları arasındaki seçmen sayısı farkı ile Z partisinin oy sayısındaki değişim hesaplanabilir.
III) 2008 yılında Z partisinin oy sayısı 1976 yılına göre azalmıştır.
IV) Y partisi ile Z partisi arasındaki oy farkı 2008 yılında, 1976 yılına göre azalmıştır.

B.8A. Buna göre <u>kesinlikle yanlış</u> yorum yapan grup aşağıdakilerden hangisidir?
A) I B) II C) III D)IV

B.8B. Bu durum için Fatma öğretmen kesinlikle yanlış yorum yapan guruba hatalarını nasıl açıklayabilir?

ITEM B.9. Ahmet öğretmen çocuklara sütun grafiklerini öğrettikten sonra histogram konusunu anlatacaktır. Histogram konusuna başlarken "sütun grafiklerinin her veri için uygun olmadığını, bazı durumlarda histogram grafiklerine ihtiyaç duyulabileceğini" öğrencilerin fark etmesini istiyor

Bu durumda öğretmenin verebileceği <u>en uygun örnek</u> aşağıdakilerden hangisidir?

A) Son 60 aya ait domates fiyatları

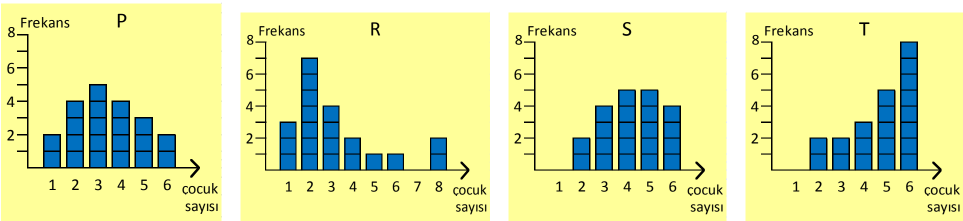
B) Son 60 ay Türkiye'de görülen 5 şiddetinin üzerindeki deprem sayısı

C) Bahçeden toplanan 60 farklı çınar yaprağının genişliği

D) 60 öğrencinin ailelerindeki çocuk sayısı

E) Sınıftan rastgele toplanan 60 kalemin renklere göre dağılımı

ITEM B.10. Öğrencilerin verinin dağılımına göre hangi merkezi eğilim ölçüsünü seçeceklerini öğretmek isteyen bir öğretmen bunu örneklerle açıklamak istiyor. Vereceği örneklerin birisinde ortalama, diğerinde ise ortancanın uygun seçim olmasını istediğine göre aşağıdaki dağılımlardan hangi ikisini kullanabilir.



<u>Ortalama</u> örneği olara hangisini seçmelidir?

A) P

B) R

C) S

D) T

<u>Ortanca</u> Örneği olarak hangisini seçmelidir?

A) P

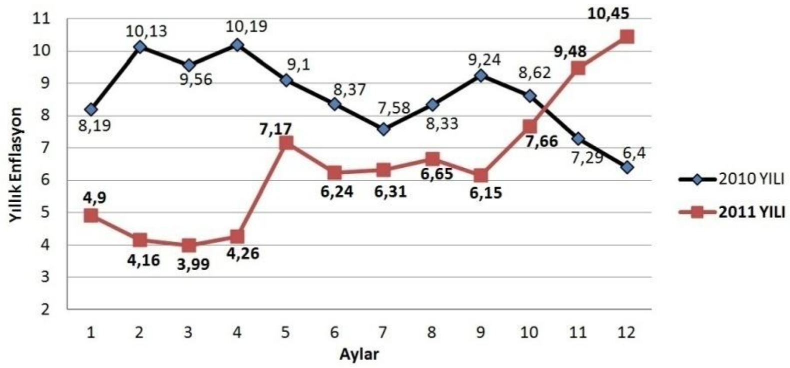
B) R

C) S

D) T

Neden bu şekilde düşündüğünüzü açıklayınız

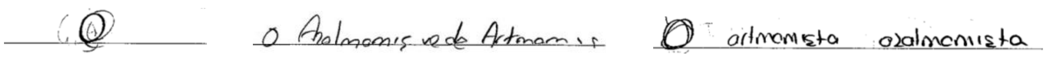
ITEM B.11. Öğretmen sınıfta çizgi grafiği ile ilgili etkinlik yapmak amacıyla yanda görülen 2010 ve 2011 yıllarının aylara göre enflasyon oranı grafiğini hazırlamış ve enflasyonun aylara göre değişimi hakkında sorular sormuştur.



Öğretmenin sorduğu sorulardan biri:

"2011'in 1. Ayındaki yıllık enflasyon bir önceki aya göre nasıl değişmiştir?" sorusudur

Bu soruya öğrencilerin büyük bir kısmından aşağıdaki gibi cevaplar gelmiştir.



Öğrencilerin bu soruya "0 (sıfır)", "Azalmamış ve artmamış" tarzında cevaplar vermesinin nedeni sizce ne olabilir?

ITEM B.12. Aşağıdaki veriyi öğrencilerin nasıl daire grafiğine dönüştürebilecekleri ile ilgili **iki farklı yöntem gösteriniz.**

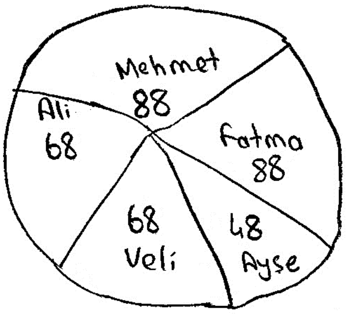
Öğrenci	**Harçlık Miktarı (TL)**
Ayşe	10
Ali	15
Veli	15
Mehmet	20
Fatma	20
TOPLAM	80

1. YÖNTEM

2. YÖNTEM

ITEM B.13. Bir öğrenci aşağıdaki tablo için yanda görülen daire grafiğini oluşturmuştur.

Öğrenci	**Harçlık Miktarı (TL)**
Ayşe	10
Ali	15
Veli	15
Mehmet	20
Fatma	20
TOPLAM	80



Öğrencinin çözümü için yaptığı açıklama:

Toplamları 80'di. 80'nin 360'a en yakın katı 320'dir. Ben de bu sayıları 320'ye tamamlamak için her sayıyı 4 ile çaptım. Sonra 360-320=40 kaldı. 5 sayı olduğu için 40'ı 5 e böldüm, 8 çıktı. Sonra tüm sayılara 8 ekledim.

Öğrencinin cevabını analiz ederek doğru veya yanlış yönlerini belirleyiniz?

ITEM B.14. Bir öğretmen öğrencilerinden histogram konusu için veri toplamalarını istemiştir. Ayşe de veri olarak sınıfındaki 30 öğrencinin ağırlığını evinden getirdiği dijital baskül ile ölçmüştür. Aşağıda bu otuz öğrenciye ait ölçümler görülmektedir.

16,1	16,5	16,6	16,7	16,8	17,6	17,6	17,8	18,0	18,1	19,5	19,8	20,0	20,1	21,8
21,9	22,0	22,1	22,1	22,2	22,4	22,7	22,7	23,5	23,5	23,6	24,5	24,5	24,7	26,8

Bu veriden yararlanarak 5 gruptan oluşan bir histogram oluşturunuz

APPENDIX B

PERMISSION LETTER FROM JUAN ANTONIO GARCÍA CRUZ

Permission Letter sent to Juan Antonio GARCÍA CRUZ and Alexandre Joaquim GARRETT in electronic format:

Requesting Permission for "Understanding the Arithmetic mean: A study with Secondary and University Students" article

Oktay Mercimek <oktaymercimek@gmail.com> Sun, Dec 18, 2011 at 3:59 PM
To: jagcruz@ull.es

Dear JUAN ANTONIO GARCÍA CRUZ and ALEXANDRE JOAQUIM GARRETT,

I am Oktay MERCİMEK, Ph.D. student from Middle East Technical University, TURKEY. I am preparing an instrument for measuring teachers' statistics knowledge. For this instrument, I would like to use some questions and results from your article "Understanding the Arithmetic mean: A study with Secondary and University Students (Garcia Cruz & Garrett, 2008)" and I would greatly appreciate if you give me permission to use some of them in my thesis.

Thank you for your cooperation,

Sincerely,

Oktay MERCİMEK

oktaymercimek@gmail.com

e159845@metu.edu.tr

omercimek@kastamonu.edu.tr

+90 505 807 3700

Response Letter from Juan Antonio GARCIA CRUZ in electronic format:

Requesting Permission for "Understanding the Arithmetic mean: A study with Secondary and University Students" article

jagcruz@ull.es <jagcruz@ull.es> Tue, Dec 20, 2011 at 11:01 AM
To: Oktay Mercimek <oktaymercimek@gmail.com>

Dear Otkay you can use whatever you want and remember to quote us precisely

APPENDIX C

COMPLETE MKT-S INSTRUMENT FOR PILOT IMPLEMENTATION

ITEMS

ITEM P.1. Bir sporcunun 100 metrelik bir mesafeyi koşma süresi, on bir öğrenci tarafından birbirinden bağımsız olarak aynı anda kaydedilmektedir ve her bir öğrenci kendi yöntemini kullanmaktadır. Bu on bir öğrencinin bulduğu süreler saniye türünden aşağıdaki gibidir.

15,05	14,97	13	15	14,98	15	14,93	15,06	14,96	15	14,96

P.1A. Bu verinin **ortancası nedir? …….**

P.1B. Bu verinin **modu nedir? …….**

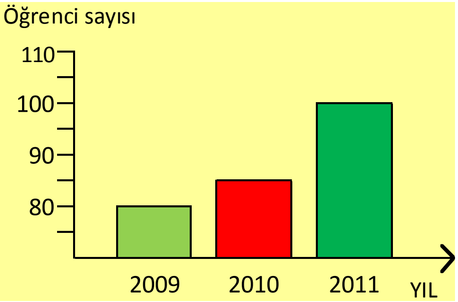
P.1C. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: ……

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız:

P.1D. Sizce 13 saniye değeri bu veriye ne kadar uygundur? Bu değer hakkında ne düşünüyorsunuz?

ITEM P.2. Ayşe öğretmen derse başlarken öğrencilerine, okullarına yeni kayıt yaptıran öğrencilerin son üç yıla göre dağılımı gösteren yandaki grafiği vermiş ve öğrencilerin bu grafiği yorumlamalarını istemiştir.



Bu örnek aşağıdaki konulardan hangisi için **tipik bir örnektir?**

A) Örneklemin seçilme yönteminin popülasyon hakkında karar verirken önemli olduğu konusuna
B) Grafiklerin bazı durumlarda yanlış anlamalara yol açabileceği konusuna
C) Hangi yayılma ölçüsünün (standard sapma vb.) veriyi daha iyi temsil edebileceği konusuna
D) Hangi eğilim ölçüsünün (ortalama vb.) veriyi daha iyi temsil edebileceği konusuna

Neden bu şekilde düşündüğünüzü açıklayınız?

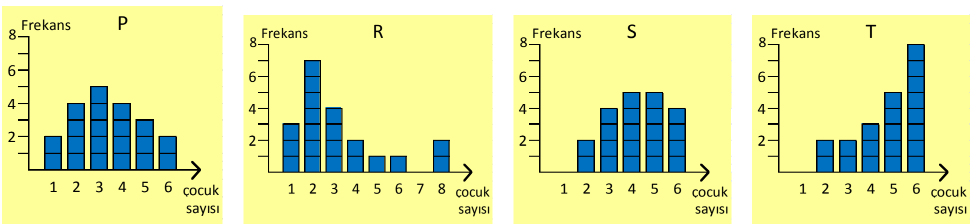
ITEM P.3. Fatma öğretmenin öğrencileri geçen sene sınıfta yandaki nesne grafiğini oluşturmuşlardır.

Fatma öğretmen bu yıl öğreteceği grafik konularının hangisinde veya hangilerinde aynı verinin kullanılabileceğini düşünmektedir.

Bu durumda Fatma öğretmene yardım etmek için aşağıdaki ifadeleri değerlendiriniz?

	Doğru	Yanlış
I. Bu nesne grafiğindeki verinin Daire Grafiği olarak da ifade edilmesi anlamlıdır.	(D)	(Y)
II. Bu nesne grafiğindeki verinin Sütun grafiği olarak da ifade edilmesi anlamlıdır.	(D)	(Y)
III. Bu nesne grafiğindeki verinin Çizgi Grafiği olarak da ifade edilmesi anlamlıdır.	(D)	(Y)
IV. Bu nesne grafiğindeki verinin Histogram olarak da ifade edilmesi anlamlıdır.	(D)	(Y)

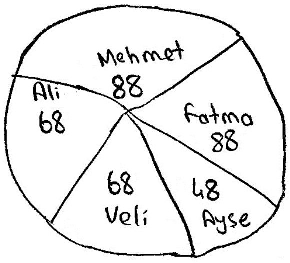
ITEM P.4. Bir öğretmen, ortalama ve ortanca ölçülerinin hangi durumlar için uygun bir merkezi eğilim ölçüsü olduğunu örneklerle açıklamak istiyor ve bir örnek ortalamanın uygun olduğu durumlar için bir örnek de ortancanın uygun olduğu durumlar için verecektir. Bu amaçla aşağıdaki grafikleri incelemektedir.



Bu öğretmenin yerinde siz olsanız bu iki örnek için hangi grafiği veya grafikleri seçersiniz? Nedeniyle birlikte açıklayız?

ITEM P.5. Bir öğrenci aşağıdaki tablo için yanda görülen daire grafiğini oluşturmuştur.

Öğrenci	Harçlık Miktarı (TL)
Ayşe	10
Ali	15
Veli	15
Mehmet	20
Fatma	20
TOPLAM	80



Öğrencinin çözümü için yaptığı açıklama:

Toplamları 80'di. 80'nin 360'a en yakın katı 320'dir. Ben de bu sayıları 320'ye tamamlamak için her sayıyı 4 ile çaptım. Sonra 360-320=40 kaldı. 5 sayı olduğu için 40'ı 5 e böldüm, 8 çıktı. Sonra tüm sayılara 8 ekledim.

Öğrencinin cevabını analiz ederek doğru veya yanlış yönlerini belirleyiniz?

ITEM P.6. Ayşe öğretmen öğrencilerine ödev olarak ortanca (medyan) ile ilgili aşağıdaki soruyu sormuştur.

ÖDEV SORUSU

Aşağıdaki tablo toplam 21 paket için her paketteki kırık bisküvi sayısını göstermektedir. Bu veri için ortancayı hesaplayınız.

Bisküvi paketlerindeki kırık bisküvi sayısı	Frekans
2	5
4	6
6	7
9	2
11	1
Toplam	21

Bir gün sonra ödev kağıtlarını toplayarak öğrencilerin bulduğu cevapları not almıştır. Bu sırada doğru cevap 4 olmasına rağmen birçok öğrencinin 5 veya 6 cevabını verdiğini görmüştür. Daha sonra ise öğrencilerin ödev kağıtlarına, neden hata yaptıklarına dair yorum yazmıştır.

Aşağıdaki soruları yukarıda verilen bilgiler doğrultusunda yanıtlayınız.

P.6A. İşlem hatası olmadığını varsayarsak, birçok öğrencinin bu soruya <u>5</u> cevabını vermesinin sebebi ne olabilir?

P.6B. İşlem hatası olmadığını varsayarsak bu soruya 6 cevabını veren öğrencilerin ödev kağıdına hatalarını düzeltebilmeleri için (neden hata yaptıklarını da göz önüne alarak) nasıl bir yorum yapardınız?

ITEM P.7. Ahmet öğretmen çocuklara sütun grafiklerini öğrettikten sonra histogram konusunu anlatacaktır. Histogram konusuna başlarken "sütun grafiklerinin her veri için uygun olmadığını, bazı durumlarda histogram grafiklerine ihtiyaç duyulabileceğini" öğrencilerin fark etmesini istiyor. Bu durumda öğretmenin verebileceği en uygun örnek aşağıdakilerden hangisidir?

A) Son 60 aya ait kuruş türünden domates fiyatları
B) Son 60 ay Türkiye'de görülen 5 şiddetinin üzerindeki deprem sayısı
C) Bahçeden toplanan 60 farklı çınar yaprağının cm türünden genişliği
D) 60 öğrencinin ailelerindeki çocuk sayısı
E) Sınıftan rastgele toplanan 60 kalemin renklere göre dağılımı

Neden bu şekilde düşündüğünüzü açıklayınız

ITEM P.8. Daire grafiklerini oluşturmak için iki temel yöntem vardır. Aşağıdaki veriyi öğrencilerin daire grafiğine nasıl dönüştürebilecekleri ile ilgili bu iki yöntemi gösteriniz.

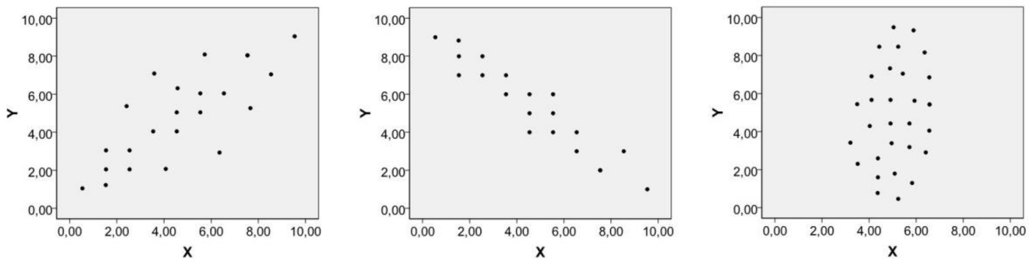
Öğrenci	**Harçlık Miktarı (TL)**
Ayşe	6
Ali	8
Veli	4
Mehmet	12
Fatma	10
TOPLAM	40

1. YÖNTEM

2. YÖNTEM

ITEM P.9. Aşağıda üç farklı durum için serpilme (saçılma) diyagramları verilmiştir?

(I) (II) (III)



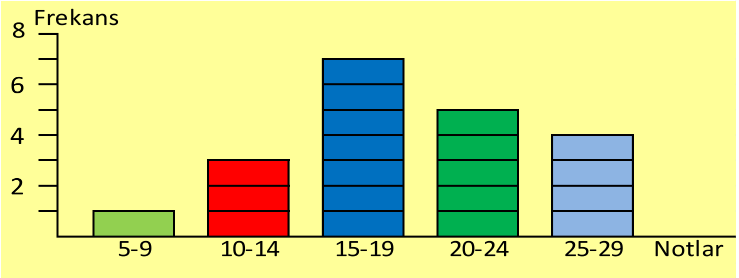
Birinci grafikteki korelasyon katsayısı r_1, ikinci grafikteki korelasyon katsayısı r_2, ve üçüncü grafikteki korelasyon katsayısı r_3 olmak üzere aşağıdaki karşılaştırmalardan hangisi doğrudur?
NOT: | | : mutlak değer anlamında kullanılmıştır.

A) $|r_1| < |r_2| < |r_3|$ B) $|r_1| < |r_3| < |r_2|$ C) $|r_2| < |r_1| < |r_3|$

D) $|r_2| < |r_3| < |r_1|$ E) $|r_3| < |r_1| < |r_2|$ F) $|r_3| < |r_2| < |r_1|$

ITEM P.10.

Aralık	Frekans
5-9	1
10-14	3
15-19	7
20-24	5
25-29	4



Bir sınıfın 30 soruluk bir testteki doğru cevaplarının sayısı için bulunan veri öğretmen tarafından oluşturulan yukarıda soldaki tabloda gösterilmiştir. Sağda ise bir öğrencinin bu gruplandırılmış veri için çizdiği grafik görülmektedir. Öğrencinin **bu grafikte** yaptığı <u>hata veya hatalar nelerdir?</u> (açıklayınız)

ITEM P.11. Bir öğretmen öğrencilerinden histogram konusu için veri toplamalarını istemiştir. Ayşe de veri olarak sınıfındaki 30 öğrencinin ağırlığını evinden getirdiği dijital baskül ile ölçmüştür. Aşağıda bu otuz öğrenciye ait ölçümler görülmektedir.

16,1	16,5	16,6	16,7	16,8	17,6	17,6	17,8	18,0	18,1	19,5	19,8	20,0	20,1	21,8
21,9	22,0	22,1	22,1	22,2	22,4	22,7	22,7	23,5	23,5	23,6	24,5	24,5	24,7	26,8

Bu veriden yararlanarak 5 gruptan oluşan bir histogram oluşturunuz

ITEM P.12. Mehmet öğretmen öğrencilerine sayıların gruplandırılması konusunu anlatırken aşağıdaki çalışma kağıdını hazırlamış ve ders başlangıcında öğrencilerine gruplar halinde çalışmalarını söyleyerek dağıtmıştır. Mehmet öğretmen öğrencilerinden veriyi 5 gruba ayırmasını istemiştir.

ÇALIŞMA KAĞIDI

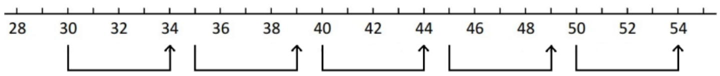
Aşağıda geçen yıl 32 öğrenciden oluşan sekizinci sınıf öğrencilerinin ağırlıkları görülmektedir. Bu veriyi 5 gruba ayırınız.

30	33	34	34	35	35	35	36	37	38	38	39	39	40	40	40
41	41	41	42	45	45	45	45	46	46	46	47	48	50	51	52

Mehmet Öğretmen kendisine de aşağıdaki cevap anahtarını hazırlamıştır.

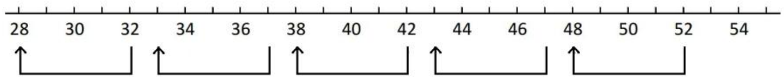
Mehmet Öğretmenin cevap anahtarı:

Aralık	Frekans
30-34	4
35-39	9
40-44	7
45-49	9
50-54	3



Mehmet öğretmen daha sonra sınıfı gezerek grupların çalışmalarını izlemiş ve her gruba yorumlar yapmıştır. Bu sırada gruplardan birinin farklı bir şekilde çalışmaya başladığını, aralıkları sondan başa doğru belirleyerek farklı frekans değerlerine ulaştıklarını görmüştür. Bu grubun yöntemi aşağıdaki gibidir.

Aralık	Frekans
28-32	1
33-37	8
38-42	11
43-47	8
48-52	4

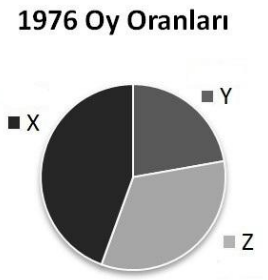


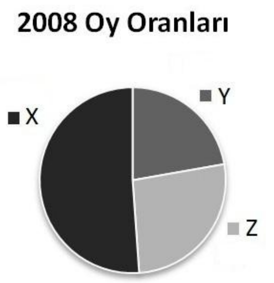
Bu durumda öğretmen siz olsanız ve bu grubun yöntemini 4 puan üzerinden değerlendirmeniz gerekirse aşağıdaki puanlardan hangisini verirsiniz? Nedenini açıklayınız?

a) Tamamen Doğrudur (4 Puan)
b) Büyük Oranda Doğrudur (3 Puan)
c) Yarı Yarıya Doğrudur (2 Puan)
d) Büyük Oranda Yanlıştır (1 Puan)
e) Tamamen Yanlıştır (0 Puan)

Neden bu şekilde düşündüğünüzü açıklayınız?

ITEM P.13. Fatma öğretmen öğrencilerini dört gruba ayırarak, bir Avrupa ülkesinde 1976 ve 2008 yıllarında yapılan seçimlerde üç partinin oy dağılımlarını gösteren aşağıdaki grafikleri incelemesini istemiştir. Ayrıca bu ülkede seçmen sayısının her geçen yıl arttığı bilgisini vererek 2008 yılındaki durumu, 1976 yılı ile nasıl karşılaştırdıklarında hangi sonuçlara ulaşılabileceğini sormuştur.





Fatma öğretmen sınıfı gezerek gruplardan aşağıdaki yorumları almıştır ve **kesinlikle yanlış yorum yapan** gruba müdahale ederek neden yanlış yaptıklarını açıklamıştır.

Grupların Yorumları:

I) Y partisi ile Z partisi arasındaki oy farkı 2008 yılında, 1976 yılına göre azalmıştır.
II) 2008 yılı seçmen sayısı bilinirse Y partisinin oyunu ne kadar arttırdığı bulunabilir.
III) 2008 yılında Z partisinin oy sayısı 1976 yılına göre azalmıştır.
IV) 2008 yılında X partisinin oy sayısı 1976 yılına göre artmıştır.

P.13A. Buna göre kesinlikle yanlış yorum yapan grup aşağıdakilerden hangisidir?
A) I B) II C) III D)IV

P.13B. Bu durum için siz olsanız kesinlikle yanlış yorum yapan gruba hatalarını düzeltebilmeleri için (neden hata yaptıklarını da göz önüne alarak) nasıl bir yorum yapardınız?

APPENDIX D

BLUEPRINT AND COMPLETE MKT-S INSTRUMENT FOR FINAL IMPLEMENTATION

BLUEPRINT

Table D.1.

	CONTENT KNOWLEDGE			MATHEMATICAL PEDAGOGICAL CONTENT KNOWLEDGE					
	K	A	R	PCK_1	PCK_2	PCK_3	PCK_4	PCK_5	PCK_6
MEAN	F.1A F.1B		F.1C F.1D			F.3		F.5A	F.5B
GRAPHICS		F.8	F.10	F.2	F.7 F.11	F.6	F.4	F.9	F.12

ITEMS

DEMOGRAFİK SORULAR

ÜNİVERSİTE:

CİNSİYET: Bayan Erkek

SINIF: 3 4

İstatistik ve Olasılık 1 dersini geçme (harf) notunuz: …..

Matematik Öğretimi veya İstatistik Öğretimi ile İlgili Seçmeli Ders Aldınız mı?

Bu Dersleri Yazınız:

İSTATİSTİK BİLGİ YAPILARINI BELİRLEME ÖLÇEĞİ

ITEM F.1. Bir sporcunun 100 metrelik bir mesafeyi koşma süresi, on bir öğrenci tarafından birbirinden bağımsız olarak aynı anda kaydedilmektedir ve her bir öğrenci kendi yöntemini kullanmaktadır. Bu on bir öğrencinin bulduğu süreler saniye türünden aşağıdaki gibidir.

15,05 14,97 13 15 14,98 15 14,93 15,06 14,96 15 14,96

F.1A. Bu verinin **ortancası nedir? …….**

F.1B. Bu verinin **modu nedir? …….**

F.1C. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: ……

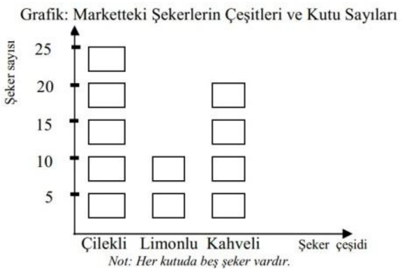
Bu sayıyı nasıl tahmin ettiğinizi açıklayınız:

F.1D. Sizce 13 saniye değeri bu veriye **ne kadar uygundur? Bu değer hakkında ne düşünüyorsunuz?**

ITEM F.2. Fatma öğretmenin öğrencileri geçen sene sınıfta yandaki nesne grafiğini oluşturmuşlardır.

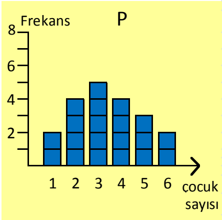
Fatma öğretmen bu yıl öğreteceği grafik konularının hangisinde veya hangilerinde aynı verinin kullanılabileceğini düşünmektedir.

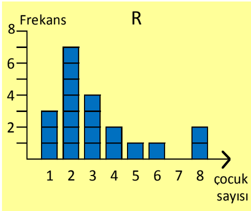
Bu durumda Fatma öğretmene yardım etmek için aşağıdaki ifadeleri değerlendiriniz?

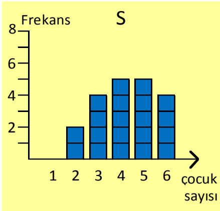


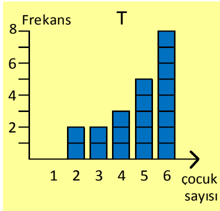
	Doğru	Yanlış
I. Bu nesne grafiğindeki verinin Daire Grafiği olarak da ifade edilmesi anlamlıdır.	(D)	(Y)
II. Bu nesne grafiğindeki verinin Sütun grafiği olarak da ifade edilmesi anlamlıdır.	(D)	(Y)
III. Bu nesne grafiğindeki verinin Çizgi Grafiği olarak da ifade edilmesi anlamlıdır.	(D)	(Y)
IV. Bu nesne grafiğindeki verinin Histogram olarak da ifade edilmesi anlamlıdır.	(D)	(Y)

ITEM F.3. Bir öğretmen, ortalama ve ortanca ölçülerinin hangi durumlar için uygun bir merkezi eğilim ölçüsü olduğunu örneklerle açıklamak istiyor ve bir örnek ortalamanın uygun olduğu durumlar için bir örnek de ortancanın uygun olduğu durumlar için verecektir. Bu amaçla aşağıdaki grafikleri incelemektedir.





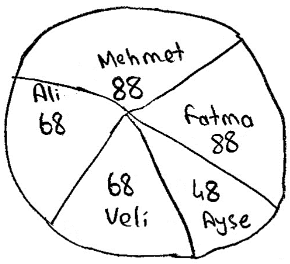




Bu öğretmenin yerinde siz olsanız bu iki örnek için hangi grafiği veya grafikleri seçersiniz? Nedeniyle birlikte açıklayız?

ITEM F.4. Bir öğrenci aşağıdaki tablo için yanda görülen daire grafiğini oluşturmuştur.

Öğrenci	**Harçlık Miktarı (TL)**
Ayşe	10
Ali	15
Veli	15
Mehmet	20
Fatma	20
TOPLAM	80



Öğrencinin çözümü için yaptığı açıklama:

Toplamları 80'di. 80'nin 360'a en yakın katı 320'dir. Ben de bu sayıları 320'ye tamamlamak için her sayıyı 4 ile çaptım. Sonra 360-320=40 kaldı. 5 sayı olduğu için 40'ı 5 e böldüm, 8 çıktı. Sonra tüm sayılara 8 ekledim.

Öğrencinin cevabını analiz ederek doğru veya yanlış yönlerini belirleyiniz?

ITEM F.5. Ayşe öğretmen öğrencilerine ödev olarak ortanca (medyan) ile ilgili aşağıdaki soruyu sormuştur.

ÖDEV SORUSU

Aşağıdaki tablo toplam 21 paket için her paketteki kırık bisküvi sayısını göstermektedir. Bu veri için ortancayı hesaplayınız.

Bisküvi paketlerindeki kırık bisküvi sayısı	Frekans
2	5
4	6
6	7
9	2
11	1
Toplam	21

Bir gün sonra ödev kağıtlarını toplayarak öğrencilerin bulduğu cevapları not almıştır. Bu sırada doğru cevap 4 olmasına rağmen birçok öğrencinin 5 veya 6 cevabını verdiğini görmüştür. Daha sonra ise öğrencilerin ödev kağıtlarına, neden hata yaptıklarına dair yorum yazmıştır.

Aşağıdaki soruları yukarıda verilen bilgiler doğrultusunda yanıtlayınız.

F.5A. İşlem hatası olmadığını varsayarsak bu soruya 5 cevabını veren öğrenciler ne düşünerek bu cevaba ulaşmış olabilirler?

F.5B. İşlem hatası olmadığını varsayarsak bu soruya 6 cevabını veren öğrencilerin ödev kağıdına hatalarını düzeltebilmeleri için (neden hata yaptıklarını da göz önüne alarak) nasıl bir yorum yapardınız?

ITEM F.6. Ahmet öğretmen çocuklara sütun grafiklerini öğrettikten sonra histogram konusunu anlatacaktır. Histogram konusuna başlarken "sütun grafiklerinin her veri için uygun olmadığını, bazı durumlarda histogram grafiklerine ihtiyaç duyulabileceğini" öğrencilerin fark etmesini istiyor.
Bu durumda öğretmenin verebileceği en uygun örnek aşağıdakilerden hangisidir?
A) Son 60 aya ait kuruş türünden domates fiyatları
B) Son 60 ay Türkiye'de görülen 5 şiddetinin üzerindeki deprem sayısı
C) Bahçeden toplanan 60 farklı çınar yaprağının cm türünden genişliği
D) 60 öğrencinin ailelerindeki çocuk sayısı
E) Sınıftan rastgele toplanan 60 kalemin renklere göre dağılımı

Neden bu şekilde düşündüğünüzü açıklayınız

ITEM F.7. Daire grafiklerini oluşturmak için iki temel yöntem vardır. Aşağıdaki veriyi öğrencilerin daire grafiğine nasıl dönüştürebilecekleri ile ilgili bu iki yöntemi gösteriniz.

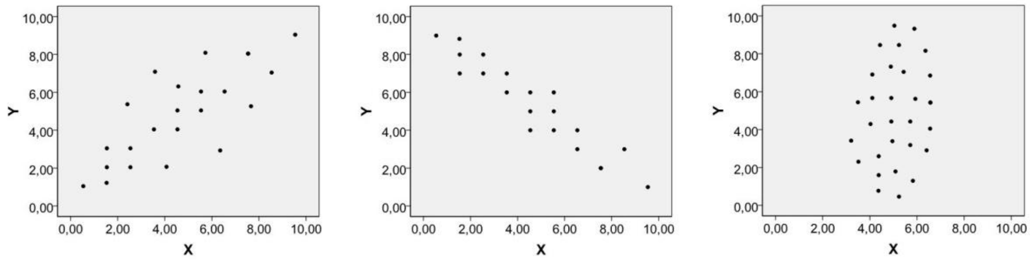
Öğrenci	**Harçlık Miktarı (TL)**
Ayşe	6
Ali	8
Veli	4
Mehmet	12
Fatma	10
TOPLAM	40

1. YÖNTEM

2. YÖNTEM

ITEM F.8. Aşağıda üç farklı durum için serpilme (saçılma) diyagramları verilmiştir?

(I) (II) (III)



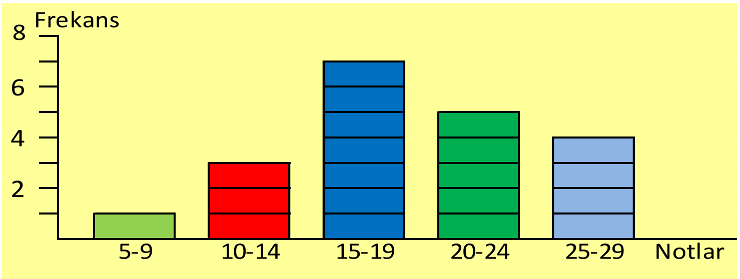
Birinci grafikteki korelasyon katsayısı r_1, ikinci grafikteki korelasyon katsayısı r_2, ve üçüncü grafikteki korelasyon katsayısı r_3 olmak üzere aşağıdaki karşılaştırmalardan hangisi doğrudur?
NOT: | | : mutlak değer anlamında kullanılmıştır.

A) $|r_1| < |r_2| < |r_3|$ B) $|r_1| < |r_3| < |r_2|$ C) $|r_2| < |r_1| < |r_3|$

D) $|r_2| < |r_3| < |r_1|$ E) $|r_3| < |r_1| < |r_2|$ F) $|r_3| < |r_2| < |r_1|$

ITEM F.9.

Aralık	Frekans
5-9	1
10-14	3
15-19	7
20-24	5
25-29	4



Bir sınıfın 30 soruluk bir testteki doğru cevaplarının sayısı için bulunan veri öğretmen tarafından oluşturulan yukarıda soldaki tabloda gösterilmiştir. Sağda ise bir öğrencinin bu gruplandırılmış veri için çizdiği grafik görülmektedir. Öğrencinin **bu grafikte** yaptığı hata veya hatalar nelerdir? (açıklayınız)

ITEM F.10. Bir öğretmen öğrencilerinden histogram konusu için veri toplamalarını istemiştir. Ayşe de veri olarak sınıfındaki 30 öğrencinin ağırlığını evinden getirdiği dijital baskül ile ölçmüştür. Aşağıda bu otuz öğrenciye ait ölçümler görülmektedir.

16,1	16,5	16,6	16,7	16,8	17,6	17,6	17,8	18,0	18,1	19,5	19,8	20,0	20,1	21,8
21,9	22,0	22,1	22,1	22,2	22,4	22,7	22,7	23,5	23,5	23,6	24,5	24,5	24,7	26,8

Bu veriden yararlanarak 5 gruptan oluşan bir histogram oluşturunuz

ITEM F.11. Mehmet öğretmen öğrencilerine sayıların gruplandırılması konusunu anlatırken aşağıdaki çalışma kağıdını hazırlamış ve ders başlangıcında öğrencilerine gruplar halinde çalışmalarını söyleyerek dağıtmıştır. Mehmet öğretmen öğrencilerinden veriyi 5 gruba ayırmasını istemiştir.

ÇALIŞMA KAĞIDI

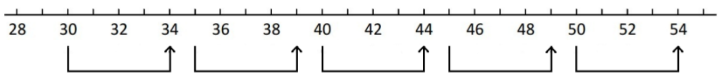
Aşağıda geçen yıl 32 öğrenciden oluşan sekizinci sınıf öğrencilerinin ağırlıkları görülmektedir. Bu veriyi 5 gruba ayırınız.

30	33	34	34	35	35	35	36	37	38	38	39	39	40	40	40
41	41	41	42	45	45	45	45	46	46	46	47	48	50	51	52

Mehmet Öğretmen kendisine de aşağıdaki cevap anahtarını hazırlamıştır.

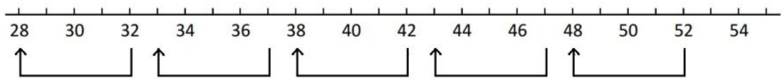
Mehmet Öğretmenin cevap anahtarı:

Aralık	Frekans
30-34	4
35-39	9
40-44	7
45-49	9
50-54	3



Mehmet öğretmen daha sonra sınıfı gezerek grupların çalışmalarını izlemiş ve her gruba yorumlar yapmıştır. Bu sırada gruplardan birinin farklı bir şekilde çalışmaya başladığını, aralıkları sondan başa doğru belirleyerek farklı frekans değerlerine ulaştıklarını görmüştür. Bu grubun yöntemi aşağıdaki gibidir.

Aralık	Frekans
28-32	1
33-37	8
38-42	11
43-47	8
48-52	4

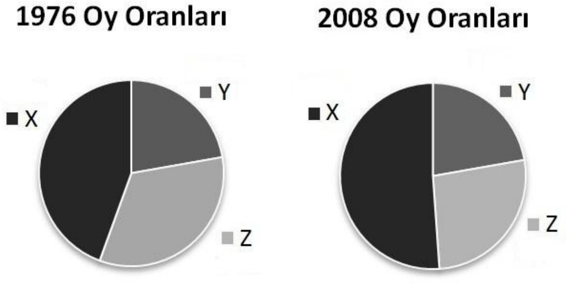


Bu durumda öğretmen siz olsanız ve bu grubun yöntemini 4 puan üzerinden değerlendirmeniz gerekirse aşağıdaki puanlardan hangisini verirsiniz? Nedenini açıklayınız?

a) Tamamen Doğrudur (4 Puan)
b) Büyük Oranda Doğrudur (3 puan)
c) Yarı Yarıya Doğrudur (2 puan)
d) Büyük Oranda Yanlıştır (1 puan)
e) Tamamen Yanlıştır (0 puan)

Neden bu şekilde düşündüğünüzü açıklayınız?

ITEM F.12. Fatma öğretmen öğrencilerini dört gruba ayırarak, bir Avrupa ülkesinde 1976 ve 2008 yıllarında yapılan seçimlerde üç partinin oy dağılımlarını gösteren aşağıdaki grafikleri incelemesini istemiştir. Ayrıca bu ülkede seçmen sayısının her geçen yıl arttığı bilgisini vererek 2008 yılındaki durumu, 1976 yılı ile nasıl karşılaştırdıklarında hangi sonuçlara ulaşılabileceğini sormuştur.



Fatma öğretmen sınıfı gezerek gruplardan aşağıdaki yorumları almıştır ve **kesinlikle yanlış yorum yapan** gruba müdahele ederek neden yanlış yaptıklarını açıklamıştır.

Grupların Yorumları:

I) Y partisi ile Z partisi arasındaki oy farkı 2008 yılında, 1976 yılına göre azalmıştır.
II) 2008 yılı seçmen sayısı bilinirse Y partisinin oyunu ne kadar arttırdığı bulunabilir.
III) 2008 yılında Z partisinin oy sayısı 1976 yılına göre azalmıştır.
IV) 2008 yılında X partisinin oy sayısı 1976 yılına göre artmıştır.

Bu durum için siz olsanız kesinlikle yanlış yorum yapan gruba hatalarını düzeltebilmeleri için (neden hata yaptıklarını da göz önüne alarak) nasıl bir yorum yapardınız?

APPENDIX E

MODEL FIT INFORMATION FOR MODEL I AND MODEL II USING MLR

MODEL FIT INFORMATION FOR MODEL I USING MLR

```
Number of Free Parameters                        68

Loglikelihood

          H0 Value                        -8449.108
          H0 Scaling Correction Factor       1.0279
            for MLR

Information Criteria

          Akaike (AIC)                    17034.216
          Bayesian (BIC)                  17339.585
          Sample-Size Adjusted BIC        17123.683
            (n* = (n + 2) / 24)

MODEL RESULTS
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 MKT_S    BY
    F1A                0.780      0.186      4.200      0.000
    F1B                1.188      0.331      3.592      0.000
    F1C                0.295      0.134      2.199      0.028
    F1D                0.386      0.119      3.249      0.001
    F2                 0.228      0.113      2.027      0.043
    F3                 0.379      0.138      2.741      0.006
    F4                 0.578      0.148      3.914      0.000
    F5A                0.892      0.230      3.873      0.000
    F5B                1.191      0.334      3.562      0.000
    F6                 0.486      0.163      2.981      0.003
    F7                 0.399      0.152      2.628      0.009
    F8                 0.287      0.137      2.089      0.037
    F9                 0.615      0.189      3.260      0.001
    F10                0.458      0.157      2.908      0.004
    F11                0.500      0.178      2.811      0.005
    F12                0.340      0.181      1.883      0.060

 Thresholds
    F1A$1             -0.377      0.091     -4.130      0.000
    F1B$1             -2.104      0.234     -9.006      0.000
    F1C$1              0.601      0.083      7.211      0.000
    F1C$2              0.746      0.085      8.733      0.000
    F1C$3              0.831      0.087      9.589      0.000
    F1C$4              3.945      0.279     14.159      0.000
    F1D$1             -0.138      0.081     -1.713      0.087
    F1D$2              0.806      0.090      8.965      0.000
    F1D$3              1.525      0.108     14.127      0.000
    F1D$4              4.945      0.451     10.954      0.000
    F2$1              -1.623      0.105    -15.434      0.000
    F2$2               0.024      0.079      0.299      0.765
    F2$3               1.520      0.103     14.735      0.000
```

F3$1	0.836	0.089	9.353	0.000
F3$2	0.888	0.091	9.786	0.000
F3$3	2.219	0.133	16.697	0.000
F3$4	2.887	0.176	16.368	0.000
F4$1	0.031	0.084	0.375	0.707
F4$2	0.084	0.084	0.994	0.320
F4$3	0.129	0.084	1.538	0.124
F4$4	1.133	0.101	11.218	0.000
F5A$1	-0.139	0.114	-1.213	0.225
F5A$2	-0.092	0.115	-0.799	0.424
F5A$3	1.181	0.157	7.534	0.000
F5A$4	1.425	0.170	8.368	0.000
F5B$1	0.720	0.150	4.784	0.000
F5B$2	0.814	0.158	5.164	0.000
F5B$3	3.163	0.349	9.059	0.000
F5B$4	4.348	0.478	9.102	0.000
F6$1	0.385	0.107	3.606	0.000
F6$2	1.188	0.127	9.320	0.000
F7$1	-2.211	0.168	-13.176	0.000
F7$2	-1.963	0.148	-13.252	0.000
F7$3	2.472	0.199	12.425	0.000
F7$4	2.645	0.214	12.373	0.000
F8$1	0.437	0.082	5.352	0.000
F9$1	0.386	0.112	3.459	0.001
F9$2	0.444	0.113	3.929	0.000
F9$3	0.680	0.117	5.806	0.000
F9$4	3.449	0.288	11.994	0.000
F10$1	0.663	0.088	7.522	0.000
F10$2	1.243	0.101	12.291	0.000
F10$3	1.564	0.109	14.301	0.000
F10$4	1.769	0.117	15.139	0.000
F11$1	-1.542	0.143	-10.792	0.000
F11$2	-0.690	0.122	-5.667	0.000
F11$3	-0.185	0.113	-1.627	0.104
F11$4	0.934	0.119	7.851	0.000
F12$1	-0.868	0.123	-7.038	0.000
F12$2	-0.765	0.121	-6.316	0.000
F12$3	1.600	0.140	11.418	0.000
F12$4	3.608	0.328	11.000	0.000
Variances				
MKT_S	1.000	0.000	999.000	999.000

STANDARDIZED MODEL RESULTS

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
MKT_S BY				
F1A	0.395	0.079	4.977	0.000
F1B	0.548	0.107	5.132	0.000
F1C	0.161	0.071	2.258	0.024
F1D	0.208	0.061	3.396	0.001
F2	0.125	0.061	2.059	0.039
F3	0.205	0.072	2.861	0.004
F4	0.304	0.070	4.311	0.000
F5A	0.441	0.092	4.808	0.000
F5B	0.549	0.108	5.099	0.000
F6	0.259	0.081	3.195	0.001
F7	0.215	0.078	2.755	0.006
F8	0.156	0.073	2.141	0.032
F9	0.321	0.088	3.635	0.000
F10	0.245	0.079	3.093	0.002
F11	0.266	0.088	3.025	0.002

F12	0.184	0.095	1.949	0.051
Thresholds				
F1A$1	-0.191	0.045	-4.209	0.000
F1B$1	-0.971	0.060	-16.306	0.000
F1C$1	0.327	0.045	7.251	0.000
F1C$2	0.406	0.046	8.795	0.000
F1C$3	0.452	0.047	9.665	0.000
F1C$4	2.147	0.154	13.965	0.000
F1D$1	-0.074	0.044	-1.709	0.088
F1D$2	0.434	0.047	9.227	0.000
F1D$3	0.822	0.056	14.807	0.000
F1D$4	2.667	0.243	10.993	0.000
F2$1	-0.888	0.057	-15.452	0.000
F2$2	0.013	0.043	0.299	0.765
F2$3	0.831	0.056	14.921	0.000
F3$1	0.451	0.047	9.585	0.000
F3$2	0.479	0.048	10.056	0.000
F3$3	1.198	0.070	17.176	0.000
F3$4	1.558	0.092	16.938	0.000
F4$1	0.017	0.044	0.376	0.707
F4$2	0.044	0.044	0.996	0.319
F4$3	0.068	0.044	1.541	0.123
F4$4	0.595	0.050	11.953	0.000
F5A$1	-0.069	0.057	-1.204	0.229
F5A$2	-0.045	0.057	-0.795	0.427
F5A$3	0.584	0.064	9.094	0.000
F5A$4	0.705	0.068	10.429	0.000
F5B$1	0.332	0.059	5.588	0.000
F5B$2	0.375	0.060	6.219	0.000
F5B$3	1.458	0.098	14.804	0.000
F5B$4	2.004	0.154	12.993	0.000
F6$1	0.205	0.056	3.653	0.000
F6$2	0.633	0.064	9.838	0.000
F7$1	-1.190	0.089	-13.379	0.000
F7$2	-1.057	0.080	-13.156	0.000
F7$3	1.331	0.099	13.447	0.000
F7$4	1.424	0.106	13.422	0.000
F8$1	0.238	0.044	5.375	0.000
F9$1	0.202	0.058	3.484	0.000
F9$2	0.232	0.058	3.970	0.000
F9$3	0.355	0.060	5.942	0.000
F9$4	1.801	0.145	12.460	0.000
F10$1	0.354	0.046	7.681	0.000
F10$2	0.665	0.052	12.884	0.000
F10$3	0.836	0.056	14.981	0.000
F10$4	0.946	0.059	15.900	0.000
F11$1	-0.819	0.073	-11.259	0.000
F11$2	-0.367	0.063	-5.851	0.000
F11$3	-0.098	0.060	-1.639	0.101
F11$4	0.496	0.064	7.748	0.000
F12$1	-0.471	0.065	-7.282	0.000
F12$2	-0.414	0.064	-6.515	0.000
F12$3	0.867	0.077	11.333	0.000
F12$4	1.955	0.174	11.241	0.000

MODEL FIT INFORMATION FOR MODEL II USING MLR

```
Number of Free Parameters                        69

Loglikelihood

          H0 Value                        -8441.154
          H0 Scaling Correction Factor       1.0274
            for MLR

Information Criteria

          Akaike (AIC)                    17020.309
          Bayesian (BIC)                  17330.169
          Sample-Size Adjusted BIC        17111.092
            (n* = (n + 2) / 24)


MODEL RESULTS

                                                    Two-Tailed
                     Estimate       S.E.  Est./S.E.    P-Value

 CK       BY
    F1A                1.021      0.239      4.266      0.000
    F1B                1.589      0.421      3.773      0.000
    F1C                0.324      0.166      1.956      0.050
    F1D                0.411      0.151      2.733      0.006
    F8                 0.331      0.156      2.127      0.033
    F10                0.520      0.170      3.068      0.002

 PCK      BY
    F2                 0.292      0.114      2.569      0.010
    F3                 0.301      0.137      2.193      0.028
    F4                 0.581      0.149      3.906      0.000
    F5A                1.065      0.218      4.892      0.000
    F5B                1.606      0.412      3.894      0.000
    F6                 0.433      0.159      2.725      0.006
    F7                 0.447      0.159      2.808      0.005
    F9                 0.715      0.212      3.367      0.001
    F11                0.596      0.214      2.783      0.005
    F12                0.388      0.190      2.045      0.041

 PCK      WITH
    CK                 0.561      0.120      4.666      0.000

 Thresholds
    F1A$1             -0.403      0.101     -3.982      0.000
    F1B$1             -2.373      0.335     -7.089      0.000
    F1C$1              0.604      0.084      7.168      0.000
    F1C$2              0.749      0.086      8.669      0.000
    F1C$3              0.835      0.088      9.526      0.000
    F1C$4              3.954      0.279     14.160      0.000
    F1D$1             -0.140      0.081     -1.725      0.085
    F1D$2              0.808      0.092      8.820      0.000
    F1D$3              1.530      0.111     13.797      0.000
    F1D$4              4.955      0.454     10.910      0.000
    F2$1              -1.634      0.106    -15.364      0.000
    F2$2               0.024      0.080      0.297      0.766
    F2$3               1.530      0.104     14.710      0.000
    F3$1               0.826      0.087      9.442      0.000
    F3$2               0.877      0.089      9.891      0.000
    F3$3               2.199      0.131     16.847      0.000
    F3$4               2.864      0.174     16.461      0.000
    F4$1               0.030      0.084      0.362      0.717
```

F4$2	0.083	0.084	0.979	0.327
F4$3	0.128	0.084	1.520	0.129
F4$4	1.132	0.100	11.277	0.000
F5A$1	-0.140	0.120	-1.168	0.243
F5A$2	-0.091	0.121	-0.751	0.453
F5A$3	1.248	0.165	7.553	0.000
F5A$4	1.501	0.178	8.417	0.000
F5B$1	0.820	0.183	4.476	0.000
F5B$2	0.928	0.193	4.799	0.000
F5B$3	3.539	0.455	7.781	0.000
F5B$4	4.805	0.593	8.100	0.000
F6$1	0.379	0.106	3.587	0.000
F6$2	1.173	0.126	9.335	0.000
F7$1	-2.227	0.170	-13.113	0.000
F7$2	-1.978	0.150	-13.211	0.000
F7$3	2.487	0.203	12.251	0.000
F7$4	2.661	0.217	12.258	0.000
F8$1	0.440	0.083	5.330	0.000
F9$1	0.394	0.115	3.442	0.001
F9$2	0.453	0.116	3.907	0.000
F9$3	0.695	0.122	5.720	0.000
F9$4	3.501	0.298	11.768	0.000
F10$1	0.673	0.090	7.519	0.000
F10$2	1.260	0.103	12.194	0.000
F10$3	1.583	0.113	14.041	0.000
F10$4	1.790	0.120	14.869	0.000
F11$1	-1.573	0.152	-10.360	0.000
F11$2	-0.707	0.128	-5.535	0.000
F11$3	-0.194	0.117	-1.648	0.099
F11$4	0.946	0.121	7.786	0.000
F12$1	-0.878	0.125	-6.995	0.000
F12$2	-0.773	0.123	-6.276	0.000
F12$3	1.608	0.141	11.367	0.000
F12$4	3.622	0.334	10.860	0.000
Variances				
CK	1.000	0.000	999.000	999.000
PCK	1.000	0.000	999.000	999.000

STANDARDIZED MODEL RESULTS

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
CK BY				
F1A	0.491	0.087	5.618	0.000
F1B	0.659	0.099	6.669	0.000
F1C	0.176	0.087	2.019	0.044
F1D	0.221	0.077	2.874	0.004
F8	0.180	0.082	2.198	0.028
F10	0.276	0.083	3.321	0.001
PCK BY				
F2	0.159	0.060	2.635	0.008
F3	0.164	0.073	2.253	0.024
F4	0.305	0.071	4.306	0.000
F5A	0.506	0.077	6.580	0.000
F5B	0.663	0.095	6.947	0.000
F6	0.232	0.081	2.880	0.004
F7	0.239	0.080	2.979	0.003
F9	0.367	0.094	3.890	0.000
F11	0.312	0.101	3.084	0.002
F12	0.209	0.098	2.139	0.032

```
PCK      WITH
   CK                 0.561      0.120      4.666      0.000
Thresholds
   F1A$1             -0.194      0.046     -4.183      0.000
   F1B$1             -0.984      0.060    -16.284      0.000
   F1C$1              0.328      0.045      7.246      0.000
   F1C$2              0.406      0.046      8.791      0.000
   F1C$3              0.453      0.047      9.670      0.000
   F1C$4              2.146      0.154     13.969      0.000
   F1D$1             -0.075      0.044     -1.721      0.085
   F1D$2              0.434      0.047      9.196      0.000
   F1D$3              0.823      0.056     14.795      0.000
   F1D$4              2.664      0.242     11.016      0.000
   F2$1              -0.889      0.057    -15.487      0.000
   F2$2               0.013      0.043      0.297      0.766
   F2$3               0.833      0.056     14.959      0.000
   F3$1               0.449      0.047      9.586      0.000
   F3$2               0.477      0.047     10.064      0.000
   F3$3               1.196      0.070     17.096      0.000
   F3$4               1.557      0.092     16.849      0.000
   F4$1               0.016      0.044      0.362      0.717
   F4$2               0.043      0.044      0.981      0.327
   F4$3               0.067      0.044      1.523      0.128
   F4$4               0.595      0.050     11.960      0.000
   F5A$1             -0.067      0.058     -1.158      0.247
   F5A$2             -0.043      0.058     -0.746      0.456
   F5A$3              0.593      0.064      9.223      0.000
   F5A$4              0.714      0.067     10.600      0.000
   F5B$1              0.338      0.060      5.607      0.000
   F5B$2              0.383      0.061      6.253      0.000
   F5B$3              1.461      0.096     15.151      0.000
   F5B$4              1.983      0.148     13.435      0.000
   F6$1               0.203      0.056      3.626      0.000
   F6$2               0.629      0.064      9.783      0.000
   F7$1              -1.192      0.089    -13.429      0.000
   F7$2              -1.059      0.080    -13.223      0.000
   F7$3               1.332      0.099     13.478      0.000
   F7$4               1.424      0.106     13.478      0.000
   F8$1               0.238      0.044      5.372      0.000
   F9$1               0.202      0.058      3.476      0.001
   F9$2               0.232      0.059      3.961      0.000
   F9$3               0.357      0.060      5.926      0.000
   F9$4               1.796      0.143     12.569      0.000
   F10$1              0.357      0.046      7.716      0.000
   F10$2              0.668      0.052     12.955      0.000
   F10$3              0.839      0.056     15.035      0.000
   F10$4              0.948      0.059     15.982      0.000
   F11$1             -0.824      0.073    -11.313      0.000
   F11$2             -0.370      0.063     -5.844      0.000
   F11$3             -0.101      0.061     -1.670      0.095
   F11$4              0.495      0.064      7.723      0.000
   F12$1             -0.473      0.065     -7.299      0.000
   F12$2             -0.417      0.064     -6.524      0.000
   F12$3              0.867      0.076     11.362      0.000
   F12$4              1.953      0.173     11.285      0.000
```

APPENDIX F

COMPLETE LIST OF RESULTS FOR ITEM F.1C

Table F.1.

Answer given by preservice teacher	N	%
Missing	148	22,46
0,03	1	0,15
2,06	1	0,15
6,1	1	0,15
7	2	0,30
9,09	1	0,15
13	2	0,30
13,4	1	0,15
13,5	1	0,15
13,92	1	0,15
14	6	0,91
14,03	3	0,46
14,5	12	1,82
14,6	2	0,30
14,62	1	0,15
14,7	8	1,21
14,75	5	0,76
14,8	9	1,37
14,81	**108**	**16,39**
14,82	1	0,15
14,83	1	0,15
14,85	5	0,76
14,86	1	0,15
14,87	1	0,15
14,88	2	0,30
14,89	1	0,15
14,9	52	7,89
14,905	1	0,15
14,91	2	0,30

Note: continued on next page.

Table F.1. (continued)

Answer given by preservice teacher	N	%
14,92	4	0,61
14,925	1	0,15
14,93	19	2,88
14,935	1	0,15
14,95	16	2,43
14,96	21	3,19
14,965	1	0,15
14,97	10	1,52
14,975	3	0,46
14,98	53	8,04
14,99	13	1,97
15	**133**	**20,18**
15,01	1	0,15
15,06	1	0,15
15,6	1	0,15
16,1	1	0,15
Total	659	100,00

APPENDIX G

ORIGINAL SAMPLE ANSWERS FOR ITEM F.1C

PART 1

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

14,81 15 14,93

15
14,91
14,98

Tahmin Ettiğiniz Değer: 14,81

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız: Aritmatik ortalama daha gerçekçi bir sonuç verdiği için

Figure G.1.

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: 14,81

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız:
Eldeki verilerin ortalaması olduğu için

Figure G.2.

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: 14,81

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız: Her zaman verinin ortalaması güvenilirdir.

Figure G.3.

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: ..14,81

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız:

Günkü ortalamayı kullanarak tüm verileri dikkate almış olurum.

Figure G.4.

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: 14,81

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız: Ortalama iyi tahmin edicidir.

Figure G.5.

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: 14,81

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız: Aritmetik ortalamayı kulladım çünkü bütün değerleri temsil eden en orta noktadır.

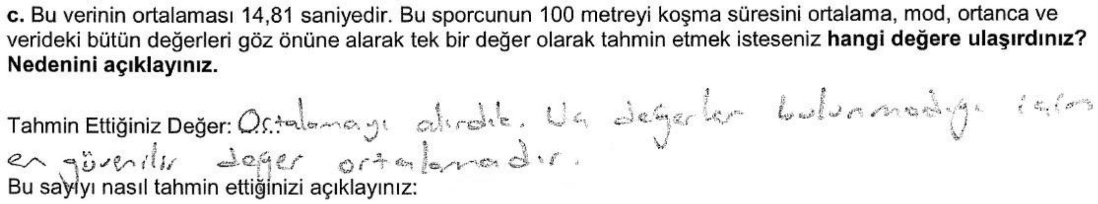
Figure G.6.

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: 14,81

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız: Ortalama zaten en geneldir. Yani hepsinin en genellenmiş hali ortalamadır.

Figure G.7.

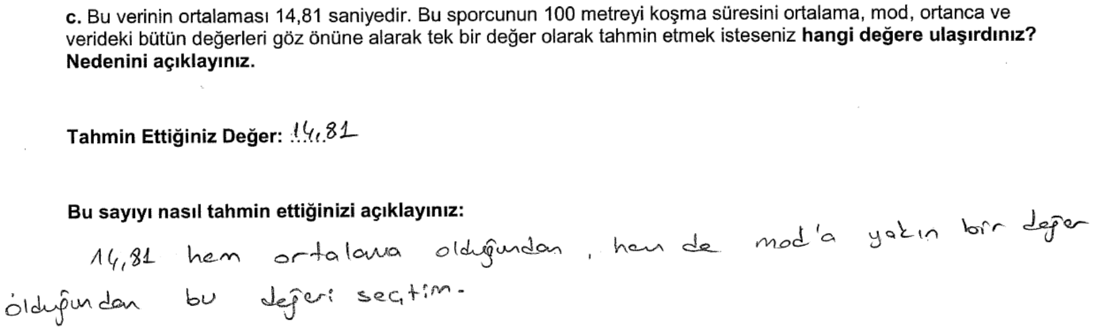


c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: Ortalamayı alırdık. Uç değerler bulunmadığı için en güvenilir değer ortalamadır.

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız:

Figure G.8.



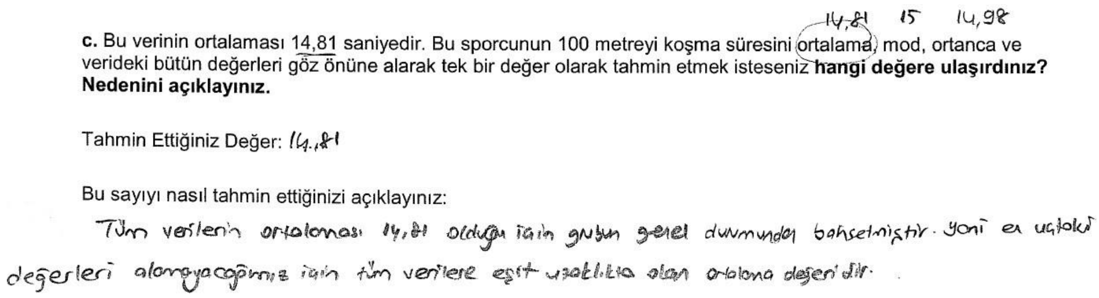
c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: 14,81

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız:

14,81 hem ortalama olduğundan, hem de mod'a yakın bir değer olduğundan bu değeri seçtim.

Figure G.9.



14,81 15 14,98

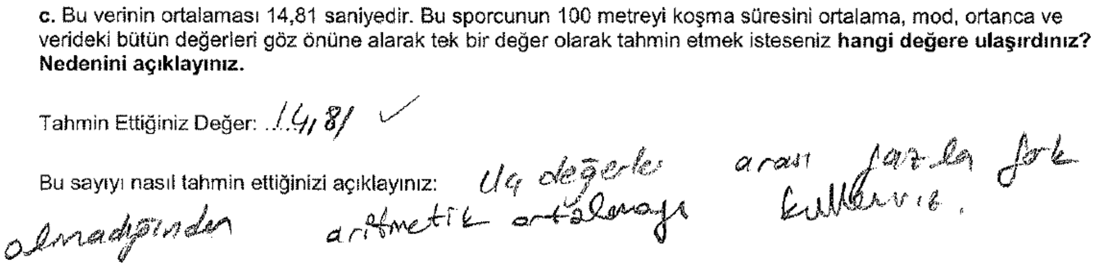
c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: 14,81

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız:

Tüm verilerin ortalaması 14,81 olduğu için grubun genel durumundan bahsetmiştir. Yani en uçtaki değerleri alamayacağımız için tüm verilere eşit uzaklıkta olan ortalama değeridir.

Figure G.10.



c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: 14,81 ✓

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız: Uç değerler arası fazla fark olmadığından aritmetik ortalamayı kullanırız.

Figure G.11.

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: ..14,81

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız:

mod=medyan ve ortalama birbirine çok yakın, dolayısıyla tüm verilerden etkilenen ortalamayı kullanırım.

Figure G.12.

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: .14,81

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız: Ortalama değer doğru alınırsa diğer değerlere de yakın olacaktır.

Figure G.13.

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: ...14,81

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız:

Değerler birbirine yakın olduğu için ortalamayı kullandım.

Figure G.14.

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Herkesin hata yapma payını göz önüne alırsak ortalama en uygun olur

Tahmin Ettiğiniz Değer: ..14,81

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız:

Ortalama.

Figure G.15.

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: 14,81

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız: Ortalama mod ve medyana göre daha sağlamdır

Figure G.16.

PART 2

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: 14,03

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız: en küçük, en büyük $\frac{13+15}{2} = 14$

Figure G.17.

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: ..14. değerine

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız: çünkü sayılar 13 ile 15 arasında değişiyor en kısa yoldan böyle düşünürdüm. 13 ile 15 arası ve birçok sayı varsa 14 dedim

Figure G.18.

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

13 14 – 15 1/15

Tahmin Ettiğiniz Değer: ...14,5

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız:

Çünkü 14,9 a yaklaşık tam kısmı 14 olan sayılar ve 15 olan sayılardanda 3 tane var
13 dende 1 tane var
14,5'a daha yakın

Figure G.19.

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: ..14,90

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız:

Genel dağılımı 14,90'nın üzerinde ama daha düşük bir değer var 13 bu değerde düşmesek 14.90 olabilir diye düşündüm.

Figure G.20.

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: 14,80

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız: Veriler genellikle 15 saniye değerinde yoğunlaşmaktadır. Ancak 13 değeri verinin ortalamasını 14'e doğru çekmektedir. 15'teki değerler de 3 tane olduğu için 15'ten az bir değer olması gerekir.

Figure G.21.

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: ..14,80

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız: Ölçümler genelde 14,95 – 15,05 arası Ancak 13 verisi bu ortalamayı düşürebilir.

Figure G.22.

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: ..14, 50

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız: 13 ile 15 arasa fakat 15 daha yakın bir sayı tahmin etmem gerekirdi..

Figure G.23.

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: .14.,80

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız:
Sayılar 13 ve 15 arasında ve mod 15 olduğu için değer 15'e daha yakın olmalı.

Figure G.24.

PART 3

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: 14,98

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız: 13 gibi bir uç değer olduğu için ortalama bundan etkilenecektir. Bundan dolayı ortancayı tercih ederdim. Moddan daha güvenilir olduğu için.

Figure G.25.

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: ...14,9 - 15 arası bir şey.

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız: 13'ü hariç bırakıp diğerlerini toplar ve 20'e bölerek bulurum. 13'ün olasılığını pek düşünmüyorum

Figure G.26.

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

Tahmin Ettiğiniz Değer: 14.99

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız: ortanca ve mod değerlerinin ortalamasını aldım. aynı zamanda verilerin ortalaması 14.81 sn ancak bu değer bazı uç değerler yüzünden düşmüş 13 bulunan değer de hata yapılmış olabilir.

Figure G.27.

c. Bu verinin ortalaması 14,81 saniyedir. Bu sporcunun 100 metreyi koşma süresini ortalama, mod, ortanca ve verideki bütün değerleri göz önüne alarak tek bir değer olarak tahmin etmek isteseniz **hangi değere ulaşırdınız? Nedenini açıklayınız.**

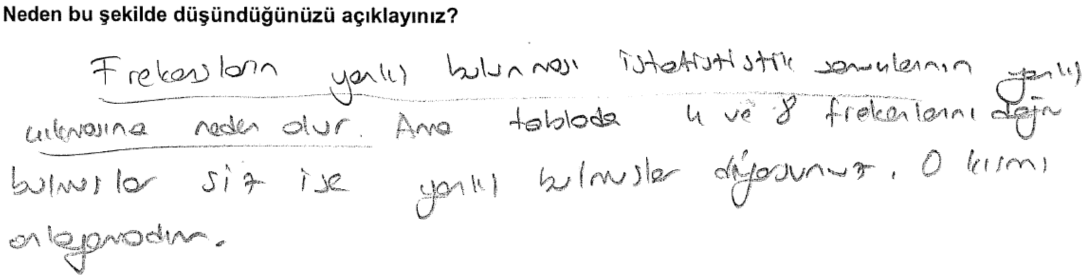
Tahmin Ettiğiniz Değer: ...14,98 veya 15 derdim. — 14,99

Bu sayıyı nasıl tahmin ettiğinizi açıklayınız: 13 ü çok önemsemezdim. Çünkü çok açıkça görülüyor ki diğerlerinden çok düşük gerçekliği tartışılır. diğerleri arasında ortalama olanı seçerdim.

Figure G.28.

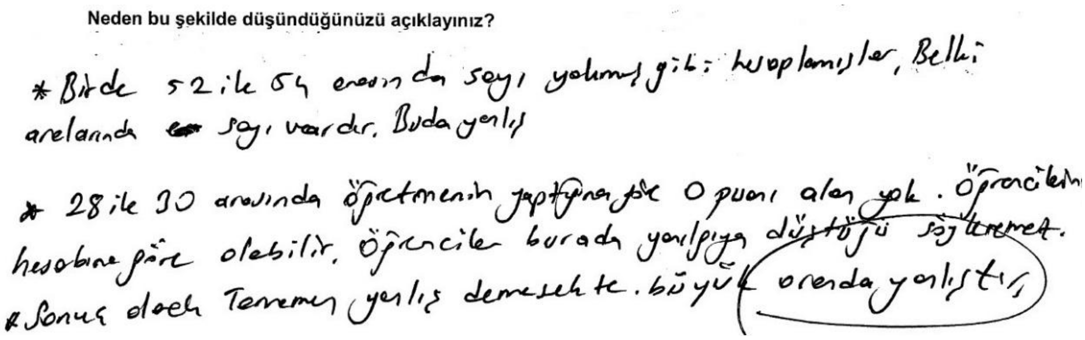
APPENDIX H

ORIGINAL SAMPLE ANSWERS FOR PCK-2


Neden bu şekilde düşündüğünüzü açıklayınız?

Frekansların yanlış bulunması istatistiksel sonuçların yanlış çıkmasına neden olur. Ama tabloda 4 ve 8 frekanslarını doğru bulmuşlar siz ise yanlış bulmuşlar diyorsunuz. O kısmı anlayamadım.

Figure H.1.


Neden bu şekilde düşündüğünüzü açıklayınız?

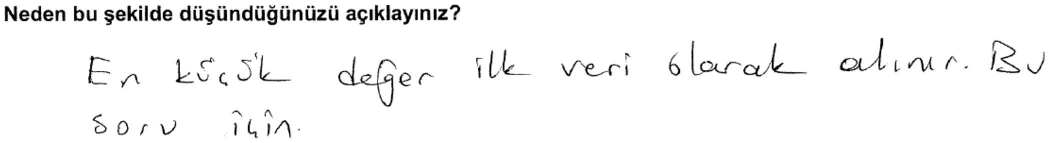
* Birde 52 ile 54 arasında sayı yokmuş gibi hesaplamışlar, Belki aralarında sayı vardır. Buda yanlış
* 28 ile 30 arasında öğretmenin yaptığına göre 0 puanı alan yok. Öğrencilerin hesabına göre olabilir. Öğrenciler burada yanlışa düştüğü söylenemez.
* Sonuç olarak Tamamen yanlış demesek te. büyük oranda yanlıştır.

Figure H.2.


Neden bu şekilde düşündüğünüzü açıklayınız?

54'ten başlamaları gerekirdi. Öyle yapsalardı doğru sonuca ulaşırlardı.

Figure H.3.


Neden bu şekilde düşündüğünüzü açıklayınız?

En küçük değer ilk veri olarak alınır. Bu soru için.

Figure H.4.

Neden bu şekilde düşündüğünüzü açıklayınız?

En son ulaştığı 28-32. Ama verilen ağırlıklar içerisinde 28,29 değerleri yoktur. Yanlış gruplama yapılmış olur.

Figure H.5.

Neden bu şekilde düşündüğünüzü açıklayınız?

Çünkü veriler küçükten büyüğe doğru dizilir ve işlem buna göre düzenlenir.

Figure H.6.

Neden bu şekilde düşündüğünüzü açıklayınız?

Öğrencide kavram yanılgısı bulunmakta. Konuyu tam bilmemektedir. Cevap tamamen yanlıştır.

Figure H.7.

Neden bu şekilde düşündüğünüzü açıklayınız?

Yığmalı frekansa baktığımızda Mehmet öğretmende 4-10-... şeklinde giderken öğrencide daha farklı olacaktır.

Figure H.8.

Neden bu şekilde düşündüğünüzü açıklayınız?

Çünkü biz böyle öğrendik. Ben de bunu sorguladım ama alt gruptan başlanılarak alınabileceği söylendi.

Figure H.9.

Neden bu şekilde düşündüğünüzü açıklayınız?

Çünkü grup ağırlıkları 30'dan başlıyor bu yüzden 30'dan başlanmalı. Çünkü 30'dan başlanmazsa eğer dışında eleman kalabilir. Yukarıda tesadüfi olarak dışarıda eleman kalmamıştır. Ama bu her zaman böyle olacağı anlamına gelmez.

Figure H.10.

Neden bu şekilde düşündüğünüzü açıklayınız?

Çünkü 28 ve 30 arasında hiç bir veri olmayıp 52 ile 54 arasında veri olabilir. Bu şekilde gruplandırma veri kaybına neden olabilir.

Figure H.11.

Neden bu şekilde düşündüğünüzü açıklayınız?

Burada oluşturulan tabloda hatalar oluşmuştur.
Bu hatadan sonra çizilen grafikte de hatalar olacaktır.
Daha sonra yapılan yorumun da büyük oranda yanlış olması beklenir. Çünkü ilerleme hataların üzerinde yapılıyor. Ve hata bir sonraki adımda sürekli daha fazla büyüyor.

Figure H.12.

Neden bu şekilde düşündüğünüzü açıklayınız?

Grup öğretmenin bulduğu sonuçların çoğunu bulamamıştır. Buldukları son değerler öğretmenin sonucuna yakın olduğu için tamamen yanlış değildir.

Figure H.13.

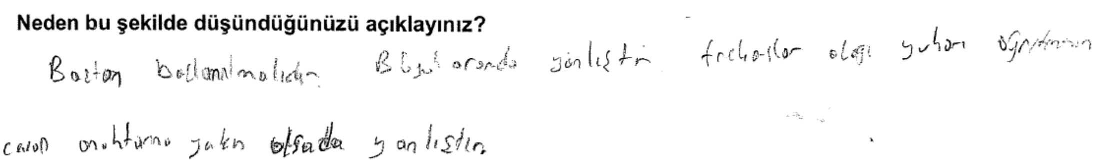


Figure H.14.

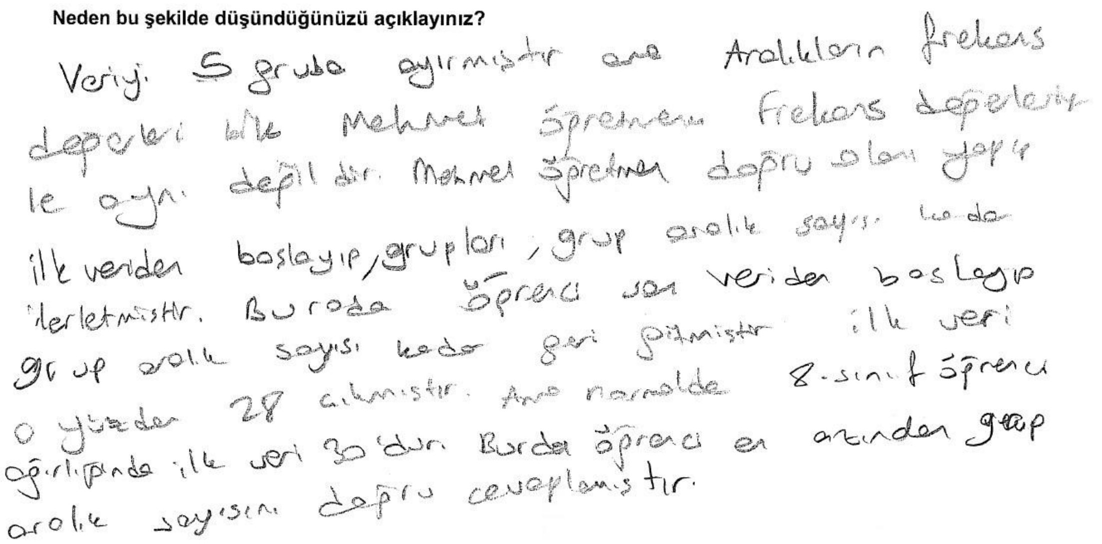


Figure H.15.

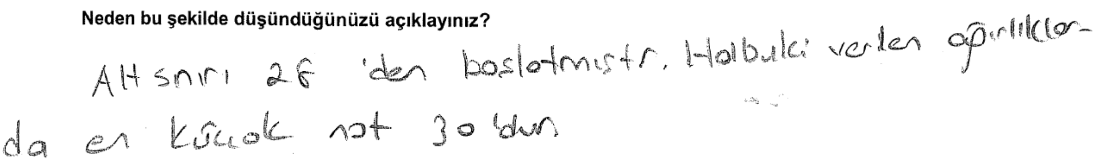


Figure H.16.

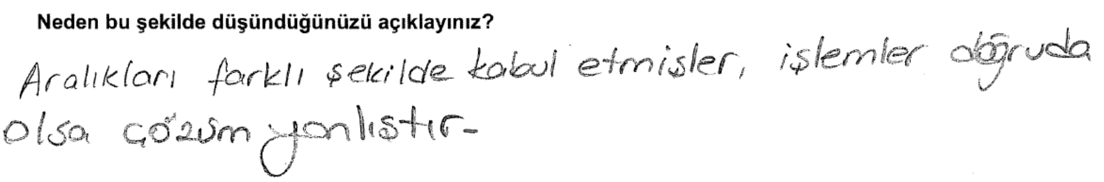
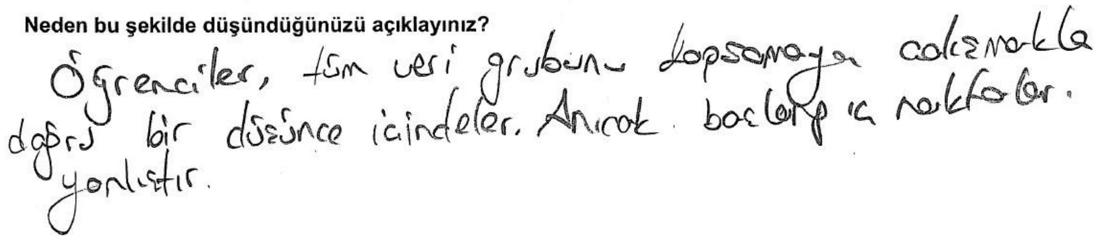
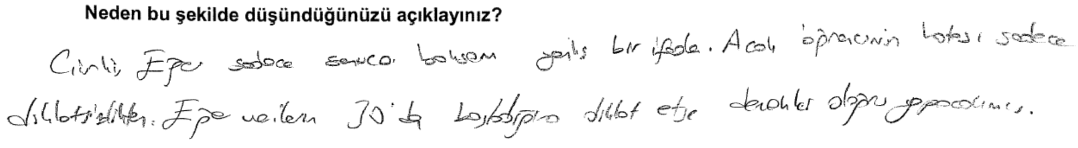


Figure H.17.


Neden bu şekilde düşündüğünüzü açıklayınız?

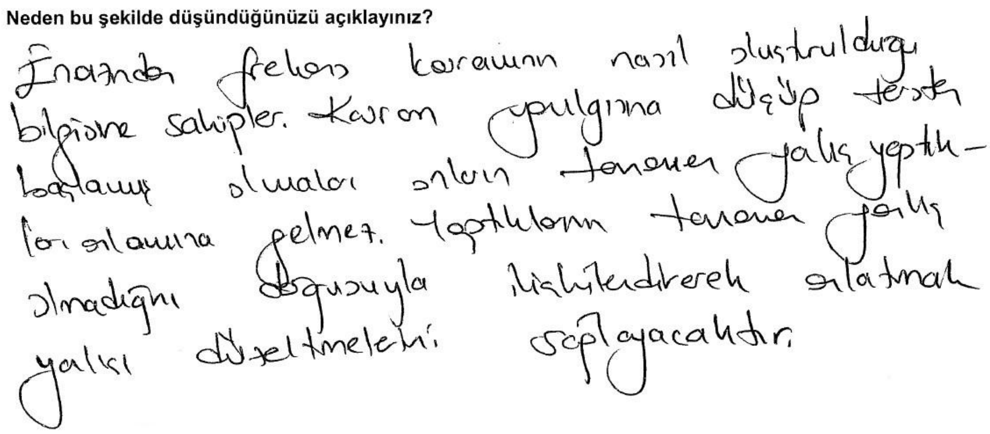
Öğrenciler, tüm veri grubunu kapsamaya çalışmakla doğru bir düşünce içindeler. Ancak başlangıç noktaları yanlıştır.

Figure H.18.


Neden bu şekilde düşündüğünüzü açıklayınız?

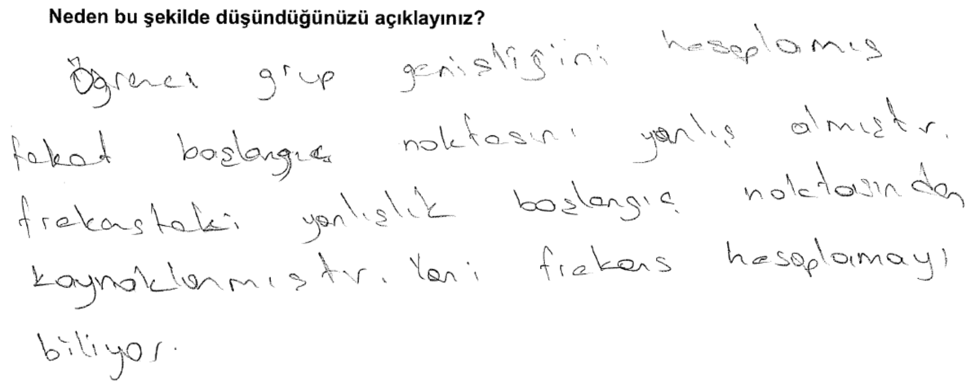
Çünkü, Ege sadece sonuca bakarak yanlış bir ifade. Ancak öğrencinin hatası sadece dikkatsizlikten. Ege verilen 30'da başlangıca dikkat etse doğru olarak yapacaktır.

Figure H.19.


Neden bu şekilde düşündüğünüzü açıklayınız?

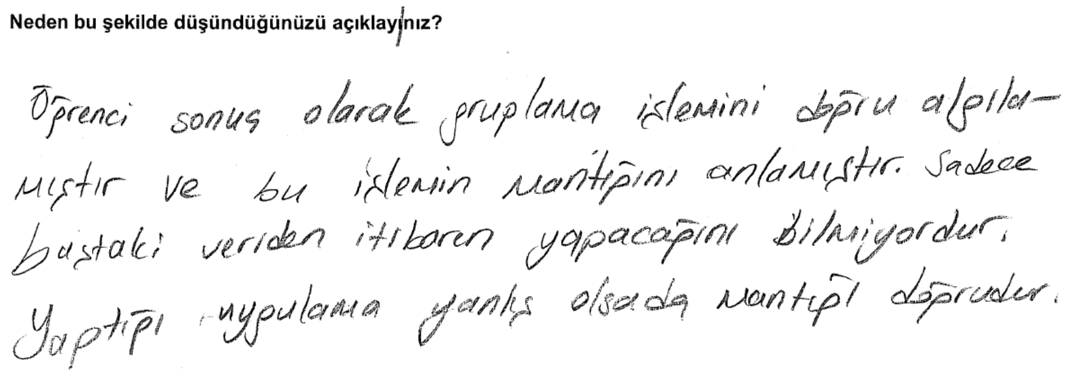
Öğrenciler frekans kavramının nasıl oluşturulduğu bilgisine sahipler. Kavram yanılgısına düşüp tersten başlamış olmaları onların tanımı yanlış yaptıkları anlamına gelmez. Yaptıklarının tanımı yanlış olmadığını doğrusuyla ilişkilendirerek anlatmak yanlışı düzeltmelerini sağlayacaktır.

Figure H.20.


Neden bu şekilde düşündüğünüzü açıklayınız?

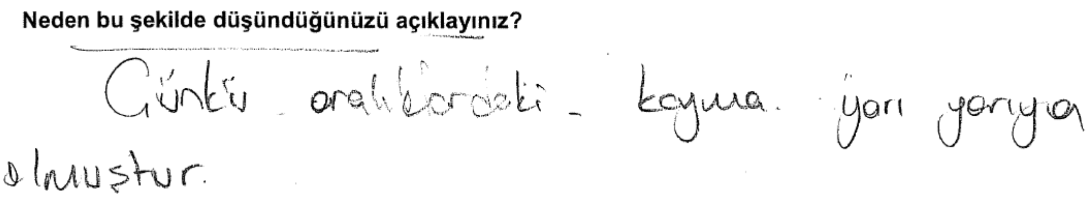
Öğrenci grup genişliğini hesaplamış fakat başlangıç noktasını yanlış almıştır. Frekanstaki yanlışlık başlangıç noktasından kaynaklanmıştır. Yani frekans hesaplamayı biliyor.

Figure H.21.


Neden bu şekilde düşündüğünüzü açıklayınız?

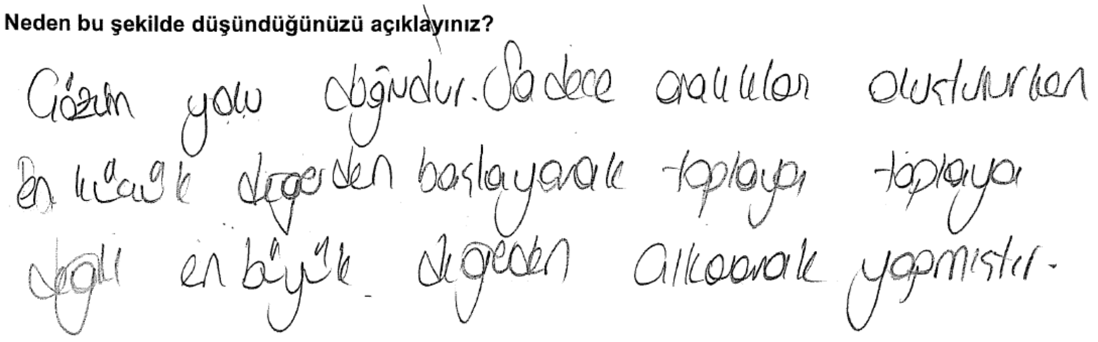
Öğrenci sonuç olarak gruplama işlemini doğru algılamıştır ve bu işlemin mantığını anlamıştır. Sadece baştaki veriden itibaren yapacağını bilmiyordur. Yaptığı uygulama yanlış olsada mantığı doğrudur.

Figure H.22.


Neden bu şekilde düşündüğünüzü açıklayınız?

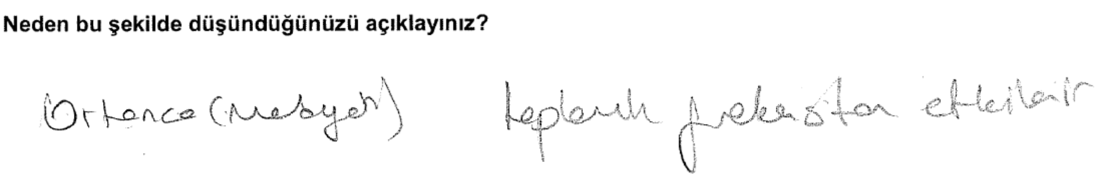
Çünkü aralıklardaki koyma yarı yarıya olmuştur.

Figure H.23.


Neden bu şekilde düşündüğünüzü açıklayınız?

Çözüm yolu doğrudur. Sadece aralıklar oluştururken en küçük değerden başlayarak toplaya toplaya değil en büyük değerden alarak yapmıştır.

Figure H.24.


Neden bu şekilde düşündüğünüzü açıklayınız?

Ortanca (Medyan) hesaplanırken frekanstan etkilenir

Figure H.25.

Neden bu şekilde düşündüğünüzü açıklayınız?

Sadece bir kaç tane veri grup dışında bulunuş tamamen yanlış sayılamaz

Figure H.26.

Neden bu şekilde düşündüğünüzü açıklayınız?

Çünkü üst yöntemde de alt yöntemde ihmal edilen sayılar mutlaka var sadece ve genellikle baştan başladığı için sonuçlar farklı çıkabilir.

Figure H.27.

Neden bu şekilde düşündüğünüzü açıklayınız?

Aralıkların genişliği değişmesede, aralıkların alt değeri ve üst değeri değiştiği için, yani farklı değerler alındığı için frekanslar değişmiştir. Fakat yöntem büyük oranda doğrudur.

Figure H.28.

Neden bu şekilde düşündüğünüzü açıklayınız?

Sadece başlangıç değeri farklı almışlar.

Ancak değerler bundan etkilenecektir

Figure H.29.

Neden bu şekilde düşündüğünüzü açıklayınız?

doğrudur ama kullanışsız dır.

Figure H.30.

Neden bu şekilde düşündüğünüzü açıklayınız?

Çünkü önemli olan belirtediği aralıklarda öğrenci olup olmamasıdır. Aralık değerlerinin farklı olması, sınıflandırmayı farklı yaptığını gösterir.

Figure H.31.

Neden bu şekilde düşündüğünüzü açıklayınız?

Çünkü İstatistik konusunda %100 bir sonuç yoktur. Verileri kapsayacak şekilde aralık alındığı için bu sonuçta doğrudan önemli olan seçilecek aralığın içinde verilerin bulunmasıdır.

Figure H.32.

APPENDIX I

ORIGINAL SAMPLE ANSWERS FOR PCK-3

Neden bu şekilde düşündüğünüzü açıklayınız?

ikisinde gösterebilmem için grafiğin normal dağılıma
yakın olması gerekir.

Figure I.1.

Neden bu şekilde düşündüğünüzü açıklayınız?

- Ortanca için simetrik olmasına dikkat eder ve (P) dağılımını kullanırım.
- Ortalama için boş veri grubu olmamasını ister ve (P) dağılımını kullanırım

Figure I.2.

Neden bu şekilde düşündüğünüzü açıklayınız?

Ortalama için P seçilmesi oranların birbirine
yakın degerler taşımasından dolayıdır.

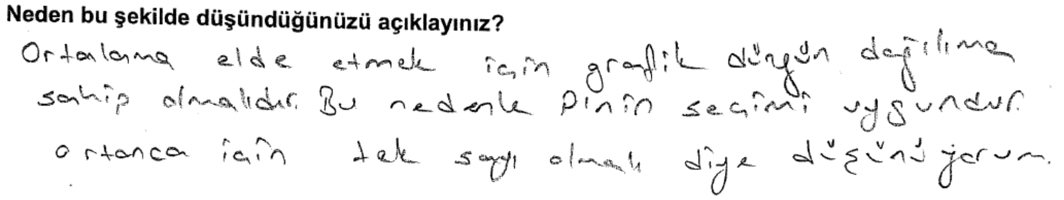
Ortanca için R seçilirse küçükten büyüğe kolay
sıralanabilir.

Figure I.3.

Neden bu şekilde düşündüğünüzü açıklayınız?

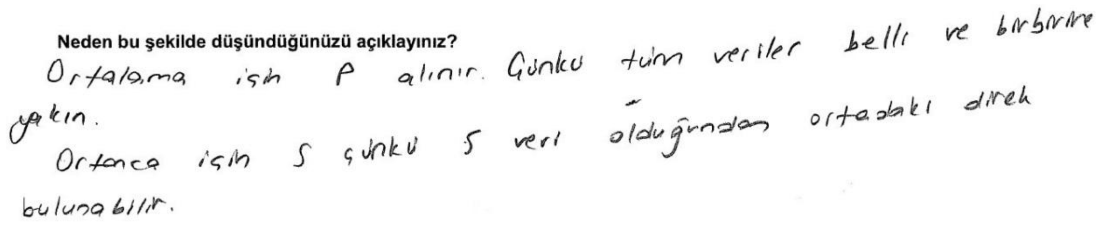
Unutmuşuz hocam ya........
Normal dağılıma yakın olduğu için P'yi seçtim
Ortanca için S'yi seçtim. Çünkü 4've 5 arasında olmasını
göstermek için.

Figure I.4.


Neden bu şekilde düşündüğünüzü açıklayınız?

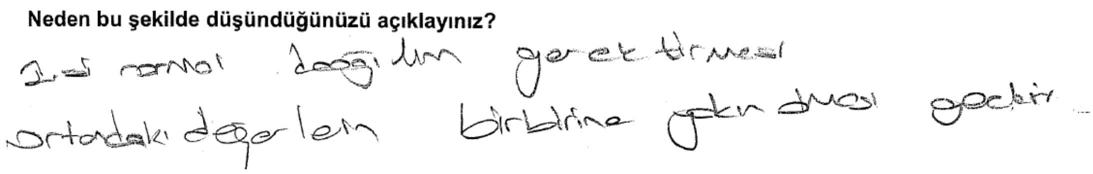
Ortalama elde etmek için grafik düzgün dağılıma sahip olmalıdır. Bu nedenle P'nin seçimi uygundur.
ortanca için tek sayı olmalı diye düşünüyorum.

Figure I.5.


Neden bu şekilde düşündüğünüzü açıklayınız?

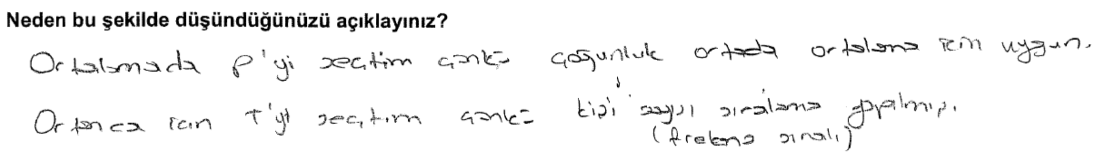
Ortalama için P alınır. Çünkü tüm veriler belli ve birbirine yakın.
Ortanca için S çünkü S veri olduğundan ortadaki direk bulunabilir.

Figure I.6.


Neden bu şekilde düşündüğünüzü açıklayınız?

1.si normal dağılım gerektirmesi
ortadaki değerlerin birbirine yakın olması gerekir

Figure I.7.


Neden bu şekilde düşündüğünüzü açıklayınız?

Ortalamada P'yi seçtim çünkü çoğunluk ortada ortalama için uygun.
Ortanca için T'yi seçtim çünkü kişi sayısı sıralama yapılmış. (frekans sıralı)

Figure I.8.

Neden bu şekilde düşündüğünüzü açıklayınız?

Ortanca için en çok tekrar eden elemanı yani frekansı en yüksek olanı göstermemiz gerekir. Bunun için T'yi seçtim.

Ortalama için de birbirine yakın değerli grafik gösterirsem daha rahat anlayabileceklerini düşündüm öğrencilerin.

Figure I.9.

Neden bu şekilde düşündüğünüzü açıklayınız?

Ortalama için; Ortalamayı hesapladım ve değer öğrenci sayısının çok olduğu yere denk geldi.

Ortanca; Veriler sıralandığında ortadaki değerin yeri daha kolay bulunuyor.

Figure I.10.

Neden bu şekilde düşündüğünüzü açıklayınız?

Ortalamada sayıların birbirine yakın olması gerekir.

Ortacada en ortadaki değer olduğundan bu değerler olmalı

Figure I.11.

Neden bu şekilde düşündüğünüzü açıklayınız?

Frekansların yakın olduğu durumlarda ortalama örneğini, düzenli azalış-artış gösteren grafiklerde ortanca örneğinin uygun olduğunu düşündüm.

Figure I.12.

Neden bu şekilde düşündüğünüzü açıklayınız?

→ Ortalama için P'yi seçerim. Çocuk grafikten önce görsel olarak tahmin eder. Sonra doğruya nasıl ulaşacağını buldururum.

→ Ortanca için S'yi seçerim. 2 tane olması ilginç gelir ve tahmin isterim. Oradan doğrusunu buldururum.

Figure I.13.

Neden bu şekilde düşündüğünüzü açıklayınız?

Ortalama uç değerlerden etkilendiği için R
Ortanca çarpık dağılımlarda daha çok fikir verdiği için T

Figure I.14.

Neden bu şekilde düşündüğünüzü açıklayınız?

Ortancayı anlatırken sıkıntı olur. Çünkü R, S ve T de ikitane mod vardır.

Figure I.15.

Neden bu şekilde düşündüğünüzü açıklayınız?

Çünkü hem ortalamanın kaç çocuk olduğunu, hemde ortanca terimi bulmak P grafiğinde daha nettir.

Figure I.16.

Neden bu şekilde düşündüğünüzü açıklayınız?

P de frekansı en yüksek olan ortancadır.

Figure I.17.

Neden bu şekilde düşündüğünüzü açıklayınız?

Ortalama bulurken aynı frekans bulunmamasına dikkat ettim
Ortancada'da frekans değer çeşidinin tek sayı (7) olmasına dikkat ettim.

Figure I.18.

Neden bu şekilde düşündüğünüzü açıklayınız?

R-) Normal dağılım olduğu için ortanca kullanılır.

R→. Sağa çarpık bir grafik çıkar. Ortalama kullanılır.

Figure I.19.

Neden bu şekilde düşündüğünüzü açıklayınız? Ortalamanın daha uçuk değerler gösteren grafikte daha iyi anlaşılabilir.
ortancanın bütün frekansları farklı olması daha ayırt edici olabilir.

Figure I.20.

Neden bu şekilde düşündüğünüzü açıklayınız?

Aradaki kopukluk olduğu için verilen verilerden bunu öğretmek daha kalıcı olur

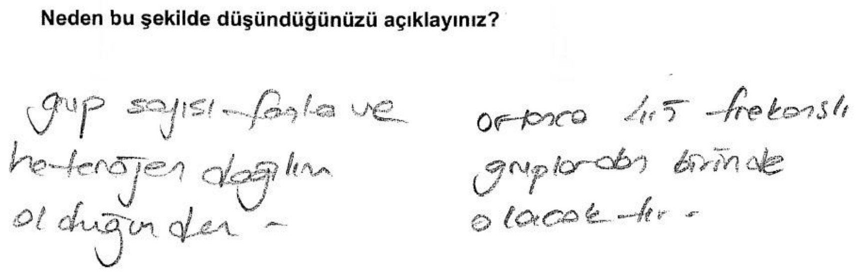
Figure I.21.

Neden bu şekilde düşündüğünüzü açıklayınız?

R deki grafikte veriler dengesiz dağılmıştır, bu yüzden ortalama kullanılabilir.

S de veriler düzgün dağıldığı için ortanca değer seçilebilir

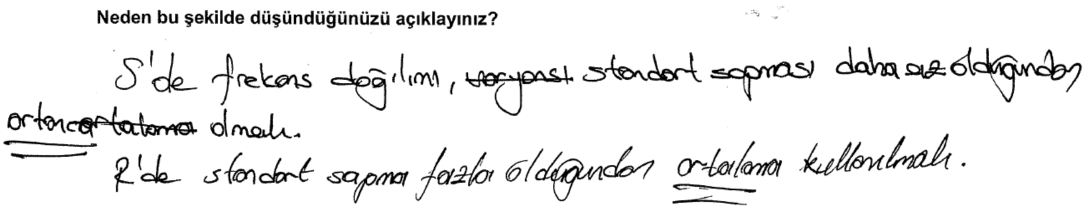
Figure I.22.



Neden bu şekilde düşündüğünüzü açıklayınız?

grup sayısı fazla ve heterojen dağılım olduğundan – ortanca 4,5 frekanslı grupların birinde olacaktır.

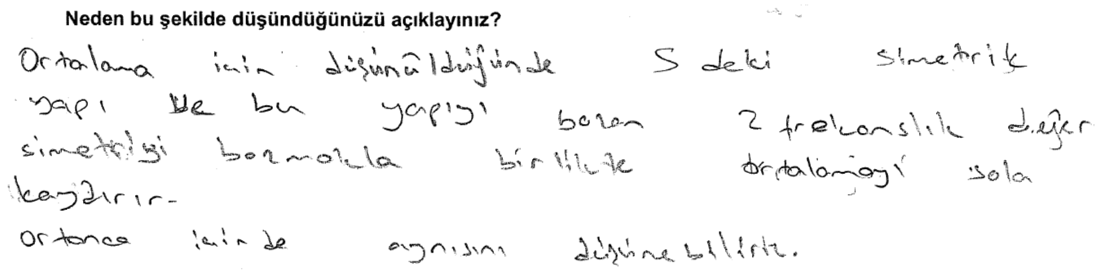
Figure I.23.



Neden bu şekilde düşündüğünüzü açıklayınız?

S'de frekans dağılımı, ~~varyansı~~ standart sapması daha ~~az~~ olduğundan ortanca ~~ortalama~~ olmalı.
R'de standart sapma fazla olduğundan ortalama kullanılmalı.

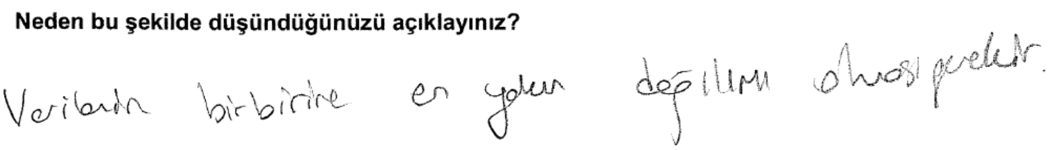
Figure I.24.



Neden bu şekilde düşündüğünüzü açıklayınız?

Ortalama için düşünüldüğünde S deki simetrik yapı ile bu yapıyı bozan 2 frekanslık değer simetriyi bozmakla birlikte ortalamayı sola kaydırır.
Ortanca için de aynısını düşünebilirim.

Figure I.25.



Neden bu şekilde düşündüğünüzü açıklayınız?

Verilerin birbirine en yakın dağılımı olması gerekir.

Figure I.26.

Neden bu şekilde düşündüğünüzü açıklayınız?

Ortalama ve Ortanca aynı aralığa düştüğü için.

Figure I.27.

Neden bu şekilde düşündüğünüzü açıklayınız?

Grafik düzenli olarak arttığı için, daha anlaşılır olduğu için T grafiğinde ~~ortanca~~ ortalamayı hesaplatmak daha kolay olacaktır.

P grafiği normal dağılıma yakın olduğu için ortanca daha kolay buldurulabilir. Baştan ve sondan eşit silmeler yaparak.

Figure I.28.

Neden bu şekilde düşündüğünüzü açıklayınız?

Uç değerler fazla olduğu için ve ortalama Değerlerden etkilendiği için

Toplamalı frekansları yukarıdaki gibi frekans toplarını tek sayı alınız onun rahat bulunmasını sağla

Figure I.29.

Neden bu şekilde düşündüğünüzü açıklayınız?

ortalama ve ortancanın en büyüğünü seçtim.

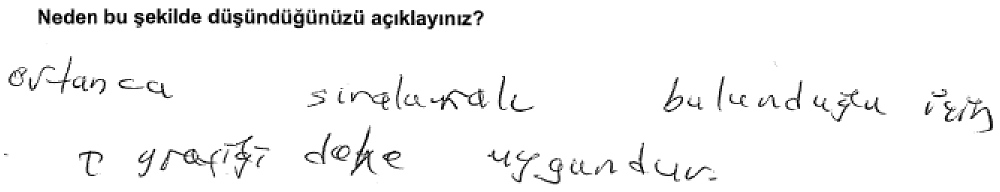
Figure I.30.

Neden bu şekilde düşündüğünüzü açıklayınız?

ortalama için T veriler arası fark çok

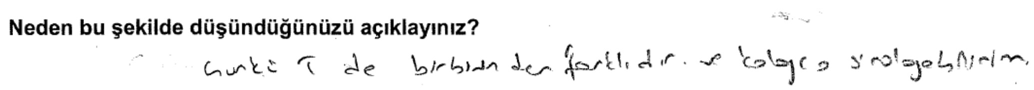
ortanca için S fark az

Figure I.31.


Neden bu şekilde düşündüğünüzü açıklayınız?

ortanca sıralamalı bulunduğu için T grafiği daha uygundur.

Figure I.32.


Neden bu şekilde düşündüğünüzü açıklayınız?

çünkü T de birbirinden farklıdır. ve kolayca sıralayabilirim.

Figure I.33.

APPENDIX J

ORIGINAL SAMPLE ANSWERS FOR PCK-4

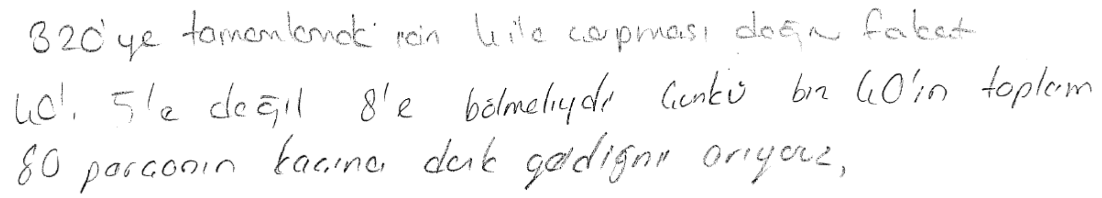


Figure J.1.

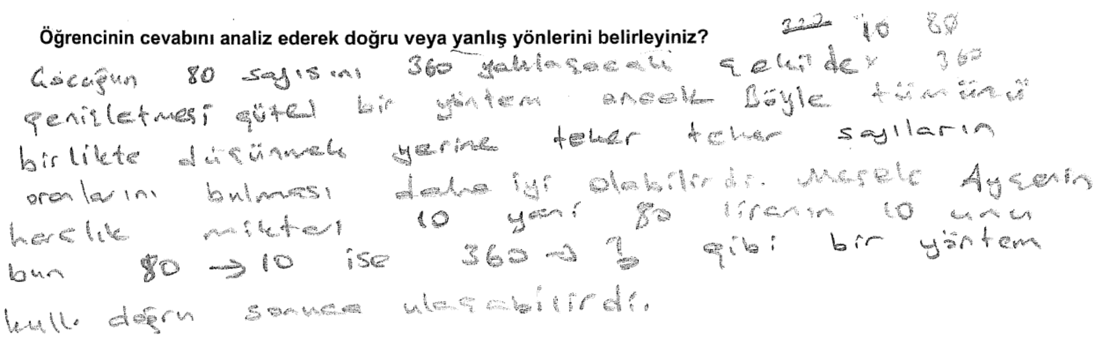


Figure J.2.

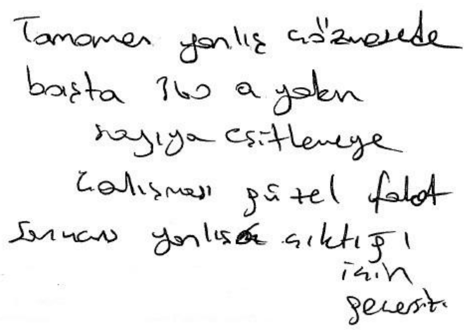


Figure J.3.

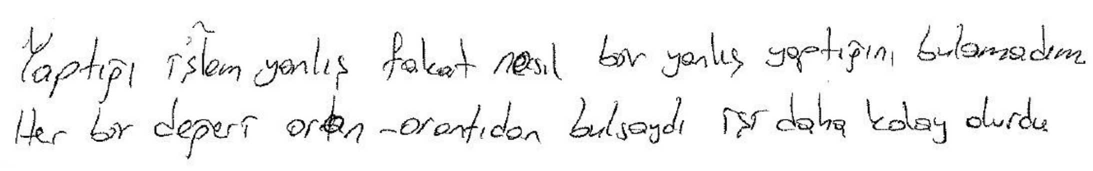


Figure J.4.

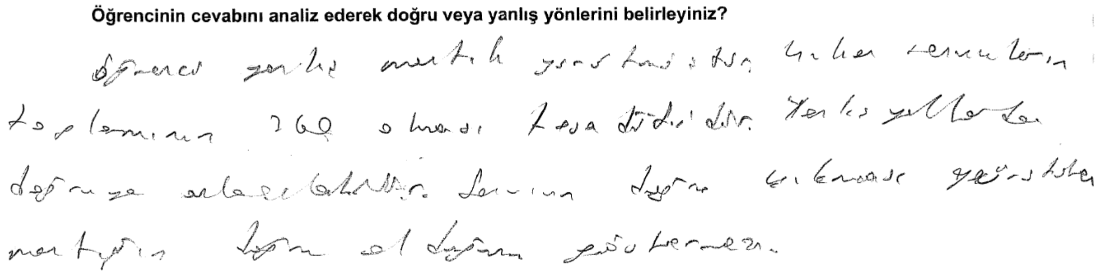


Figure J.5.

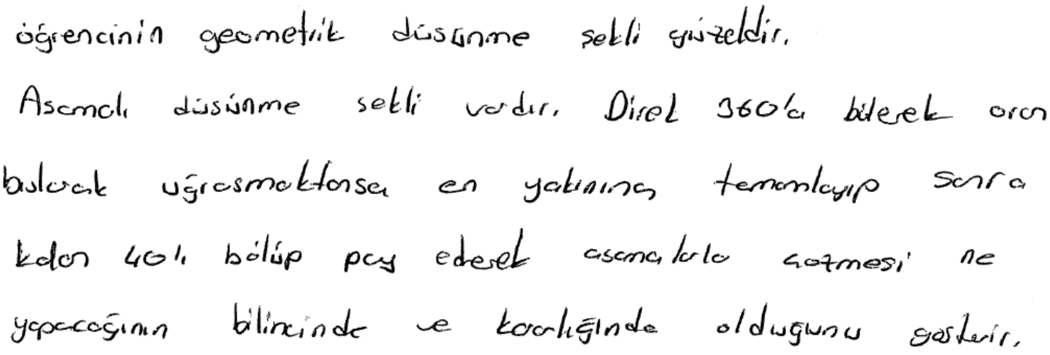


Figure J.6.

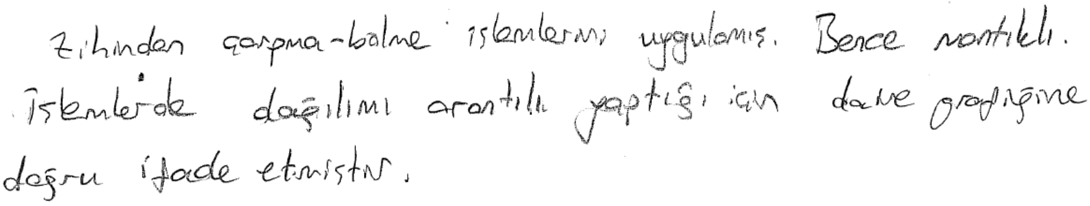


Figure J.7.

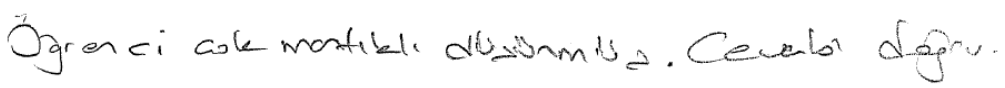


Figure J.8.

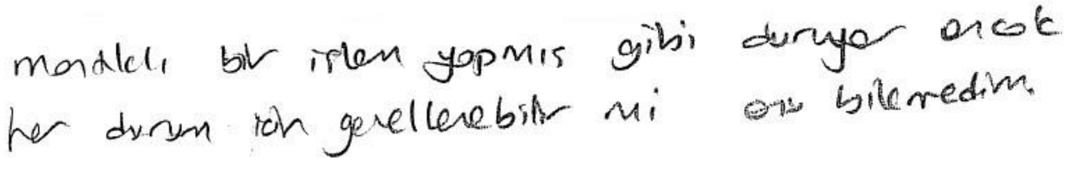


Figure J.9.

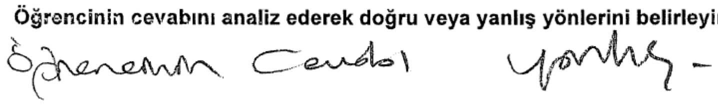


Figure J.10.

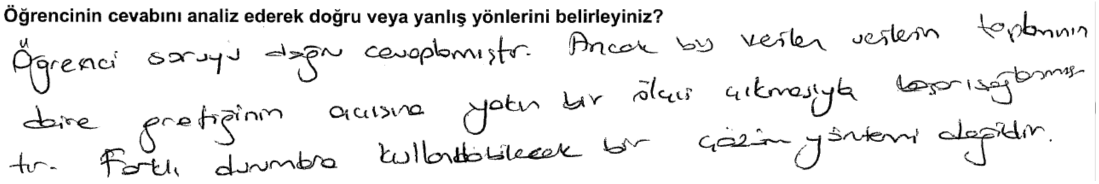


Figure J.11.

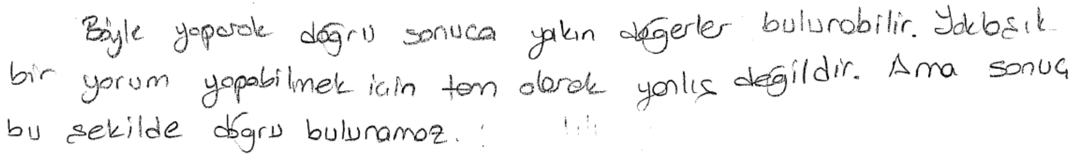


Figure J.12.

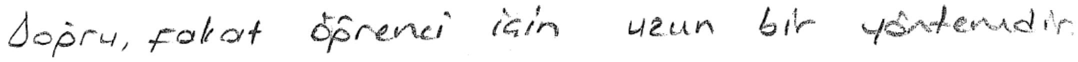


Figure J.13.

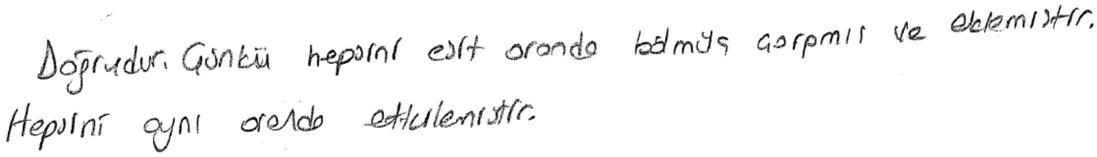


Figure J.14.

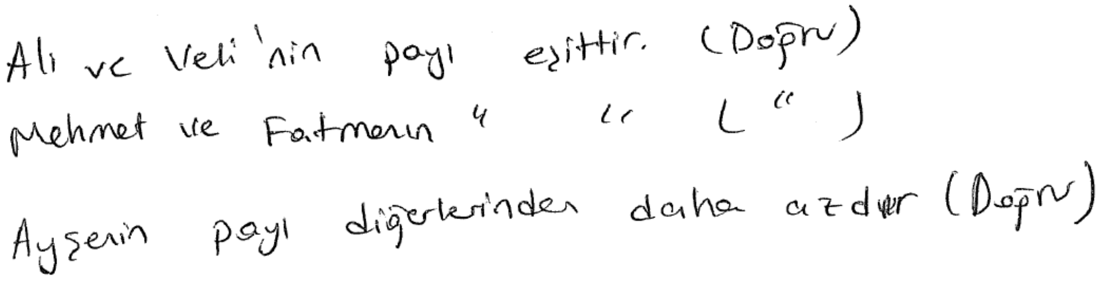


Figure J.15.

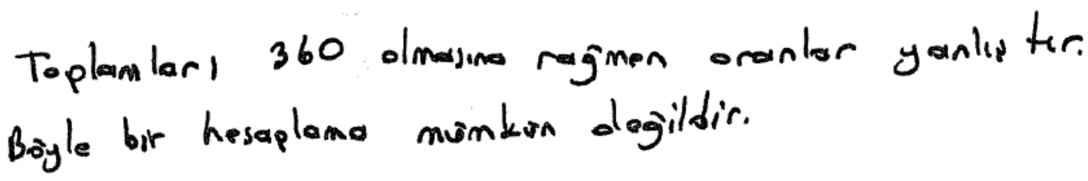


Figure J.16.

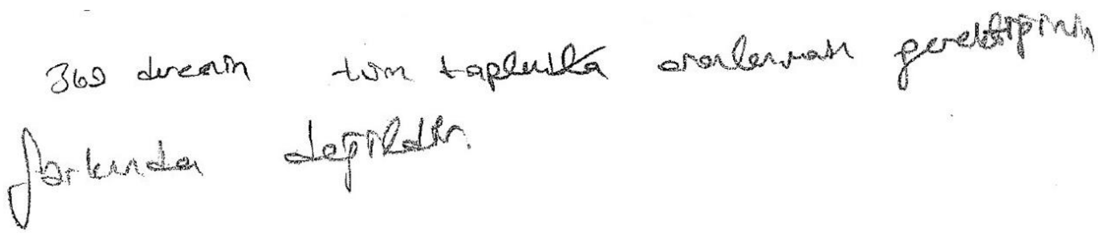


Figure J.17.

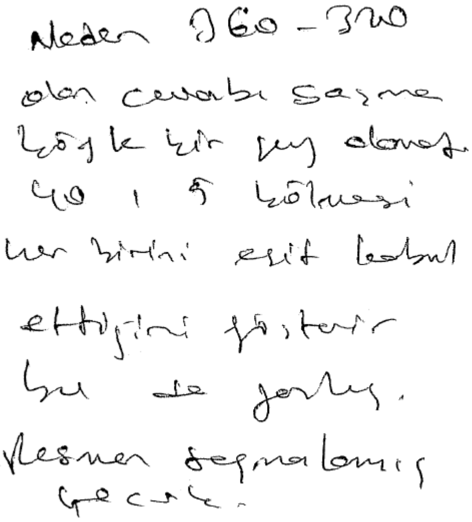


Figure J.18.

APPENDIX K

ORIGINAL SAMPLE ANSWERS FOR PCK-5

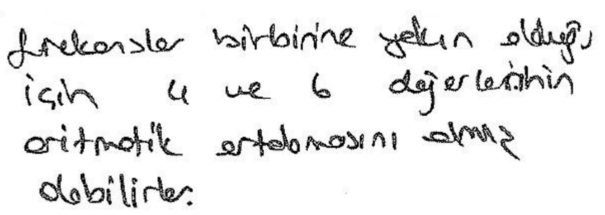

frekanslar birbirine yakın olduğu için 4 ve 6 değerlerinin aritmetik ortalamasını almış olabilirler.

Figure K.1.

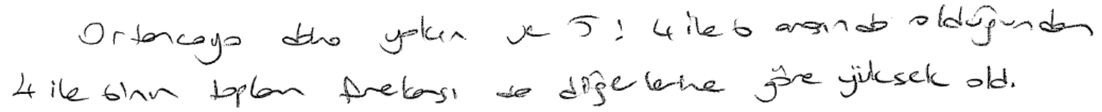

Ortancaya daha yakın ve 5! 4 ile 6 arasında olduğundan 4 ile 6'nın toplam frekansı da diğerlerine göre yüksek old.

Figure K.2.

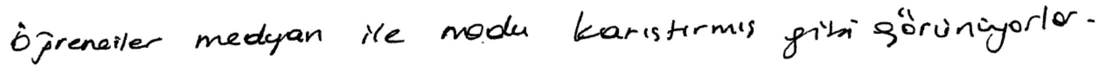

Öğrenciler medyan ile modu karıştırmış gibi görünüyorlar.

Figure K.3.

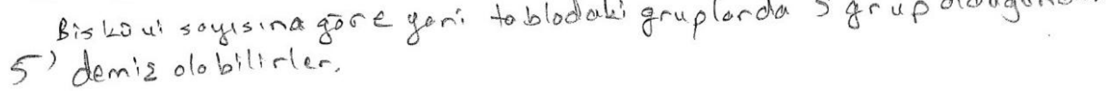

Bisküvi sayısına göre yani tablodaki gruplarda 5 grup olduğundan 5'demiş olabilirler.

Figure K.4.

$\frac{11-2}{2} = 4,5 \approx 5$ demiş olabilirler Çünkü 2, 4, 6, 9, 11 sayıları var

Bunların ortasındaki sayıyı bulmuş olabilirler.

Figure K.5.

Medyanın verilerin (yani) frekansın orta değerini olduğunu bilmiyor bence

Figure K.6.

Ortanca ile frekans kavramlarını idrak edememiş olabilir.

Figure K.7.

Kırık bisküvi sayılarında 5 diye bir sayı yok. Aslında burada niye 5 yazdıklarını kağıtlarına bakarak anlayabiliriz. Buradan bir yorum yapmak zor.

Figure K.8.

Paketlerdeki kırık bisküvi sayısı en çok 11 bunu 2'ye bölmüş olabilir.

Figure K.9.

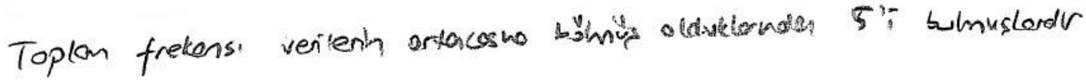


Figure K.10.

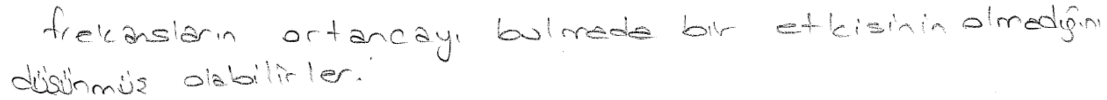


Figure K.11.

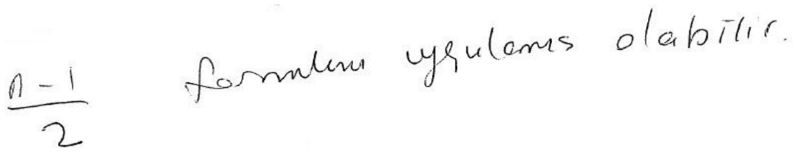


Figure K.12.

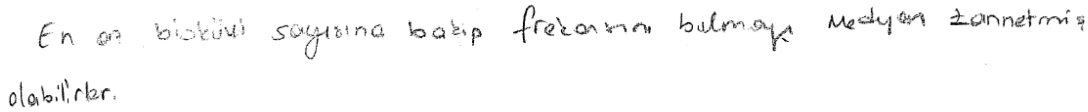


Figure K.13.

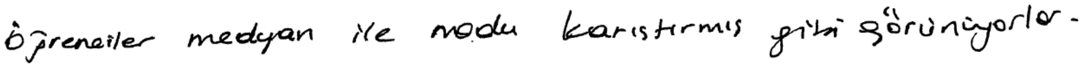


Figure K.14.

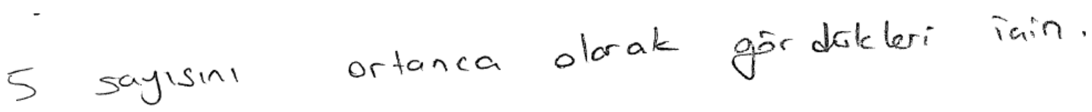


Figure K.15.

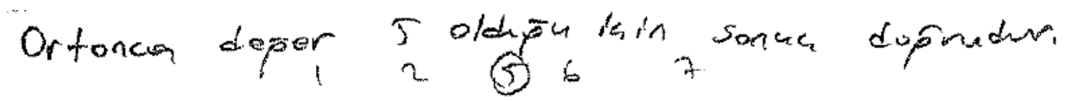


Figure K.16.

Figure K.17.

APPENDIX L

ORIGINAL SAMPLE ANSWERS FOR PCK-6

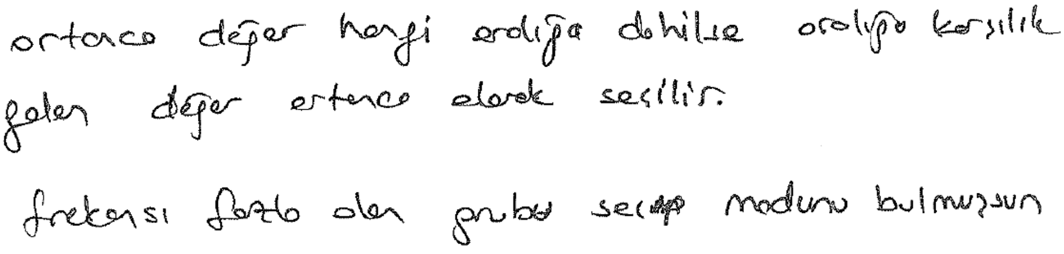


Figure L.1.

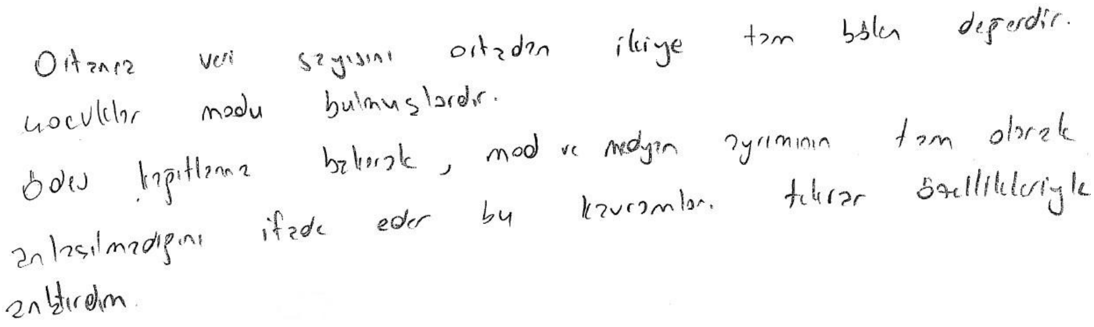


Figure L.2.

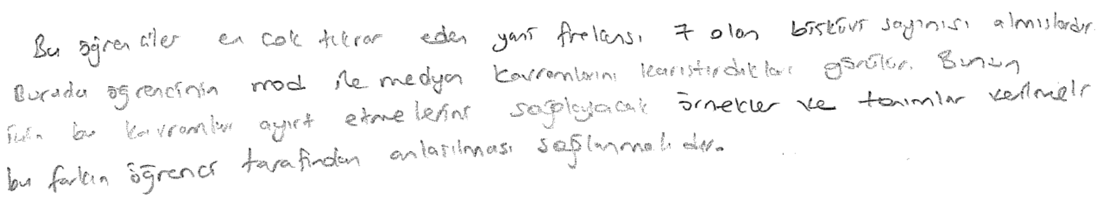


Figure L.3.

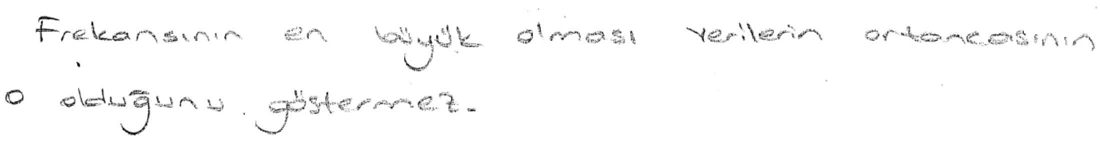


Figure L.4.

Hatanın nedeni büyük ihtimalle en fazla tekrar ediyor ve ortanca da orada bir yerde olur diye düşünmüş
Yorum ise her zaman frekansı çok olduğu yerde olmak zorunda değil

Figure L.5.

frekansı en büyük 7 olduğu için 6 bisküvi kırılık geldiği cevabı vermiş olabilir. Bu da onların ortancanın mantığını kavrayamadıklarını gösterir.

Figure L.6.

Ortanca tanımını tekrar gözden geçirmesini isterim.

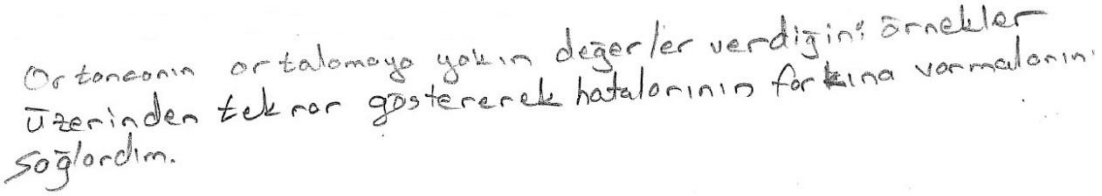
Figure L.7.

Tekrar yapmalarını söyler ve daha dikkatli okumalarını isterdim

Figure L.8.

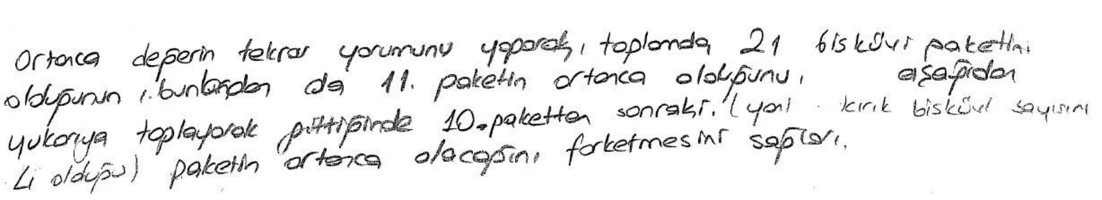
frekansın ne anlama geldiğini bilip bilmediğini sorardım
Daha sonra frekansları kadar kırık bisküvi sayısı olduğunu birde bu şekilde bakmalarını sağlardım

Figure L.9.



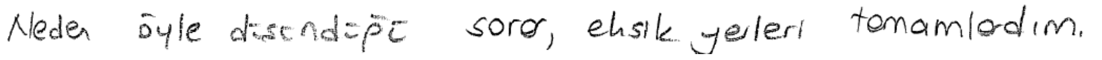
Ortancanın ortalamaya yakın değerler verdiğini örnekler üzerinden tekrar göstererek hatalarının farkına varmalarını sağladım.

Figure L.10.



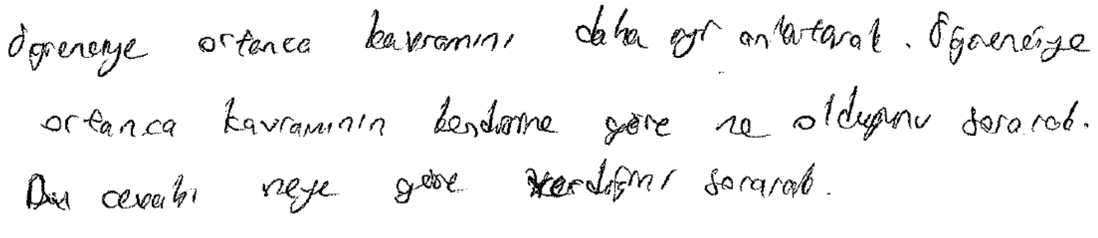
Ortanca değerin tekrar yorumunu yaparak, toplamda 21 bisküvi paketi olduğunun, bunlardan da 11. paketin ortanca olduğunu, aşağıdan yukarıya toplayarak gittiğinde 10. paketten sonraki (yani kırık bisküvi sayısını 4 olduğu) paketin ortanca olacağını farketmesini sağlarım.

Figure L.11.



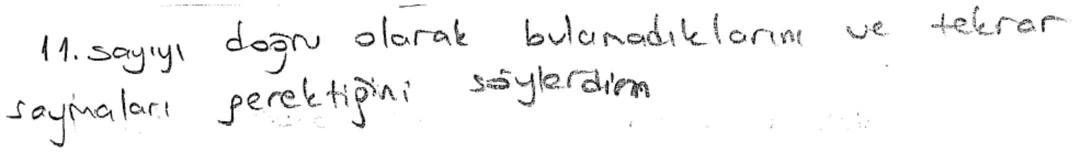
Neden öyle düşündüğü sorar, eksik yerleri tamamlardım.

Figure L.12.



öğrenciye ortanca kavramını daha iyi anlatarak. Öğrenciye ortanca kavramının kendisine göre ne olduğunu sorarak. Bu cevabı neye göre verdiğini sorarak.

Figure L.13.

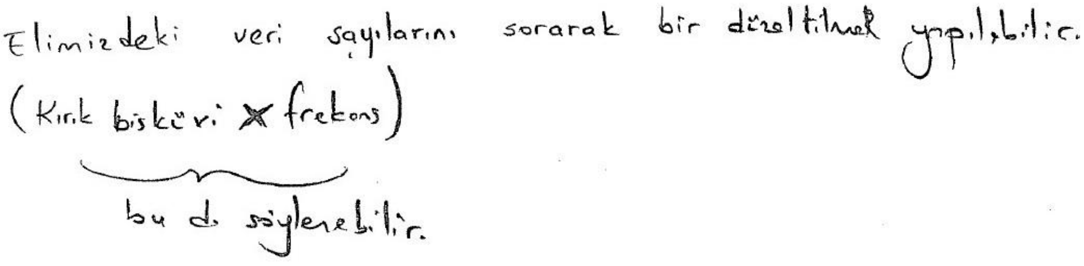


11. sayıyı doğru olarak bulamadıklarını ve tekrar saymaları gerektiğini söylerdim

Figure L.14.

- Buradaki ortanca kırık bisküvilerin olduğu paketlerin ortancasıdır
Frekansların orta noktası değil

Figure L.15.



Elimizdeki veri sayılarını sorarak bir düzeltilmek yapılabilir.
(Kırık bisküvi × frekans)
bu da söylenebilir.

Figure L.16.

Medyan frekanslar ile işlem yapılarak bulunur.
Aritmetik ortalama ile karıştırmayınız

Figure L.17.

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Mercimek, Oktay
Nationality: Turkish (TC)
Date and Place of Birth: 19 August 1980, Adana
Marital Status: Married
Phone: +90 506 530 6265
E-mail: oktaymercimek@gmail.com

EDUCATION

Degree	**Institution**	**Year of Graduation**
M.Ed.	University of Georgia, USA Mathematics Education	2007
BA	Çukurova University Mathematics	2003

WORK EXPERIENCE

Year	**Place**	**Enrollment**
2007- Present	Kastamonu University	Research Assistant
2003- 2004	Private Sector	Mathematics Teacher

PUBLICATIONS

1. Mercimek, O. & Pektaş, M. (2013). İstatistiğe Yönelik Öz-Yeterlik Ölçeğinin Geliştirilmesi. *Kastamonu Eğitim Dergisi*, *21*(2), 759-771.

RESEARCH INTERESTS

- Mathematics teacher knowledge
- Professional development of preservice and inservice mathematics teachers
- Statistics education