ASSESSING PRESERVICE MATHEMATICS TEACHERS' STATISTICAL REASONING

by
Rabia Karatoprak
B.S., Secondary School Science and Mathematics Education, Boğaziçi University, 2011

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Secondary School Science and Mathematics Education
Boğaziçi University
2014

ASSESSING PRESERVICE MATHEMATICS TEACHERS' STATISTICAL REASONING

APPROVED BY:

Assist. Prof. Gülseren Karagöz Akar (Thesis Supervisor)	....
Prof. Ayşenur Yontar Toğrol	....
Assist. Prof. Bengü Börkan	....

DATE OF APPROVAL: 11.07.2014

ACKNOWLEDGEMENTS

I would firstly thank my supervisor Asisst. Prof. Gülseren Karagöz for her support and patience during the research. She has been a strong influence on my intellectual development.

I also would like to thank my thesis committee member, Assist. Prof. Bengü Börkan for her guidance throughout the study.

I am grateful to Assist. Prof. Sevil Akaygün for providing me opportunities to gain research experiences during my undergraduate years. Without her contributions, I would not reach this point in academic life.

I also would like to thank Asisst. Prof. Fatma Aslan Tutak for her unconditional support to my academic life. She has been always there for help and guidance.

I would like to thank deeply Assist. Prof. Hülya Kılıç for her confidence in me and support during the Ph.D. applications. I am grateful to her providing me chance to work with her.

I also would like to thank Prof. Ayşenur Yontar Toğrol, Prof. Ali Baykal, Asisst. Prof. Emine Adadan, Assoc. Prof. Buket Yakmacı Güzel, Asisst. Prof. Fatih Çağlayan Mercan. Each one of them as a researcher, provided me invaluable perspectives.

I also would like to thank Prof. Yavuz Akpınar, Assist. Prof. Aylin Ünaldı, Assist. Prof. Engin Ader, Assoc. Prof. Seyfi Kenan, Assoc. Prof. Ali Delice, Assoc. Prof. Emin Aydın, Assist. Prof. Güneş Yavuz, Assist. Prof. Çiğdem Arslan, Assist. Prof. Yeşim İmamoğlu for their support throughout the research. Without their help, this research could not be completed.

My sincere thanks go to my family for their endless support and love. Especially I would like to thank my father, Mehmet Karatoprak for teaching me how to be successful in my educational life. His encouragement means a lot to me.

My special thanks go to İlyas Erşen making me laugh and feel better. I cannot imagine how these two years could be without him.

Lastly, I would like to thank my friends, İclal, Eray, Nurseda, Münevver, Güneş, Abdullah, Şükrü, Ceyda, Esra, İlker, Şule and Şerife for being just besides me whenever I need them.

ABSTRACT

ASSESSING PRESERVICE MATHEMATICS TEACHERS' STATISTICAL REASONING

Statistical reasoning is to understand and reason with statistical information and make interpretations based on sets of data (Garfield, 2002). In this study, statistical reasoning types of preservice mathematics teachers studying in İstanbul were investigated. Participants were 91 secondary and 82 elementary school preservice mathematics teachers studying in the last semester of their program. Instrument was Statistical Reasoning Assessment – SRA (Garfield, 2003). It was translated to Turkish and then validity and reliability was examined before applying to preservice mathematics teachers. Data was analyzed descriptively. Results showed that both secondary and elementary school preservice mathematics teachers were successful in interpreting probability, understanding independence, interpreting two-way tables and understanding importance of large samples. On the other hand, there was lack of understanding in selecting appropriate average, correctly computing probability, sampling variability, distinguishing between correlation and causation. Also results showed that they had law of small numbers misconception and equiprobability bias. Furthermore a comparison between secondary and elementary school preservice mathematics teachers' statistical reasoning was done. Results showed that there was no statistical difference between their reasoning types. In this regard, the results suggest that in order to address and develop their understanding in concepts and misconceptions, statistics courses are suggested to be given importance in mathematics education programs. Results, portraying the statistical reasonig types of preservice mathematics teachers in Istanbul, might also be taken as a basis for further resarch investigating preservice mathematics teachers' statiatistical reasoning types in Turkey.

ÖZET

MATEMATİK ÖĞRETMEN ADAYLARININ İSTATİKSEL AKIL YÜRÜTMESİNİN ÖLÇÜLMESİ

Garfield (2002) istatiksel akıl yürütmeyi istatiksel bilgiyi anlama, kullanma ve veriye dayalı çıkarımda bulunabilmek olarak tanımlanmıştır. Bu çalışmanın amacı matematik öğretmen adaylarının istatiksel akıl yürütmelerini incelemektir. Katılımcılar İstanbul'da bölümlerinin son sınıfında okuyan 91 ortaöğretim ve 82 ilköğretim matematik öğretmeni adayıdır. Çalışmada İstatiksel Akıl Yürütme Testi (Statistical Reasoning Assessment – SRA) kullanılmıştır (Garfield, 2003). Test Türkçe'ye çevrilmiş, geçerlilik ve güvenilirlik çalışmaları yapılmıştır. Testin Türkçe hali elde edildikten sonra, öğretmen adaylarına uygulanmıştır. Veri betimsel olarak analiz edilmiştir. Bulgulara göre hem ortaöğretim hem ilköğretim matematik öğretmeni adayları olasılığı yorumlamada, bağımsızlığı anlamada, iki yönlü tabloları anlamlandırmada ve büyük örneklemlerin önemini anlamada başarı gösterdiler. Fakat ortalamayı seçmede, olasılık hesaplamada, örneklem çeşitliliğinde ve ilişki ve nedenselliği ayırt etmede zorlandıkları bulunmuştur. Ayrıca, küçük sayılar kavram yanılgısına ve eşit olasılık yanlılığına sahip oldukları görülmüştür. Ek olarak ortaöğretim ve ilköğretim matematik öğretmeni adaylarının istatiksel akıl yürütmesi karşılaştırılmış ve istatiksel olarak anlamlı bir fark bulunmamıştır. Bu durumda, bulgulara göre istatiksel akıl yürütme kavramlarının geliştirilmesi ve kavram yanılgılarının ortadan kaldırılabilmesi için matematik eğitimi programlarında istatistik dersine önem verilmesi tavsiye edilmektedir. İstanbul'daki öğretmen adayları hakkında elde edilen sonuçlar, Türkiye'deki diğer öğretmen adaylarının istatiksel akıl yürütmesinin incelenmesine dayanak oluşturmaktadır.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF SYMBOLS

$\bar{x}$	Mean
N	Number of participants
ρ	Spearman rho correlation coefficient

LIST OF ACRONYMS/ABBREVIATIONS

H	Head
MoNE	Ministry of National Education
NCTM	National Council of Teachers of Mathematics
PCK	Pedagogical Content Knowledge
SD	Standard Deviation
SRA	Statistical Reasoning Assessment
T	Tail

1. INTRODUCTION

In society many people face with issues such as new increasing threats to the environment or changing needs for aging population. People have to make choices about events depending on chance or risk analysis such as buying lottery tickets, insurance policies, or comprehending medical advice. Also, people need to participate in debate or community action and to be aware of some phenomena such as crime rates, population growth, spread of diseases, industrial production, educational achievement, employment trends (Gal, 2002; Watson & Callingham, 2003). Since these daily issues have statistical components, it is important to strengthen statistical reasoning on the individual's part (Gal, 2002; Watson & Callingham, 2003; Garfield & Ben-Zvi, 2007).

In addition, National Council of Teachers of Mathematics(NCTM) calls attention to the fact that an individual's reasoning statistically is not always intuitive (NCTM, 2005). Thus, statistical reasoning needs to be developed as an outcome of schooling (Gal, 2002; Watson & Callingham, 2003; Garfield & Ben-Zvi, 2007). As explained under the next heading, although some concepts of statistics was integrated to the middle school National Mathematics Curriculum of Turkey in 2005 and probability and statistics was included as a branch in 2011, there seems to be almost any research done on neither school students' nor preservice and in-service teachers' statistical reasoning. Furthermore, it seems that not all of the preservice mathematics teachers take statistics course in Turkey. For instance, 7 of the 22 programs include statistics courses in their programs when the official web sites of the secondary school mathematics education programs in Turkey were examined. On the other hand at elementary level, statistics course is a "must" course at all of the accessible programs. However what teachers know is what students know (Fennema & Franke, 1992; Heaton & Mickelson, 2002; Yolcu, 2012; Uçar & Akdoğan, 2009). Therefore preservice mathematics teachers' reasoning need to be examined (Shaugnessy, 2007; Bulut, 2001).

In this regard, the purpose of this study is to investigate the statistical reasoning types of pre service mathematics teachers since it is imperative for both preservice and in-service mathematics teachers to have statistical reasoning in order to be able to educate individuals who can reason statistically (Garfield, 2002; Yolcu, 2012). Considering the fact that, there is an attempt at preparing preservice teachers for teaching statistical reasoning, determining preservice mathematics teachers' statistical reasoning types might further assist teacher preparation programs to include statics courses for teaching in their programs.

2. LITERATURE REVIEW

2.1. Statistics Content Area in National Curriculum

High School National Mathematics Curriculum of Turkey (2013b) included particular skills related with statistics education for the students in grades five through twelve. Some of these skills follow:

(i) to be able to draw tables and graphics related with a phenomenon and explain mathematical reasoning through examining it
(ii) to be able to construct hypotheses explaining relationships within the phenomenon
(iii) to be able to draw inferences regarding their hypotheses .

These are all relate to statistical reasoning such awareness of which must be raised on the students' part (NCTM, 2005). Thus, it becomes important that teachers have statistical reasoning prior to their teaching for it (Heaton & Mickelson, 2002; Yolcu, 2012).

In the following headings, first I briefly mention to what extent statistics is included in the National Mathematics Curriculum of Turkey. Then, I explain statistical literacy, statistical reasoning and statistical thinking based on the previous research. After words, I provide a brief analysis on statistics included in the Turkish Curriculum in terms of the cognitive demands (Delmas, 2002) required of each domain. This might provide lenses as to how these domains are applied in Turkish Curriculum and the hidden or explicit expectations from both preservice elementary and secondary teachers.

2.1.1. Elementary School Statistics in Curriculum

Elementary School National Mathematics Curriculum (2013a) also includes data processing and probability as a branch such that data processing exists in each grade. Within the scope of this content area, students are expected

(i) to construct research questions, and show data applicable to these questions as tables and graphics,
(ii) to make comparisons and analyses by calculating average and range,
(iii) to find mean, median and mode by looking at data.

Again, the purpose is to provide background towards the development of statistical reasoning on students' part. In this regard, it also becomes significant for elementary school mathematics teachers to have statistical reasoning.

2.1.2. High School Statistics in Curriculum

Secondary School National Mathematics Curriculum of Turkey (2013b) involves data, counting, and probability as a branch. Within the Data, counting, and probability concepts area students are expected for instance to learn central tendency and distribution concepts as well as presenting data as graphics. As can been seen, the importance of statistical reasoning becomes more apparent at high school level and the teachers' being knowledgeable about the statistics concepts become more valuable.

2.2. Statistical Literacy, Reasoning and Thinking

Although statistical reasoning is expected as an outcome of schooling (Gal, 2002; Watson & Callingham, 2003; Garfield & Ben-Zvi, 2007), there has been research on statistical literacy and thinking as related areas of statistics education. Because of their interrelatedness, in this section, based on the previous research, statistical literacy, statistical reasoning and thinking are compared and contrasted regarding the aim of this research. Following that, research on statistics education is discussed.

Statistical literacy, reasoning and thinking are the intended outcomes of statistics education at different levels of education ranging from primary school to university. These domains are unique but they overlap with each other at some points (Garfield & Ben-Zvi, 2007; Delmas, 2002; Rumsey, 2002; Chance, 2002; Garfield, 2002). Although these points cannot be distinguished clearly, these domains differ from each other in terms of the cognitive processes they require of an individual. More than that there is a hierarchy between these domains such that statistical literacy is the foundation for statistical reasoning and thinking, and statistical thinking is higher order thinking than statistical reasoning (Garfield & Ben-Zvi, 2007). Required cognitive processes for each of them and the structure of hierarchy are explained in the following sections.

2.2.1. Statistical Literacy

What statistical literacy is, has been answered by many authors (Gal, 2002; Watson & Callingham, 2003; Wallman, 1993) but Rumsey (2002) and Gal (2002) provided a much more detailed definition.

From Rumsey's (2002) perspective, an individual, who is statistically literate and has researcher skills, at least need to understand and use statistical ideas and terms such as mean, median etc.. As a further step they should be able to question, probe, compare and contrast, explain, decide, judge, evaluate and make decisions about the information. In order to do these, Rumsey (2002) stated that people should have statistical competence, which is also a foundation for statistical reasoning and thinking.

Statistical competence involves *data awareness*, *understanding of basics* such as measure of center etc., *data production and results*, *interpretation at a basic level* such as descriptive statistics, and *basic communication skills.* Related aspects with data awareness is that *data exists at every part of the world*, *data might be misleading*, and *decisions based on data have strong implications on our lives.* For instance, when airbags are produced according to average man, they might kill children and women. Regarding *understanding of basics*, Rumsey (2002) stated that it means to be able to relate a statistical idea with other concepts, to be able to explain the idea and use it in other contexts, and to be able to answer questions about it. In terms of understanding *data production and results*, Rumsey claimed that data production involves writing research questions and locating data as appropriate to the research questions from given data set. Similarly, for *interpretation at a basic level*, interpretation requires reaching a conclusion and explaining what it means. Finally, without communication skills the information inferred from data cannot be transferred to the others and thus people cannot see the others' viewpoint. When all these aspects of statistical competence being a part of statistical literacy and a foundation for statistical reasoning and thinking are considered, Rumsey argued that statistical literacy should be an outcome of introductory statistics courses. From her perspective, these courses need to aim at raising students as statistically literate individuals and make them gain researcher skills.

Another detailed definition is made by Gal (2002) who has been considered as a prominent researcher regarding statistical literacy (Shaughnessy, 2007). He proposed a model about statistically literate persons. His model consisted of both the knowledge elements and the dispositional elements. Knowledge elements include literacy skills, statistical knowledge, mathematical knowledge, context knowledge and critical questions. Dispositional elements involve beliefs and attitudes towards statistics.

Knowledge elements stand for people's ability to comprehend, interpret and evaluate the statistical messages. A brief explanation on the components of knowledge elements follow: *Literacy skills* refer to make sense of messages in any format such as lists, indexes, texts or graphs. *Statistical knowledge* contains five elements: knowing why data are needed and how data can be produced, familiarity with basic terms and ideas related to descriptive statistics, familiarity with graphical and tabular displays and their interpretation, understanding basic notions of probability, and knowing how statistical conclusions or inferences are reached (Gal, 2002). *Mathematical knowledge* is to have the ability to understand the procedures applied in statistics in order to interpret the data in a meaningful way. Also interpreting the data depends on the individual's familiarity with the *context*. Data is not independent from the context. *Critical question* refers to ask on your own spontaneously the worry questions. Questions support critical evaluation since they are about the reasonableness of the data and the same data can be used to support or oppose a claim. Since statistically literate people are expected to be able to determine about truth of the claim, they can ask worry questions.

Dispositional aspects refer to the tendency of an individual share and communicate with others about the thoughts on a claim (s) resulted from raising worry questions. Here the "worry questions" refer to someone's raising concerns related to the validness of the inferences made from the analysis of the data.

As the aforementioned explanations indicate, Gal and Rumsey thought of statistical literacy as making the core of statistical reasoning and thinking. In the following part statistical reasoning is explained since it requires higher level cognitive engagement than statistical literacy according to hierarchy between these three domains.

2.2.2. Statistical Reasoning

Garfield (2002) defines statistical reasoning as to understand and reason with statistical information and make interpretations based on sets of data. Although there is no consensus about the development of statistical reasoning or how to determine the level and correctness of the reasoning, it is suggested that statistical reasoning is developed at the end of statistics courses. Therefore, students who have statistical reasoning are expected to (i) apply their reasoning in different contexts and (ii) make correct judgments and interpretations (Garfield, 2002).

Garfield (2003) developed an instrument, called Statistical Reasoning Assessment (SRA), according to the correct reasoning types learners are expected to gain and to the incorrect reasoning types learners should not develop. The correct reasoning skills involve reasoning about; *data*, *representations of data*, *statistical measures*, *uncertainty*, *samples*, and, *association*.

Incorrect reasoning types involve misconceptions regarding both probability concept and some statistical measures. For instance, among are *misconceptions about averages*, *outcome orientation*, *bias about good sample*, *the law of small numbers*, *representativeness misconception* and *equiprobability bias*.

<u>2.2.2.1. Correct Reasoning Types.</u> In the following paragraphs, each of the correct reasoning types are explained.

Reasoning about data: Statistics is based on data (Moore, 1990; Garfield & Ben-Zvi, 2008). How data is produced is the most important thing since conclusions are directly related to data analysis and thus data production (Garfield & Ben-Zvi, 2008). When investigating a phenomenon, variables are determined (Mickelson & Heaton, 2004) and data are values of variables (Garfield & Ben-Zvi, 2008). Furthermore data are only numbers without context (Moore, 1990). Within the context, type of data as qualitative or quantitative can be determined. Then inferences can be drawn based on type of data (Garfield & Ben-Zvi, 2008). However organizing or understanding data changes according to different representations and different representations show different characteristics of data (Konold, Higgins, Russell, & Khalil, 2003). Therefore different types of data require different representations. For instance, qualitative data are shown as pie graphs or bar charts while quantitative data are shown as histograms or dot plots (Garfield & Ben-Zvi, 2008). In this regard reasoning about data involves being able to identify the variables within the context in which the data is produced, being able to determine the type of data, and being able to both know the types of representations regarding different types of data and show them. Therefore, students are expected to first be aware of these aspects of data.

Reasoning about representations of data: It refers to being able to read and understand a graph; to be able to modify it in order to interpret the data or recognize the general characteristics (Garfield, 2003). Curcio (1987) investigated the predictors of graph comprehension. In her study, skills of graph comprehension are categorized as (i) reading the data, (ii) reading between the data and (iii) reading beyond the data. Also, Friel, Curcio and Bright (2001) examined factors affecting graph comprehension. Paralled to Curcio's skills, Friel *et al.* (2001) categorized the indicators of graph comprehension as translation, interpretation, and extrapolation/interpolation. All these researchers agree upon the fact that , graph comprehension requires at first level to read the graph literally such as to be able to understand what the graph is about, at a second level to interpret the presented data such as comparison of values or finding median, and at the highest level, to make inferences from the given data such as answering different questions not obvious in present data or establishing the relationship between the variables.

Reasoning about statistical measures: It deals with knowing what measures of center, spread and position tell about data; when and how to use them in interpreting data (Garfield, 2003).

Measures of central tendency are representatives of the group of scores (Gay, Mills, & Airasian, 2009). Mode, median and mean are the most frequently used ones for describing nominal, ordinal and interval or ratio data, respectively. Mode is the most frequent score observed in the data and not determined through calculation. Median is the midpoint of the scores ordered with respect to their value and it does not have to be one of the actual scores. Mean is the arithmetic average of the scores. It is the most preferred and stable measure in describing the data since each participant's score is included in the mean (Gay *et al.*, 2009). However, in case of existing extreme scores in data, mean may not show the typical score of the data. In these cases median and mode should be considered to be used in order to describe the data.

Measures of central tendency sometimes may not be sufficient in describing data.

In these cases measures of spread should be taken into account especially in comparing data sets (Garfield, 2003). For instance, while comparing two sets of data, the data sets might have the same mean yet their spread might be totally different from each other. This means then that these data sets are not the same. Therefore, it becomes important to consider the measures of spread when examining the data sets relative to each other. Range, quartile deviation and standard deviation are the measures of spread used for nominal, ordinal and interval or ratio data, respectively (Gay *et al.*, 2009). Range is the difference between the highest and the lowest data scores. Standard deviation is the square root of variance that tells how far each score is from the mean (Gay *et al.*, 2009). Similar to the mean, standard deviation is the most stable and preferred measure of spread since each score is taken into account. Standard deviation and mean together can give information about the scores of the data and thus data sets can be compared (Gay *et al.*, 2009).

Reasoning about uncertainty: It is to understand and use the ideas of randomness, chance, and likelihood such as to know the ways of determining the probability of events (Garfield, 2003). Outcome of an event cannot be determined certainly even if possible outcomes are definite (Moore, 1990). For instance, if a fair coin is tossed, possible outcomes are head and tail but which one is going to be shown is uncertain. These possible outcomes are in a pattern since showing of something except from head and tail is not possible. At the same time, each outcome is uncertain. Thus the phenomenon under investigation is called as random (Moore, 1990) and its likelihood could be measured.

Reasoning about samples: It requires being able to know the part-whole relationship between sample and population (Watson & Moritz, 2000). Sample is a subset of the population and by examining this subset; inferences can be made about the population. That is to say results of the sample can be generalized to population. To be able to generalize the results randomness, representativeness and bias are critical notions to be considered (Watson & Moritz, 2000). In order to select a representative sample, members should be selected randomly since each member of the population has the equal chance to be selected (Gay *et al.*, 2009). That is, selection based on human judgment is intentionally abandoned since it is less possible to gather persistently outrageous scores of the population with randomness (Moore, 1990). However just random selection does not guarantees the representativeness since sample size effects the results (Tversky & Kahneman, 1971). Small samples even if it is randomly selected might not show all of the properties of the population since variability exists within the population (Watson & Moritz, 2000). If sample size is big enough to include the variability within the population and is a small scale replication, unbiased inferences can be drawn. Therefore sample size is important in drawing inferences.

Reasoning about association: It refers to being able to interpret the relationship between variables (Garfield, 2003). Whether there is a relationship between variables or to what degree the relationship exists are the concerning questions. Even if two variables are related to each other, it does not mean that causation exists between these variables (Gay *et al.*, 2009). For instance long hours of TV watching and low school achievement level might be related but it does not mean that low school achievement is the result of watching TV for long hours.

Researchers concluded that once the students have these reasoning skills, they might be considered as having statistical reasoning (Garfield, 2002, 2003; Garfield & Gal, 1999). However, they also contend that for someone to be able to reason statistically, they should not have some particular misconceptions. In the following paragraphs, I explain each of these incorrect statistical reasoning types.

2.2.2.2. Incorrect Reasoning Types. In the following paragraphs, each of the incorrect reasoning types are explained.

Misconceptions involving averages: Misconceptions involving averages might vary. For instance, some might think that averages are the most frequent number. In addition, some other people might consider of mean and median as the same. Moreover, some might argue that groups are compared according to their averages and average is calculated with add-divide procedure regardless of the outliers. Researchers stated that students should not be using these types of (incorrect) reasoning while analyzing statistical information (Garfield, 2002, 2003; Garfield & Gal, 1999).

The outcome orientation: It refers to the tendency to decide by looking at just a single event rather than series of the events (Konold, Pollatsek, Well, Lohmeier, & Lipson, 1993). Therefore, outcome orientation seems to be related to being able to think about uncertainty. For instance, as can be seen in SRA, people who have outcome orientation see 70% likelihood of rain in ten days as it should rain in each of the ten days (Garfield, 2003). However, there is still a 30% chance of not to rain for each day. On the other hand, people with this misconception decide for each day independently rather than looking at the whole ten days. According to their reasoning, 95% means it rains, 50% means it cannot be known and 10% means it does not rain. The other percent values are decided according to their proximity to these values. That is for instance 70% chance means it rains while 15% means it does not rain at all.

Good samples have to represent a high percentage of the population: It means that if a sample size is a large percentage of a population then it is a good sample (Garfield & Ben-Zvi, 2008; Garfield, 2003). People who have this misconception think that if population size increases then sample size should also increase. However for a reliable estimation absolute size of the sample is important rather than the sample size relative to the population (Smith, 2004). A well chosen sample can provide sound estimations even if the sample is not a high percentage of the population (Garfield & Ben-Zvi, 2008) since sample is randomly chosen. In other words, as stated earlier, random choice of samples can provide representativeness and unbiased results. For instance probability of heads can be obtained accurately from a thousand tosses although population size is infinite and sample size is not a particular percentage of the population (Smith, 2004).

The law of small numbers: It refers to the fact that small samples resemble the populations in making inferences (Garfield, 2003). People believe that a randomly drawn any sample should have the same characteristics of the population without considering sample size (Tversky & Kahneman, 1971). Thus randomly drawn two samples are similar to each other since they are randomly chosen. However these two samples might be different from each other since random samples varies, especially the small ones (Garfield & Ben-Zvi, 2008). In this regard, this misconception might be thought as complementary of the third misconception. That is, people having the third misconception does not take into account the random choice of the sample while thinking of the sample size, on the other hand, people having this misconception ignore the sample size while thinking of the randomness. For instance as can be seen in SRA, people with the intuition of law of small numbers tend to produce sequences with equal numbers of heads and tails even in a small number of tosses. Although this is valid for very large sample according to the law of large numbers (very large samples highly represents population), people intuitively apply this law to the small samples too.

The representativeness misconception: It means that the likelihood of sample depends on how it resembles to the population (Kahneman & Tversky, 1974). People who have representativeness misconception tend to use heuristics instead of probability principles. For instance, as a result of flipping a fair coin 6 times, obtaining HTHTHT, is seen as less likely than obtaining HTTTHT (Kazak, 2008; Kazak, 2009). People who have this misconception reason that HTHTHT is not random since there is a pattern. Therefore they see it as less likely. However theoretically, probabilities of both events are the same, which is $(\frac{1}{2})^6$. Thus, this misconception is related to the reasoning about sampling and uncertainty.

The equiprobability bias: It means that the likelihood of events are tended to be seen as the same since it happens by chance (Lecoutre, 1992). Thus they are equiprobable by nature. For instance when two dice are simultaneously thrown, the possibility of obtaining two fives and the possibility of obtaining one five and one six on two rolls are seen as equally likely (Kazak, 2008; Kazak, 2009). However one five and one six can be obtained in two ways {56, 65} while there is just one way of obtaining two fives because sample space equals {55}. Therefore the chance of getting one five and one six double the possibility of getting two fives. Thus, people having this misconception do not take into account the sampling.

After statistical reasoning is explained above, lastly statistical thinking is presented. Among statistical literacy, statistical reasoning and statistical thinking domains, statistical thinking is at the highest level in terms of cognitive engagement (Garfield & Ben-Zvi, 2007).

2.2.3. Statistical Thinking

Statistical thinking can be defined separately from statistical reasoning and literacy but assessing and teaching overlaps with them (Chance, 2002). Statistical thinking requires deep understanding of statistical processes and inferences (Garfield & Ben-Zvi, 2007) such that in the simplest way, it refers to thinking and behaving like a statistician (Shaughnessy, 2007; Delmas, 2002; Chance, 2002; Pfannkuch & Wild, 2000; Wild & Pfannkuch, 1999).

Processes of statistical thinking include summarizing data, solving a particular problem, reasoning through a procedure and explaining the conclusion. Unique to the statistical thinking is to see the processes as a whole, understand variation, explore the data beyond the given prescriptions and generate new questions beyond the asked ones (Rumsey, 2002).

Pfannkuch and Wild (2000) constructed a theory on statistical thinking. They interviewed statisticians. The logic behind their research was that if the processes and strategies of statisticians were known, then the type of thinking styles, which should be developed in learners of statistics, could have been determined. Results showed that there are four elements of statistical thinking. These are, taking (i) variation into account, (ii) transnumeration, (iii) constructing and reasoning from models and (iv) integration or synthesis of problem context-matter and statistical understandings. The aspects of variation as (i) being omnipresent, informative and inescapable, (ii) noticing variation existing around us and (iii) wondering why it was placed to the center of the statistical thinking were found to be the key indicators of statistical thinking. By the same token, transnumeration was to be found another important aspect of statistical thinking. Transnumeration refers to change data representations in order to understand the situation. It is to change the constructs into measurable form or change the form of data so that different perspectives can be inferred from the data. Similarly, constructing and reasoning from models were profound indicators of statistical thinking. It refers to reasoning about the data through mathematical constructs such as confidence intervals or standard errors in order to say that the results are reliable. Finally, integration or synthesis of problem context-matter and statistical understandings is to combine both context knowledge and statistical knowledge in order to get information from data such that without the context knowledge, statistical knowledge is not sufficient to get real conclusions.

All in all, although statistical literacy, reasoning and thinking seem to include similar characteristics, they differ from each other in terms of cognitive engagement (Delmas, 2002). The cognitive engagement related to each domain, is shown in Table 2.1.

Table 2.1. Tasks Distinguishing Three Domains.

BASIC LITERACY	Identify, Describe, Rephrase, Translate, Interpret, Read
REASONING	Why?, How?, Explain (the process)
THINKING	Apply, Critique, Evaluate, Generalize

These verbs given above are keys in instruction for distinguishing these domains (Delmas, 2002). For instance statistically literate person can know when to use mean, median and mode or critically evaluate the statistical statement. The abilities to question, compare and contrast such as to be able to explain the add-divide procedure in finding mean, why we divide n-1 in finding standard deviation or to be able to relate the concepts with each other such as independence of an outcome and representativeness misconception belong to statistical reasoning. Lastly, applying the ideas to the new problems and asking questions of your own such as choosing the best way to analyze data or diagnosing weaknesses in the statistics procedures shows us statistical thinking.

2.2.4. The Analysis of the National Curriculum

Both the Secondary and Elementary School National Mathematics Curriculum of Turkey (2013b) was analyzed in order to examine whether the existing objectives regarding the statistics content are parallel to the cognitive tasks provided by DelMass (2002) written above. This is important because as mentioned earlier, the statistics education was integrated into the National Mathematics curriculum lately. The expectations of the curriculum developers seem to indicate that they suppose students to develop statistical reasoning at the end of their high school education. For instance these two sentences are quoted from the National Curriculum:

> “Comparison of more than two data sets are covered.”
> “Explain box-plot, draw and interpret box-plot of a data set and use box-plot while comparing data sets.”
> “Discrete and continuous data are defined and differences between them are emphasized.”

Students are expected to compare the groups and explain usage of box-plot while comparing these groups. Also difference between discrete and continuous data is emphasized in national curriculum. Therefore students are expected to engage in tasks of statistical reasoning as given in Table 2.1. This suggests teachers to have at least statistical reasoning to be able teach students for such a purpose. Therefore, determining preservice teachers’ reasoning types becomes important so that the education they get prior to their actual teaching might be continued or might change accordingly. Therefore in the next section, past research results found in the literature is discussed. According to these results, statistical reasoning types of preservice mathematics teachers are tried to be anticipated.

2.3. Research in Statistics Education

In this part, related research with statistical reasoning is described in both global and Turkish context. First of all, research conducted with preservice teachers in global context was described. Then research related to statistics education in Turkey is presented.

Related to reasoning about statistical measures, Groth and Bergner (2006) conducted a study describing preservice elementary school mathematics teachers' conceptual and procedural knowledge of mean, median and mode. They were asked to compare and contrast use of measures of center. Results showed that very few of them (3 out of 46) could talk about the hypothetical situations in which one of the mean, median and mode might be a better measure of center although all of them could discuss the procedures of these measures. So this result can be interpreted as the fact that their lack of conceptual knowledge prevent them to understand when to use these measures of center.

Another study related to statistical measures, secondary school preservice teachers' discussion of variation in comparing two data sets was examined by Makar and Confrey (2005). Their use standart statistical language was weak and they tended to use informal words while describing variation and distribution. Although their use of these words implied their understanding of measures of spread, it emerged from shape of the distribution. So this might suggest that their reasoning about measures of distribution was weak.

Moreover there were studies related to preservice teachers' use of statistical measures in comparing distributions. One of them was conducted by Leavy (2006). Leavy (2006) aimed to find out statistical concepts of elementary school preservice teachers used in analyzing and comparing distributions of data. Elementary school preservice teachers compared the data sets with just descriptive statistics. Moreover they focused on measures of center too much and ignore the variation. Another study investigating elementary school preservice teachers' and middle schools students' reasoning about distributions in comparing two data sets was conducted by Canada (2008). Similarly, Canada (2008) found that elementary school preservice teachers compare the data sets based on means of the data and ignore the variability. These studies can be interpreted as preservice teachers are more comfortable in using measures of center than measures of spread.

Related to another statistical reasoning type which is sampling, Groth and Bergner (2005) investigated preservice elementary school mathematics teachers' concept of sample through metaphors. They were asked to construct a metaphor about sample and discuss. When their metaphors were analyzed few of them (10 out of 54) could give metaphors pointing characteristic of sample which is representative part of the whole. An important part of the participants (20 out of 54) just included the notion of part-whole relationship between sample and population in their metaphors. However the rest of the group could not even mention that relationship. According to these resullts, sampling might be difficult to understand for preservice teachers as one of the indispensible component of statistical reasoning (Garfield, 2003).

Furthermore Heaton and Mickelson (2002) examined elementary school preservice teachers' integration statistical investigation into elementary school curricula. They were asked to conduct statistical investigation in their practicum class and help children in their practicum class to apply statistical investigation. It was found that preservice teachers understand the difficulty and importance of formulating meaningful questions and data collection. Moreover they used their data to reach conclusions. However they ignored the critical evaluation of their conclusions and representativeness of their data. By the same token, they thought reasoning with data was just about teaching and they could not appreciate the meaning of statistical investigation. Additionally children in their practicum class could not go beyond technical components of graphing data and there was lack of deep interpretation of data. This situation was connected to teachers' lack of understanding of statistical investigation and the fact that teaching depends on what teachers know and understood. Therefore it was suggested that preservice teachers should be supported pedagogically.

Additionally, there exists studies conducted with both elementary and secondary school preservice mathematics teachers. Jendraszek (2010) investigated preservice tecahers' probability misconceptions. The participants of the study were intented to teach at elementary, secondary or college level. It is found that preservice teachers have equiprobability bias and misconception of law of small numbers. Also, it was found that their performance was related with their probability education especially if it happens at college. So, participants' background in probability and statistics education, is an important factor.

Taking all of these into consideration, these research points correct reasoning skills mentioned previously in Section 2.2.2. According to these research findings; preservice teachers have procedural knowledge rather than conceptual understanding in measures of center (Groth & Bergner, 2006). Despite that, they overemphasize measures of center in comparing data sets and do not take variation into account (Leavy, 2006; Canada, 2008). Even though they might have informal notions about variation, this stems from their focus on shape instead of measures of spread (Makar & Confrey, 2005). Thus their comparison of data sets is shallow (Leavy, 2006). In fact comparison of data sets revealed that they have difficulties in understanding data as aggregate (Leavy, 2006; Makar & Confrey, 2005). Furthermore their reasoning about data is far from deep interpretation and so do their students (Heaton & Mickelson, 2002). They do not find statistical investigation meaningful and do not appreciate it (Heaton & Mickelson, 2002). Lastly there are missing points in their concept of sample and they ignore the critical evaluation of representativeness (Groth & Bergner, 2005; Heaton & Mickelson, 2002). Since their teaching is about what they know, their students' learning will be the same (Heaton & Mickelson, 2002). Therefore preservice mathematics teachers' statistical reasoning need to be investigated so that necessary educational support could be given in order to develop students' statistical reasoning (Heaton & Mickelson, 2002).

2.3.1. Research in Statistics Education in Turkey

There are only a few studies conducted in the statistics area in Turkey (Ulutaş & Ubuz, 2008). Also, in these research studies, participants were dominantly selected from elementary school students. Still these studies are presented since preservice teachers' and students' conceptions are qualitatively similar to each other (Canada, 2008; Makar, 2004). So conceptions of students can lead us toward Turkish preservice teachers' conceptions. Related to elementary school mathematics, Uçar and Akdoğan (2009) focused solely on the concept of average in elementary school level, Kaynar and Halat (2012) examined 8^{th} grades' statistics subjects, and Yolcu (2012) investigated statistical literacy of 8^{th} grades. Also Özen (2012) examined preservice elementary school mathematics teachers' statistical and probabilistic knowledge.

Related to statistical measures, the study of Uçar and Akdoğan (2009) aimed at investigating how elementary school students understand average. They found out that, half of the students, out of 18 could understand average as a representative value of data whereas the other half could not. Additionally, most of the students think average as arithmetic mean and students mostly attribute algorithmic add-divide procedure as the meaning of average rather than understanding it conceptually. Also Kaynar and Halat (2012) examined the statistics concepts of 8^{th} grade students. It was found that students' overall success level at solving central tendency and dispersion questions were low. Simlarly, almost all of them (92.7%) of them could not solve the standard deviation question.

Additionally, Yolcu (2012) investigated 8th grade students' statistics concepts, existing at Elementary Mathematics National Curriculum, according to Watson's three tiers (1997) in statistical thinking. Where the first tier refers to the understanding statistical terminology without context and context-free calculation of measures of tendency and spread; the second tier involves understanding and interpretation of statistics and terminology embedded in context (Watson, 1997). And the third tier includes adopting critical stance toward statistical results and to be able to challenge them. Yolcu (2012) found that pair wise differences existed between tiers. That is students performed highest at the second tier, lowest at the third tier and slightly higher at the first tier. Yolcu attributed the lowest and highest performance of students at the third and second tier respectively to the National Curriculum whereas low performance of students at first tier was associated with the mathematics teachers' incorrect usage of statistical terminology. Additionally, Yolcu emphasized those mathematics teachers' attitudes and critical evaluation of statistical knowledge might have affected the students' performance at third tier.

Regarding studies with preservice teachers, Özen (2012) investigated preservice elementary school mathematics teachers' usage of critical thinking skills with statistical and probabilistic knowledge in media contexts. The findings indicated that participants use sample size as a single aspect of the statistics and they failed to relate it with the context. Also results revealed that they were more familiar with data collection methods than data analysis techniques. Finally, participants did not question the results and quality of the research processes while thinking about the results and the conclusions of the research contexts provided to them.

General implications of these previously discussed studies can be addressed as follows. Firstly students do not use statistical terminology correctly (Yolcu, 2012). Secondly they have misunderstanding in some concepts (Yolcu, 2012; Kaynar & Halat, 2012; Uçar & Akdoğan, 2009). For instance, they fail to understand average as a representative value of data (Uçar & Akdoğan, 2009). Also they tend to see mean and average as the same and they underestimate the meaning of average as add-divide procedure. Thirdly they do not have critical stance toward statistical conclusions (Yolcu, 2012). More importantly, possible reasons that generate these results are linked to the teachers' gaps in their statistical content knowledge (Yolcu, 2012; Kaynar & Halat, 2012; Uçar & Akdoğan, 2009). Furthermore findings of Özen (2012) and Yolcu (2012) correspond with each other. They indicate that both students and preservice elementary level mathematics teachers do not have critical stance toward statistical conclusions. They both contended that their findings need to be investigated further because teachers' knowledge influences their students' knowledge (Yolcu, 2012; Uçar & Akdoğan, 2009). Therefore, these results suggest that preservice teachers' statistical reasoning levels need to be investigated in order to have a good grasp of the nature of their reasoning so that both their education prior to their actual teaching and their students' statistical reasoning can be developed properly.

2.4. Significance of the Study

Recent studies done in Turkey revealed that statistics is one of the subjects which needs to be studied further (Ulutaş & Ubuz, 2008; Uçar & Akdoğan, 2009). Also, research suggested that university students do not have much interest towards statistics and Turkish people are not used to applying statistics knowledge in everyday life contexts (Şahin, 2012). However, research indicated that students who blindly believe the claims in everyday life context are in need of statistics education (Watson, 1997). By the same token, research results suggested that preservice teacher education should emphasize development of statistics concepts because teachers' knowledge has an influence on students' performance (Fennema & Franke, 1992; Heaton & Mickelson, 2002). Yet, the studies of statistics education conducted in the global context are mostly about students' understanding and there are fewer studies with teachers than with students (Mickelson & Heaton, 2004; Shaughnessy, 2007).

In this regard, this study aimed at investigating statistical reasoning types of preservice mathematics teachers. An additional purpose is to provide the Turkish of the Statistical Reasoning Assessment (SRA) (Garfield, 2003) for mathematics education field in Turkey. This is especially important since the new curricula involve statistics education of students at both elementary and secondary level. With the Turkish version, both teacher educators and teachers might assess their students' statistical reasoning. This is because although SRA was developed and used for high school students initially (Garfield & Gal, 1999), later on, it was also used with university students (Garfield, 2003).

A further aim of this study was to scrutinize whether a significant difference existed between the elementary and the secondary preservice teachers' statistical reasoning types. The reason follows: First, this research is conducted with preservice teachers educated in all of the universities in Istanbul. As mentioned earlier, when the teacher education programs was examined, not many of them (7 out of 22) already include statistics courses for secondary school although there is a must course in elementary school mathematics education programs. If the results of this study show that there is a difference between these teachers' statistical reasoning types, these courses' contents might be further investigated towrads a standardized course in teaching statistics for secondary/elementary mathematics. If there is no difference, then this might provide a description of the situation among preservice teachers who are supoosedly known to enter the universities in Istanbul with the relatively highest scores in Turkey. Therefore, this might create an awareness on the teacher educators' part to include such learning opportunities in the preservice teachers' knowledge reportoire.

3. STATEMENT OF THE PROBLEM

The purpose of the study is to investigate statistical reasoning types of the preservice elementary and secondary school mathematics teachers in İstanbul.

3.1. Research Questions and Hypotheses

Research questions of this study are given below:

(i) What are the statistical reasoning types of elementary school preservice mathematics teachers?
(ii) What are the statistical reasoning types of secondary school preservice mathematics teachers?
(iii) Is there any difference between statistical reasoning types of preservice secondary school mathematics teachers and elementary school mathematics teachers?

The research hypotheses for the third research question, "Is there any difference between statistical reasoning types of preservice secondary school mathematics teachers and pre-service elementary school mathematics teachers?" is

- $H_0$3: There is not any significant difference between statistical reasoning types of preservice secondary school mathematics teachers and preservice elementary school mathematics teachers in.

3.2. Variables

In this study there are two variables: statistical reasoning types and teaching level of preservice mathematics teachers. These variables are explained below.

Statistical reasoning types: In this study Garfield's (2003) definition of statistical reasoning is adopted. According to her, there are eight correct reasoning skills learners are expected to gain and eight incorrect reasoning types learners should not develop in order to be counted as having statistical reasoning. These correct reasoning skills and misconceptions are showed in Table 3.1 and Table 3.2.

Table 3.1. Correct Reasoning Skills.

Correct Reasoning Skills
Correctly interpret probabilities
Understand how to select an appropriate average
Correctly compute probability
Understand independence
Understand sampling variability
Distinguish between correlation and causation
Correctly interpret two way tables
Understand importance of large samples

Table 3.2. Misconceptions.

Misconceptions
Misconceptions involving averages
Outcome orientation
Good samples represents high percentage of the population
Law of small numbers
Representativeness
Correlation implies causation
Equiprobability bias
Groups can be compared if they are the same size

Teaching level of preservice mathematics teachers: preservice mathematics teachers were the students studying in the last term of mathematics education program at secondary and elementary level. Mathematics teaching are taught in two different programs in Turkey as secondary and elementary school mathematics teaching. Elementary school preservice mathematics teachers are expected to teach 5^{th}, 6^{th}, 7^{th} and 8^{th} grades. Secondary school preservice mathematics teachers are expected to teach 9^{th}, 10^{th}, 11^{th} and 12^{th} grades.

4. METHODOLOGY

Aim of this study is to assess statistical reasoning of preservice secondary and elementary school mathematics teachers. In parallel with the purpose of the study; participants, research design, instrument, procedure and data analysis were explained in this chapter. Sample section includes sampling method and description about the sample.Then method of the study is introduced in research design section. Afterwards data collection instrument and obtaining its Turkish version is explained. Also how data was collected is explained in the procedure part. Lastly data analysis is described.

4.1. Sample

Target population of the study is preservice secondary and preservice elementary school mathematics teachers studying in the mathematics teaching programs in İstanbul. Preservice teachers who participated the teaching methods course at the time of data collection constituted the sample. Although all of the population members were aimed to be reached, sample consists of 80% of the population of preservice secondary school mathematics teachers and 50% of the population of preservice elementary school mathematics teachers.

Three secondary school mathematics teaching program exists as one private and two state universities in Istanbul. So population of this study at secondary level were 109 preservice teachers as 30 of them studying in private and 79 of them studying in state universities . Sample size is 91 since data were collected from 24 out of 30, 25 out of 29 and 42 out of 50 preservice teachers at secondary level. So at least 80% of the preservice secondary school mathematics teachers population for each university were reached.

Four elementary school mathematics teaching program exists as one private and three state universities in Istanbul. So population of this study at elementary level were 170 preservice teachers 15 of whom studying in private and 155 of whom studying in state universities. Sample size is 82 since data were collected from 27 out of 50, 30 out of 60, 21 out of 45 and 4 out of 15 preservice teachers at elementary level. So about 50% of the preservice elementary school mathematics teachers population for each university were reached.

Participants were asked whether they took statistics course or not. 85.7% of secondary school preservice teachers took statistics course and 90.2% of elementary school preservice teachers took a course to teach probability and statistics.

4.2. Research Design

Research design is descriptive since the situation, which is statistical reasoning of preservice mathematics teachers in this case, is aimed to be described with quantitative data and to determine existing differences between secondary and elementary school preservice mathematics teachers (Gay *et al.*, 2009). Groups are established according to the grouping variable which is the teaching level of preservice mathematics teachers. Statistical reasoning of preservice teachers according to their teaching level are compared.

4.3. Instrument

Researchers and teachers use different methods such as performance assessments or interviewing students while assessing teaching, learning or the development of statistical reasoning. However these types are not practical to administer in large scale assessments (Garfield, 2003; Tempelaar, 2004). Statistical Reasoning Assessment – SRA, on the other hand, is a paper-pencil test which is easy to administer and score large scale assessments. Also it is helpful for the design of instructions since responses provide information about correct reasoning skills and misconceptions (Sundre, 2003). By the same token, when the National Mathematics Curriculum is examined, the same statistical reasoning types such as understanding how to select an appropriate average or misconception of comparing groups based on average are also included in the secondary and elementary school national mathematics curriculum. Therefore Statistical Reasoning Assessment – SRA (Garfield, 2003) was used in this study.

Table 4.1. Correct Reasoning Subscales.

Scales	**Items and Alternatives**
1- Correctly interpret probabilities	2d, 3d
2-Understand how to select an appropriate average	1d, 4c, 17c
3-Correctly computes probability a: Understands probabilities as ratios b: Uses combinatorial reasoning	 8c 13a, 18b, 19a, 20b
4-Understand independence	9e, 10cdf, 11e
5-Understand sampling variability	14b, 15d
6-Distinguish between correlation and causation	16c
7-Correctly interpret two way tables	5,1d
8-Understand importance of large samples	6b, 12b, 7ef

SRA is developed firstly to evaluate the effect of US high school statistics curriculum on students' statistical reasoning in 1998 and was published in 2003. In many studies, it was used and accepted as a valid and reliable instrument (Tempelaar, 2004; Liu, 2002). Also it was translated into French, Spanish and Chinese (Garfield & Gal, 1999). It includes 20 multiple choice items. Each item is about probability or statistics. Choices of items are particular reasoning statements which show correct reasoning, misconception or simply false instances. Therefore there are two main categories as correct reasoning skills and misconceptions. These two main categories have 8 scales. Measuring items and alternatives for each scale belonging correct reasoning skills and misconceptions are given below (See Table 4.1 and Table 4.2).

Table 4.2. Misconceptions Subscales.

Scales	**Items and Alternatives**
1-Misconceptions involving averages	
a: Averages are the most common number	1a, 4a, 17e
b: Fails to take outliers into consideration when computing the mean	1c, 4b
c: Compares groups based on their averages	15bf
d: Confuses mean with median	17a
2-Outcome orientation	2e, 3ab, 11abd, 12c, 13b
3-Good samples have to represent high percentage of the population	7bc, 16ad
4-Law of small numbers	12a, 14c
5-Representativeness	9abd, 10e, 11c
6-Correlation implies causation	16be
7-Equiprobability bias	13c, 18a, 19d, 20d
8-Groups can only be compared if they are the same size	6a

4.3.1. Turkish Version of Instrument

Firstly Turkish version of SRA was obtained (See Appendix A). Adaptation of an instrument include processes of translation, examining validity, determining equivalence of the items, and examining reliability of the instrument (Hambleton & Patsula, 1998). Therefore, the same procedures were followed in order to obtain Turkish version of SRA. In the following paragraphs firstly how SRA translated to Turkish is explained. Then expert-opinion about the construct validity of the items was described. Then processes of examining equivalence of the items and reliability of the test were allocated. Data analysis procedure and results were given place in both equivalence and reliability parts.

Translation: Instrument is translated by a professional translator, a graduate student studying at a Mathematics Education Master Program and the researcher of this study. Aspects of familiarity with the content of the instrument and competency at both of the languages are taken into consideration while choosing translators. Only forward translation was applied since back translation method might have some disadvantages (Hambleton & Kanjee, 1993). For instance, in back translation instrument is translated into target language and then it is translated back to source language in order to compare the differences with the original one. Since comparison is made between the source language and back translated version, the possible problems appeared in target language cannot be detected. Additionally, lack of exactly the same expressions or similar language structure in both of the languages might be responsible for poor translation although back translated version seems to be the same as the original version. After translation, the researcher and thesis advisor revised the translated instruments independently and after reaching agreement, final version was constructed. Also a Turkish teacher controlled the grammar of the test.

Validity: Whether the questions measure the same construct both in the Turkish version and in the original instrument were evaluated by three experts in the area other than board members. These experts were researchers in mathematics education, foreign language education and computer and educational technologies area. Also they previously had similar studies. Their evaluation resulted as the questions measures the same construct in both versions. Also there were minor suggestions related to Turkish context.Therefore the adapted version is revised by the researcher and then it was checked by the researcher and thesis advisor twice according to experts' comments. After final version was reached, the translated version of the test was administered for the equivalence and reliability studies.

Equivalence: In this part firstly sample and procedure of equivalence study is explained. Then data analysis and results are presented.

In order to establish empirical evidence for linguistic equivalence bilingual or monolingual participants can participate in the study, but finding sufficient number of bilingual participants might be difficult (Hambleton & Kanjee, 1993; Rapp & Allalouf, 2003). Therefore monolingual university students participated in the equivalence study. These participants were competent at both languages since their native language was Turkish but they have been studying for four years in a university in which instruction language is English. The number of participants should be at least 30 for this equivalence study because of the parametric analysis requirement (Gay *et al.*, 2009). In this part, 61 university students as two groups took firstly the original instrument. At least two weeks should be left between test administrations as a precaution (Aksayan & Gözüm, 2002). If shorter than two weeks, participants might remember the items and higher scores might appear. Therefore three weeks later they retook the Turkish version of the instrument.

Analysis of the equivalence of the forms was done item by item. Responses were coded as 1 for correct answer and 0 for incorrect answers. Also blank answers were coded as 0. So data is ordinal type of data. Therefore McNemar test as one of non-parametric tests was used in data analysis (Gay *et al.*, 2009). Its assumptions (Baştürk, 2010) and how these assumptions were ensured are given below:

(i) Test should be administered to same people at two different times: McNemar Test is used in analysis of matched-pair data (Durkalski, Palesch, Lipsitz, & Rust, 2003; Baştürk, 2010). In this part of the study, same groups of students took the English version of the test and 3 weeks later they took the Turkish version of the test. Therefore the assumption of administration at two different times to the same people was ensured.

(ii) Variable of interest should be dichotomous: Scores are coded as dichotomous since responses are scored as 1 for correct and 0 for incorrect or blank ones. For instance a case cannot be coded as both 1 and 0; it should be either 1 or zero.

(iii) Data can be presented in 2x2 table: Since data are dichotomous and test administered at two different times to the same participants, data can be presented in 2x2 table. How data can be placed is showed above in Table 4.3. In this table, *true* and *false* refer to the number of correct answers and incorrect answers respectively. Turkish and English refer to the administration of Turkish and original version of the instrument respectively.

Table 4.3. Presentation of Data.

		English Version	
		True	False
Turkish Version	True		
	False		

Cross-tables of McNemar Test for each item are shown in Appendix B. The results (N=61, p value 2-tailed) as a whole is given below (See Table 4.4). Insignificant p-value means “there is no difference between the Turkish and English forms” because of the hypothesis of McNemar Test.

Significant difference was not appeared in all of the questions, except 1^{st}, 2^{nd}, 4^{th},10^{th}-c, 11^{th}, 13^{th} and 17^{th}.

Firstly translations of these questions was checked but language of these questions were not complicated. Then these questions were translated back to English by a professional translator who are familiar with the statistics and probability. Then back translated version and original version were compared by a Ph.D. student in mathematics education who are about to have Ph.D. degree and competent in both languages. Since she determined as there is no difference between two versions of these questions which might affect the results, cross-tables of McNemar Test were examined. According to McNemar analysis of these questions, significant result stems from the difference between incorrect answers in administration of original version and correct answers in administration of translated version. It is seen that most of the participants answered incorrectly in administration of original instrument whereas they answered

Table 4.4. Results of McNemar Test.

Questions	**p-value**
1	.007
2	.008
3	1
4	.031
5-1	.581
5-2	.815
6	.815
7-e	.078
7-f	.049
8	.125
9	1.
10-c	.001
10-d	.690
10-f	.664
11	.021
12	.791
13	.013
14	.227
15	.754
16	.664
17	.007
18	1.
19	.687
20	.219

correctly in administration of Turkish version (See Appendix B). Furthermore 1^{st}, 4^{th}, 17^{th} questions assess selecting an appropriate average and 2^{nd}, 10^{th}-c, 11^{th} and 13^{th} questions assess outcome orientation misconception. Since these questions are related to two subjects and their answers were incorrect in original version while correct in Turkish version, it is accounted for the data that significant difference might have resulted from their learning of the these subjects.

In order to examine correlation between results of Turkish and English administrations for total correct reasoning skills score and misconceptions score, participants' total score were found. How their scores were found, explained below (See *Scoring*).

Scoring: The original scoring of the instrument was applied. Each person has total correct reasoning skills score, total misconceptions score and 16 sub-scales scores.

Instead of scoring each item as correct or incorrect, responses were analyzed as signaling a correct or incorrect type of reasoning. Responses contributed two main categories as score of correct reasoning skills and score of misconceptions. These two main categories have 8 sub-scales. Therefore 16 scores are obtained from these sub-scales. There are particular alternatives of the particular items contributing to the score of these sub-scales. The list of alternatives and items for each scale are also given previously in Section 4.3 (See Table 4.1 and Table 4.2).

Correct reasoning skills score is obtained per person in the following way: Someone selecting the correct alternative(choice) gets 1 point otherwise 0 point. These scores of particular responses of the items contributing each scale are added and then divided by the number of items since each scale includes different number of responses. In this way, scores of each sub-scale change on a scale of 0 to 2. For instance the scale of *Correctly Interprets Probabilities* is composed by alternative d of 2^{nd} and 3^{rd} items. If someone answers correctly one of them and incorrectly the other, s/he gets .5 point for *Correctly Interprets Probabilities* sub-scale since $\frac{(1+0)}{2}$ equals .5. The other correct reasoning subscales scores are found by the same way. Then, after scores of each subscales generating correct reasoning skills score per person are found, they are added in order to obtain a total correct reasoning score. This procedure is repeated in order to find each participant' score.

The same procedure is applied for obtaining "Misconceptions" score but at this part if someone selects the alternative signaling a misconception, then s/he gets 1 point. Therefore 0 means for misconception part, s/he does not have the misconception. For instance the scale of *Representativeness Misconception* is composed by alternative a, b, d of 9^{th}, e of 10^{th} and c of 11^{th} items. If someone selects all of these alternatives for these questions, s/he gets 1.67 point for *Representativeness Misconception* sub-scale since $\frac{(1+1+1+1+1)}{3}$ equals 1.67. The other misconceptions subscales scores are found by the same way. Then, after scores of each miconceptions subscales scores are found, they are added in order to obtain a total misconceptions score. This procedure is repeated in order to find each participant' score.

In order to found correlations between results of Turkish and English administrations for total correct reasoning skills score and misconceptions score Pearson-r correlation coefficient was intented to be calculated. However to calculate Pearson-r correlation coefficient, both variables should be continuous and both groups should be normally distributed (Büyüköztürk, 2010). Variables become continuous in calculating total scores and how to obtain these total scores were explained below. Then normal distribution was checked by Kolmogorov-Smirnov Test (Büyüköztürk, 2010).

Table 4.5. Test of Normality- Equivalence of Correct Reasoning Skills Scale.

	Kolmogorov-Smirnov		
	Statistic	df	Sig.
Original Version	.070	62	.200
Turkish Version	.092	62	.200

As can be seen from Table 4.5, normal distribution assumption for total correct

Table 4.6. Equivalence-Correct Reasoning Skills Scale.

	Original Version	Turkish Version
Original Version Spearman's ρ N=(61)	1	.649**
Turkish Version Spearman's ρ N=(61)	.649**	1

reasoning skills score were failed. Therefore Spearman-rho correlation coefficient was calculated. Significant and high correlation between results of Turkish and English administration for correct reasoning skills score were found, $\rho = .649$, $p < 0.01$ (See Table 4.6).

Table 4.7. Test of Normality-Equivalence of Misconceptions Scale.

	Kolmogorov-Smirnov		
	Statistic	df	Sig.
English Version	.073	61	.200
Turkish Version	.067	61	.200

As can be seen from Table 4.7, normal distribution assumption for misconceptions score were failed. Therefore Spearman-rho correlation coefficient was calculated. Significant and low correlation between results of Turkish and English administration for misconceptions score were found, $\rho = .285$, $p < 0.05$ (See Table 4.8).

Reliability: In this part firstly how the sampling was done and procedure of reliability study is explained. Then analysis of the data and results of the reliability analysis are presented.

After equivalence was established between the forms of the instrument, reliability

Table 4.8. Correlation-Equivalence of Misconceptions Scale.

		Original Version	Turkish Version
Original Version N=(61)	Spearman's ρ	1	.285*
Turkish Version N=(61)	Spearman's ρ	.285*	1

analysis was conducted. In original version of the test, internal consistency was found as quite low. This result was considered as the test does not measure one trait or ability (Garfield, 2003). Therefore test-retest reliability coefficient was calculated in original study. In congruent to this, test-retest reliability coefficient was calculated in this study.

Table 4.9. Test of Normality-Reliability of Correct Reasoning Skills Scale.

	Kolmogorov-Smirnov		
	Statistic	df	Sig.
1^{st} administration	.064	62	.200
2^{nd} administration	.073	62	.200

Adapted version was re-administered to the same participants (N=61) again 3 weeks after the first administration of the adapted version was done. In order to match the results of the same participants, they were asked to write their student number. Consistency between two administrations was controlled by test-retest reliability analysis. Pearson-r Correlation coefficient was intented to be calculated for correct reasoning skills scale and misconceptions scale (Büyüköztürk, 2010). However to calculate Pearson-r correlation coefficient, both variables should be continuous and both groups should be normally distributed as it is explained in *equivalence* part (Büyüköztürk, 2010). Normal distribution was checked by Kolmogorov-Smirnov Test (Büyüköztürk,

Table 4.10. Reliability of Correct Reasoning Skills Scale.

	1^{st} administration	2^{nd} administration
1^{st} administration Spearman's ρ N=(62)	1	.764**
2^{nd} administration Spearman's ρ N=(62)	.764**	1

2010) and normal distribution assumption was failed Correct Reasoning Skills Scale as can be seen in Table 4.9. Therefore Spearman-rho correlation coefficient was calculated. Significant and high correlation between results of two Turkish administration for correct reasoning skills score were found, $\rho = .764$, $p < 0.01$ (See Table 4.10).

Table 4.11. Test of Normality-Reliability of Misconceptions Scale.

	Kolmogorov-Smirnov		
	Statistic	df	Sig.
1^{st} administration	.069	62	.200
2^{nd} administration	.111	62	.054

Normal distribution was checked by Kolmogorov-Smirnov Test (Büyüköztürk, 2010) and normal distribution assumption was failed for Misconceptions Scale as can be seen in Table 4.11. Significant and high correlation between results of two Turkish administration for misconceptions score were found, $\rho = .666$, $p < 0.01$ (See Table 4.12). Since reliability of .70 for the correct reasoning skills scale and .75 for the misconceptions scale was found in original study, reliability results are parallel with the original study.

Table 4.12. Reliability of Misconceptions Scale.

		1^{st} administration	2^{nd} administration
1^{st} administration N=(62)	Spearman's ρ	1	.666**
2^{nd} administration N=(62)	Spearman's ρ	.666**	1

4.4. Procedure

After the instrument was translated into Turkish and validity-reliability studies were completed, instrument was administered to preservice teachers in order to investigate the research questions.

Firstly the universities in İstanbul in which mathematics education at secondary and elementary level are offered were determined. Then, head of the programs were determined and their consent were requested to administer the instrument. After the permissions were obtained, the instructors of the teaching methods course of each university were reached since preservice teachers take methods courses during their last year of undergraduate education. Also permissions of these instructors were asked so that instrument could be administered during their method course. After preservice teachers were asked to volunteer for the research, instrument is administered. This process was repeated for all of the secondary school preservice teachers in the sample. However 13 elementary school preservice teachers of two particular university, had to take the test in online since their classes finished earlier and instrument could not be administered during their method courses. Instead they answered the questions in online. Since the results of these 13 participants were not different from the others, they were included to the sample. The rest of the elementary school preservice teachers in the sample took the test in their method courses. Administration took approximately 20 to 30 minutes.

After the data collection, data were subscribed into SPSS and analyzed quantitatively.

4.5. Data Analysis

In this section, participants' correct reasoning skills score, total misconceptions score and 16 sub-scales scores were used. How these scores were obtained, were explained in the Section 4.3 (see *scoring*).

For the first and second research question: The reasoning skills of preservice teachers are tried to be described. Firstly scores of each sub-scale, then correct reasoning skills scale and misconceptions scale score was calculated as it was explained above. Mean and standard deviation for each groups were calculated separately. Also frequency of the most common answer for each question is presented.

For the third research question: After describing the data, in order to compare the group results for each sub-scale, correct reasoning skills score and misconceptions score; independent samples t-test or Mann Whitney U Test was administered. Assumptions of independent samples t-test are that

(i) dependent variable should be interval or ratio data,
(ii) samples should be independent from each other,
(iii) both groups should be normally distributed (Büyüköztürk, 2010).

Dependent variables are the scores and they are in continuous form. Also groups are independent from each other; each individual belongs to only one of the groups. In order to check for normal distribution, Kolmogorov-Smirnov Test was applied. When the groups are normally distributed, independent samples t-test can be conducted. However if the groups are not normally distributed, then non-parametric test of Mann-Whitney U was applied to compare the scores of the groups (Baştürk, 2010). Assumptions of Mann-Whitney U Test are that

(i) dependent variable should be at least ordinal type of data,

(ii) samples should be independent from each other (Büyüköztürk, 2010).

Also there are two sub-scale (*distinguishes between correlation and causation*, *groups can only be compared if they are the same size*). Independent Samples t-test or Mann-Whitney U Test could not administered for these subscales. They are constituted by just one question. So scores of these subscales are dichotomous since answers of the question were coded as 0 or 1. Therefore chi-square test was administered in order to examine differences between groups' scores in these subscales.

5. RESULTS

There are three main sections. These sections include subsections demonstrating results of correct reasoning skills subscales, results of misconceptions subscales and total scores. Findings related to each research question are presented under each of these subsections. These subsections include descriptive statistics.These statistics are given separately for both elementary school and secondary school preservice mathematics teachers related to first and second research questions. Also these subsections include comparisons of results for each groups related to third research question.

5.1. Correct Reasoning Skills

In this section results for each correct reasoning skills subscales presented. These sub-scales were given in Section 4.3 previously as Table 4.1 (See Instrument Section). Results of each subscale includes descriptive statistics which were provided for both groups separately.

In order to examine whether there exists a difference between subscale scores of secondary school preservice mathematics teachers (N=91) and elementary school preservice mathematics teachers (N=82) firstly Kolmogorov-Smirnov normality test was administered for each subscale except *distinguishes between correlation and causation* subscale. Significant value of normality test for each subscale was found .00 in both groups. Thus normal distribution assumption was failed for each subscale scores of both groups. Therefore results of the groups were compared with Mann-Whitney U Test. Statistics of both of these analysis were given as tables under each subscale subsection.

5.1.1. Correctly interprets probabilities

This subscale consists answers to 2^{nd} and 3^{rd} questions. Scores of this subscale is calculated over 1. According to data analysis, scores changed between 0 and 1. So there were participants who could not answer correctly any of the questions and who could answer correctly both of the questions.

Table 5.1. Descriptive Statistics for Correctly interprets probabilities.

	Secondary Level	**Elementary Level**
N	91	82
$\bar{x}$	.747	.683
SD	.273	.309
2d	94.5%	90.2%
3d	54.9%	46.3%
Sig. Value of Normality Test	.00	.00

When Table 5.1 is examined, descriptive statistics for secondary school preservice mathematics teachers (N=91) are found as $\bar{x}$= .747 and SD=.273. For elementary school preservice mathematics teachers (N=82), $\bar{x}$= .683 and SD=.309 were found.

Table 5.2. Mann-Whitney U Test for Correctly interprets probabilities.

Groups	**N**	**Mean Rank**	**Sum of Ranks**	**U**	**Z**	**p**
Elementary Level	82	82.43	6759	3356	-1.288	.198
Secondary Level	91	91.12	8292			

Results of Mann-Whitney U test is given in Table 5.2. Results show that, a significant difference between correctly interprets probabilities subscale scores of between secondary school preservice mathematics teachers and elementary school preservice mathematics teachers was not found (U= 3356; $p > 0.05$). When mean rank is considered, scores of secondary school preservice mathematics teachers are higher than elementary school preservice mathematics teachers but this score difference is not sufficient for statistically significant difference.

5.1.2. Understands how to select an appropriate average

This subscale consists answers to 1^{st}, 4^{th} and 17^{th} questions. Scores of this subscale is calculated over 1. According to data analysis, scores changed between 0 and 1. So there were participants who could not answer correctly any of the questions and who could answer correctly all of the questions.

Table 5.3. Descriptive Statistics for Understands how to select an appropriate average.

	Secondary Level	**Elementary Level**
N	91	82
$\bar{x}$	.454	.447
SD	.312	.311
1d	30.8%	34.1%
4c	46.2%	36.6%
17c	59.3%	63.4%
Sig. Value of Normality Test	.00	.00

As Table 5.3 is examined, descriptive statistics for secondary school preservice mathematics teachers (N=91) are found as $\bar{x}$=.454 and SD=.312. For elementary school preservice mathematics teachers (N=82), $\bar{x}$=.447 and SD=.311 are found.

Results of Mann-Whitney U test is given in Table 5.4. When the results are examined, a significant difference in *understands how to select an appropriate average*

Table 5.4. Mann-Whitney U Test for Understands how to select an appropriate average.

Groups	N	Mean Rank	Sum of Ranks	U	Z	p
Elementary Level	82	86.34	7079.50	3676.5	-.174	.862
Secondary Level	91	87.60	7971.50			

subscale scores of between secondary and elementary school preservice mathematics teachers was not found (U= 3676.5; $p > 0.05$). When mean rank is considered, scores of secondary school preservice mathematics teachers are higher than elementary school preservice mathematics teachers but this score difference is not sufficient for statistically significant difference.

5.1.3. Correctly computes probability.

Table 5.5. Descriptive Statistics for Correctly computes probability.

	Secondary Level	Elementary Level
N	91	82
$\bar{x}$	.314	.337
SD	.179	.211
8c	95.6%	93.9%
13a	22.0%	31.7%
18b	11.0%	14.6%
19a	12.1%	13.4%
20b	16.5%	14.6%
Sig. Value of Normality Test	.00	.00

This subscale consists answers to the 8^{th}, 13thth, 18^{th}, 19^{th} and 20^{th} questions. Scores of this subscale is calculated over 1. According to data analysis, scores changed between .20 and 1. So all of the participants answered at least one question which is most likely 8^{th} since its most common answer is the correct answer and its percentage is 95.6 and 93.9 for secondary and elementary level respectively. Also there were participants who could answer correctly all of the questions of this subscale.

When Table 5.5 is examined, descriptive statistics for secondary school preservice mathematics teachers (N=91) were found as $\bar{x}$= .314 and SD=.179. For elementary school preservice mathematics teachers (N=82), $\bar{x}$=.337 and SD=.211 were found.

Table 5.6. Mann-Whitney U Test for Correctly computes probability.

Groups	**N**	**Mean Rank**	**Sum of Ranks**	**U**	**Z**	**p**
Elementary Level	82	88.88	7288	3577	-.538	.591
Secondary Level	91	85.31	7763			

Results of Mann-Whitney U test is given in Table 5.6. When the results are examined, a significant difference in *correctly computes probability* subscale scores of between secondary school and elementary school preservice mathematics teachers was not found (U= 3577; $p > 0.05$). When mean rank is considered, scores of elementary school preservice mathematics teachers are higher than secondary school preservice mathematics teachers but this score difference is not sufficient for statistically significant difference.

5.1.4. Understands independence

This subscale consists answers to the 9^{th}, 11^{th} and alternative c, d, f of 10^{th} questions. Scores of this subscale is calculated over 1.67. According to data analysis, scores changed between 0 and 1.67. So there were participants who could not answer correctly any of the questions and who could answer correctly all of the questions.

When Table 5.7 is examined, descriptive statistics for secondary school preservice

Table 5.7. Descriptive Statistics for Understands independence.

	Secondary Level	**Elementary Level**
N	91	82
$\bar{x}$	1.187	1.150
SD	.345	.426
9e	93.4%	92.7%
10c	61.5%	65.9%
10d	31.9%	35.4%
10f	76.9%	63.4%
11e	92.3%	87.8%
Sig. Value of Normality Test	.00	.00

mathematics teachers (N=91) were found as $\bar{x}$=1.187 and SD=.345. For elementary school preservice mathematics teachers (N=82), $\bar{x}$= 1.150 and SD=.426 were found.

Table 5.8. Mann-Whitney U Test for Understands independence.

Groups	**N**	**Mean Rank**	**Sum of Ranks**	**U**	**Z**	**p**
Elementary Level	82	85.95	7047.50	3644.5	-.273	.785
Secondary Level	91	87.95	8003.50			

Results of Mann-Whitney U test is given in Table 5.8. When the results are examined, a significant difference in *understands independence* subscale scores of between secondary and elementary school preservice mathematics teachers was not found (U=3644.5; $p > 0.05$). When mean rank is considered, scores of secondary school preservice mathematics teachers are higher than elementary school preservice mathematics teachers but this score difference is not sufficient for statistically significant difference.

5.1.5. Understands sampling variability

This subscale consists answers to the 14^{th} and 15^{th} questions. Scores of this subscale is calculated over 1. According to data analysis, scores changed between 0 and 1. So there were participants who could not answer correctly any of the questions and who could answer correctly both of the questions.

Table 5.9. Descriptive Statistics for Understands sampling variability.

	Secondary Level	**Elementary Level**
N	91	82
$\bar{x}$	.148	.122
SD	.253	.243
14b	24.2%	14.6%
15d	5.5%	9.8%
Sig. Value of Normality Test	.00	.00

When Table 5.9 is examined, descriptive statistics for secondary school preservice mathematics teachers (N=91) were found as $\bar{x}$=.148 and SD=.253. For elementary school preservice mathematics teachers (N=82), $\bar{x}$=.122 and SD=.243 were found.

Table 5.10. Mann-Whitney U Test for Understands sampling variability.

Groups	**N**	**Mean Rank**	**Sum of Ranks**	**U**	**Z**	**p**
Elementary Level	82	84.57	6935	3532	-.805	.421
Secondary Level	91	89.19	8116			

Results of Mann-Whitney U test is given in Table 5.10. When the results are examined, a significant difference in *understands sampling variability* subscale scores of between secondary and elementary school preservice mathematics teachers was not found (U= 3532; $p > 0.05$). When mean rank is considered, scores of secondary school preservice mathematics teachers are higher than elementary school preservice mathematics teachers but this score difference is not sufficient for statistically significant difference.

5.1.6. Distinguishes between correlation and causation

This subscale consists answers to the only 16^{th} question. Scores of this subscale is calculated over 1. According to data analysis, scores changed between 0 and 1.

Table 5.11. Descriptive Statistics for Distinguishes between correlation and causation.

	Secondary Level	**Elementary Level**
N	91	82
$\bar{x}$	.319	.232
SD	.469	.425
16c	31.9%	23.2%

When Table 5.11 is examined, descriptive statistics for secondary school preservice mathematics teachers (N=91) were found as $\bar{x}$=.319 and SD=.469. For elementary school preservice mathematics teachers (N=82), $\bar{x}$=.232 and SD=.425 were found.

Table 5.12. Chi-square Test for Distinguishes between correlation and causation.

	16					
	True	False	Total	X^2	df	p
Elementary School	19	63	82	1.628	1	.202
Secondary School	29	62	91			
Total	48	125	173			

Normality test was not administered since subscale includes only one question such that data is dichotomous. Therefore Chi-Square test was administered in order to examine whether there exists a difference between *distinguishes between correlation and causation* subscale scores of secondary school preservice mathematics teachers (N=91) and elementary school preservice mathematics teachers (N=82). Results of Chi-Square test is given in Table 5.12. When the results are examined, a significant difference between *distinguishes between correlation and causation* subscale scores of secondary and elementary school preservice mathematics teachers was not found (X^2= 1.628; $p > 0.05$).

5.1.7. Correctly interprets two-way tables

This subscale consists answer to the 5^{th} question. Scores of this subscale is calculated over 2 since it has two parts. According to data analysis, scores changed between 0, 1 and 2. So there were participants who could not understand how the two-way table works and interpret the table in comparing the groups. When Table C.1 and Table C.2 are examined, 53.7% and 71.4% of elementary and secondary school preservice mathematics teachers could answered this question correctly respectively.

Table 5.13. Descriptive Statistics for Correctly interprets two-way tables.

	Secondary Level	**Elementary Level**
N	91	82
$\bar{x}$	1.406	1.207
SD	.856	.885
51d	71.4%	51.2%
Sig. Value of Normality Test	.00	.00

When Table 5.13 is examined, descriptive statistics for secondary school preservice mathematics teachers (N=91) were found as $\bar{x}$= 1.406 and SD=.856. For elementary school preservice mathematics teachers (N=82), $\bar{x}$=1.207 and SD=.885 were found.

Table 5.14. Mann-Whitney U Test for Correctly interprets two-way tables.

Groups	N	Mean Rank	Sum of Ranks	U	Z	p
Elementary Level	82	81.29	6666	3263	-1.613	.107
Secondary Level	91	92.14	8385			

Results of Mann-Whitney U test is given in Table 5.14. When the results are examined, a significant difference between *correctly interprets two-way tables* subscale scores of secondary and elementary school preservice mathematics teachers was not found (U= 3263; $p > 0.05$). When mean rank is considered, scores of secondary school preservice mathematics teachers are higher than elementary school preservice mathematics teachers but this difference is not sufficient for statistically significant difference.

5.1.8. Understands importance of large samples

Table 5.15. Descriptive Statistics for Understands importance of large samples.

	Secondary Level	Elementary Level
N	91	82
$\bar{x}$	.861	.833
SD	.352	.356
6b	63.7%	63.4%
7e	41.8%	41.5%
7f	83.5%	73.2%
12b	69.2%	72%
Sig. Value of Normality Test	.00	.00

This subscale consists answers to the 6^{th}, 12^{th} and alternatives e and f of 7^{th} questions. Scores of this subscale is calculated over 1.33. According to data analysis, scores changed between 0 and 1.33. So there were participants who could not answer correctly any of the questions and who could answer correctly all of the questions.

As Table 5.15 is examined, descriptive statistics for secondary school preservice mathematics teachers (N=91) were found as $\bar{x}$=.861 and SD=.352. For elementary school preservice mathematics teachers (N=82), $\bar{x}$=.833 and SD=.356 were found.

Table 5.16. Mann-Whitney U Test for Understands importance of large samples.

Groups	**N**	**Mean Rank**	**Sum of Ranks**	**U**	**Z**	**p**
Elementary Level	82	85.05	6974.50	3571.5	-.505	.614
Secondary Level	91	88.75	8076.50			

Results of Mann-Whitney U test is given in Table 5.16. When the results are examined, a significant difference between *understands importance of large samples* subscale scores of secondary and elementary school preservice mathematics teachers was not found (U=3571; $p > 0.05$). When mean rank is considered, scores of secondary school preservice mathematics teachers are higher than elementary school preservice mathematics teachers but this difference is not sufficient for statistically significant difference.

5.2. Misconceptions

In this section descriptive statistics of each misconception subscales are presented for both groups separately. These sub-scales were given in Section 4.3 previously as Table 4.2 (See Instrument Section). Then results of the groups are compared. Statistics of both of these analysis are given as tables for each subscale.

In order to examine whether there exists a difference between misconception subscales' scores of secondary school preservice mathematics teachers (N=91) and elementary school preservice mathematics teachers (N=82) firstly Kolmogorov-Smirnov normality test was administered for each subscale except the *Groups can only be compared if they are the same size* subscale. Significant value of normality test of each subscale was found .00 for both groups. Thus normal distribution assumption was failed for each subscale scores of both groups. Therefore results of the groups were compared with Mann-Whitney U Test. Statistics of both of these analysis are given as tables under each subscale subsection.

5.2.1. Misconceptions involving averages

Table 5.17. Percentages for Misconceptions involving averages.

	Secondary Level	**Elementary Level**
1a	30.8%	35.4%
1c	30.8%	20.7%
4a	19.8%	18.3%
4b	30.8%	40.2%
15b	6.6%	7.3%
15f	63.7%	54.9%
17a	6.6%	6.1%
17e	7.7%	7.3%

This subscale consists answers to alternatives a and c of 1^{st}, alternatives a and b of 4^{th}, alternatives b and f of 15^{th} and alternatives a and e of 17^{th} questions. Scores of this subscale is calculated over 2. According to data analysis, scores changed between 0 and 2. So there were participants who did not have any misconceptions involving averages and who had all of the misconceptions.

When Table 5.18 is examined, descriptive statistics for secondary school preservice mathematics teachers (N=91) were found as $\bar{x}$=.480 and SD=.224. For elementary school preservice mathematics teachers (N=82), $\bar{x}$=.475 and SD=.256 were found.

Table 5.18. Descriptive Statistics for Misconceptions involving averages.

	Secondary Level	**Elementary Level**
N	91	82
$\bar{x}$	.480	.475
SD	.224	.256
Sig. Value of Normality Test	.00	.00

Table 5.19. Mann-Whitney U Test for Misconceptions involving averages.

Groups	**N**	**Mean Rank**	**Sum of Ranks**	**U**	**Z**	**p**
Elementary Level	82	87.32	7160	3705	-.083	.934
Secondary Level	91	86.71	7891			

Results of Mann-Whitney U test is given in Table 5.19. When the results are examined, a significant difference between *misconceptions involving averages* subscale scores of secondary and elementary school preservice mathematics teachers was not found (U=3705; $p > 0.05$). When mean rank is considered, scores of elementary school preservice mathematics teachers are higher than secondary school preservice mathematics teachers but this difference is not sufficient for statistically significant difference.

5.2.2. Outcome orientation misconception

This subscale consists answers to alternative e of 2^{nd}, alternative a and b of 3^{rd}, alternative a, b and d of 11^{th}, alternative c of 12^{th}, and alternative b of 13^{th} questions. Scores of this subscale is calculated over 1.60. According to data analysis, scores changed between 0 and .60. So there were participants who did not have outcome orientation misconception and there was no one who selected all of these alternatives.

Table 5.20. Percentages for Outcome orientation misconception.

	Secondary Level	**Elementary Level**
2e	2.2%	1.2%
3a	34.1%	35.4%
3b	0%	7.3%
11a	0%	1.2%
11b	0%	1.2%
11d	0%	2.4%
12c	11%	6.1%
13b	45.1%	37.8%

When Table 5.21 is examined, descriptive statistics for secondary school preservice mathematics teachers (N=91) were found as $\bar{x}$=.185 and SD=.164. For elementary school preservice mathematics teachers (N=82), $\bar{x}$=.185 and SD=.156 were found.

Table 5.21. Descriptive Statistics for Outcome orientation misconception.

	Secondary Level	**Elementary Level**
N	91	82
$\bar{x}$	.185	.185
SD	.164	.156
Sig. Value of Normality Test	.00	.00

Results of Mann-Whitney U test is given in Table 5.22. When the results are examined, a significant difference between *outcome orientation misconception* subscale scores of secondary and elementary school preservice mathematics teachers was not found (U=3711.5; $p > 0.05$). When mean rank is considered, scores of elementary school preservice mathematics teachers are higher than secondary school preservice mathematics teachers but this difference is not statistically significant.

Table 5.22. Mann-Whitney U Test for Outcome orientation misconception.

Groups	N	Mean Rank	Sum of Ranks	U	Z	p
Elementary Level	82	87.24	7153.50	3711.5	-.063	.950
Secondary Level	91	86.79	7897.50			

5.2.3. Good samples have to represent a high percentage of the population

This subscale consists answers to alternatives b and c of 7^{th} and alternative a and d of 16^{th} questions. Scores of this subscale is calculated over 2. According to data analysis, scores changed between 0 and 2. So there were participants who had the misconception strongly and who did not have at all.

Table 5.23. Descriptive Statistics for Good samples have to represent a high percentage of the population.

	Secondary Level	Elementary Level
N	91	82
$\bar{x}$	.588	.470
SD	.520	.523
7b	30.8%	25.6%
7c	44%	36.6%
16a	18.7%	15.9%
16d	24.2%	15.9%
Sig. Value of Normality Test	.00	.00

When Table 5.23 is examined, descriptive statistics for secondary school preservice mathematics teachers (N=91) were found as $\bar{x}$=.588 and SD=.520. For elementary school preservice mathematics teachers (N=82), $\bar{x}$=.470 and SD=.523 were found.

Table 5.24. Mann-Whitney U Test for Good samples have to represent a high percentage of the population.

Groups	N	Mean Rank	Sum of Ranks	U	Z	p
Elementary Level	82	80.27	6582.50	3179.5	-1.763	.078
Secondary Level	91	93.06	8468.50			

Results of Mann-Whitney U test is given in Table 5.24. When the results are examined, a significant difference between *good samples have to represent a high percentage of the population* subscale scores of secondary and elementary school preservice mathematics teachers was not found (U=3179.5; $p > 0.05$). When mean rank is considered, scores of secondary school preservice mathematics teachers are higher than elementary school preservice mathematics teachers but this difference is not statistically significant.

5.2.4. Law of small numbers

Table 5.25. Descriptive Statistics for Law of small numbers.

	Secondary Level	Elementary Level
N	91	82
$\bar{x}$	.335	.402
SD	.342	.309
12a	16.5%	19.5%
14c	50.5%	61%
Sig. Value of Normality Test	.00	.00

This subscale consists answers to alternatives a of 12^{th} and alternative c of 14^{th} questions. Scores of this subscale is calculated over 1. According to data analysis, scores changed between 0 and 1. So there were participants who had the misconception strongly and who did not have at all.

When Table 5.25 is examined, descriptive statistics for secondary school preservice mathematics teachers (N=91) were found as $\bar{x}$=.335 and SD=.342. For elementary school preservice mathematics teachers (N=82), $\bar{x}$=.402 and SD=.309 were found.

Table 5.26. Mann-Whitney U Test for Law of small numbers.

Groups	N	Mean Rank	Sum of Ranks	U	Z	p
Elementary Level	82	92.55	7589	3276	-1.531	.126
Secondary Level	91	82	7462			

Results of Mann-Whitney U test is given in Table 5.26. When the results are examined, a significant difference between *law of small numbers* subscale scores of secondary and elementary school preservice mathematics teachers was not found (U=3276; $p > 0.05$). When mean rank is considered, scores of elementary school preservice mathematics teachers are higher than secondary school preservice mathematics teachers but this difference is not statistically significant.

5.2.5. Representativeness misconception

This subscale consists answers to alternatives a, b and d of 9^{th}, alternative e of 10^{th}, and alternative c of 11^{th} questions. Scores of this subscale is calculated over 1.67. According to data analysis, scores changed between 0 and .67. So there were participants who did not have the misconception and who did not select all of the alternatives written above.

When Table 5.27 is examined, descriptive statistics for secondary school preservice mathematics teachers (N=91) were found as $\bar{x}$=.044 and SD=.142. For elementary school preservice mathematics teachers (N=82), $\bar{x}$= .053 and SD=.143 were found.

Table 5.27. Descriptive Statistics for Representativeness misconception.

	Secondary Level	**Elementary Level**
N	91	82
$\bar{x}$	.044	.053
SD	.142	.143
9a	1.1%	1.2%
9b	1.1%	1.2%
9d	2.2%	3.7%
10e	2.2%	3.7%
11c	6.6%	6.1%
Sig. Value of Normality Test	.00	.00

Table 5.28. Mann-Whitney U Test for Representativeness misconception.

Groups	**N**	**Mean Rank**	**Sum of Ranks**	**U**	**Z**	**p**
Elementary Level	82	88.51	7258	3607	-.680	.497
Secondary Level	91	85.64	7793			

Results of Mann-Whitney U test is given in Table 5.28. When the results are examined, a significant difference between *representativeness misconception* subscale scores of secondary and elementary school preservice mathematics teachers was not found (U=3607; $p > 0.05$). When mean rank is considered, scores of elementary school preservice mathematics teachers are higher than secondary school preservice mathematics teachers but this difference is not statistically significant.

5.2.6. Correlation implies causation

This subscale consists answers to alternatives b and e of 16^{th} questions. Scores of this subscale is calculated over 2. According to data analysis, scores changed between 0 and 2. So there were participants who had the misconception strongly and who did not have at all.

Table 5.29. Descriptive Statistics for Correlation implies causation.

	Secondary Level	**Elementary Level**
N	91	82
$\bar{x}$	.780	.927
SD	.772	.813
16b	35.2%	46.3%
16e	42.9%	46.3%
Sig. Value of Normality Test	.00	.00

When Table 5.29 is examined, descriptive statistics for secondary school preservice mathematics teachers (N=91) were found as $\bar{x}$=.780 and SD=.772. For elementary school preservice mathematics teachers (N=82), $\bar{x}$=.927 and SD=.813 were found.

Table 5.30. Mann-Whitney U Test for Correlation implies causation.

Groups	**N**	**Mean Rank**	**Sum of Ranks**	**U**	**Z**	**p**
Elementary Level	82	91.44	7498	3367	-1.181	.237
Secondary Level	91	83.00	7553			

Results of Mann-Whitney U test is given in Table 5.30. When the results are examined, a significant difference between *correlation implies causation* subscale scores of secondary and elementary school preservice mathematics teachers was not found

(U=3711.5; $p > 0.05$). When mean rank is considered, scores of elementary school preservice mathematics teachers are higher than secondary school preservice mathematics teachers but this difference is not statistically significant.

5.2.7. Equiprobability bias

Table 5.31. Descriptive Statistics for Equiprobability bias.

	Secondary Level	**Elementary Level**
N	91	82
$\bar{x}$	.668	.637
SD	.289	.312
13c	29.7%	23.2%
18a	75.8%	70.7%
19d	84.6%	81.7%
20d	76.9%	79.3%
Sig. Value of Normality Test	.00	.00

This subscale consists answers to alternative c of 13^{th}, alternative a of 18^{th}, alternative d of 19^{th} and alternative d of 20^{th} questions. Scores of this subscale is calculated over 1. According to data analysis, scores changed between 0 and 1. So there were participants who had the misconception strongly and who did not have at all. Also when Table C.1 and Table C.2 is examined, it is seen that these alternatives signaling equiprobability bias are the most common answer of 18^{th}, 19^{th} and 20^{th} questions for both groups with a percentage of 70 and over.

When Table 5.31 is examined, descriptive statistics for secondary school preservice mathematics teachers (N=91) were found as $\bar{x}$=.668 and SD=.289. For elementary school preservice mathematics teachers (N=82), $\bar{x}$=.637 and SD=.312 were found.

Results of Mann-Whitney U test is given in Table 5.32. When the results are

Table 5.32. Mann-Whitney U Test for Equiprobability bias.

Groups	N	Mean Rank	Sum of Ranks	U	Z	p
Elementary Level	82	84.85	6958	3555	-.585	.558
Secondary Level	91	88.93	8093			

examined, a significant difference between *equiprobability bias* subscale scores of secondary and elementary school preservice mathematics teachers was not found (U=3555; $p > 0.05$). When mean rank is considered, scores of secondary school preservice mathematics teachers are higher than elementary school preservice mathematics teachers but this difference is not statistically significant.

5.2.8. Groups can only be compared if they are the same size

Table 5.33. Descriptive Statistics for Groups can only be compared if they are the same size.

	Secondary Level	Elementary Level
N	91	82
$\bar{x}$	.330	.329
SD	.473	.473
6a	33%	33%

This subscale consists answers to alternative a of 6^{th} question. Scores of this subscale is calculated over 1. According to data analysis, scores changed between 0 and 1. So there were participants who had the misconception and who did not have at all.

When Table 5.33 is examined, descriptive statistics for secondary school preservice mathematics teachers (N=91) were found as $\bar{x}$=.330 and SD=.473. For elementary school preservice mathematics teachers (N=82), $\bar{x}$=.329 and SD=.473 were found. Also when Table C.1 and Table C.2 is examined, it is seen that correct answer is the most common answer of 6^{th} question for both groups with a percentage of around 63.

Table 5.34. Chi-square Test for Groups can only be compared if they are the same size.

	6		Total	X^2	df	p
	True	False				
Elementary School	27	55	82	.00	1	.996
Secondary School	30	61	91			
Total	57	116	173			

Normality test was not administered since this subscale includes only one question with dichotomous type of data. Therefore Chi-Square test was administered in order to examine whether there exists a difference between *Groups can only be compared if they are the same size* subscale scores of secondary school preservice mathematics teachers (N=91) and elementary school preservice mathematics teachers (N=82). Results of Chi-Square test is given in Table 5.34. When the results are examined, a significant difference between *Groups can only be compared if they are the same size* subscale scores of secondary and elementary school preservice mathematics teachers was not found (X^2=.00; $p > 0.05$).

5.3. Total Scores

In this section descriptive statistics of total correct reasoning skills and misconceptions are given for both groups separately. Then results of the groups are compared. Statistics of both of these analysis are given as tables.

5.3.1. Correct Reasoning Skills Score

Scores of each subscale generating correct reasoning skills score are added in order to obtain Correct Reasoning Skills Score as it is explained in Section 4.5. Scores of this scale is calculated over 10 that is if someone have all of these correct reasoning, then s/he can get at most 10 points for total Correct Reasoning Skills score. Results showed that, scores changed between 2.03 and 8.07 for secondary level. For elementary level scores changed between 1.87 and 8.37. So there was not anyone who answered all of the questions correctly and also incorrectly.

Table 5.35. Descriptive Statistics for Correct Reasoning Skills Score.

	Secondary Level	**Elementary Level**
N	91	82
$\bar{x}$	5.437	5.011
SD	1.507	1.609
Min	2.03	1.87
Max	8.07	8.37
Sig. Value of Normality Test	.012	.200

When Table 5.35 is examined, descriptive statistics for secondary school preservice mathematics teachers (N=91) were found as $\bar{x}$=5.437 and SD=1.507. For elementary school preservice mathematics teachers (N=82), $\bar{x}$=5.011 and SD=1.609 were found.

Table 5.36. Mann-Whitney U Test for Correct Reasoning Skills Score.

Groups	**N**	**Mean Rank**	**Sum of Ranks**	**U**	**Z**	**p**
Elementary Level	82	79.67	6533	3130	-1.827	.068
Secondary Level	91	93.60	8518			

In order to examine whether there exists a difference between *Correct Reasoning Skills* scale scores of secondary school preservice mathematics teachers (N=91) and elementary school preservice mathematics teachers (N=82), firstly Kolmogorov-Smirnov normality test was administered. As Table 5.35 shows that significant value of normality test is .012 and .200 for secondary and elementary level groups respectively. So normal distribution assumption was failed for secondary school preservice mathematics teachers. Therefore Mann-Whitney U Test was administered. Results of Mann-Whitney U test are given in Table 5.36. When the results are examined, a significant difference between *Correct Reasoning Skills* scale scores of secondary and elementary school preservice mathematics teachers was not found (U=3711.5; $p > 0.05$). When mean rank is considered, scores of secondary school preservice mathematics teachers are higher than elementary school preservice mathematics teachers but this difference is not statistically significant.

5.3.2. Misconceptions Scores

Table 5.37. Descriptive Statistics for Misconceptions Score.

	Secondary Level	**Elementary Level**
N	91	82
$\bar{x}$	3.410	3.479
SD	1.148	1.277
Min	.45	1.00
Max	6.67	6.70
Sig. Value of Normality Test	.200	.043

Scores of each subscales generating Misconceptions score are added in order to obtain Misconceptions Score as it is explained in Section 4.5. Scores of this scale are calculated over 12.27 that is if someone have all of these misconceptions, then s/he can get at most 12.27 points for total misconceptions score. According to results, scores changed between .45 and 6.67 for secondary level. For elementary level scores changed between 1.00 and 6.70. So there was not anyone who had all of the misconceptions and who did not have any of them.

When Table 5.37 is examined, descriptive statistics for secondary school preservice mathematics teachers (N=91) were found as $\bar{x}$=3.410 and SD=1.148. For elementary school preservice mathematics teachers (N=82), $\bar{x}$=3.479 and SD=1.277 were found.

Table 5.38. Mann-Whitney U Test for Misconceptions Score.

Groups	**N**	**Mean Rank**	**Sum of Ranks**	**U**	**Z**	**p**
Elementary Level	82	87.57	7181	3684	.143	.886
Secondary Level	91	86.48	7870			

In order to examine whether there exists a difference between *Misconceptions* scale scores of secondary school preservice mathematics teachers (N=91) and elementary school preservice mathematics teachers (N=82), firstly Kolmogorov-Smirnov normality test was administered. As Table 5.37 shows that significant value of normality test is .200 and .043 for secondary and elementary level groups respectively. So normal distribution assumption was failed for elementary school preservice mathematics teachers. Therefore Mann-Whitney U Test was administered. Results of Mann-Whitney U test is given in Table 5.38. When the results are examined, a significant difference between *Misconceptions* scale scores of secondary and elementary school preservice mathematics teachers was not found (U=3684.5; $p > 0.05$). When mean rank is considered, scores of elementary school preservice mathematics teachers are higher than secondary school preservice mathematics teachers but this difference is not statistically significant.

6. CONCLUSION AND DISCUSSION

This chapter includes four sections. First results are discussed considering the research questions. Secondly limitations of the study is presented. Lastly, implications for teaching and further research regarding conclusions and limitations are discussed.

Research questions of the study as follows:

(i) What are the statistical reasoning types of elementary school preservice mathematics teachers?
(ii) What are the statistical reasoning types of secondary school preservice mathematics teachers?
(iii) Is there any difference between statistical reasoning types of preservice secondary school mathematics teachers and elementary school mathematics teachers?

Conclusions regarding first and second research questions that is correct and incorrect reasoning types of elementary and secondary school preservice teachers are discussed under each subscale according to results and previous researches. After these correct and incorrect reasoning types are discussed separately, whether is there a difference between statistical reasoning types of preservice secondary school mathematics teachers and elementary school mathematics teachers or not are discussed based on the results and previous researches. Lastly, correct reasoning skills and misconceptions are discussed as a whole.

6.1. Correct Reasoning Skills

In this section conclusions for each correct reasoning skills subscales were presented. These sub-scales were given in Section 4.3 previously as Table 4.1 (See Instrument Section).

6.1.1. Correctly interprets probabilities

Answers to the 2^{nd} and 3^{rd} questions constitue this subscale scores. If each participant had answered these questions correctly, the mean would have been 1. Results show that both elementary ($\bar{x}$=.747) and secondary ($\bar{x}$=.683) school preservice mathematics teachers got relatively high score in correctly interpreting probabilities. Although 2^{nd} questions answered correctly by over 90% of both groups, participants were not successful in 3^{rd} question as they were in 2^{nd} question. Therefore the difference of the group means and 1 might be resulted from 3^{rd} question which assess also outcome orientation. It is discussed in detail under *Outcome orientation misconception* subscale in Section 6.2 part. Interpretation requires them making sense of a message, which includes a quantitative statement, in order to use as a cretarion for an opinion or judgment (Gal, 2005). For instance, in our case they were expected to interpret 15% chance of getting rash in the 2^{nd} question. They are expected to understand probabilistic statement of 15% as there is a set of 100 people using a medication and 15 of them get rash. Accordingto results, it seems that both elementary and secondary school preservice mathematics teachers could understand the probabilistic statement correctly.

6.1.2. Understands how to select an appropriate average

Answers to the 1^{st}, 4^{th} and 17^{th} questions constitue this subscale scores. If each participant had answered these questions correctly, the mean would have been 1. However means were .447 for elementary and .454 for secondary school preservice mathematics teachers. So, data show that, there is lack of understanding among both elementary and secondary school preservice teachers regarding how to select an appropriate average. When the frequencies of the alternatives of these questions were examined, the percentage of correct answers given to 1^{st} and 4^{th} questions is low. However most of the elementary (63.4%) and secondary (59.3%) school preservice mathematics teachers could answer 17^{th} question correctly. In 1^{st} and 4^{th} questions, participants were asked to find appropriate average of a given data set but in 17^{th} question mean of the data set was given to participants and they were asked to interpret it. Determinig how to find an average of a data set rather than interpreting average of a data set might be more difficult for both elementary and secondary school preservice mathematics teachers. Similarly, Groth and Bergner (2006) found that preservice elementary school mathematics teachers had difficulties in determining appropriate measures of center according to given data. Therefore, taking all of these into consideration, preservice teachers' understanding how to select an appropriate average should be regarded cautiously.

6.1.3. Correctly computes probability

Answers to the 8^{th}, 13thth, 18^{th}, 19^{th} and 20^{th} questions constitue this subscale scores. If each participant had answered these questions correctly, the mean would have been 1. However, data showed that, both elementary ($\bar{x}$=.337) and secondary ($\bar{x}$=.314) school preservice teachers' reasoning in computing probability was very low. In particular, although the 8^{th} question was answered correctly by 93.9% of elementary and 95.6% of secondary school preservice mathematics teachers, for the other questions assessing combinatorial reasoning, very few of the participants (less than 15%) in both groups provided correct answers. High scores in the 8^{th} question might be related with the context of the question. The context involved drawing out marbles from two boxes. This context is very similar to the probability context used in probabilty questions in Turkish mathematics books. On the other hand, questions assessing combinatorial reasoning required participants to distinguish between for instance the sample space of rolling two 5s and one 5 and one 6. Corresponding with earlier research results (Batanero, Godino, & Cañizares, 2005), both elementary and secondary prospective teachers failed to reason correctly in these problems. Therefore, their relatively low scores in combinatorial reasoning questions took their subscale mean scores to .337 and .314 for elementary and secondary level respectively. Although ressults also showed that there was not statistically significant difference between secondary and elementary school preservice teachers' scores of this subscale, the results regarding each question alerts for the need in terms of combinatorial reasoning to be taken into consideration. That is, results indicate that both elementary and secondary prospective teachers have difficulty in reasoning on problems regarding combinatorial reasoning and this sugests that their education might include their preparation for combinatorial reasoning prior to their teaching.

6.1.4. Understands independence

Answers to the 9^{th}, 11th^{th}, and alternative c, d, f of 10^{th} questions constitue this subscale scores. If each participant had answered these questions correctly, the mean would have been 1.67. Still, results show that both elementary ($\bar{x}$=1.150) and secondary ($\bar{x}$=1.187) school preservice mathematics teachers got high scores in understanding independence. This subscale assesses that for instance in an event of flipping a fair coin many times, outcome of a trial is independent from each other and do not affect the next one (Walle, Karp, & Bay-Williams, 2007). It seems that both elementary and secondary school preservice matematics teachers are quite successful in understanding this notion. As Table C.1 and Table C.2 show, the most common alternatives chosen by both elementary and secondary school preservice teachers for these questions are correct. As a matter of fact, the percentage of these most common answers are higher than 60% except alternative d of 10^{th} question. Alternative d of 10^{th} question was answered correctly by 35.4% of elementary and 31.9 of secondary school preservice teachers. Since the other questions answered correctly are at high percentages, the mean sores are not affected much by this low percentage of alternative d of 10^{th} question. Besides, data shows that there was not statistically significant difference was not found between scores of secondary and elementary school preservice mathematics teachers.

6.1.5. Understands sampling variability

Answers to the 14^{th} and 15^{th} questions constitues this subscale scores. If each participant had answered these questions correctly, the mean would have been 1. However their mean were found as .122 and .148 for elementary and secondary school preservice teachers respectively. So data shows that both elementary and secondary school preservice mathematics teachers' understanding of sampling variability is very low. When Table 5.9 is examined, precentages of both elementary and secondary school preservice teachers who chose correct answer are very low. Furthermore the most common alternatives chosen by them signal misconceptions (see Table C.1 and Table C.2). The misconception in 14^{th} question is law of small numbers. This was discussed in detail under *law of small numbers* subscale in Section 6.2 part. In 15^{th} question, data also shows that they did not take into account variation and compare the groups based on average. This result is parallel with findings of Leavy (2006) such that preservice mathematics teachers focus too much on average and do not consider variation in their comparisons. Furthermore, preservice teachers might tend to compare data sets through shape of the distributions (Makar & Confrey, 2005). In 15^{th} question the most common answer indicates that both elemantary and secondary school preservice teachers might have compared the data sets according to shape of the distribution. Therefore preservice teachers' reasoning in comparing data sets needs to be improved so that their understanding in sampling variability could be developed.

6.1.6. Distinguishes between correlation and causation

Answers to the 16^{th} question constitues this subscale scores. If each participant had answered the question correctly, the mean would have been 1. However group means were found as .232 and .319 for elementary and secondary school preservice teachers respectively. When Table C.1 and Table C.2 is examined, 23.2% of elementary and 31.9% of secondary school preservice teachers chose correct answer. According to data, the number of both elementary and secondary school preservice mathematics teachers distinguishing between correlation and causation is low. So most of them could not discriminate between variables relating to each other and variables causing each other. In the context of the question few of them could realized that even though there exists a relationship between variables, it does not mean that these variables cause each other. Furthermore in both groups the most common answer chosen by participants signaling correlation implies causation misconception.

6.1.7. Correctly interprets two-way tables

Answer to the 5^{th} question constitues this subscale scores. 5^{th} question has two parts. First part assesses to be able to read the two-way table and second part assess interpreting the two-way table. Therefore if each participant had answered this question correctly, the mean would have been 2. Data shows that both elementary ($\bar{x}$=1.207) and secondary ($\bar{x}$= 1.406) school preservice teachers are quite successful in interpreting two way tables. Both of the groups most commonly chose the correct answer. Although percentage of secondary school preservice teachers (71.4%) who answered correctly is higher than elementary school preservice teachers (51.2%), there is not statistically significant difference between their scores of this subscale. However these results are inconsistent with Estrada's findings (2006). In the study of Estrada (2006), it was found that preservice teachers have difficulties in understanding two-way tables.

6.1.8. Understands importance of large samples

Answers to the 6^{th}, 12thth, and alternative e and f of 7^{th} questions constitues this subscale scores. If each participant had answered these questions correctly, the mean would have been 1,33. Results show that both elementary ($\bar{x}$=.833) and secondary ($\bar{x}$=.861) school preservice teachers are quite successful in understanding importance of large samples. Sample size should be large enough to guarantee the representativeness of the population since representativeness is important to make unbiased inferences about the population (Watson & Moritz, 2000). In particular percentage of correct answers to these questions are quite high. For instance, precentage of both elementary (63.4%) and secondary (63.7%) school preservice mathematics teachers answered correctly to 6^{th} question. Correct answer to 6^{th} question implies that both elementary and secondary school preservice mathematics teachers are able to focus sample size while making inferences. Also answers to 12^{th} question shows that quite high percentage of elementary (72%) and secondary (69.2%) school preservice mathematics teachers can make their decisions according to scientific results rather than anecdotal events. So both elementary and secondary school preservice mathematics teachers are aware of data beats anecdotes (Garfield & Ben-Zvi, 2008).

Furthermore data shows that they are aware of the critical notions to generalize the results to the population to some extent. They are more successful in considering representativeness but not in randomness. Randomness and representativeness have to be considered in order to generalize the results to the population (Watson & Moritz, 2000). Related to representativeness, selecting alternative f of 7^{th} question requires them to take into account representativeness. Percentage of both elementary (73.2%) and secondary (83.5%) school preservice teachers considering representativeness in generalizing results to the population are quite high. However this contradicts with the findings of Groth and Bergner (2005). This contradiction might result from the research context. In the study of Groth and Bergner (2005), elementary school preservice mathematics teachers were asked to write metaphors. They might have had difficulties in writing metaphor. In case of randomness, alternative e of 7^{th} question assess considering randomness in constructing sample. However the percentage of both elementary (41.5%) and secondary (41.8%) school preservice teachers who select alternative e of 7^{th} question are quite low. Moreover, when the alternatives of 6^{th} question examined, alternative c was chosen by quite number of elementary (47.6%) and secondary (57.1%) school preservice mathematics teachers. This alternative written below, state that sampling should have done nonrandomly:

> The patients should not have been randomly put into groups, because the

most severe cases may have just by chance ended up in one of the groups.

So they think that sampling should be based on human judgment which is contrary to generalization of results (Moore, 1990; Watson & Moritz, 2000). So, the number of both elementary and secondary school preservice teachers who overlooked the notion of randomness in sampling cannot be unnoticed. Therefore both elementary and secondary school mathematics teacher education might need to focus on the notion of randomness more.

To sum up, correct reasoning skills of elementary and secondary school preservice teachers are tried to be described. They are successful in interpreting probability, understanding independence, interpreting two-way tables and understanding importance of large samples. However there lack of understanding in selecting appropriate average, correctly computing probability, sampling variability, distinguishing between correlation and causation. Therefore these reasoning skills should be regarded cautiously.

6.2. Misconceptions

In this section conclusions for each misconception subscales are presented. These subscales were given in Section 4.3 previously as Table 4.2 (See Instrument Section).

6.2.1. Misconceptions involving averages

This subscale consists answers of alternatives a and c of 1^{st}, alternatives a and b of 4^{th}, alternatives b and f of 15^{th} and alternatives a and e of 17^{th} questions. If each participant had selected these alternatives, the mean would have been 1. Data shows that there exists both elementary ($\bar{x}$=.475) and secondary ($\bar{x}$=.480) school preservice mathematics teachers who have misconceptions involving averages. Athough the mean scores are not so high, it cannot be overlooked. For instance the percentages of both elementary and secondary school preservice mathematics teachers who chose misconceptions in 1^{st} and 4^{th} questions are quite higher than the percentages of correct answers given to 1^{st} and 4^{th} question. So there are preservice teachers who fail to recognize outlier and who think averages are the most common number. Moreover these two questions especially important since they assess how to find appropriate average and they are not so successful in finding appropriate average. Also results are consistent with the results of *understands how to select an appropriate average* presented above. They do not have complete understanding of selecting appropriate average. Still they have misconceptions about averages even though averages concept is the most familiar topic for preservice mathematics teachers (Heaton & Mickelson, 2002; Leavy, 2006; Canada, 2008). Furthermore, as findings of Groth and Bergner (2006) are considered, their misconceptions might have resulted from lack of conceptual understanding of averages. Therefore, both elementary and secondary preservice mathematics teachers' misconceptions involving averages should be regarded cautiously.

6.2.2. Outcome orientation misconception

Answers to the 2^{nd}, 3^{rd}, 11^{th}, 12^{th}, and 13^{th} questions constitue this subscale scores. If each participant had selected these alternatives, the mean would have been 1.60. Results shows that both elementary ($\bar{x}$=.185) and secondary ($\bar{x}$=.185) school preservice mathematics teachers do not have outcome orientation misconception. When Table 5.20 examined, percentages of these alternatives chosen by both elementary and secondary school preservice teachers is very low. However, alternatives of the 3^{rd} and 13^{th} questions which signal outcome orientation are relatively high compared to the other questions. For instance in 13^{th} question the percentages for both elementary and secondary preservice mathematics teachers is around 40%, in 11^{th} question it is around 1%. Also findings of Batanero *et al.* (2005) are inconsistent with these results. Therefore outcome orientation might be better to examine deeply among both elementary and secondary preservice mathematics teachers especially because of the difference in percentages of questions.

6.2.3. Good samples have to represent a high percentage of the population

Answers to the alternatives b and c of 7^{th} and alternative a and d of 16^{th} questions constitue this subscale scores. If each participant selected these alternatives, the mean would be 2. According to data, both elementary ($\bar{x}$=.470) and secondary ($\bar{x}$=.588) school preservice mathematics teachers do not have misconception of good samples have to represent a high percentage of the population. This misconception As Table 5.23 examined, percentages of these alternatives chosen by both elementary and secondary school preservice teachers is low. Still, the percentage of both elementary (36.6%) and secondary (44%) school preservice mathematics teachers who think that

> The sample of 2,050 teenagers is too small to permit drawing conclusions about the entire country.

cannot be overlooked although they are aware of representativeness as it is discussed in *understanding importance of large samples*. Therefore both elementary and secondary school preservice teachers' conceptions about good samples have to represent a high percentage of the population might need to be investigated deeply.

6.2.4. Law of small numbers

Answers to the 12thth and 14^{th} questions as alternative a and c respectively constitue this subscale scores. If each participant had selected these alternatives, the mean would have been 1. According to data it seems that both elementary ($\bar{x}$=.402) and secondary ($\bar{x}$=.335) school preservice mathematics teachers do not have misconception of law of small numbers. Consistently, the percentage of both elementary and secondary school preservice mathematics teachers who select the alternative a of 12thth question is very low. However the situation is not the same for alternative c of 14thth question. On the contrary, percentages are quite high. 61% of elementary and 50.5% of secondary school teachers chose the alternative c. 14^{th} question can be seen below,

> Half of all newborns are girls and half are boys. Hospital A records an average

of 50 births a day. Hospital B records an average of 10 births a day. On a particular day, which hospital is more likely to record 80% or more female births?

a. Hospital A (with 50 births a day)
b. Hospital B (with 10 births a day)
c. The two hospitals are equally likely to record such an event.

In the 14^{th} question sample of Hospital A is bigger than the Hospital B. According to Law of Large Numbers, as sample size gets bigger, results get close to the population statistics but results deviates more in small samples (Well, Pollatsek, & Boyce, 1990). Therefore the proportion of girls varies in the smaller hospital than the bigger one. So the two hospitals are not equally likely to record such an event. However selecting alternative c shows that sample size is not a matter of fact for both elementary and secondary school preservice teachers. These results are consistent with the findings of Jendraszek (2010). Both elementary and secondary school preservice teachers are not much aware of the effect of sample size. Also both elementary and secondary school preservice mathematics teachers' understanding of sampling variability are discussed above. Consistently, their reasoning about sampling variability is very low. So it is meaningful that they could not realize effect of sample size in $14th^{th}$ question.

6.2.5. Representativeness misconception

Answers of alternatives a, b and d of 9^{th}, alternative e of 10^{th}, and alternative c of 11^{th} questions constitues this subscale scores. If each participant had selected all of these alternatives, the mean would be 1.67. Results shows that both elementary ($\bar{x}$=.053) and secondary ($\bar{x}$=.044) school preservice mathematics teachers do not have representativeness misconception. Also percentages of these alternatives chosen by both elementary and secondary school preservice teachers is very low (see Table 5.27). People who have this misconception fail to understand independence. For instance they tend to see HTHTHT is less likely than THTTTH since there is a pattern in outcome of HTHTHT. However each of these Hs or Ts are independent from each other and thus the probability of these outcomes are equal. In our case, both elementary and secondary school preservice mathematics teachers' understanding of independence is high as it is discussed in *Understands independence* section. Although these results shows that both elementary and secondary school preservice mathematics teachers do not have representativeness misconception, this might stem from their familiarity to these questions because of the university entrance exam on Turkey. While students study for this high stake exam, they have to solve similar kind of questions. So preservice teachers might be familiar with these questions. Therefore it might be better to examine deeply representativeness misconception among preservice mathematics teachers so that their reasoning behind their choices might be well understood.

6.2.6. Correlation implies causation

Answers to the alternatives b and e of 16^{th} question constitues this subscale scores. If each participant had selected these alternatives, the mean would have been 2. According to data, it seems that both elementary ($\bar{x}$=.927) and secondary ($\bar{x}$=.780) school preservice mathematics teachers do not have misconception of correlation implies causation strongly. However alternatives b and e are the most common answer among elementary and secondary school preservice mathematics teachers. When the frequencies of this subscale scores are examined, the percentage of elementary (36.6%) and secondary (42.9%) school preservice mathematics teachers who did not chose any of the two alternatives is not high. So these show that attributing a causal relationship between variables quite common. However in the context of the question, research was a kind of gathering information of students' tv watching hours and report cards. To be able to assign a causal relationship between variables, there should be experiment results. It seems that preservice teachers are not aware of that. Also results of *Distinguishes between correlation and causation* subscale support that they are not successful in distinguishing between correlation and causation. Therefore both elementary and secondary preservice mathematics teachers' misconception of correlation implies causation need more attention.

6.2.7. Equiprobability bias

Answers to the alternative c of 13^{th}, alternative a of 18^{th}, alternative d of 19^{th} and alternative d of 20^{th} questions constitues this subscale scores. If each participant had selected these alternatives, the mean would have been 1. Both elementary ($\bar{x}$=.668) and secondary ($\bar{x}$=.637) school preservice mathematics teachers get quite high scores for this subscale. In discussion of *Correctly computes probability* above, it is discussed as they are not successful in computing probability questions requiring combinatorial reasoning. More than that, according results they have equiprobability bias as consistent with findings of Batanero *et al.* (2005). It seems that they see events as equally likely since they happens by chance (Lecoutre, 1992). For instance as a results of rolling two dice at the same time, obtaining two 5s and obtaining one 5 and one 6 are seen as equally likely. These events are seen as happening by chance so that probability of both events are $\frac{1}{36}$. However one 5 and one 6 are combinations of two events which are {56, 65}. All in all, when their high scores in *Equiprobability bias* subscale and low scores in *Correctly computes probability* are considered in juxtaposition to each other, both elementary and secondary school preservice teachers might need to take courses focusing on equiprobability bias.

6.2.8. Groups can only be compared if they are the same size

Answers to the 6^{th} question as alternative a constitue this subscale scores. If each participant had selected this alternative, the mean would be 1. Data shows that both elementary ($\bar{x}$=.330) and secondary ($\bar{x}$=.329) school preservice mathematics teachers do not have misconception of groups can only be compared if they are the same size. While comparing groups measures of center and distribution should be taken into consideration so that data can be seen as aggregate (Garfield, 2003). As discussed earlier in *Understands sampling variability* section, both elementary and secondary preservice mathematics teachers' reasoning in sampling variability is very low. Therefore their reasoning in comparison of data sets should be examined deeply and improved although they do not have misconception of groups can only be compared if they are the same size.

To sum up, misconceptions of elementary and secondary school preservice mathematics teachers are tried to be described. Other than equiprobability bias, they do not have these misconceptions strongly. Nevertheless, these misconceptions should be regarded cautiously. For instance, there are particular questions which include alternatives signaling some misconceptions such as law of small numbers assessed in 14^{th} question. Also these misconceptions are consistent with correct reasoning skills. Implications of these conclusions related to correct reasoning skills combined with misconceptions are presented in Section 6.3 part.

Both elementary and secondary school preservice mathematics teachers had lack of reasoning about statistical measures. Related to measures of center, it seems that they were good at interpreting average but they have difficulties in selecting appropriate average. Their ability to interpret average might stem from their procedural knowledge in finding mean with add-divide algorithm. Literature supported this fact that they have procedural knowledge of add-divide algorithm commonly (Groth & Bergner, 2006). Also their difficulties in selecting appropriate average might stem from their lack of reasoning about data since use of average is determined according to type of data (Gay *et al.*, 2009). Furthermore when we look at the misconceptions they had, it was appeared that they did not have misconceptios strongly related to measures of center. Even though their misconceptions was not strong, they still could not select appropriate average. Related to measures of spread, data showed that they did not take into account variation. They failed to consider variation in comparing distributions of two data sets and compared the groups according to averages of the data sets. Although they do not have the misconception that groups can only be compared if they are the same size, they still lacked reasoning about measures of spread. Therefore, their reasoning about data and then reasoning about statistical measures need to be developed prior to their actual teaching.

Data suggested that both elementary and secondary school preservice mathematics teachers' notions about sampling should be regarded cautiously, too. They had reasoning about the importance of large samples that is sample should be big enough to represent population (Watson & Moritz, 2000). Also they were aware of the fact that good samples do not have to be a high percentage of the population in order to be representative. What matters is the sampling method providing representative results of the population (Gay *et al.*, 2009). However, they might have applied the same reasoning of large samples to the small samples. Quite high number of them ignored the effect of small samples on results that is results varies more in small samples (Well *et al.*, 1990). Besides they could not take into account variation in sampling. This might be related to their lack of reasoning about measures of spread. Furthermore, they lacked the notion of randomness which is critical in sampling. Randomness and representativeness are important notions in sampling (Watson & Moritz, 2000). They have reasoning related to representativeness but not in randomness. Therefore randomness might need to be emphasized more in teacher education. Additionally, their lack of reasoning in measures of spread seems to affect their reasoning about samples. Thus measures of spread also need to be focused more in mathematics teacher education.

Both elementary and secondary school preservice mathematics teachers understood independence. Consistently they do not have representativeness misconception. However they could not compute probability required them to use combinatorial reasoning. Consistently, they had equiprobability bias since they did not have combinatorial reasoning. On the other hand they could interpret probability although there were some participants who have outcome orientation. According to Bulut (2001), probability is a difficult subject for Turkish preservice mathematics teachers. Therefore, probability might need to be focused more.

Regarding *third research question*, If there is any difference between statistical reasoning types of preservice secondary school mathematics teachers and elementary school mathematics teachers; there is not any significant difference between scores of secondary and elementary school preservice teachers for any of the subscales. According to findings of Jendrazsek (2010) success of preservice mathematics teachers' on probability or statistics is related with taking courses during their entire education. In our case, preservice teachers took statistics courses as either must or elective course in their university years. Also it is the contention that the both groups might have been educated under the same curricula since almost the same high school curriculum is applied in all over the country. Therefore, consisted with findings of Jendrazsek (2010), it becomes meaningfull that this research findings suported the fact that there was not any statistical difference between secondary and elementary school preservice teachers' probability or statistics education.

6.3. Limitations and Implications for Further Research

Aim of this study was to assess statistical reasoning of preservice mathematics teachers in İstanbul. Results obtained from this research yield some possible suggestions for further research.

Among statistical measures, measures of central tendency are the most familiar topic for preservice teachers (Heaton & Mickelson, 2002; Leavy, 2006; Canada, 2008). Nevertheless their performances are not high in selecting appropriate average according to given data. They have difficulties in recognizing outliers while finding mean and differentiating use of mean from mode. Related to measures of spread, their preformance is much worse. They ignore the variation and compare groups based on averages. Therefore, results suggest that prior to their actual teaching, they need to gain conceptual understanding in statistical measures especially in variation.

It seems that they have difficulties in sampling concept. They are not aware of the importance of randomness for unbiased inferences. Also they have the law of small numbers misconception which cause them to ignore effect of sample size in results. So research findings indicated that sampling concept, especially notion of randomness, need to be improved on the preservice teachers' part.

According to earlier researches, probability is a difficult subject for Turkish preservice mathematics teachers (Bulut, 2001). Our study results also shows that their preformance is very low in computing probability which is related to lack of combinatorial reasoning. Consistently, they have equiprobability bias strongly.

As literature suggests, teachers' knowledge affect students' knowledge (Heaton & Mickelson, 2002; Yolcu, 2012).Therefore, if preservice teachers have difficulties in understanding these concepts such as statistical measures, sampling and combinatorial reasoning or misconceptions such as law of small numbers and equiprobability bias, students will have lack of understanding in these concepts or these misconceptions. So preservice teachers' correct reasoning of these concepts need to be improved. In order to develop correct reasonings, mathematics education programs were recommended to include teaching probability and statistics course (Bulut, 2001). However in our study elementary school preservice teachers (90.2%) took a course to teach probability and statistics. Secondary school preservice teachers (85.7%) took statistics course. Nevertheless, as the results of this study showed, preservice teachers had limitations in their reasoning on selecting appropriate average, correctly computing probability, sampling variability, distinguishing between correlation and causation. Also they had law of small numbers misconception and equiprobability bias. So, these misconceptions or correct reasoning types of preservice teachers might be examined deeply and by the same token, contents of statistics courses in teacher education programs might be developed taking the the results of this study into consideration. Moreover previous research showed that the more courses preservice teachers take, the higher their performance is in probability (Jendraszek, 2010). This research findings, showed that preservice mathematics teachers still had limitations in their reasoning although most of them participated in a statistics course. Therefore, the number of courses to teach statistics might be increased in mathematics education programs so that their correct reasoning skills develop.

Additionally, participants of this study are elementary and secondary school preservice mathematics teachers studying in İstanbul. Also participants were not selected randomly because of practical reasons. So our results cannot be generalized to other preservice teachers. Therefore more studies with different participants especially from other regions of Turkey, need to be conducted in order to have an idea about statistical reasoning of Turkish preservice teachers.

Lastly, instrument was obtained from SRA which was not developed originally for Turkish context. For instance there are correctly answered questions by more than 90% of the participants in both groups. These questions might be replaced with new ones. For instance 8^{th} question which involved the context of drawing out marbles from two boxes. Also a new instrument might be developed regarding context of Turkish education system especially university entrance exam. These questions might be easy for them. That might result from university entrance exam because during the preparation for this exam, similar questions and much more difficult questions have been asked. Therefore, their familiarity might result in high scores in these questions.

APPENDIX A: Turkish Version of the Instrument

(i) Cinsiyet: K ( ) E ( )

(ii) Yaş:.....................

(iii) Bölüm: İlköğretim - Matematik Öğretmenliği ( )

Ortaöğretim - Matematik Öğretmenliği ( )

(iv) İstatistik dersi aldınız mı?

Hayır ( ) Evet ( )

Evetse hangi bölümden aldınız?

Matematik Bölümünden aldım. ( )

Eğitim Fakültesinden aldım. ( )

Diğer...

İSTATİKSEL AKIL YÜRÜTME TESTİ

Bu araştırmanın amacı istatiksel bilgiyi günlük hayatta nasıl kullandığınızı belirlemektir. Aşağıda verilen sorular çeşitli durumları okumanızı ve bu durumlar hakkında düşünmenizi gerektirmektedir. Eğer ne yapmanız istendiğinden emin değilseniz, lütfen yardım almak için elinizi kaldırınız.

İlerleyen sayfalar, istatistik ve olasılık ile ilgili çoktan seçmeli sorular içermektedir. Herhangi bir cevabı işaretlemeden önce soruyu dikkatlice okuyunuz.

1. Bir Fen Bilgisi dersinde küçük bir nesnenin kütlesi dokuz farklı öğrenci tarafından aynı ölçü birimiyle ölçülüyor. Her bir öğrencinin kaydettiği kütleler (gram cinsinden) aşağıda verilmiştir.

   6,2 6,0 6,0 15,3 6,1 6,3 6,2 6,15 6,2

   Öğrenciler mümkün olduğunca hatasız bir şekilde nesnenin gerçek kütlesini belirlemek istemektedir. Aşağıdaki yöntemlerden hangisini kullanmalarını tavsiye ederdiniz?

   a) En çok tekrar eden sayıyı, yani 6,2'yi kullanmalarını
   b) En hassas ölçüm olduğu için 6,15'i kullanmalarını
   c) Verilen 9 sayıyı toplayıp, 9'a bölmelerini
   d) 15,3'ü atıp, diğer 8 sayıyı toplayıp 8'e bölmelerini

2. Reçeteyle satılan bir şişe ilacın üzerinde aşağıdaki yönerge yazılıdır.

   "UYARI: Cilde uygulandığında kızarıklık oluşma ihtimali %15'tir. Eğer kızarıklık oluşursa doktorunuza danışınız."

   Aşağıdakilerden hangisi bu uyarının en iyi yorumudur?

   a) İlacı cildinizde kullanmayın, kızarıklık oluşma ihtimali yüksektir.
   b) Cilde uygularken önerilen dozun sadece %15'ini uygulayın.
   c) Eğer kızarıklık oluşursa muhtemelen cildin sadece %15'inde oluşacaktır.
   d) Bu ilacı kullanan 100 kişiden yaklaşık 15'inde kızarıklık oluşur.
   e) Bu ilacı kullanırken kızarıklık oluşma ihtimali neredeyse hiç yoktur.

3. Meteoroloji Genel Müdürlüğü hava tahminlerinin doğruluğunu belirlemek istiyordu. Yağmur yağma olasılığının %70 olarak bildirildiği günlerin raporlarını incelediler. Bu tahminleri söz konusu günlerde gerçekten yağmur yağıp yağmadığını gösteren kayıtlarla karşılaştırdılar.

   Yağmur yağan günlerin yüzdesi aşağıdakilerden hangisi olursa %70 yağmur yağma olasılığı, <u>çok</u> isabetli bir tahmin olarak kabul edilebilir?

   a) %95 - %100

   b) %85 - %94

   c) %75 - %84

   d) %65 - %74

   e) %55 - %64

4. Bir öğretmen, öğrencilerinin söz alma sayısını arttırmak amacıyla sınıftaki oturma düzenini değiştirmek ister. İlk önce mevcut oturma düzeninde öğrencilerin kaç kere söz aldığını belirlemeye karar verir. Buna göre aşağıdaki tablo, 8 öğrencinin bir ders saatindeki söz alma sayısını göstermektedir.

Öğrenci İsimlerinin Baş Harfleri	A.A.	R.F.	A.G.	J.G.	C.K.	N.K.	J.L.	A.V.
Yorum Sayısı	0	5	2	22	3	2	1	2

Öğretmen o gün yapılan tipik söz alma sayısını hesaplayarak bu veriyi özetlemek ister.

Aşağıdaki yöntemlerden hangisini kullanmasını önerirsiniz?

a) En çok tekrar eden sayıyı, yani 2'yi kullanmasını

b) Verilen 8 sayıyı toplayıp 8'e bölmesini

c) 22'yi atıp diğer 7 sayıyı toplayıp 7'ye bölmesini

d) 0'ı atıp diğer 7 sayıyı toplayıp 7'ye bölmesini

5. Bir cilt hastalığı olan egzemanın tedavisinde kullanılan yeni bir ilaç, etkinliğini belirlemek için test edilmektedir. Çalışmaya katılması için 30 egzema hastası seçilir. Hastalar rastgele iki gruba ayrılır. Deney grubundaki 20 hasta ilacı alırken kontrol grubundaki 10 hasta ilaç almaz. 2 ay sonra alınan sonuçlar aşağıda gösterilmiştir.

	Deney Grubu (ilaç verilen)	Kontrol Grubu (ilaç verilmeyen)
İyileşme var.	8	2
İyileşme yok.	12	8

Veriye göre, bence ilaç;

1.biraz etkilidir.

2.temelde etkisizdir.

<u>Eğer birinci seçeneği seçtiyseniz</u> aşağıdaki açıklamalardan hangisi düşüncenizi <u>en iyi</u> şekilde belirtir?

a) Deney grubundaki insanların %40'ı (8/20) iyileşme gösterdi.

b) Kontrol grubunda sadece 2 insan iyileşme gösterirken deney grubunda 8 insan iyileşme gösterdi.

c) Deney grubundaki iyileşme gösteren insan sayısı, iyileşme göstermeyen insan sayısından sadece 4 eksikken (12-8) kontrol grubunda bu fark 6 (8-2)'dır.

d) Deney grubundaki hastaların %40'ı (8/20) iyileşme gösterirken kontrol grubundaki hastaların sadece %20'si (2/10) iyileşme gösterdi.

<u>Eğer ikinci seçeneği seçtiyseniz</u> aşağıdaki açıklamalardan hangisi düşüncenizi <u>en iyi</u> şekilde belirtir?

a) Kontrol grubunda, 2 kişi ilaç verilmediğinde bile iyileşme gösterdi.

b) Deney grubunda, iyileşme göstermeyenlerin sayısı iyileşme gösterenlerden daha çok (12'ye 8)'tur.

c) İyileşme göstermeyenlerle iyileşme gösterenlerin farkı her iki grupta da yaklaşık olarak aynı (4'e 6)'dır.

d) Deney grubundaki hastaların sadece %40'ı (8/20) iyileşme gösterdi.

6. 5. soruda ifade edilen deney sonuçlarının sorgulanmasına yönelik birkaç sebep aşağıda listelenmiştir. Katıldığınız her sebebi işaretleyiniz.

a) İki grubu karşılaştırmak uygun olmaz çünkü iki gruptaki hasta sayısı farklıdır.
b) 30 kişilik bir örneklem sonuca varmak için çok küçüktür.
c) Hastalar gruplara rastgele yerleştirilmemeliydi çünkü çok ciddi vakalar şans eseri aynı grupta toplanmış olabilir.
d) Doktorların hastaların iyileşme gösterip göstermediğine nasıl karar verdiği hakkında yeterli bilgi verilmiyor. Doktorlar kararlarında taraflı davranmış olabilirler.
e) Bu ifadelerin hiçbirine katılmıyorum.

7. Bir araştırma şirketinden 13 -19 yaşlardaki gençlerin kaset, CD, plak, MP3, MP4 gibi müzik kayıtlarına ne kadar para harcadıklarını belirlemesi istenmiştir. Şirket ülke genelinde rastgele 80 alışveriş merkezi seçmiştir. Bir saha araştırmacısı alışveriş merkezinde merkezi bir yerde durup oradan geçenler arasında yaşı uygun görünenlerden anketi doldurmasını istemiştir. Toplamda 2050 anket gençler tarafından doldurulmuştur. Bu araştırmaya göre, araştırma şirketi bu ülkedeki ortalama bir gencin kayıtlı müziğe yılda 155 TL harcadığını rapor etmiştir.

Aşağıda bu araştırma hakkında birkaç ifade verilmiştir. Katıldığınız her ifadeyi işaretleyiniz.

a) Ortalama, gençlerin harcadıkları miktara yönelik tahminlerine dayalıdır. Bu yüzden gençlerin gerçekte ne kadar harcadıklarından oldukça farklı olabilir.
b) Eğer ülke genelindeki gençler için ortalama bulunmak isteniyorsa araştırma, 80'den fazla alışveriş merkezinde yapılmalıydı.
c) 2050 gençten oluşan örneklem bütün ülke geneli hakkında sonuca varmak için çok küçüktür.
d) Anket, müzik mağazalarından çıkan gençlere uygulanmalıydı.
e) Gençler anketi doldurmaları için rastgele seçilmediğine göre bu ortalama, tüm gençlerin harcamalarının zayıf bir göstergesi olabilir.
f) Sadece alışveriş merkezindeki gençler örnekleme dahil edildiğine göre bu ortalama, tüm gençlerin harcamalarının zayıf bir göstergesi olabilir.
g) Bu durumda ortalama hesaplamak uygun değildir çünkü gençlerin harcadığı miktarlarda çok varyasyon var.
h) Bu ifadelerin hiçbirine katılmıyorum.

8. A ve B olarak etiketlenmiş iki kutu aşağıda belirtilen miktarlarda kırmızı ve mavi bilyelerle doldurulur.

Kutu	**Kırmızı**	**Mavi**
A	6	4
B	60	40

Her bir kutu iyice sallanır. Kutulardan birini seçtikten sonra elinizi içine sokup bakmadan bir bilye seçeceksiniz. Eğer bilye maviyse 50 TL kazanacaksınız.

Mavi bilye seçme olasılığı hangi kutuda daha yüksektir?

a) A Kutusu (6 kırmızı ve 4 mavi)

b) B Kutusu (60 kırmızı ve 40 mavi)

c) Her iki kutuda da olasılık aynıdır.

9. Hilesiz bir madeni paranın 5 kez yazı-tura atılması sonucunda aşağıdaki sıralamalardan hangisinin gelme olasılığı en yüksektir? (Y: yazı, T: tura)

   a) TTTYY
   b) YTTYT
   c) YTYYY
   d) TYTYT
   e) 4 sıralamanın herbirinin de gerçekleşme olasılığı aynıdır.

10. 9. Soruya verdiğiniz cevap için aşağıdaki açıklamalardan bir veya daha fazlasını seçiniz.

   a) Madeni para hilesiz olduğu için hemen hemen eşit sayıda yazı ve tura gelir.
   b) Yazı-tura atışı rastgele olduğu için yazıdan turaya ve turadan yazıya sıkça geçiş olur.
   c) Sıralamalardan herhangi biri gerçekleşebilir.
   d) Eğer bir parayı arka arkaya 5 defa atarsan, bu sıralamaların her biri diğer sıralamaların herhangi biri kadar sık gerçekleşir.
   e) Eğer art arda bir çift tura gelirse, bir sonraki atışta yazı gelme olasılığı artar.
   f) 5 atışın her sıralamasının gerçekleşme olasılığı tam olarak aynıdır.

11. 9. sorudaki Y (yazı) ve T (tura) sıralamalarının aynısı aşağıda verilmiştir. Hilesiz bir madeni paranın 5 kez yazı-tura atılması sonucunda aşağıdaki sıralamalardan hangisinin gelme olasılığı en düşüktür?

   a) TTTYY
   b) YTTYT
   c) YTYYY
   d) TYTYT
   e) 4 sıralamanın herbirinin de gerçekleşmeme olasılığı eşittir.

12. Yılmaz Ailesi yeni bir araba almak istiyordu ve seçeneklerini “A” ve “B” olarak sınırlandırdı. Öncelikle çeşitli arabaların tamir ücretlerini karşılaştıran Tüketici Raporu’nu incelediler. Her çeşit için 400 araba üzerinde yapılan tamir kayıtları “B”nin “A”ya göre daha az mekanik sorun çıkardığını gösteriyordu.

   Sonra Yılmaz Ailesi ikisi “A”, biri eskiden “B” sahibi üç tanıdığı ile konuştu. İki “A” sahibi, önemli olmayan birkaç mekanik sorundan bahsetti. Fakat “B” sahibi, arabasından memnun olup olmadığı sorulduğunda köpürüverdi:

   “İlk önce yakıt enjeksiyonu bozuldu, 250TL ona gitti. Sonra arka tamponla sorun yaşamaya başladım ve onu değiştirmek zorunda kaldım. Şanzıman bozulduktan sonra sonunda arabayı satmaya karar verdim. Bir daha asla “B” almam.”

   Yılmaz Ailesi, yüksek masraflı tamir çıkarma olasılığı daha az olan arabayı almak istiyor. Şu an bildiklerini göz önüne alırsak hangi arabayı almalarını tavsiye edersiniz?

   a) Özellikle arkadaşının “B” ile yaşadığı sorunlardan dolayı “A” almalarını tavsiye ederim. “A” hakkında benzer korkunç hikayeler duymadıkları için “A”yı almalılar.
   b) Arkadaşının kötü deneyimine rağmen “B” almalarını tavsiye ederim. Bu sadece bir durumken Tüketici Raporu’nda verilen bilgi birçok duruma dayanıyor. Ve oradaki veriye göre “B” daha az tamir gerektiriyor.
   c) Onlara hangi arabayı aldıklarının bir önemi olmadığını söylerim. Modellerden birinin tamir gerektirmesi diğerine göre daha muhtemel olmasına rağmen, yine de şans eseri çok tamir isteyen bir arabaya denk gelebilirler. Karar vermek için yazı-tura atsalar da olur.

13. Hilesiz bir zarın beş yüzü siyaha, bir yüzü beyaza boyanır. Zar altı kez atılır.
Aşağıdaki sonuçlardan hangisinin gelme olasılığı daha yüksektir?

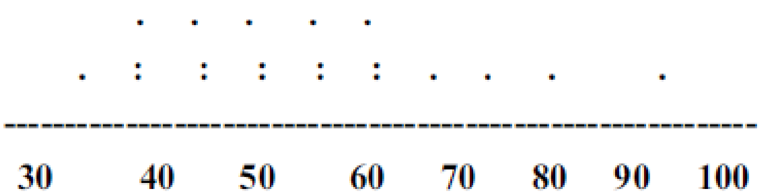
a) 5 defa siyah yüz, bir defa beyaz yüz üste gelir.
b) 6 atışın hepsinde de siyah yüz üste gelir.
c) a ve b'nin gerçekleşme ihtimali aynıdır.

14. Yenidoğan tüm bebeklerin yarısı kız, yarısı erkektir. "A" Hastanesi'nde günde ortalama 50 doğum, "B" Hastanesi'nde günde ortalama 10 doğum kaydediliyor.
Belirli bir günde hangi hastanenin %80 veya daha fazla kız doğumu kaydetmesi daha olasıdır?

a) "A" Hastanesi (günde 50 doğum)
b) "B" Hastanesi (günde 10 doğum)
c) İki hastanenin de böyle bir olayı kaydetme olasılığı eşittir.

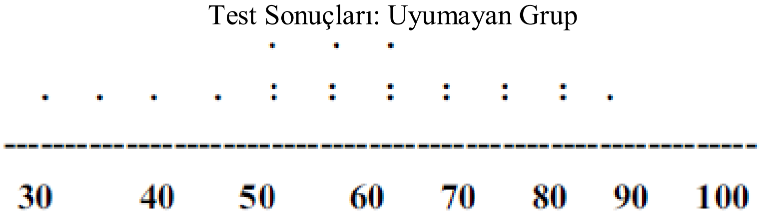
15. 40 üniversite öğrencisi, uykunun test sonuçlarına etkisi üzerine yapılan bir çalışmaya katılmışlardır. 20 öğrenci testten bir önceki gece uyumayıp çalışmak için gönüllü olmuştur (uyumayan grup). Diğer 20 öğrenci (kontrol grubu) testten bir önceki gece 23:00'te yatmıştır. Her bir grup için test sonuçları aşağıdaki grafiklerde gösterilmiştir. Grafikteki her nokta belirli bir öğrencinin puanını temsil etmektedir. Örneğin, II. grafikte 80 üzerindeki iki nokta, uyuyan gruptaki 2 öğrencinin testten 80 aldığını göstermektedir.

I. Grafik



Test Sonuçları: Uyumayan Grup

II. Grafik



Test Sonuçları: Uyuyan Grup

İki grafiği de dikkatlice inceleyiniz. Sonra aşağıda verilen 6 muhtemel sonuçtan <u>en çok</u> katıldığınızı seçiniz.

a) Uyumayan grup daha iyi yapmıştır, çünkü bu öğrencilerin hiçbiri 40'ın altında puan almamıştır ve en yüksek puanı alan öğrenci bu gruptadır.
b) Uyumayan grup daha iyi yapmıştır, çünkü bu grubun ortalaması uyuyan grubun ortalamasından biraz daha yüksek görünmektedir.
c) İki grup arasında fark yoktur, çünkü iki grubun sonuçları büyük ölçüde örtüşmektedir.
d) İki grup arasında fark yoktur, çünkü puanlardaki varyasyon miktarıyla kıyaslandığında grupların ortalamaları arasındaki fark küçüktür.
e) Uyuyan grup daha iyi yapmıştır, çünkü bu grupta daha çok öğrenci 80 ve üzeri puan almıştır.
f) Uyuyan grup daha iyi yapmıştır, çünkü bu grubun ortalaması uyumayan grubun ortalamasından biraz daha yüksek görünmektedir.

16. 500 ilköğretim öğrencisi bir ay boyunca her gün, televizyon izlemeye ayırdıkları vaktin kaydını tutmuştur. Bir haftalık televizyon izleme sürelerinin ortalaması 28 saat olarak bulunmuştur. Ayrıca, çalışmayı yapan araştırmacılar her öğrencinin karnesine ulaşmışlardır. Araştırmacılar okulda başarılı öğrencilerin, başarısı düşük öğrencilere göre televizyon izlemeye daha az vakit ayırdığını bulmuşlardır. Bu çalışmanın sonuçları hakkında olası yorumlar aşağıda verilmiştir. Katıldığınız her yorumu işaretleyiniz.

a) 500 kişilik örneklem sonuç çıkarmak için çok küçüktür.
b) Eğer bir öğrenci televizyon izlemeye ayırdığı vakti azaltırsa okul başarısı artar.
c) Başarılı öğrenciler daha az televizyon izlemiş olsa bile bu televizyon izlemenin okul performansına zarar verdiği anlamına gelmez.
d) Öğrencilerin televizyon izlemeye yaklaşık kaç saat harcadıklarını berlirlemek için bir ay, yeterince uzun bir süre değildir.
e) Bu araştırma, televizyon izlemenin daha zayıf okul performansına neden olduğunu gösterir.
f) Bu yorumların hiçbirine katılmıyorum.

17. Küçük bir köyün okul komisyonu köydeki ev başına düşen ortalama çocuk sayısını belirlemek istemiştir. Köydeki toplam çocuk sayısını toplam ev sayısı olan 50'ye bölmüşlerdir. Eğer ev başına düşen ortalama çocuk sayısı 2,2 ise aşağıdaki yorumların hangisi doğru olmalıdır?

a) Köydeki evlerin yarısında 2'den fazla çocuk var.
b) Köydeki 3 çocuklu evlerin sayısı, 2 çocuklu evlerin sayısından daha çoktur.
c) Köyde toplam 110 çocuk vardır.
d) Köydeki her bir yetişkin başına 2,2 çocuk düşer.
e) Evlerde en çok rastlanan çocuk sayısı 2'dir.
f) Yukarıdakilerden hiçbiri.

18. İki zar aynı anda atıldığında aşağıdaki 2 sonuçtan birisinin gerçekleşmesi mümkündür.
Sonuç 1: 5 ve 6 gelir. **Sonuç 2**: iki tane 5 gelir.
En çok katıldığınız cevabı seçiniz.

a) Bu sonuçların herbirini elde etme olasılığı eşittir.
b) Sonuç 1'i elde etme olasılığı daha yüksektir.
c) Sonuç 2'yi elde etme olasılığı daha yüksektir.
d) Cevap vermek imkansızdır. (Lütfen nedenini açıklayınız.)…………………………………
……………………………………………………………………………………………………..

19. Üç zar aynı anda atıldığında, aşağıdaki sonuçların hangisinin elde edilme olasılığı EN YÜKSEKTİR?

a) Sonuç 1: "bir tane 5, bir tane 3 ve bir tane 6"
b) Sonuç 2: "üç tane 5"
c) Sonuç 3: " iki tane 5 ve bir tane 3"
d) Üç sonucun da gerçekleşme olasılığı eşittir.

20. Üç zar aynı anda atıldığında, aşağıdaki sonuçların hangisinin elde edilme olasılığı EN DÜŞÜKTÜR?

a) Sonuç 1: "bir tane 5, bir tane 3 ve bir tane 6"
b) Sonuç 2: "üç tane 5"
c) Sonuç 3: " iki tane 5 ve bir tane 3"
d) Üç sonucun da gerçekleşmeme olasılığı eşittir.

APPENDIX B: Results of McNemar Test

Table B.1. Mc Nemar Test-Item 1.

		English Version	
		False	True
Turkish Version	False	23	2
	True	13	23

Table B.2. Mc Nemar Test Statistics-Item 1.

	-Turkish & -English
N	61
Exact Sig. (2-tailed)	.007

Table B.3. Mc Nemar Test-Item 2.

		English Version	
		False	True
Turkish Version	False	4	0
	True	8	49

Table B.4. Mc Nemar Test Statistics-Item 2.

	-Turkish & -English
N	61
Exact Sig. (2-tailed)	.008

Table B.5. Mc Nemar Test-Item 4.

		English Version	
		False	True
Turkish Version	False	21	4
	True	14	22

Table B.6. Mc Nemar Test Statistics-Item 4.

	-Turkish & -English
N	61
Exact Sig. (2-tailed)	.031

Table B.7. Mc Nemar Test-Item 10c.

		English Version	
		False	True
Turkish Version	False	8	4
	True	21	28

Table B.8. Mc Nemar Test Statistics-Item 10c.

	-Turkish & -English
N	61
Exact Sig. (2-tailed)	.001

Table B.9. Mc Nemar Test-Item 11.

		English Version	
		False	True
Turkish Version	False	3	1
	True	9	48

Table B.10. Mc Nemar Test Statistics-Item 11.

	-Turkish & -English
N	61
Exact Sig. (2-tailed)	.021

Table B.11. Mc Nemar Test-Item 13.

		English Version	
		False	True
Turkish Version	False	34	14
	True	3	10

Table B.12. Mc Nemar Test Statistics-Item 13.

	-Turkish & -English
N	61
Exact Sig. (2-tailed)	.013

Table B.13. Mc Nemar Test-Item 17.

		English Version	
		False	True
Turkish Version	False	11	4
	True	17	29

Table B.14. Mc Nemar Test Statistics-Item 17.

	-Turkish & -English
N	61
Exact Sig. (2-tailed)	.007

APPENDIX C: Table of The Most and Second Most Common Alternatives

Table C.1. Frequency of Alternatives-Elementary Level.

Item	Most common			Second most common		
	Alternative	%		Alternative	%	
1	a	35.4	misconception	d	34.1	correct
2	d	90.2	correct	none	-	-
3	d	46.3	correct	a	35.4	misconception
4	b	40.2	misconception	c	36.6	correct
5	1-d	51.2	correct	none	-	-
6	b	63.4	correct	c	47.6	false
7	f	73.2	correct	a / e	41.5	false/ correct
8	c	93.9	correct	none	-	-
9	e	92.7	correct	none	-	-
10	c	65.9	correct	f	63.4	correct
11	e	87.8	correct	none	-	-
12	b	72	correct	a	19.5	false
13	b	37.8	misconception	a	31.7	correct
14	c	61	misconception	a	18.3	false
15	f	54.9	misconception	d / e	9.8	correct / false
16	b and e	46.3	misconception	c	23.2	correct
17	c	63.4	correct	e	7.3	misconception
18	a	70.7	misconception	b	14.6	correct
19	d	81.7	misconception	a	13.4	correct
20	d	79.3	misconception	b	14.6	correct

Table C.2. Frequency of Alternatives-Secondary Level.

Item	**Most common**			**Second most common**		
	Alternative	%		Alternative	%	
1	a, c	30.8	misconception	none	-	-
1	d	30.8	correct	none	-	-
2	d	94.5	correct	none	-	-
3	d	54.9	correct	a	34.1	misconception
4	c	46.2	correct	b	30.8	misconception
5	1-d	71.4	correct	none	-	-
6	b	63.7	correct	c	57.1	false
7	f	83.5	correct	c	44	misconception
8	c	95.6	correct	none	-	-
9	e	93.4	correct	none	-	-
10	f	76.9	correct	c	61.5	correct
11	e	92.3	correct	none	-	-
12	b	69.2	correct	a	16.5	false
13	b	45.1	misconception	c	29.7	misconception
14	c	50.5	misconception	a/b	24.2	false/correct
15	f	63.7	misconception	b	6.6	misconception
16	e	42.9	misconception	b	35.2	misconception
17	c	59.3	correct	b, f	7.7	false
18	a	75.8	misconception	b	11	correct
19	d	84.6	misconception	a	12.1	correct
20	d	76.9	misconception	b	16.5	correct

REFERENCES

Aksayan, S., and S., Gözüm, 2002, "Kültürlerarası Ölçek Uyarlaması için Rehber I: Ölçek Uyarlama Aşamaları ve Dil Uyarlaması", *Hemşirelik Araştırma Dergisi*, Vol. 4, No. 1, pp. 9-14.

Aksayan, S., and S., Gözüm, 2003, "Kültürlerarası Ölçek Uyarlaması için Rehber II: Ölçek Uyarlama Aşamaları ve Dil Uyarlaması", *Hemşirelik Araştırma Dergisi*, Vol. 5, No. 1, pp. 3-14.

Batanero, C., J. D., Godino, and M. J., Cañizares, 2005, "Simulation as a Tool to Train Preservice School Teachers". In Proceedings of ICMI First African Regional Conference, http://www.ugr.es/~batanero/ARTICULOS/CMIRCr.pdf, [Accessed July 2014].

Baştürk, R., 2010, *Bütün Yönleriyle SPSS Örnekli Nonparametrik İstatiksel Yöntemler*, Anı Yayıncılık, Ankara.

Bulut, S., 2001, "Matematik Öğretmen Adaylarının Olasılık Performanslarının İncelenmesi", *Hacettepe Üniversitesi Egitim Fakültesi Dergisi* , Vol. 20, pp. 33-39.

Büyüköztürk, S., 2006, *Sosyal Bilimler İçin Veri Analizi El Kitabı*, Pegem Yayınları, Ankara.

Canada, D. L., 2008, "Conceptions of Distribution Held by Middle School Students and Preservice Teachers", In C. Batanero, G. Burrill, C. Reading, and A. Rossman (Eds.), Proceedings of the *ICMI Study 18 and 2008 IASE Round Table of Conference*, http://iase-web.org/documents/papers/rt2008/T2P9_Canada.pdf, [Accessed July 2014].

Chance, B., 2002, "Components of Statistical Thinking and Implications for Instruction and Assessment", *Journal of Statistics Education*, Vol. 10, No.3, www.amstat.org/publications/jse/v10n3/chance.html, [Accessed July 2014].

Curcio, F. R., 1987, "Comprehension of Mathematical Relationships Expressed in Graphs", *Journal for Research in Mathematics Education*, pp. 382-393.

Ben-Zvi, D., and J., Garfield, 2005, *The Challenge Of Developing Statistical Literacy, Reasoning and Thinking*, Springer Science + Business Media, Netherlands.

Delmas, R. C., 2002, "Statistical Literacy, Reasoning, and Learning: A Commentary", *Journal of Statistics Education*, Vol. 10, No.3.

Durkalski, V. L., Y. Y., Palesch, S. R., Lipsitz, and P. F., Rust, 2003, "Analysis Of Clustered Matched-Pair Data", *Statistics in Medicine*, Vol. 22, No. 15, pp. 2417-2428.

Estrada, A., and C. Diaz, 2006, "Computing Probabilities from Two Way Tables: An Exploratory Study with Future Teachers", In Proceedings of *Seventh international conference on teaching of statistics*, Salvador, International Association for Statistical Education, Bahia.

Fennema, E., and M. L., Franke, 1992, "Teachers' Knowledge and Its Impact", In D. A. Grouws (ed.), *Handbook of research on mathematics teaching and learning: A project of the national council of teachers of mathematics*, pp. 147-165, Macmillan Publishing Co, Inc.

Friel, S. N., F. R., Curcio, and G. W., Bright, 2001, "Making Sense of Graphs: Critical Factors Influencing Comprehension and Instructional Implications", *Journal For Research In Mathematics Education*, Vol. 32, No. 2, pp. 124-158.

Gal, I., 2002, "Adults' Statistical Literacy: Meanings, Components, Responsibilities", *International Statistical Review*, Vol. 70, No. 1, pp. 1-25.

Gal, I., 2005, “Towards Probability Literacy for All Citizens: Building Blocks and Instructional Dilemmas”, In G. A. Jones (Ed.), *Exploring Probability in School*, pp. 39-63, Springer.

Garfield, J., 2002, “The Challenge of Developing Statistical Reasoning”, *Journal of Statistics Education*, Vol. 10, No. 3, pp. 1-11.

Garfield, J., 2003, “Assessing Statistical Reasoning”, *Statistics Education Research Journal*, Vol. 2, No. 1, pp. 22-38.

Garfield, J., and D. Ben-Zvi, 2007, “How Students Learn Statistics Revisited: A Current Review Of Research on Teaching and Learning Statistics”, *International Statistical Review*, Vol. 75, No. 3, pp. 372-396.

Garfield, J. B., and D. Ben-Zvi, 2008, *Developing Students' Statistical Reasoning. Connecting Research and Teaching Practice*, Netherlands: Springer, Dordrecht.

Garfield, J., and I. Gal, 1999, “Teaching and Assessing Statistical Reasoning”, *NCTM 1999 Yearbook: Mathematical Reasoning*, pp. 1-17, Https://Apps3.Cehd.Umn.Edu/Artist/Articles/Garfield01.Pdf, [Accessed February 2014].

Gay, L. R., G. E., Mills, and P. W., Airasian, 2009, *Educational Research: Competencies For Analysis and Applications*, Upper Saddle River, NJ, USA: Pearson Education.

Groth, R., and J. A., Bergner, 2005, “Preservice Elementary School Teachers' Metaphors for The Concept of Statistical Sample”, *Statistics Education Research Journal*, Vol. 4, No. 2, pp. 27-42.

Groth, R., and J. A., Bergner, 2006, “Preservice Elementary Teachers' Conceptual And Procedural Knowledge of Mean, Median, and Mode”, *Mathematical Thinking and Learning*, Vol. 8, No.1, pp. 37-63.

Hambleton, R. K., and L. Patsula, 1999, “Increasing the Validity of Adapted Tests: Myths To Be Avoided and Guidelines for Improving Test Adaptation”, *Practices Journal of Applied Testing Technology*, Vol. 1, No. 1, pp. 1-30.

Hambleton, R. K., and L. Patsula, 1998, “Adapting Tests for Use in Multiple Languages and Cultures”, *Social Indicators Research*, Vol. *45*, *No.* 1-3, pp. 153-171.

Heaton, R. M., and W. T., Mickelson, 2002, “The Learning and Teaching of Statistical Investigation in Teaching and Teacher Education” *Journal of Mathematics Teacher Education*, Vol. 5, No. 1, pp. 35-59.

Jendraszek, P., 2010, “An Analysis of a Misconception of Probability Among Future Mathematics Teachers”, *Journal of Mathematics Education at Teachers College*, Vol. 1, No. 1.

Kahneman, D., and A. Tversky, 1974, “Subjective Probability: A Judgment of Representativeness”, In *The Concept of Probability In Psychological Experiments*, Springer, Netherlands.

Kaynar, Y., and E. Halat, 2012, İlköğretġm II. Kademe Matematġk Öğretim Programının Olasilik ve İstatistik Alt Öğrenme Alanının İstatistik Boyutunun Incelenmesi, *X. Ulusal Fen Ve Matematik Eğitimi Kongresi*, Niğde.

Kılıçer, K., and F. H., Odabaşı, 2010, “Bireysel Yenilikçilik Ölçeği (BYÖ): Türkçeye Uyarlama, Geçerlik Ve Güvenirlik Çalışması”, *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, Vol. *38*, pp. 150-164.

Konold, C., T. Higgins, S. J., Russell, and K. Khalil, 2003, “Data Seen Through Different Lenses”, In *Educational Studies in Mathematics*, Springer.

Konold, C., A. Pollatsek, A. Well, J. Lohmeier, and A. Lipson, 1993, “Inconsistencies in Students' Reasoning about Probability”, *Journal for Research in Mathematics Education*, Vol. 24, pp. 392-414.

Leavy, A. M., 2006, “Using Data Comparison to Support a Focus on Distribution: Examining Preservice Teacher's Understandings of Distribution When Engaged in Statistical Inquiry”, *Statistics Education Research Journal*, Vol. 5, No. 2, pp. 89-114.

Lecoutre, M. P., 1992, “Cognitive Models and Problem Spaces in Purely Random Situations”, *Educational Studies in Mathematics,* Vol. 23, *No.* 6, pp. 557-568.

Liu, H. J. C., 2002, “Translation of Instruments for Cross-Cultural Research”, *Journal Of Da-Yeh University*, Vol. 11, No. 2, pp. 79-88.

Liu, H. J. C., and J. Garfield, 2002, “Sex Differences In Statistical Reasoning”, *Bulletin Of Educational Psychology*, Vol. 34, No. 1, pp. 123-137.

Makar, K., 2004, *Developing Statistical Inquiry: Prospective Secondary Mathematics and Science Teachers' Investigations of Equity and Fairness Through Analysis of Accountability Data*, Unpublished Doctoral Dissertation, University of Texas at Austin.

Makar, K., and J. Confrey, 2005, “Variation-talk: Articulating Meaning in Statistics”, *Statistics Education Research Journal*, Vol. 4, No. 1, pp. 27-54.

Mickelson, W. T., and R. M., Heaton, 2004, “Primary Teachers’ Statistical Reasoning About Data”, In D. Ben-Zvi, and J., Garfield (ed.) *The Challenge Of Developing Statistical Literacy, Reasoning and Thinking*, pp. 327-352, Springer, Netherlands.

Milli Eğitim Bakanlığı, 2013a, *İlköğretim Matematik Dersi 6-8. Sınıflar Öğretim Programı ve Kılavuzu,* Ankara.

Milli Eğitim Bakanlığı, 2013b, *Ortaöğretim Matematik Dersi 9-12. Sınıflar Öğretim Programı ve Kılavuzu,* Ankara.

Moore, D. S., 1990, “Uncertainty”, In L. A. Steen, (Ed.). *On The Shoulders of Giants: New*

Approaches to Numeracy, pp. 96-137, National Academies Press.

National Council of Teachers of Mathematics, 2005, *Principles and Standards for School Mathematics*, Reston, VA, USA: National Council of Teachers of Mathematics.

Özen, M., 2012, *Investigation of Preservice Mathematics Teachers' Critical Thinking Processes through Statistical and Probabilistic Knowledge in The Context Of Popular Media Texts*, M. S. Thesis, Middle East Technical University.

Pfannkuch, M., and C. J., Wild, 2000, "Statistical Thinking and Statistical Practice: Themes Gleaned from Professional Statisticians", *Statistical Science,* Vol. 15, No. 2, pp. 132-152.

Rapp, J., and A, Allalouf, 2003, "Evaluating Cross-Lingual Equating", *International Journal of Testing*, Vol. 3, No. 2, pp. 101-117.

Rumsey, D. J., 2002, "Statistical Literacy as a Goal for Introductory Statistics Courses" *Journal of Statistics Education*, Vol. 10, No. 3, pp. 6-13.

Shaughnessy, J. M., 2007, "Research on Statistics Learning", In F. K. Lester, (ed.). *Second Handbook of Research on Mathematics Teaching and Learning*, pp. 957-1009, IAP.

Shulman, L. S., 1987, "Knowledge and Teaching: Foundations of The New Reform", *Harvard Educational Review*, Vol. 57, No. 1, pp. 1-23.

Smith, M. H., 2004, "A Sample/Population Size Activity: Is it the sample size of the sample as a fraction of the population that matters?", *Journal Of Statistics Education*, Vol. 12, No. 2.

Sundre, D. L., 2003, "Assessment of Quantitative Reasoning to Enhance Educational Quality", In *American Educational Research Association Meeting*, Chicago.

Tempelaar, D., 2004, “Statistical Reasoning Assessment: An analysis of the SRA instrument”, In *The ARTIST Roundtable Conference on Assessment in Statistics*, Appleton, Wisconsin, Citeseer.

Şahin, F., 2012, *A Study for Development of Statistical Literacy Scale for Undergraduate Students*, M. S. Thesis, Boğaziçi University.

Tversky, A., and D. Kahneman, 1971, “Belief in the Law of Small Numbers” *Psychological Bulletin*, Vol. 76, No. 2, pp. 105-110.

Ucar, Z., and E. N., Akdoğan, 2009, “Middle School Students’ Understanding Of Average”, *İlköğretim Online*, Vol. 8, No. 2, pp. 391-400.

Ulutaş, F., & B. Ubuz, 2008, “Research and Trends in Mathematics Education: 2000 To 2006”, *Elementary Education Online*, Vol. 7, No. 3, pp. 614-626.

Van De Walle, J. A., K. S., Karp, and J. M., Bay-Williams, 2007, *Elementary and Middle School Mathematics: Teaching Developmentally*, Pearson Allyn and Bacon.

Watson, J. M., 1997, “Assessing Statistical Thinking Using the Media”, In I. Gal, J. Garfield, and Y. Gal (eds.), *The Assessment Challenge In Statistics Education*, pp. 107-121, IOS Press.

Watson, J. M., and J. B., Moritz, 2000, “Developing Concepts of Sampling”, *Journal For Research in Mathematics Education*, Vol. 31, No. 1, pp. 44-70.

Well, A. D., A. Pollatsek, and S. J., Boyce, 1990, “Understanding The Effects of Sample Size on The Variability of The Mean”, *Organizational Behavior And Human Decision Processes*, Vol. 47, No. 2, pp. 289-312.

Wild, C. J., and M. Pfannkuch, 1999, “Statistical Thinking in Empirical Enquiry”, *International Statistical Review*, Vol. 67, No. 3, pp. 223-248.

Yolcu, A., 2012, *An Investigation of Eighth Grade Students' Statistical Literacy, Attitudes towards Statistics and Their Relationship*, M. S. Thesis, Middle East Technical University.