ASSESSING PRESERVICE MATHEMATICS TEACHERS STATISTICAL REASONING AND EXAMINING THEIR REFLECTIONS TO TEACH STATISTICS

by
Fatıma Halenur Öztürk
B.S., Secondary School Mathematics Education, Boğaziçi University, 2021

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

Graduate Program in Mathematics and Science Education
Boğaziçi University
2025

ACKNOWLEDGEMENTS

To begin with, I would like to thank my supervisor, Assoc. Prof. Kemal Akoğlu, for his unshakable and unwavering support, patience, encouragement, and invaluable guidance throughout this research. His insights into statistics, statistical education, and statistical reasoning significantly enhanced my understanding in these fields. This study would not have been accomplished without his assistance.

I also wish to extend my thanks to Assoc. Prof. Yeşim İmamoğlu and Assoc. Prof. Sibel Kazak, who served on my thesis committee, for their steadfast support, constructive feedback, and invaluable counsel over the duration of this study.

I am thankful to Assoc. Prof. Sevda Yerdelen Damar for her assistance with course selection and for helping create the schedule during my master's program.

My heartfelt appreciation goes to Mervenur Belin for always being there when I needed her guidance and support, and for her unwavering presence since our undergraduate days.

I would like to thank Fatih Cava for making life happier and more comfortable and for always being there for me when I needed him.

Lastly, I wish to convey my sincere gratitude to my beloved family for their limitless support and patience throughout this undertaking. For their unwavering love and belief in me over the years, I am particularly thankful to my mother, Ayşe, and my father, Bilal. I also appreciate my brother for being a companion and study partner. I consider myself lucky to have such a lovely family.

ABSTRACT

ASSESSING PRESERVICE MATHEMATICS TEACHERS STATISTICAL REASONING AND EXAMINING THEIR REFLECTIONS TO TEACH STATISTICS

Statistical reasoning is understanding statistical information, making interpretations considering the data sets, and reasoning statistically (Garfield, 2002). This study examines the statistical reasoning skills of preservice mathematics teachers studying at Bogazici University and examines their reflections on statistical knowledge to teach statistics and experiences in learning statistics. The research includes a quantitative part assessing participants correct reasoning skills and misconceptions using Statistical Reasoning Assessment (Garfield, 2003) and a qualitative part examining their reflections on teaching statistics according to statistical knowledge and experiences with semi-structured interviews. The sample includes 29 preservice mathematics teachers (20 primary, 9 secondary). Strengths in computing probability, understanding independence, and understanding the significance of large samples were shown by the analysis of SRA responses. Conversely, there was an inadequate amount of understanding in selecting averages, computation of probability, sample variability, differentiation of correlation and causation, and interpreting two-way tables. They had misconceptions including averages, the law of small numbers, and equiprobability bias, and there was a statistically significant difference between the groups who took the SCED354 course or who did not take the course in total misconception scores. Interviews provided insights into participants' reflections on their statistical knowledge and teaching approaches. This study revealed critical gaps in preservice teachers' knowledge of some statistical concepts and their pedagogical approaches. In mathematics education, emphasis need to be placed on conceptual understanding and effective teaching methods in mathematics. Longitudinal studies to investigate the long-term effect of teaching probability and statistics courses on the development of preservice mathematics teachers are among the recommendations for future research.

ÖZET

MATEMATİK ÖĞRETMEN ADAYLARININ İSTATİSTİKSEL MUHAKEME BECERİLERİNİN DEĞERLENDİRİLMESİ VE İSTATİSTİK ÖĞRETME YANSIMALARININ İNCELENMESİ

İstatistiksel muhakeme, istatiksel bilgiyi anlama, veri setlerini dikkate alarak yorumlama ve istatistiksel olarak düşünme anlamına gelir (Garfield, 2002). Bu çalışmada Boğaziçi Üniversitesi'nde öğrenim gören matematik öğretmen adaylarının istatistiksel muhakeme becerilerini incelemekte ve istatistik öğretmek için gerekli olan istatistiksel bilgiye ve deneyimlerine dair yansımaları araştırmaktadır. Araştırma, katılımcıların doğru muhakeme becerilerini ve kavram yanılgılarını değerlendirmek için kullanılan İstatistiksel Muhakeme Testi (SRA-2003) ile nicel bir bölüm ve istatistik öğretimine dair bilgi ve deneyimlerini inceleyen yarı yapılandırılmış görüşmelerle nitel bir bölüm içermektedir. Örneklem, 20 ortaokul ve 9 lise düzeyinden olmak üzere toplam 29 öğretmen adayından oluşmaktadır. SRA yanıtlarının analizi, öğretmen adaylarının olasılığı yorumlama, bağımsızlığı anlama ve büyük örneklemlerin önemini kavrama konusunda başarılı olduklarını göstermiştir. Ancak, uygun ortalama seçme, olasılığı doğru hesaplama, örneklem çeşitliliğinde, ilişki ve nedensellik arasındaki farkı ayırt etme ve iki yönlü tabloları yorumlama konularında anlamada eksiklik bulunmuştur. Sonuçlar, katılımcıların ortalamalar, küçük sayılar yasası ve eşit olasılık yanlılığı ile ilgili kavram yanılgılarına sahip olduğunu ve SCED 354 dersini alan ve almayan gruplar arasında toplam kavram yanılgıları puanı açısından istatistiksel olarak anlamlı bir fark olduğunu göstermiştir. Bu çalışma, öğretmen adaylarının bazı istatistiksel kavramlarda ve pedagojik yaklaşımlarda önemli bilgi eksikliklerini ortaya koymaktadır. Matematik öğretimi programlarında kavramsal ve etkili öğretim yöntemlerine odaklanılmasının gerekliliğini vurgulamaktadır. Gelecek araştırmalar için öneriler arasında, öğretmen adaylarının istatistik ve olasılık derslerinin uzun vadeli etkilerini inceleyen çalışmalar yer almaktadır.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF SYMBOLS

df	Degrees of Freedom
N	Number of Participants
P	P-Value
t	T-Score
U	Uniform Distributions
$\bar{x}$	Mean
Z	Z-Score

LIST OF ACRONYMS/ ABBREVIATIONS

ASA	American Statistical Association
CCK	Common Content Knowledge
CCSM	Common Core State Standards for Mathematics
GAISE	Guidelines for Assessment and Instruction in Statistics Education
H	Head
KCS	Knowledge of Content and Students
KCT	Knowledge of Content and Teaching
LOCUS	Levels of Conceptual Understanding in Statistics
MKT	Mathematical Knowledge for Teaching
NCTM	National Council of Teachers of Mathematics
PCK	Pedagogical Content Knowledge
PPDAC	Problem, Plan, Data, Analysis, Conclusion
SCK	Specialized Content Knowledge
SD	Standard Deviation
SKT	Statistical Knowledge for Teaching
SRA	Statistical Reasoning Assessment
T	Tail

1. INTRODUCTION

People often have to make decisions in certain situations depending on their experiences or chance and analyze those situations, such as selecting health insurance policies, making investments, or monitoring the efficiency of their daily work. In order to be conscious of certain circumstances, including crime rates, increase in population, the spread of disease, employment developments, national and international political developments, employment trends, industrial production, and educational achievement, they must also engage in community actions and discussions as part of their civic duties (Gal, 2002; Watson & Callingham, 2003). It is crucial to talk about enhancing people's statistical thinking because all of these everyday problems involve statistical elements (Garfield & Ben-Zvi, 2007).

According to the National Council of Teachers of Mathematics (NCTM, 2000) and the American Statistical Association (ASA, 2007), students need to learn how to reason quantitatively, which calls for acquiring and using a set of skills in order to become knowledgeable citizens, astute consumers, successful workers, contented and healthy individuals. Turkey's Middle School National Mathematics Curriculum began to incorporate certain statistics ideas in 2005, and in 2011, statistics and probability were added as a section. Students were expected to create research questions, present data using tables and graphs, perform comparisons and analyses by finding average and range, and determine mean, median, and mode by examining the data as part of the Elementary School National Mathematics Curriculum (2013a), which covered probability and data. The Turkish Secondary School National Mathematics Curriculum (2013b) incorporated probability, data, and counting as a section. Students were supposed to learn about distribution and central tendency in this subject area, as well as how to present data.

Until 2018, not every preservice mathematics teacher took statistics courses in Turkey. With the regulation of the Council of Higher Education (2018), probability and statistics courses were added to both primary and secondary school mathematics education programs in Turkey. However, teaching probability and statistics courses were only added as compulsory courses in primary school mathematics education programs. In other words, there is a difference between primary and secondary school preservice mathematics teachers in terms of taking probability and statistics courses and, therefore, their content and pedagogical content knowledge. However, it is crucial that teachers possess statistical reasoning before they teach because what they know is related to what pupils know (Heaton & Mickelson, 2002; Uçar & Akdoğan, 2009; Yolcu, 2012). Thus, it is necessary to analyze the statistical reasoning of preservice mathematics teachers (Bulut, 2001; Shaughnessy, 2007).

The aim of this study is to examine the scores of statistical reasoning skills of preservice mathematics teachers studying at Bogazici University and to examine their reflections on statistical knowledge to teach statistics and experiences in learning statistics. To teach people who can reason statistically, preservice and in-service mathematics teachers at the primary and secondary levels must develop statistical reasoning skills (Garfield, 2002; Yolcu, 2012). Determining the statistical reasoning skills of preservice mathematics teachers and their current situation may help teacher preparation programs include probability and statistics courses or teaching probability and statistics courses, as this is an attempt to prepare them for teaching statistics.

2. LITERATURE REVIEW

2.1. Nature of Statistics vs Nature of Mathematics

Even if statistics is considered one of the parts of mathematics, it has significant features that are not found in mathematics. Moore (1988) stated that statistics is a general intellectual method applied wherever data, variation, and chance appear. He suggested that it is an independent discipline with its own set of fundamental concepts rather than a subfield of mathematics since chance, data, and variation are all around us in life. Therefore, statistics is a distinct discipline that is separate from mathematics. Also, Moore (1990) stated that statistics is more than a discipline since it has some assertion to be basic method of inquiry; in other words, it is a general way of thinking. According to him, this different thinking method is more important than any other specific technique making the discipline.

Statistics is a methodological discipline including a coherent set of ideas and tools for dealing with data. Circumstances around the individual's words vary, and the person needs to deal with such a mass of data. Statistics helps to provide meanings to deal with such data by considering the omnipresence of variability. Cobb and Moore (1997) stated that the demand for statistics emerged from the omnipresence of variability. In mathematics, an exact answer can be found inductively or deductively at the conclusion of the process. Conversely, in statistics, inferences can be varied because of the variability of data. Therefore, the omnipresence of variability is a critical point in understanding the nature of statistics. In other words, Cobb and Moore (1997), DelMas (2004), and Pereira-Mendoza (2002) stated the differences between statistics and mathematics by considering the reasoning under uncertainty and the deterministic nature of mathematics.

Statistics is distinct from mathematics since it requires the consideration of numbers within context, not only the numbers. For instance, 33,31,30,31,31 are the numbers, and the pattern has no meaning without context. These numbers are the weather report for Istanbul between the 15th and 19th of July 2023 (https://www.mgm.gov.tr/tahmin/il-ve-ilceler.aspx?il=İstanbul). The pattern becomes meaningful when the data context is given or explained. Cobb and Moore (1997) explain that context provides meaning in statistics, especially in analysis of data, since it examines whether the patterns have any value. On the other hand, context blacks out the structure in mathematics, and it is considered irrelevant since the models are abstractly presented and computational algorithms are done in transition from models to methods and from methods to results processes.

As Cobb and Moore (1997) assert, statistical thinking is not a mere extension of mathematical thinking but a unique approach. It is rooted in “the need for the data”, “the impotence of data production”, “the omnipresence of variability”, and “the quantification and explanation of variability” in the given context. In contrast, mathematical thinking is characterized by abstraction, deduction, and symbolic representations and its use of mathematical tools and strategies to solve diverse problems.

2.2. Mathematical and Statistical Knowledge for Teaching

The distinctions between mathematics and statistics are discussed in the previous section. Because of their distinct nature, the knowledge needed to teach them is likely different from each other (Moore, 1988). Doing statistics includes basic nonmathematical processes including explaining the meaning of data by looking at the context, selecting relevant study designs to make inferences, and interpreting the results by considering the meaning of the data (Groth, 2007). Because of their distinct nature, statistics knowledge for teaching (SKT) is not directly corresponding to mathematics knowledge for teaching (MKT). On the other hand, preservice mathematics teachers need to develop both mathematical and statistical knowledge for teaching statistics. In this manner, frameworks for such knowledge types existing in the literature are reviewed.

2.2.1. Mathematical Knowledge for Teaching

Until Shulman's (1986) framework, the main emphasis for the teachers' knowledge was on content knowledge, and teachers completed such courses including the content knowledge at the university level. With Shulman's (1986) framework, a change began since the emphasis shifted to include both content and pedagogical knowledge to teach mathematics. A framework for teacher subject matter knowledge, comprising curriculum knowledge, pedagogical content knowledge (PCK), and content knowledge, was proposed by Shulman. "The amount and organization of knowledge per se in the mind of the teacher" (p.9) is how he (1986) described the first component. Knowledge "which extends beyond knowledge of subject matter per se to the dimension of subject matter knowledge for teaching" is the definition of the second one, a new construct for teacher knowledge (Shulman, 1986, p.9). Curriculum knowledge, the last component, refers to both vertical and horizontal knowledge of a topic's curriculum. According to Hill *et al*. (2005), this framework served as the cornerstone for teacher knowledge. Ball and her colleagues focused on building mathematical knowledge for teaching in the light of Shulman's work (MKT; Ball & Bass, 2000; Ball *et al*., 2001; Hill *et al*., 2005). Six elements of mathematical knowledge for teaching (MKT) were identified by Ball *et al*. (2008). This approach separates teachers' knowledge into two categories: subject matter knowledge and pedagogical content knowledge. Common content knowledge (CCK), specialized content knowledge (SCK), and horizon content knowledge are the three categories of subject matter knowledge. Knowledge of content and students (KCS), knowledge of content and teaching (KCT), and knowledge of content and curriculum are the three categories of pedagogical content knowledge. Ball and colleagues defined specialized content knowledge as mathematics knowledge rather than pedagogy, in contrast to Shulman's framework (Hill *et al*., 2005). Table 2.1 includes Ball and colleagues' (2008) six components of teachers' knowledge.

Table 2.1. Six components of Ball et al. (2008) framework.

Component	**Definition**
Common content knowledge	Knowledge that teachers need to teach their students
Specialized content knowledge	Knowledge of various representations and unusual student strategies
Horizon knowledge	Knowledge of how the mathematical topics are over-related to the span of mathematics
Knowledge of content and students	Knowledge of common students' difficulties with content and their thinking patterns
Knowledge of content and teaching	Knowledge of content-specific teaching strategies
Knowledge of content and curriculum	Vertical and horizon knowledge of mathematics and curriculum

2.2.2. Statistical Knowledge for Teaching

Since mathematics and statistics have distinct natures, mathematical knowledge for teaching (MKT) is different from statistical knowledge for teaching (SKT). When preparing preservice teachers, it is important to take account the distinction between MKT and SKT (Groth, 2007). The SKT required to effectively teach statistics has been the subject of research after (Burgess, 2006; Groth, 2007, 2013; Pfannkuch & Ben-Zvi, 2011).

Burgess (2006) suggested the first step in constructing a foundation for SKT. In order to teach statistical investigation, his framework includes both SKT and statistical thinking components. Burgess' (2006) framework is an eight-by-four two-dimensional matrix with the following columns: common knowledge of content, specialized knowledge of content, knowledge of content and students, and knowledge of content of teaching. (p. 4) The rows are the components of statistical thinking and empirical inquiry as outlined by Wild and Pfannkuch (1999): “need for data”, “transnumeration”, “variation”, “reasoning with models”, “integration of statistical and contextual thinking”, “investigative cycle”, “interrogative cycle”, and “dispositions” (p. 4) as well as the four types of knowledge defined by Ball and Hill (2015).

By examining the distinctions between MKT and SKT as well as the concepts of Hill, Schilling, and Ball's (2004), Groth (2007) hypothesizes a framework. Using this framework, Groth (2007) describes elements of common and specialized knowledge for statistical problem-solving using the stages: question formulation, data collection, data analysis, and result interpretation. Groth (2013) extended his framework for SKT by incorporating the elements of pedagogical content knowledge from Ball *et al*. (2008), the concept of key developmental understandings (Simon, 2006), and pedagogically powerful ideas (Silverman & Thompson, 2008). He recategorized knowledge of content and students from subject matter knowledge to pedagogical content knowledge. Additionally, he incorporated elements from Ball *et al*. (2008) that were absent from his earlier approach. Also, he proposed that the basis for advancing statistical knowledge in instruction is the development of a fundamental understanding of statistical knowledge. In other words, he believes that teachers can develop pedagogical content knowledge as pedagogically powerful ideas after they have firsthand experience with key developmental understanding.

In addition to mathematics, mathematics teachers need to have statistical common and specialized content knowledge in order to teach statistics. Additionally, they need to think about the advantages of teaching statistics and mathematics together, as well as design certain pedagogical ways to teach statistics, particularly the statistical "investigative cycle" (Wild & Pfannkuch, 1999, p. 225). To be ready to teach statistics, mathematics teachers need to master both MKT and SKT, even if they differ from one another.

Groth (2013) states a developmental structure for statistical knowledge for teaching. The core of his framework is that teachers need to have a key developmental understanding and knowledge of the subject matter (common content knowledge) before developing pedagogical statistical knowledge. The content knowledge teachers need to teach is based on the Guidelines for Assessment and Instruction in Statistics Education (GAISE) K-12 report (Franklin *et al*., 2007). The four stages of the statistical investigative cycle and the three developmental levels make up the two dimensions of the framework for the statistics education curriculum in grades K-12 that is included in the GAISE report. The sophistication and abstraction levels rise one after the other, even in the absence of clear definitions for each level. The information in level A pertains to themes for inexperienced statistics learners, level B represents more sophisticated statistical content, and level C comprises more advanced content. Each level is specified for specific content (Franklin *et al*., 2007). Preservice mathematics teachers need to have statistical knowledge of the four phases of the investigative cycle at all GAISE levels in order to be equipped to teach statistics.

This study focuses on preservice mathematics teachers' content knowledge to teach statistics rather than their pedagogical knowledge to teach statistics. They develop minimal experience in this knowledge type until they enter the classroom (Stuart & Thurlow, 2000), therefore the main focus is given to the statistical content knowledge of preservice mathematics teachers.

2.3. Statistical Reasoning

From primary school to university, statistics education aims to develop students' statistical literacy, reasoning, and thinking skills. Despite their uniqueness, these domains might be misused due to their interdependencies and overlap (Chance, 2002; Delmas, 2002;

Garfield, 2002; Garfield & Ben-Zvi, 2007; Rumsey, 2002). These domains are distinct from one another in terms of a person's cognitive processes, despite the fact that they are difficult to differentiate. Because statistical literacy is the basis for both statistical reasoning and statistical thinking, and because statistical thinking is deeper in thought than statistical reasoning, Garfield and Ben-Zvi (2007) argue a hierarchy between both areas. The next sections provide a hierarchy between the cognitive processes for each domain and compare them.

2.3.1. Statistical Literacy, Reasoning, and Thinking

2.3.1.1. Statistical Literacy. Numerous scholars have provided explanations of statistical literacy such as Wallman (1993), Gal (2002), and Watson and Callingham (2003). On the other hand, Gal (2002) and Rumsey (2002) explain with more detailed information. Gal (2002) proposes a model for statistically literate persons consisting of "knowledge and dispositional elements" (p. 4). Literacy skills, statistical and mathematical knowledge, content knowledge, and critical questions are among the knowledge elements. Beliefs and attitudes regarding statistics are among the dispositional elements. Knowledge elements describe a person's ability to understand, analyze, and assess statistical concepts. This model defines literacy abilities as the ability to understand messages presented in a specific format, such as lists, texts or graphs; knowing why data are needed and how to generate it, being familiar with the fundamental concepts and terms of descriptive statistics, being familiar with the representations of graphs and tables and how to interpret them, comprehending the fundamental parts of probability, and understanding of making inferences or conclusions are all components of statistical knowledge. Mathematical knowledge is the capacity to comprehend the procedure in the statistical investigation in order to interpret the data meaningfully. Critical questioning involves asking specific questions to assess the reasonableness of the data and determine whether data support or contradict the assertions (Gal, 2002). Additionally, since context affects data and a statistically literate person can ask probing question to ascertain the veracity of a claim, context and skepticism are the hidden components of this model. Second, the term "disposal elements" describes a person's capacity to express and interact with others regarding their opinions on the argument or arguments that arise from skeptic inquiries. Concerns regarding the validity of the inferences drawn throughout the statistical investigation process are known as skeptic questions.

Rumsey (2002) explains that being a statistically literate person and having research skills are the ability of understanding and using statistical terms such as mean, median, and standard deviation. In other words, individuals should possess the ability to question, explain, argue, assess, and make decisions about the information and the data. According to Rumsey (2002), a person should have statistical competence, including data awareness, understanding of basics, data production and analysis of the results, interpretation at a basic level, and basic communication skills, which are foundations for statistical reasoning and statistical thinking. Data awareness is related to the existence of data worldwide, the misleadingness of data, and the implications of decisions in our lives. Understanding basics refers to the ability to connect statistical ideas with other concepts, explain and use them in other concepts, and answer questions about them. Data production and results include posing a research question(s) and collecting appropriate data considering the question(s). Then, interpretation at a basic level is the ability to interpret a conclusion and to explain its meaning. Effective communication skills facilitate the conveyance of information derived from data and allows one to understand the perspective of others. According to Rumsey (2002), each of these competencies is a part of statistical literacy and serves as the basis for statistical reasoning and thinking. Additionally, she discussed how introductory statistics courses should prepare students for statistical literacy and help them develop their research skills.

Statistical reasoning is the foundation of statistical thinking and statistical reasoning, according to Gal (2002) and Rumsey (2002. Because statistical reasoning necessitates a higher level of cognitive involvement that statistical literacy due to the ordered relationship between these three domains, it is discussed in the following section.

<u>2.3.1.2. Statistical Reasoning</u>. Statistical reasoning is defined by Garfield and Chance (2000) as the way people reason with statistical ideas and make sense of statistical information, which includes making interpretations based on data sets, representations of data, or statistical summaries of data. Much of statistical reasoning combines ideas about data and chance, leading to making inferences and interpreting statistical results. In addition to this explanation, Wild and Pfannkuch (1999) define engaging statistical reasoning as a process, namely "shuttling back and forth" between data and context.

According to Garfield (2002), statistical reasoning is the ability of understanding data, reasoning with data, and making interpretations from data sets. Despite the fact that there is no consensus about how to assess the level and accuracy of reasoning as well as how to create statistical reasoning, it is recommended that to reason statistically need to be improved at the conclusion of statistics education (Garfield, 2002). Students that possess skills to reason statistically are therefore expected to use their reasoning in distinct cases and arrive at accurate judgements and interpretations (Garfield, 2002; Karatoprak, 2014). In order to identify the types of correct reasoning skills that students should acquire and the types of incorrect reasoning skills that they shout not, the "Statistical Reasoning Assessment (SRA)" was developed (Garfield, 2003, p. 24). The types of correct reasoning are "reasoning about data", "reasoning about representations of data", "reasoning about statistical measures", "reasoning about uncertainty", "reasoning about samples", and "reasoning about association" (Garfield, 2003, p. 24); the incorrect reasoning skills are "misconceptions involving averages", "the outcome orientation", "good samples have to represent a high percentage of the population", "the law of small numbers", "the representativeness misconception", and "the equiprobability bias" (Garfield, 2003, p. 25). All these types of reasoning are explained clearly under the Types of Statistical Reasoning.

2.3.1.3. Statistical Thinking. Given the requirement for a deep understanding of statistical processes and inferences, thinking statistically can be distinguished from statistical literacy and statistical reasoning (Garfield & Ben-Zvi, 2007). In other words, it is the mindset and conduct of a statistician (Chance, 2002; DelMas, 2002; Pfannkuch & Wild, 2000; Wild & Pfannkuch, 1999;). According to Wild & Pfannkuch (1999), the foundation statisticians' art is statistical thinking.

Wild, a statistician, and Pfannkuch, a mathematics educator (1999), constructed a theory on statistical thinking to determine the processes and strategies used by statisticians and their type of thinking styles, which statistics learners should develop. At the end of the research, a four-dimensional framework is stated for statistical thinking in empirical enquiry. The framework involves "the investigative cycle, types of thinking, the interrogative cycle, and dispositions" (p. 226). The investigative cycle focuses on an individual's actions and thinking during the statistical investigation course. Mackay & Oldford's (1994) PPDAC model is adapted since abstracting and solving a statistical problem grounded in real life are concerned. Types of thinking are divided into two categories, fundamental types of statistical thinking: recognition of the need for data (statistical impulse), transnumeration (changing or reforming data representation for better understanding), consideration of variation (concerning uncertainty derived from omnipresent variation), reasoning by using statistical models, incorporating the contextual and statistics knowledge (forward back relationship); and general types of statistical thinking: strategic thinking (being aware of the constraints of planning), seeking explanations, modeling and applying techniques. The interrogative cycle is a recursive cycle of generating, seeking, interpreting, criticizing, and judging the activities of the thinker. Disposition involves curiosity and awareness (noticing variation and wondering why questions), imagination (examining the situations from various angles), skepticism (a constant tendency for seeking logical flaws when unfamiliar ideas and information are received), and being logical.

Snee (1990) asserts that statistical thinking is a "thought process" because variation is present in everything and everywhere. Therefore, individuals need to recognize, characterize, quantify, control, and minimize variation to create possibilities for development (Britz *et al*., 1997; Mallows, 1998; Dransfield *et al*., 1999). The American Statistical Association (ASA) has also approved Moore's (1997) list of components of statistical thinking, which include the necessity of data, the significance of data generation, the omnipresence of variability, and the measurement and modeling of variability. Chance (2002) explains that the components of statistical thinking are habits of the mind.

Garfield, DelMas, & Chance (2003) stated that statistical literacy involves the ability of organizing data, constructing and displaying tables, and working with various data representations. Also, it involves the understanding of symbols, vocabulary and concepts, as well as the understanding of probability with the idea of uncertainty. In order to evaluate statistical results and explain statistical processes, one should be able to relate concepts or combine notions about chance and data. This is known as statistical reasoning. Statistical thinking includes the understanding of the reasons behind the conclusions of statistical investigations, as well as the key concepts that underpin them, such as the omnipresent nature of variation and the appropriate method of data analysis.

Despite their similarities, to literate, to reason, and to think statistically seem to be similar, proper distinctions can be made by considering the cognitive outcomes (DelMas, 2002). According to him, there are two overlapping figures to explain literacy, reasoning, and thinking domains. The first figure is independent domains with some overlap. The second one is the reasoning and thinking domains within literacy. Latter emphasizes statistical literacy; in other words, it refers to a statistical expert as not just an individual who knows how to think statistically but a person who is entirely statistically literate. DelMas (2002) also gives the verbs to distinguish these domains in the instructions, such as for basic literacy, identify, describe, rephrase, translate, interpret, and read; for reasoning, why?, how? and explain (the process); and for thinking, apply, critique, evaluate, and generalize.

2.3.2. The Types of Statistical Reasoning

Statistical reasoning and mathematical reasoning can be confused when statistics and mathematics are considered the same (Garfield & Gal, 1999). Gal and Garfield (1997) distinguish between statistical reasoning and mathematical reasoning in four ways. Firstly, data are considered as numbers within a context, the core of meaning, the basis of making interpretations from results, and the motivation for the procedures in statistics. Secondly, the vast and uncertain data differentiate statistical investigations from being precise, in other words, the limited nature of mathematical explorations. Mathematical ideas and methods are used to solve statistical problems. However, the necessity for correct computation applications is subordinated to need for chosen, deliberate, and accurate use of technological tools or software. Finally, many statistical problems have yet to have a single solution or answer. They begin with a query, and results come from an opinion backed up by particular facts and presumptions. These responses need to be examined regarding the nature of data and evidence, method appropriateness, and reasoning quality.

In order to assess statistical reasoning, Garfield (2003) clarified the types of reasoning that students should develop, and she constructed Statistical Reasoning Assessment (SRA). According to her, the correct reasoning types are “reasoning about data, representation of data, statistical measures, uncertainty, samples, and association” (p.25). Incorrect reasoning skills are “misconceptions about some probability concepts and statistical measures, averages, outcome orientation, bias about good sample, the law of small numbers, representativeness misconception, and equiprobability bias” (Garfield, 2003, p. 26).

Reasoning About Data: Statistics is based on data since conclusions are directly affected by how data is produced and analyzed (Moore, 1990; Garfield & Ben-Zvi, 2008). Data are the values of variables, and they have meaning with context (Moore, 1990; Garfield & Ben-Zvi, 2008). The data type can be defined as either quantitative or qualitative based on the context; then, inferences can be made according to the data type. In order to understand data, different representations need to be used. For example, histograms or dot plots can be used for quantitative data, while pie graphs or bar charts can be used for qualitative data (Garfield & Ben-Zvi, 2008). Consequently, the ability to recognize variables in the context of data production, identify the type of data as quantitative or qualitative, discrete or continuous; and understand the various representations of the data as tables, graphs, or statistical measures are all components of data reasoning (Garfield, 2002).

Reasoning About Representation of Data: It is the ability to read and understand a graph first, then modify it to make interpretations and explain the general characteristics (Garfield, 2003). In one of the studies, Curcio (1987) discusses graph comparison skills in three categories as reading the data, reading between the data, and reading beyond the data. These skills respectively agree upon the ability to read the graph, to interpret the presented data, and to make inferences from the given data as understanding what a plot means to represent a sample, understanding how graphs may be modified for better representation, and seeing the beyond of data to recognize general characteristics such as shape, center and spread (Garfield, 2002).

Reasoning About Statistical Measures: Understanding of measures of center, spread, and position of the data set, determining which to use in which situations, determining the accuracy of large samples versus small samples, and understanding why a good summary of data includes both a measure of center and s a measure of spread are all covered (Garfield, 2002).

Measures of central tendency are the representativeness of the data set (Gay *et al*., 2009). The most common method for determining the frequent score obtained from the data, the midpoint of the scores arranged according to their values, and the arithmetic average of the scores are mode, median, and mean, respectively. Since the mean includes each score, it is typically chosen to describe the data. However, when characterizing data with extreme scores or outliers, the median and mode should also be taken into account. These measures need to be adequate for characterizing data. When describing and comparing data sets in such cases, a measure of spread should be considered. The most frequently used metrics are range, standard deviation, and variation. The range represents the difference between the data's highest and lowest scores. Each score's spread from the mean is shown by the standard deviation, which is the square root of variance (Gay *et al*., 2009). Standard deviation, like the mean, is preferable since all scores are considered. Mean and standard deviation give information about the scores, particularly when comparing two data sets. For instance, two data sets might differ in their spread but share the same mean. As a result, these data sets are regarded as different, and examining the data sets requires considering the measures of spread.

Reasoning About Uncertainty: It is using ideas of randomness, chance, and randomness to make judgments about uncertain events, knowing that all outcomes are equally likely, and selecting a correct method to determine the likelihood of different events such as probability tree diagram, a simulation to toss a coin, or a computer program (Garfield, 2002). For example, it is uncertain which of the two possible outcomes-head or tail-will be displayed when a fair coin is tossed. Since it is difficult to depict anything other than the head and tail, these possible outcomes follow a pattern. However, each result is unpredictable. Accordingly, the phenomena being studied is referred as random (Moore, 1990), and it is possible to quantify is likelihood.

Reasoning About Samples: It is about relating samples to a population in order to make inferences, selecting a sample to represent a population accurately, differentiating unrepresentative samples, and being skeptical about inferences that are made using small or biased samples (Garfield, 2002). The sample is one of the population subsets, and inferences are made about the population by examining this subset. To choose a representative sample, random selection should be made since there must be an equal chance for each member of the population. On the other hand, random selection does not ensure accurate representativeness since the sample size also needs to be considered. In other words, unbiased inferences can be drawn with a sample size that is large enough to comprise the variability in the population.

Reasoning About Association: It refers to making judgments and interpretations of a relationship between two variables, examining bivariate relationships such as a two-way table or scatterplot, and appreciating that a strong correlation between two variables does not mean a cause-and-effect relationship (Garfield, 2002). For example, there might be a strong correlation between working hours and achievement, but it does not imply that the achievement in high school arises from more working hours.

2.3.3. Incorrect Statistical Reasoning

Misconceptions Involving Averages: Misconceptions related to averages might vary. Garfield (2002) gives some misconception examples developed by students such that averages are viewed as the most common number (the value that occurs more often than the others) and averages are obtained by adding all the numbers and then dividing by the number of data values (regardless of outliers). In some cases, they consider mean and median to be the same thing. Also, there might be a belief to use the differences in the averages to compare groups.

The Outcome Orientation: An intuitive model of probability leads students to decide yes or no about single events instead of considering a series of events (Konold, 1989). In other words, it is related to considering uncertainty. Students undergoing outcome orientation, for instance, explain that there is 70% chance of rain in ten days, as it should occur every ten days, as observed in SRA (Garfield, 2003). However, the likelihood that there will not be any rain every single day is still a 30%. Because they concentrate on the result of a single event rather than being able to look at the sequence of occurrences, students who have this misconception make decisions for each independently rather than considering the entire ten days. By their logic, 95% indicates that it rains, 50% indicates that it is uncertain, and less than 50% indicates that it does not.

Good Samples Have To Represent A High Percentage Of The Population: Many people have strong intuitions about random sampling, and most are wrong (Kahneman *et al*., 1982). Garfield (2002) asserts that students should intuitively understand the importance of the sample size to population ratio. A sample is considered good if its size is a large percentage of the population (Garfield & Ben-Zvi, 2008). They believe that sample size ought to rise in tandem with population size. However, the sample size absolute terms is more significant than the sample size relative to the population for accurate estimation (Smith, 2004). Since the sample is selected at random, a carefully chosen sample can yield reliable estimations even if it represents a small portion of the population (Garfield & Ben-Zvi, 2008). Stated differently, samples chosen at random can yield results that are unbiased and representative of the population.

The Law of Small Numbers: People believe that samples should resemble the populations. (Garfield, 2002). Also, they believe that no matter how small, any two random samples will be more similar to each other and the population since they are randomly selected. On the other hand, random selections vary, so these two samples might differ (Garfield & Ben-Zvi, 2008). This misconception is complementary to the previous misconception. Individuals with the third misconception do not consider the sample's random selection while considering the sample size. However, individuals with this misconception disregard the sample size when evaluating randomness.

The Representativeness Misconception: Individuals estimate the likelihood of a sample based on how close it is to the population (Garfield, 2002). In this misconception, a particular sequence of *n* fair coin flips having approximately heads and tails mix is considered more likely than a sequence with more heads and fewer tails. When a fair coin is tossed six times, for instance, getting HTHTHT appears to be more likely than getting HTHHHH (Kahneman *et al*., a981). When a fair coin is tossed six times, people with this misconception consider that getting HTHTHT is less likely than getting HTTTHT because they reason that the first outcome has a pattern and is therefore less likely (Kazak 2008; 2009). On the other hand, the probability value of each event is $\left(\frac{1}{2}\right)^6$. The Gambler's Fallacy, as defined by Garfield (2002), is the belief that, in a fair coin toss, a tail is more likely to occur on the subsequent toss than a head after a long series of heads. This is fallacy because, in a fair coin, the probability of obtaining a head or a tail on the subsequent toss is equal.

The Equiprobability Bias: In this misconception, different outcomes of an experiment are considered equally likely (Garfield, 2002). For example, the likelihood of getting two fives and the likelihood of getting one five and one six on two rolls are seen as being equally likely when two dice are thrown at the same time (Kazak, 2008; Kazak, 2009). However, since sample space is equal $\{55\}$, there is only one way to obtain two fives, whereas one six and one five can be obtained in two ways $\{56, 65\}$. Therefore, the likelihood of obtaining two fives is doubled when one five and one six are obtained. This misconception prevents people from considering the sampling.

2.3.4. Statistical Reasoning Assessment (SRA)

Traditional paper-and-pencil assessment instruments frequently concentrate on computing abilities or problem-solving instead than emphasizing reasoning and understanding (Garfield, 2002). According to Garfield (2003), the majority of test questions that involve statistical information frequently lack the relevant context and primarily focus on the exactness of statistical computations, the proper use of formulas, or the correctness of graphs and charts. However, questions or tasks that include correct or incorrect responses need to accurately represent students' statistical thinking. Therefore, they provide restricted information about students' statistical reasoning process and their ability to make statistical investigations (Gal & Garfield, 1997).

Garfield and Chance (2000) suggest techniques to examine students' statistical reasoning, such as case studies or authentic tasks, concept maps, critiques of statistical ideas or issues in the news, minute papers, enhanced multiple-choice items, and statistical reasoning assessment. Additionally, during class actives, students’ statistical reasoning can be evaluated through informal methods like asking them to provide written or verbal interpretations of data, to explain of concepts, or to match different types of representations (boxplots to histograms, graphs to statistics) (Garfield, 2002). Even though statistical reasoning can best be assessed with one-to-one communication with students, interviews, observation, or in-depth student work, a well-designed paper-and-pencil instrument can be used to get information about students' reasoning (Garfield, 2002).

In order to assess the efficacy of a new statistical curriculum in high school, the NSF-funded Chance Plus Project (Garfield, 1991; Konold, 1990) constructed and validated the Statistical Reasoning Assessment (SRA), which was subsequently published in 2003. There is currently no alternative instrument to assess high school students' understanding of statistical concepts and application of statistical reasoning (Garfield, 2003). Additionally, SRA can offer a useful method for scoring and administering in large-scale assessments. In numerous studies, it was employed and acknowledged as a reliable and valid assessment tool (Liu, 2002; Tempelaar, 2004). SRA is modified and utilized in research projects in a number of English-speaking countries such as Australia and the United Kingdom (Garfield, 2003) and it has also been translated to Chinese, Spanish, and French (Garfield & Gal, 1999).

There are 20 multiple-choice questions on the SRA. Students must choose the option that best fits their reasoning for each item that presents a probability or statistics problems with multiple answers (Garfield, 2003). Choices of items are stated to examine correct reasoning, possible misconceptions, or false instances. As a result, there are two main categories: correct reasoning skills and misconceptions. There are eight scales in these two main categories. Alternatives and corresponding items for each scale relating to correct reasoning skills and misconceptions are given in the following tables.

Table 2.2. Correct reasoning skills measured by SRA and the corresponding items and alternatives (Garfield, 2003, p. 27).

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	8c
b. Uses combinatorial reasoning	13a, 18b, 19a, 20b
4. Understand independence	9e, 10cdf, 11e
5. Understand sampling variability	14b, 15d

Table 2.2. Correct reasoning skills measured by SRA and the corresponding items and alternatives (Garfield, 2003, p. 27). (cont.)

Scales	**Items and Alternatives**
6. Distinguish between correlation and causation	16c
7. Correctly interpret two-way tables	5.1d
8. Understand the importance of large samples	6b, 12b

Table 2.3. Misconceptions measured by SRA and the corresponding items and alternatives. (Garfield, 2003, p. 27).

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 means with median	17a
2. Outcome orientation	2e, 3ab, 11abd, 12c, 13b
3. Good samples have to represent a 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

Each item could be scored as correct and incorrect, and total correct scores could be obtained. The total correct scores could not adequately reflect individuals' reasoning abilities (Garfield, 2002). Therefore, a method is used in which each response to an item is considered to identify the correct or incorrect reasoning (Garfield, 2002). Eight scales of correct reasoning are stated, and eight scales of incorrect reasoning are also stated (see Table 2.2 and Table 2.3). The 7th item is not included in a correct reasoning scale since it seems effective in assessing misconception (Garfield, 2002). Each item and alternative and corresponding scales are stated in Appendix B.

2.3.5. Research on Statistical Reasoning

Groth and Bergner (2006) conducted a study outlining some aspects of the statistical content knowledge of preservice primary school teachers. They were asked to compare how the center's measures were used. They were assessed on their conceptual knowledge and procedural knowledge of mean, median, and mode. The findings indicated that 3 out of 46 participants were able to discuss hypothetical situations where the mean, median, and mode would be more appropriate measures of center. However, all the primary school preservice mathematics teachers could talk about how these measures were used. Thus, it is possible to interpret this result as a lack of conceptual knowledge that might prevent preservice mathematics teachers' understanding when to use measures of center. Alkas Ulusoy and Kayhan Altay (2017) carried out another study examining the statistical content knowledge level of preservice primary school mathematics teachers. It was discovered that although the preservice teachers could accurately calculate the standard deviation and interquartile range, they still needed to understand what these measures reveal about the variation of the data set.

Makar and Confrey (2005) examined 17 prospective secondary school mathematics and science teachers' discussion of variation in comparing two data sets. When discussing variation and distribution, they tended to utilize informal words; in other words, they did not employ proper statistical terminology very well. It showed a relationship between expression of variation and expression of distribution, despite the fact that there was a relationship between their usage of these terms and their comprehension of measures of spread. This might indicate that they needed to develop their reasoning on measures of distribution.

Leavy (2006) conducted the research on how preservice mathematics teachers compare distributions using statistical measures. To compare and analyze the data distributions, the study focused to determine the statistical concepts of primary school preservice mathematics teachers used. Participants compared the data descriptively, ignoring variation and concentrating only on measures of center. Additionally, Canada (2008) investigated the reasoning of preservice primary school mathematics teachers regarding distributions in another study. The outcomes were comparable. When comparing the distributions, these studies suggested that preservice primary school mathematics teachers typically use measures of center rather than measures of spread.

Uçar and Akdoğan (2009) investigated how primary school pupils perceive average. Half of the 18 pupils, they discovered, were able to comprehend average as a representative data value. Additionally, the majority of students mistake average for an arithmetic mean, and they primarily concentrate on calculating the arithmetic mean rather than conceptualizing it.

In a case study, Lavigne and Lajoie (2006) conducted with six 7th-grade students participating in the four inquiry stages: posing a question, collecting data, analyzing data, and representing data. They found that students used a variety of models or ways of reasoning, including "population-relevant, standardization-oriented, frequency-based, and organization-based" (p. 647). However, the majority of statistical reasoning modes contextualized specifics of statistical reasoning types, such as samples, data, statistical measures, and data representation. Additionally, they claimed that the variety of methods in which pupils reasoned and thought demonstrated the potential complexity of young minds.

These study results indicate that preservice mathematics teachers possess procedural knowledge instead of conceptual grasp of the measures of the center (Groth & Bergner, 2006). In addition, preservice mathematics teachers overemphasize measures of center, or they do not consider variation when comparing data sets (Canada, 2008; Leavy, 2006;). When comparing data sets, they do not mainly use statistical language to express distributions and variation (Makar & Confrey, 2005). Students and preservice teachers need to develop their conceptual understanding of representing the data (Uçar & Akdoğan, 2009) since the number of different ways that students think, and reason statistically influence their thinking level (Lavigne & Lajoie, 2006). Since what preservice teachers know is what students know or what students will be, preservice mathematics teachers' statistical reasoning needs to be examined to improve statistical reasoning skills of students (Heaton & Mickelson, 2002).

2.4. University Curricula in Probability and Statistics Education

In Turkey, mathematics education in universities is divided into two departments: primary and secondary school mathematics education. In these programs, courses are given to students to teach what mathematics concepts are and how they are taught to students; in other words, both content knowledge and pedagogical knowledge are introduced to teacher candidates. In the junior and senior years, students are specialized according to their areas, and they take courses related to pedagogical content knowledge, teaching methods, designing instructions, and assessment techniques.

According to the Council of Higher Education regulation (2018), probability and statistics courses are added to both primary and secondary school mathematics education departments in Turkey. However, teaching probability and statistics courses are only added as compulsory courses to primary school mathematics education programs. In other words, secondary school mathematics teachers need to take method courses to teach probability and statistics. Eighty-five universities have primary school mathematics education departments, and 13 have secondary school mathematics education programs in Turkey. Nine universities include probability and statistics courses as formal courses for the primary school mathematics education department. Only one secondary school mathematics education department includes a probability and statistics course in the program. On the other hand, this course is necessary for both groups of preservice teachers to have statistical reasoning skills and to educate individuals who can statistically reason.

2.4.1. SCED354: Teaching Probability and Statistics Course

At Bogazici University, the teaching probability and statistics course is included in the course catalog with the code SCED354. Basic concepts of probability, introduction to statistics, analysis of learning of probability and statistics concepts, descriptive statistics, correlation and regression, estimation and analysis techniques, distributions, and teaching methods in probability and data analysis are introduced to students in their sixth semester with three credits. In the 3rd or 4th grade, students in primary school mathematics teaching programs take this course as a compulsory course, and students in secondary school mathematics teaching programs can take it as an elective course. In other words, this course is included in the primary school mathematics education program and offered as an elective course in the secondary school mathematics education program.

This course is designed to bridge theory and practice of how students develop an understanding of key concepts in probability and data science. Students' understanding and misconceptions about probability and statistical concepts and teaching strategies for these concepts are discussed in detail. Students taking this course will be able to explore key concepts of probability and statistics, employ research-based instructional strategies to develop probabilistic and statistical understanding, to develop tasks, questions, and lessons that can be used to assess as well as further develop grades 5-12 students' understanding of probability and statistic; and to reflect on readings, videos, and activities to consider how to teach probability and statistical concepts best. Students participate in class discussions, group work, individual technology-based work, and readings and reflections. Throughout the course, historical perspectives of probability, randomness, distinction of probability and statistics, frequentist vs. Bayesian interpretations, conditional probability, multiplication rule, law of total probability, using tree to organize the computation, independence, Bayes' Theorem, random variables, specific distributions, arithmetic with random variables, expected value, SASI framework, statistical habits of mind, and technological tools are introduced to students. Students are assessed with assignments such as in-class and take-home, initial and final conceptual understanding assessments, midterm exams, and a final exam. In the problem-solving solutions sessions, the course assistant shares practice questions and their solutions according to the end of the covering the topics in a week. Students in the primary school mathematics education department have to take this course, and they have to pass this course in order to complete their programs. However, there is no such obligation for secondary school preservice mathematics teachers to take and pass the course.

2.5. Significance of the Study

Until a few decades ago, knowing how to do math took precedence over understanding and applying mathematical concepts in many situations. The vast amount of quantitative data available today influences people's decisions in daily life, politics, and the workplace. As a result, students need to develop their quantitative reasoning (NCTM, 2000). According to

Watson (1997), statistics education is necessary for students who blindly accept and believe the statements made in daily life. In order to educate students on how to be a statistically literate person, how to reason statistically, and how to think statistically, for teachers, it is essential to be, reason, and think statistically. In addition, they need to possess the pedagogical tools to develop and deepen students' statistics education. In other words, integrating content and pedagogical knowledge with these skills is vital to preparing tomorrow's teachers since the teachers' knowledge, reasoning, and thinking influence students' knowledge, reasoning, and thinking.

"Reasoning" is frequently used interchangeably with several forms of thinking, including brainstorming, critical thinking, problem-solving, and decision making. These activities involve common processes and mental operations, converting provided data into an inference or conclusion by drawing on prior knowledge (DelMas, 2004). On the other hand, Galotti (1989) defines reasoning for distinguishing it from other thinking forms. She claims that reasoning is mental activeness that changes information that has been provided, that is directed toward at least one objective (drawing conclusion or inferring something), that is consistent with leading premises (whether changed or not), and that is consistent with logic systems when all the assertions are stated. Additionally, she states that conclusions do not need to be deductively valid and that mental activeness does not need to be self-contained (the reasoner may alter premises). In other words, it is crucial to investigate reasoning in order to determine whether or not someone has changed the premises and to judge the reasoning's quality appropriately.

This study aims to examine the statistical reasoning skills of preservice mathematics teachers studying at Bogazici University and examine their reflections on statistical knowledge to teach statistics and experiences in learning statistics. When the teacher education programs were examined, with the Council of Higher Education regulation, probability and statistics courses were included in all the programs. On the other hand, only a few already include a formal course about teaching statistics for secondary school preservice teachers in contrast to primary school education programs. Suppose this study's result shows a difference between these preservice mathematics teachers' statistical reasoning skills. In that case, the course contents might be further examined towards a standardized course in teaching statistics for the departments of mathematics education. If there is no statistically significant difference, then this might describe the case with preservice mathematics teachers who are supposedly known to have the highest entrance exam score in Turkey.

3. STATEMENT OF THE PROBLEM

This study aims to examine the statistical reasoning skills of preservice mathematics teachers studying at Bogazici University and to examine their reflections on statistical knowledge to teach statistics and experiences in learning statistics.

3.1. Research Questions

The study includes quantitative analysis of statistical reasoning assessment to examine preservice mathematics teachers' statistical reasoning skills and qualitative analysis of interviews to examine their reflections on teaching statistics.

The following questions guide this research:

(i) What are the total correct reasoning skills scores and total misconceptions scores of preservice mathematics teachers studying at a high-ranked public university in Turkey, depending on whether they have taken a formal course on teaching probability and statistics or not?

(ii) Is there a statistically significant difference in correct reasoning skills scores and misconceptions scores between primary and secondary school preservice mathematics teachers studying at Bogazici University?

a. Is there a statistically significant difference in correct reasoning skills scores and misconceptions scores between primary and secondary school preservice mathematics teachers at Bogazici University who have taken the SCED 354: Teaching Probability and Statistics course and those who have not?

(iii) What are the reflections of primary and secondary school preservice mathematics teachers at Bogazici University regarding:

a. Their statistical knowledge to teach statistics,
b. Their experiences in learning statistics?

The research hypotheses for the second research question and its sub-question:

- H_02: There is not a statistically significant difference in correct reasoning skills scores and misconceptions scores between primary and secondary school preservice mathematics teachers studying at Bogazici University.

- H_02a: There is not a statistically significant difference in correct reasoning skills scores and misconceptions scores between primary and secondary school preservice mathematics teachers at Bogazici Universiy who have taken the SCED 354: Teaching Probability and Statistics course and those who have not.

3.2. Variables

This study has one dependent variable, statistical reasoning skills score, and two independent variables: the department of preservice mathematics teachers and whether or not they are taking the SCED354 course. These variables are explained below.

The Statistical Reasoning Skills: Garfield's (2003) research on assessing statistical reasoning is adopted in this study. She defines that there are correct reasoning skills that students should develop and misconceptions that students should not develop. Learners are expected to have correct reasoning skills and should not develop misconceptions about statistical reasoning. Table 2.2 and Table 2.3 (Statistical Reasoning-SRA Section) show the eight scales for correct reasoning skills and the eight for misconceptions. According to these scales, participants' answers are scored, and sixteen scores are used as statistical reasoning skills scores.

Preservice mathematics teachers: They were the students who study in the last year of the mathematics education program at the universities. Bogazici University has two departments in mathematics education: primary school mathematics education and secondary school mathematics education. Primary school mathematics teachers are expected to teach 5th, 6th, 7th, and 8th grades, and secondary school preservice mathematics teachers are supposed to teach 9th, 10th, 11th, and 12th grades.

SCED 354: Teaching Probability and Statistics Course: This is one of the courses included in the mathematics education programs at Bogazici University. In this course, basic concepts of probability, introduction to statistics, analysis of learning of probability and statistics concepts, descriptive statistics, correlation and regression, estimation and analysis techniques, distributions, and teaching methods in probability and data analysis are introduced to students in the second semester with three credits. Students enrolling in primary school mathematics education programs have to take this course and pass it in order to complete their program. In contrast, students in secondary school mathematics education programs can take it as an elective course. In this study, the SCED354 course is considered as a categorical variable for those who take SCED354 and those who do not.

Statistical Knowledge for Teaching Statistics: Groth's (2013) framework explains a developmental structure for statistical knowledge for teaching. The main of this framework is that teachers need to have a key developmental understanding and subject matter knowledge (common content knowledge) before developing pedagogical statistical knowledge. This study examines and explains preservice mathematics teachers’ common content knowledge and pedagogical approach.

Experience: Actual observation or firsthand familiarity with fact or events is regarded as the source of knowledge, according to the Oxford English Dictionary, Preservice mathematics teachers’ experiences in learning statistics are examined in this study.

4. METHODOLOGY

This study aims to examine the statistical reasoning skills of preservice mathematics teachers studying at Bogazici University, as well as their reflections on their statistical knowledge to teach statistics and their experiences in learning statistics. There are two methods in this research. In the initial quantitative part of the study, participants' statistical reasoning skills are assessed by SRA (Garfield, 2003). The second qualitative phase is conducted to explain participants' reflections on their statistical knowledge of teaching statistics and their experiences in learning statistics with a semi-structured interview.

This chapter explains the study's participants, research design, instruments, data collection process, and data analysis in tandem with its goal. The sample procedure and a sample description are included in the sample section. The study method is presented in the section on research design. The instruments and data collection process are then described. Finally, the analysis of data is explained.

4.1. Sample

The target population of this research is primary and secondary school preservice mathematics teachers studying at Bogazici University. The Council of Higher Education states that Bogazici University's primary school mathematics education programs are subject to a quota restriction of 41, while secondary school mathematics education programs are subject to a quota restriction of 21. The study sample was selected from senior year undergraduate mathematics education students in primary and secondary schools. 49% of primary school mathematics teachers and 43% of secondary school mathematics teachers are represented in our sample. Data was gathered from 20 out of 41 preservice teachers at the primary school level and 9 out of 21 preservice teachers at the secondary school level, yielding a sample size of 29.

Participants were asked whether they took courses related to probability and statistics or teaching probability and statistics. 11 out of 20 primary school preservice mathematics teachers and 7 out of 9 secondary school preservice mathematics teachers took probability and statistics courses. In contrast, 17 out of 20 at the primary school level and 2 out of 9 at the secondary school level took the SCED 354: Teaching Probability and Statistics course at Bogazici University.

Table 4.1. Distribution of participants according to the courses they took.

Department	**Courses**	
	Probability and Statistics	SCED354: Teaching Probability and Statistics
Preservice Mathematics Teachers at Primary School Level	11	17
Preservice Mathematics Teachers at Secondary School Level	7	2

4.2. Research Design

This study aims to examine the statistical reasoning skills of preservice mathematics teachers studying at Bogazici University and to examine their reflections on statistical knowledge to teach statistics and experiences in learning statistics. This study's research design includes quantitative and qualitative approaches to make interpretations from both data sets (Creswell, 2015). The first case, which is assessing the statistical reasoning skills scores of preservice mathematics teachers studying at Bogazici University, is aimed to be represented by examining quantitative data. The second case, preservice mathematics teachers' reflections on their statistical knowledge to teach statistics and their experiences in learning statistics, is aimed to be explained with qualitative data.

4.3. Instrument

Researchers and instructors assess teaching, learning, understanding, thinking, or the improvement of statistical reasoning through different methods such as interviews, one-to-one communication, and performance assessments. On the other hand, according to Garfield (2003) and Tempelaar (2004), these techniques could be more practical to administer in large-scale assessments. In contrast to these methods, Statistical Reasoning Assessment-

SRA (Garfield, 2003) is a paper and pencil test that is simple to conduct and grade on a broad scale assessment. It helps to guide researchers and instructors in designing their instructions since responses give an idea about the correct reasoning skills and misconceptions of the participants (Sundre, 2003). Some correct reasoning skills, which are "understanding how to select an appropriate average", "interpreting two-way tables", and "misconceptions involving averages" (Garfield, 2003, p. 27), are also stated in primary and secondary school mathematics education curricula in Turkey. Therefore, SRA (Garfield, 2003) was used for the first phase of this study to assess the scores of correct reasoning skills and the scores of misconceptions of primary and secondary school preservice mathematics teachers.

A semi-structured interview was used to examine preservice mathematics teachers' reflections on their knowledge to teach statistics and their experiences learning statistics. Interview questions are categorized into two parts. Part A is related to the statistical knowledge of preservice mathematics teachers. 1st, 15th, and 17th questions in the SRA (Garfield, 2003) are asked again to preservice mathematics teachers, but this time, they are asked what their positions and explanations would be if they asked these questions to their students. For instance, alternative d of the 1st question is related to the second scale of the correct reasoning skills, understanding how to select an appropriate average; alternatives a and c of the 1st question are related to the first scale of misconceptions, which is misconceptions involving averages. In this manner, in light of preservice mathematics teachers' explanations, their reflections are examined in terms of their statistical knowledge to teach statistics. Part B is related to preservice mathematics teachers' experiences in learning statistics and nine questions are asked to them, which are experiences in the courses related to probability and statistics; in the SCED354: Teaching Probability and Statistics course; in using digital tools, in their instruction design; in lesson plan preparation; and their self-reflections. A total of 12 questions is asked to preservice primary and secondary school mathematics teachers, and these questions are indicated in Appendix C.

4.4. Data Collection

This study includes a two-phase data collection and analysis approach, incorporating both quantitative and qualitative data. The first phase focuses on the statistical reasoning skills of preservice mathematics teachers studying at Bogazici University. The second phase explores their reflections on their statistical knowledge to teach statistics and their experiences in learning statistics.

For the first part, the Statistical Reasoning Assessment (Garfield, 2003) was administered to primary and secondary school preservice mathematics teachers studying at Bogazici University. Then, their responses were collected online. After these processes, data were subscribed into SPSS-Version 29.0.0.0(241) and analyzed quantitatively. The second part examined preservice mathematics teachers' reflections on their statistical knowledge to teach statistics and their experiences. In order to examine these reflections, 12 interview questions were selected. Then, a request e-mail for an interview was sent to twelve participants. These participants, six from primary and six from secondary school preservice mathematics teachers, were selected based on their total correct reasoning scores and misconception scores. Nine students agreed to be interviewed, but five of the interviews could be analyzed due to the participants' explanations. In Table 4.2. information about these five participants is explained. Twelve questions were asked in a period not exceeding 45 minutes. After the interviews were completed online, the recordings were converted to PDF and uploaded into Atlas.ti. Each response was considered and coded by the researcher and another master's student who graduated from Bogazici University. Coding was compared, and then they were analyzed and interpreted.

Table 4.2. Information about the interviewees.

	A	**B**	**C**	**D**	**E**
Department	Secondary Level	Primary Level	Secondary Level	Primary Level	Primary Level
Score of Correct Reasoning Skills	6.87	5.34	7.54	5.27	6.07
Score of Misconceptions	2.0	3.4	2.0	2.45	1.65
Sced354 Course	No	Yes	Yes	No	Yes

4.5. Data Analysis

Scoring: Quantitative data were analyzed for the first phase of the study. In the process, preservice mathematics teachers' eight correct reasoning skills scores, eight misconceptions scores, and total scores for correct reasoning skills and misconceptions were calculated. The two main categories of responses were the score of correct reasoning skills and the score of misconceptions. With eight scales for these two main categories, 16 scored are obtained. There are alternatives for the specific items that go into these scales' scores. The alternatives and items for each scale is listed in Table 2.2. and Table 2.3. Correct reasoning skills score was obtained per individual in the following way: Someone selecting the correct alternative or choice got 1 point; otherwise, 0 points. The scores of responses of the items in each scale were added and then divided by the number of items since each scale included a different number of responses. For example, the scale of correctly interpreting probabilities includes the alternative d of the second and third questions. If someone correctly and incorrectly answers one question, s/he gets .5 points. The other scale scores for correct reasoning skills were found, and the procedure was repeated to find each participant's score. The same procedure is applied for obtaining a misconceptions score, but at this part, if someone selects the alternative signaling a misconception, then s/he got 1 point. In other words, 0 means that for the misconception part, s/he did not have the misconception. For example, the scale of representativeness misconception includes alternatives a, b, and d of the ninth, e of the tenth, and c of the eleventh questions. If all the alternatives are selected for this scale, the score is 1.67. Other misconception scores were found, and the procedure was repeated to find each participant's score. In the end, each scale generating correct reasoning skills and misconception scores were added to obtain total scores, respectively.

For the first and second research questions: Firstly, scores of each correct reasoning skill and misconceptions were calculated as in the process that was explained above. The descriptive statistics of each scale were done separately. After describing the data, an independent samples t-test or Mann-Whitney U Test was administered to compare the scores of primary and secondary school preservice mathematics teachers and to compare the scores of preservice mathematics teachers who took SCED354: Teaching Probability and Statistics course and those who did not take the course. Assumptions of independent samples t-test are normality, random sampling, independence, and equal variances (Privitera, 2017).

Participants were selected randomly. Groups are independent of each other, and each individual belongs to only one of the groups. The Kolmogorov-Smirnov Test was applied to check normality and equal variances, but for more accurate results, distributions were examined, and their skewness values were considered. If the skewness is between -1.00 and 1.00, then it implies the existence of normal distribution. An Independent sample t-test was conducted for the groups holding these assumptions. On the other hand, for the groups in which the normality assumption failed, the non-parametric test of Mann-Whitney U was applied to compare the scores of the groups.

For the third research question: The reflections of primary and secondary school preservice mathematics teachers in terms of their statistical knowledge to teach statistics and their experiences in learning statistics are identified. After the interviews, they were converted into PDFs and uploaded to Atlas.ti, and the codes were created. Codes were created under two main groups to explain statistical knowledge to teach statistics and experiences in learning statistics. The codes are given in Table 4.3. Each participant's response was categorized based on these codes and grouped. Then, they were analyzed and interpreted.

Table 4.3. Codes to explain the reflections of preservice mathematics teachers.

Statistical Knowledge To Teach Statistics	
Statistical Concepts	Measures of central tendency, measures of dispersion, difference between mean and mode, finding the average, data representation, normal distributions vs skewness, outliers, theoretical probability vs experimental probability, omnipresence of variability, probability vs statistics
Pedagogical Approach	Instruction design, objectives, real-life examples, the role of students, technology integration, additional Sources
Experiences in Learning Statistics	
As A Learner	Other course(s), SCED 354: Teaching Probability and Statistics
As A Preservice Teacher	Willingness to teach statistics

5. RESULTS

There exist two main sections in this part. They examine the statistical reasoning skills of preservice mathematics teachers studying at Bogazici University and examine their reflections on statistical knowledge to teach statistics and experiences in learning statistics.

5.1. Statistical Reasoning Skills

There are three parts in this section. These include results of correct reasoning skills, results of misconceptions, and results of total scores. Outcomes associated with the first and second questions are presented in these parts. The first part consists of descriptive statistics for the scores of correct reasoning skills of primary and secondary school preservice mathematics teachers and descriptive statistics for the scores of correct reasoning skills of participants who took SCED354: Teaching Probability and Statistics course and who did not take it, and investigation of statistically significant differences among the groups. In the second part, the scores of misconceptions of these groups are stated to describe the groups statistically, and the differences are examined by considering the departments and taking the course. In the third part, the total scores of both groups and their differences are stated.

5.1.1. Scores of Correct Reasoning Skills

This part presents results for each scale of correct reasoning skills. The scales were previously outlined in Section 2.5 in Table 2.2. The results for each scale encompass descriptive statistics, which were provided for i. primary school preservice mathematics teachers and secondary school preservice mathematics teachers and ii. preservice mathematics teachers who have taken the SCED354: Teaching Probability and Statistics course and those who have not.

To begin with, an analysis was conducted to determine if there is a difference in statistical reasoning scores between primary school preservice mathematics teachers (N=20) and secondary school preservice mathematics teachers (N=9). The skewness values of the distributions were examined for each correct reasoning skills scale. Based on the skewness values, some scale scores, such as those related to correctly computing probability, did not meet the normal distribution assumption. Consequently, the results of the groups were compared with the use of an independent sample t-test for the groups in which scores were normally distributed or with the Mann-Whitney U test for those scores that did not meet this criterion. Following this, the same procedures were applied to evaluate whether there is a difference in statistical reasoning skills scores between preservice mathematics teachers who have taken the SCED354: Teaching Probability and Statistics course and those who have not.

Correctly Interpreting Probabilities: This scale includes responses to the 2nd and 3rd questions. Scores of this scale were calculated over 1. According to data analysis, scores changed between .5 and 1. So, there were participants who could answer one of the questions correctly and both questions correctly.

Table 5.1. Descriptive statistics for correctly interpreting probabilities based on departments.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
Primary School Level	20	.800	1.00	.251	.500	1.00	.500	-.442
Secondary School Level	9	.611	.500	.220	.500	1.00	.250	1.62

In the table, descriptive statistics for primary school preservice mathematics teachers (N=20) were found $\bar{x}$ = .800 and .251 standard deviation; for secondary school mathematics teachers (N=9), $\bar{x}$ = .611 and .220 standard deviation were found. According to the skewness values, the distribution of the scale scores for primary school preservice mathematics teachers was relatively normal. However, the normality test for the secondary school preservice mathematics teachers failed.

Table 5.2. Mann-Whitney U Test for correctly interpreting probabilities based on departments.

Groups	***N***	**Mean Rank**	**Sum of Ranks**	***U***	***Z***	***p***
Primary School Level	20	16.7	334	56.0	-1.85	.116
Secondary School Level	9	11.2	101			

The values of the Mann-Whitney U test are shown in Table 5.2. Primary and secondary school preservice mathematics teachers did not differ in interpreting probability correctly. A statistically significant difference between the groups was not found ($U = 56$; $p > 0.05$). The scores of secondary school preservice mathematics teachers were lower than those of primary school preservice mathematics teachers when mean ranks are considered. However, a statistically significant difference requires more than just this difference.

Table 5.3. Descriptive statistics for correctly interpreting probabilities based on the taking SCED354.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
SCED354 taken	19	.790	1.00	.254	.500	1.00	.500	-.348
SCED354 not taken	10	.650	.500	.242	.500	1.00	.250	1.04

In the table, descriptive statistics for preservice mathematics teachers who took the SCED354 course (N=19) were found $\bar{x}$ = .790 and .254 standard deviation; for preservice mathematics teachers who did not take the SCED354 course (N=10), $\bar{x}$ = .650 and .242 standard deviations were found. According to the skewness values, the distribution of the scale scores for the course takers was relatively normal, but the normality test for other preservice mathematics teachers failed.

Table 5.4. Mann-Whitney U Test for correctly interpreting probabilities based on the taking SCED354.

Groups	***N***	**Mean Rank**	**Sum of Ranks**	***U***	***Z***	***p***
SCED354 taken	19	16.4	312	68.5	-1.40	.160
SCED354 not taken	10	12.4	124			

The scores of the Mann-Whitney U test are stated in Table 5.4. Preservice mathematics teachers did not differ in correctly interpreting probability. A statistically significant difference between the groups was not found (U = 68.5; $p > 0.05$). When mean ranks are considered, the scores of the participants taking the SCED354 course were higher than those of participants who did not. However, this difference was not sufficient for a statistically significant difference.

Understanding How to Select an Appropriate Average: This scale includes responses to 1st, 4th, and 17th questions. Scores of this scale were calculated over 1. Based on data analysis, scores changed between 0 and 1. So, there were participants who could not answer any of the questions correctly and who could answer all the questions correctly.

Table 5.5. Descriptive statistics for understanding how to select and appropriate average based on departments.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
Primary School Level	20	.584	.670	.285	.000	1.00	.340	-.047
Secondary School Level	9	.630	.670	.262	.330	1.00	.500	.183

In the table, descriptive statistics for primary school preservice mathematics teachers (N=20) were found $\bar{x}$ = .584 and .285 standard deviation; for secondary school mathematics teachers (N=9), $\bar{x}$ = .630 and .262 standard deviation were found. According to the skewness values, distributions of the scale scores for primary and secondary school preservice mathematics teachers were relatively normal.

Table 5.6. Independent samples t-test for understanding how to select an appropriate average based on departments.

Groups	***M***	***SD***	***t***	***df***	***p***
Primary School Level	.583	.285	.416	27	.681
Secondary School Level	.630	.262			

Table 5.6. shows that primary school preservice mathematics teachers were not significantly different (p = .681) from secondary school preservice mathematics teachers on this scale.

Table 5.7. Descriptive statistics for understanding how to select an appropriate average based on the taking SCED354.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
SCED354 taken	19	.580	.670	.270	.000	1.00	.340	-.183
SCED354 not taken	10	.630	.670	.294	.330	1.00	.670	.204

In the table, descriptive statistics for preservice mathematics teachers who took the SCED354 course (N=19) were found $\bar{x}$ = .580 and .270 standard deviation; for participants who did not take SCED354 (N=10), $\bar{x}$ = .630 and .294 standard deviation were found. According to the skewness values, distributions of the scale scores for preservice mathematics teachers were relatively normal.

Table 5.8. Independent samples t-test for understanding how to select an appropriate average based on the taking SCED354.

Groups	***M***	***SD***	***t***	***df***	***p***
SCED354 taken	.580	.270	.493	27	.684

Table 5.8. Independent samples t-test for understanding how to select an appropriate average based on the taking SCED354. (cont.)

Groups	*M*	*SD*	*t*	*df*	*p*
SCED354 not taken	.630	.294			

Table 5.8. shows that preservice mathematics teachers who took the SCED354 course were not significantly different (p = .684) from preservice mathematics teachers who did not take the course on this scale.

Correctly Computing Probability: This scale includes responses to the 8th, 13th, 18th, 19th, and 20th questions. Results of this scale were calculated over 1. Scores changed between 0 and 1. The 8th question was the most common and correctly answer, and its percentage was 100 and 89 for primary and secondary school preservice mathematics teachers, respectively. Also, some participants could not correctly answer the 19th and 20th questions at the secondary school level.

Table 5.9. Descriptive statistics for correctly computing probability based on departments.

Groups	*N*	Mean	Median	Std. Deviation	Min	Max	IQR	Skewness
Primary School Level	20	.370	.200	.292	.200	1.00	.200	1.64
Secondary School Level	9	.356	.200	.296	.000	1.00	.300	1.38

In the table, descriptive statistics for primary school preservice mathematics teachers (N=20) were found $\bar{x}$ = .370 and .292 standard deviation. For secondary school mathematics teachers (N=9), $\bar{x}$ = .356 and .296 standard deviation were found. According to the skewness values, distributions of the scale scores for both primary school preservice mathematics teachers and secondary school preservice mathematics teachers failed.

Table 5.10. Mann-Whitney U Test for correctly computing probability based on departments.

Groups	*N*	Mean Rank	Sum of Ranks	*U*	*Z*	*p*
Primary School Level	20	15.0	300	90.0	.000	1.00
Secondary School Level	9	15.0	135			

The scores of the Mann-Whitney U test are stated in Table 5.10. Primary school and secondary school preservice mathematics teachers did not differ on correctly computing probability. A statistically significant difference between the groups was not found ($U = 90$; $p > 0.05$). Based on the sum of ranks, the scores of secondary school preservice mathematics teachers were lower than those of primary school. A statistically significant difference requires more than just this difference.

Table 5.11. Descriptive statistics for correctly computing probability based on the taking SCED354.

Groups	*N*	Mean	Median	Std. Deviation	Min	Max	IQR	Skewness
SCED354 taken	19	.347	.200	.304	.000	1.00	.200	1.66
SCED354 not taken	10	.400	.300	.267	.200	1.00	.400	1.41

In the table, descriptive statistics for preservice mathematics teachers who took the SCED354 (N=19) were found $\bar{x}$ = .347 and .304 standard deviation. For preservice mathematics teachers who did not take SCED354 (N=10), $\bar{x}$ = .400 and .267 standard deviation were found. According to the skewness values, distributions of the scale scores for groups were failed.

Table 5.12. Mann-Whitney U Test for correctly computing probability based on the taking SCED354.

Groups	***N***	**Mean Rank**	**Sum of Ranks**	***U***	***Z***	***P***
SCED354 taken	19	14.0	267	76.5	-.954	.340
SCED354 not taken	10	16.9	169			

The scores of the Mann-Whitney U test are shown in Table 5.12. Preservice mathematics teachers did not differ in computing probability correctly. A statistically significant difference between the groups was not found ($U = 76.5$; $p > 0.05$). If mean ranks are regarded, the scores of the participants who did not take the SCED354 course were higher than those of the participants taking the course. However, this difference was not sufficient for a statistically significant difference.

Understanding Independence: This scale includes responses to the 9th and 11th questions and alternatives c, d, and f to the 10th. Scores of this scale were calculated over 1.67. Based on data analysis, scores changed between 1 and 1.67. So, there were participants who could answer at least one of the questions correctly, which was the 9th question.

Table 5.13. Descriptive statistics for understanding independence based on departments.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
Primary School Level	20	1.30	1.33	.264	1.00	1.67	.590	.212
Secondary School Level	9	1.30	1.33	.311	1.00	1.67	.670	.277

In the table, descriptive statistics for primary school preservice mathematics teachers (N=20) were found $\bar{x} = 1.30$ and .264 standard deviation. For secondary school mathematics teachers (N=9), the same mean value and .311 standard deviation were found. According to the skewness values, distributions of the scale scores for primary and secondary school preservice mathematics teachers were relatively normal.

Table 5.14. Independent samples t-test for understanding independence based on departments.

Groups	***M***	***SD***	***t***	***df***	***p***
Primary School Level	1.30	.264	-.25	27	.980
Secondary School Level	1.30	.311			

Table 5.14. shows that primary school preservice mathematics teachers were not statistically significant (p = .980) from secondary school preservice mathematics teachers on this scale.

Table 5.15. Descriptive statistics for understanding independence based on the taking SCED354.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
SCED354 taken	19	1.33	1.33	.274	1.00	1.67	.670	.023
SCED354 not taken	10	1.23	1.17	.275	1.00	1.67	.420	.709

In the table, descriptive statistics for preservice mathematics teachers who took SCED354 (N=19) were found as $\bar{x}$ = 1.33 and .274 standard deviation. For preservice mathematics teachers who did not take SCED354 (N=10), $\bar{x}$ = 1.23 and .275 standard deviation were found. According to the skewness values, distributions of the scale scores for both preservice mathematics teachers were relatively normal.

Table 5.16. Independent samples t-test for understanding independence based on the taking SCED354.

Groups	***M***	***SD***	***t***	***df***	***p***
SCED354 taken	1.33	.274	.935	27	.736
SCED354 not taken	1.23	.275			

Table 5.16. shows that preservice mathematics teachers who took the SCED354 course were not statistically significant (p = .736) from preservice mathematics teachers who did not take the course on this scale.

Understanding Sampling Variability: This scale includes answers to the 14^{th} and 15^{th} questions. Scores of this scale were calculated over 1. According to data analysis, scores changed between 0 and 1. There were participants who could not answer any of these questions correctly and who could answer both questions correctly.

Table 5.17. Descriptive statistics for understanding sampling variability based on departments.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
Primary School Level	20	.225	.000	.302	.000	1.00	.500	1.03
Secondary School Level	9	.278	.500	.264	.000	.50	.50	-.271

In the table, descriptive statistics for primary school preservice mathematics teachers (N=20) were found $\bar{x}$ = .225 and .302 standard deviation. For secondary school mathematics teachers (N=9), $\bar{x}$ = .278 and .264 standard deviation were found. According to the skewness values, the distribution of the scale scores for secondary school preservice mathematics teachers was relatively normal.

Table 5.18. Independent samples t-test for understanding sampling variability based on departments.

Groups	***M***	***SD***	***t***	***df***	***p***
Primary School Level	.225	.302	.451	27	.655
Secondary School Level	.278	.264			

Table 5.18. shows that primary school preservice mathematics teachers were not statistically different (p = .655) from secondary school preservice mathematics teachers in understanding sampling variability.

Table 5.19. Descriptive statistics for understanding sampling variability based on the taking SCED354.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
SCED354 taken	19	.211	.000	.303	.000	1.00	.500	1.17
SCED354 not taken	10	.300	.500	.258	.000	.50	.50	-.484

In the table, descriptive statistics for preservice mathematics teachers who took SCED354 (N=19) were found $\bar{x}$ = .211 and .303 standard deviation. For preservice mathematics teachers who did not take SCED354 (N=10), $\bar{x}$ = .300 and .258 standard deviation were found. According to the skewness values, the distribution of the scale scores for preservice mathematics teachers who did not take the course was relatively normal.

Table 5.20. Mann-Whitney U Test for understanding sampling variability based on the taking SCED354.

Groups	***N***	**Mean Rank**	**Sum of Ranks**	***U***	***Z***	***p***
SCED354 taken	19	14.0	267	76.0	-.999	.318
SCED354 not taken	10	16.9	169			

The scores of the Mann-Whitney U test are stated in Table 5.20. Preservice mathematics teachers did not differ in understanding sampling variability. A statistically significant difference between the groups was not found (U = 76.0; $p > 0.05$). Based on the mean ranks, the scores of the participants who did not take the SCED354 course are higher than those participants taking the course, a statistically significant difference requires more than just this difference.

Distinguishing Between Correlation and Causation: This scale includes responses to only the 16th question. Scores of this scale were calculated over 1. According to data analysis, scores changed between 0 and 1. There were participants who could answer this question correctly or not.

Table 5.21. Descriptive statistics for distinguishing between correlation and causation based on departments.

Groups	*N*	Mean	Median	Std. Deviation	Min	Max	IQR	Skewness
Primary School Level	20	.350	.000	.489	.000	1.00	1.00	.681
Secondary School Level	9	.556	1.00	.527	.000	1.00	1.00	-.257

In the table, descriptive statistics for primary school preservice mathematics teachers (N=20) were found $\bar{x}$ = .350 and .489 standard deviation. For secondary school mathematics teachers (N=9), $\bar{x}$ = .556 and .537 standard deviation were found. According to the skewness values, distributions of the scale scores for both primary school preservice mathematics teachers and secondary school preservice mathematics teachers were relatively normal.

Table 5.22. Independent samples t-test for distinguishing between correlation and causation based on departments.

Groups	*M*	*SD*	*t*	*df*	*p*
Primary School Level	.350	.489	1.02	27	.316
Secondary School Level	.556	.527			

Table 5.22. shows that primary school preservice mathematics teachers were not statistically significant (p = .316) from secondary school preservice mathematics teachers on this scale.

Table 5.23. Descriptive statistics for distinguishing between correlation and causation based on the taking SCED354.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
SCED354 taken	19	.369	.000	.489	.000	1.00	1.00	.593
SCED354 not taken	10	.500	.500	.527	.000	1.00	1.00	-.257

In the table, descriptive statistics for preservice mathematics teachers who took the SCED354 (N=19) were found $\bar{x}$ = .369 and .489 standard deviation. For preservice mathematics teachers who did not take the SCED354 (N=10), $\bar{x}$ = .500 and .527 standard deviation were found. According to the skewness values, distributions of the scale scores for both groups were relatively normal.

Table 5.24. Independent samples t-test for distinguishing between correlation and causation based on the taking SCED354.

Groups	***M***	***SD***	***t***	***df***	***p***
SCED354 taken	.369	.489	-.665	27	.413
SCED354 not taken	.500	.527			

Table 5.24. shows that preservice mathematics teachers who took the SCED354 course were not significantly different (p = .413) from preservice mathematics teachers who did not take the course on this scale.

Correctly Interpret Two-way Tables: This scale includes answers to the 5th question. Scores of this scale were calculated over 2 since it has a sub-question. Based on data analysis, results changed between 0 and 1 and there were participants who could not answer both parts correctly.

Table 5.25. Descriptive statistics for correctly interpreting two-way tables based on departments.

Groups	*N*	Mean	Median	Std. Deviation	Min	Max	IQR	Skewness
Primary School Level	20	.550	1.00	.510	.000	1.00	1.00	-.218
Secondary School Level	9	.667	1.00	.500	.000	1.00	1.00	-.857

In the table, descriptive statistics for primary school preservice mathematics teachers (N=20) were found $\bar{x}$ = .550 and .510 standard deviation. For secondary school mathematics teachers (N=9), $\bar{x}$ = .667 and .500 standard deviation were found. According to the skewness values, the distribution of the scale scores for primary school preservice mathematics teachers was relatively normal. The distribution of the scale scores for secondary school preservice mathematics teachers was assumed to be normal.

Table 5.26. Independent samples t-test for correctly interpreting two-way tables based on departments.

Groups	*M*	*SD*	*t*	*df*	*p*
Primary School Level	.550	.510	.573	27	.571
Secondary School Level	.667	.500			

Table 5.26. shows that primary school preservice mathematics teachers were not significantly different (p = .571) from secondary school preservice mathematics teachers on this scale.

Table 5.27. Descriptive statistics for correctly interpreting two-way tables based on the taking SCED354.

Groups	*N*	Mean	Median	Std. Deviation	Min	Max	IQR	Skewness
SCED354 taken	19	.580	1.00	.507	.000	1.00	1.00	-.348

Table 5.27. Descriptive statistics for correctly interpreting two-way tables based on the taking SCED354. (cont.)

Groups	*N*	Mean	Median	Std. Deviation	Min	Max	IQR	Skewness
SCED354 not taken	10	.600	1.00	.516	.000	1.00	1.00	-.484

In the table, descriptive statistics for preservice mathematics teachers who took SCED354 (N=19) were found $\bar{x}$ = .580 and .507 standard deviation. For preservice mathematics teachers who did not take SCED354 (N=10), $\bar{x}$ = .600 and .516 standard deviation were found. According to the skewness values, the distribution of the scale scores for both groups was assumed to be normal.

Table 5.28. Independent samples t-test for correctly interpreting two-way tables based on the taking SCED354.

Groups	*M*	*SD*	*t*	*df*	*p*
SCED354 taken	.580	.507	-.106	27	.829
SCED354 not taken	.600	.516			

Table 5.28. shows that preservice mathematics teachers who took the SCED354 course were not significantly different (p = .829) from preservice mathematics teachers who did not take the course on this scale.

Understanding Importance of Large Samples: This scale consists of responses to the 6th and 12th questions. Results of this scale were calculated over 1. Based on data analysis, scores changed between 0 and 1 and there were individuals who could not answer any of the questions and who could answer all the questions.

Table 5.29. Descriptive statistics for understanding importance of large samples based on departments.

Groups	*N*	Mean	Median	Std. Deviation	Min	Max	IQR	Skewness
Primary School Level	20	.866	1.00	.352	.000	1.00	.500	-.565
Secondary School Level	9	.833	1.00	.250	.500	1.00	.500	-.436

In the table, descriptive statistics for primary school preservice mathematics teachers (N=20) were found $\bar{x}$ = .866 and .352 standard deviation. For secondary school mathematics teachers (N=9), $\bar{x}$ = .833 and .250 standard deviation were found. According to the skewness values, distributions of the scale scores for both primary school preservice mathematics teachers and secondary school preservice mathematics teachers were relatively normal.

Table 5.30. Independent samples t-test for understanding importance of large samples based on departments.

Groups	*M*	*SD*	*t*	*df*	*p*
Primary School Level	.866	.352	.253	27	.131
Secondary School Level	.833	.250			

Table 5.30. shows that primary school preservice mathematics teachers were not statistically significant (p = .131) from secondary school preservice mathematics teachers in understanding the importance of large samples.

Table 5.31. Descriptive statistics for understanding importance of large samples based on the taking SCED354.

Groups	*N*	Mean	Median	Std. Deviation	Min	Max	IQR	Skewness
SCED354 taken	19	.737	1.00	.348	.000	1.00	.500	-.998

Table 5.31. Descriptive statistics for understanding importance of large samples based on the taking SCED354. (cont.)

Groups	*N*	Mean	Median	Std. Deviation	Min	Max	IQR	Skewness
SCED354 not taken	10	.900	1.00	.211	.500	1.00	.130	-1.78

In the table, descriptive statistics for preservice mathematics teachers who took the SCED354 (N=19) were found $\bar{x}$ = .737 and .348 standard deviation. For preservice mathematics teachers who did not take the SCED354 (N=10), $\bar{x}$ = .900 and .211 standard deviation were found. According to the skewness values, distributions of the scale scores for the groups who did not take the SCED354 course failed.

Table 5.32. Mann-Whitney U Test for understanding importance of large samples based on the taking SCED354.

Groups	*N*	Mean Rank	Sum of Ranks	*U*	*Z*	*p*
SCED354 taken	19	13.8	262	72.0	-1.26	.207
SCED354 not taken	10	16.9	169			

The scores of the Mann-Whitney U test are stated in Table 5.32. Preservice mathematics teachers did not differ in understanding the importance of large samples. A statistically significant difference between the groups was not found (U = 72.0; $p > 0.05$). According to the mean ranks, the results of the participants who did not take the SCED354 course are higher than those participants taking the course.

5.1.2. Scores of Misconceptions

In this section, results for each misconception scale were presented. These scales were previously outlined in Section 2.5 in Table 2.3. Results of each scale include descriptive statistics, which were provided for i. primary school preservice mathematics teachers and secondary school preservice mathematics teachers and ii. preservice mathematics teachers who took the SCED354: Teaching Probability and Statistics course and who did not take the course.

Initially, to assess the existence of a statistically significant difference between the primary school preservice mathematics teachers (N=20) and secondary school preservice mathematics teachers (N=9), the skewness of the distributions was evaluated for each correct reasoning scale. Based on the skewness values, the assumption of normal distribution was not met for some scale scores of the groups, particularly for good samples that have to represent a high percentage of the population. Consequently, the results of the groups were compared using with the independent sample t-test for those with normally distributed scores or with the Mann-Whitney U test for those did not meet this criterion. Following this, to determine if there is a difference in misconception scores between preservice mathematics teachers who have taken the SCED354: Teaching Probability and Statistics course and those who have not, the same procedures were applied.

Misconceptions Involving Averages: This scale includes responses to questions on alternatives a and c of 1st, alternatives a and b of 4th, alternatives b and f of 15th, and alternatives a and e of 17th. Results of this scale were calculated over 2. Based on data analysis, results changed between 0 and .75 and there were some participants who had no misconceptions involving averages.

Table 5.33. Descriptive statistics for misconceptions involving averages based on departments.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
Primary School Level	20	.425	.500	.200	.000	.75	.250	-.055
Secondary School Level	9	.444	.250	.243	.250	.750	.500	.549

In the table, descriptive statistics for primary school preservice mathematics teachers (N=20) were found $\bar{x}$ = .425 and .200 standard deviation. For secondary school mathematics teachers (N=9), $\bar{x}$ = .444 and .243 standard deviation were found. According to the skewness values, distributions of the scale scores for both primary school preservice mathematics teachers and secondary school preservice mathematics teachers were assumed to be relatively normal.

Table 5.34. Independent samples t-test for misconceptions involving averages based on departments.

Groups	***M***	***SD***	***t***	***df***	***p***
Primary School Level	.912	.243	.227	27	.822
Secondary School Level	.425	.200			

Table 5.34. shows that primary school preservice mathematics teachers were not statistically significant (p = .822) from secondary school preservice mathematics teachers for misconceptions involving averages.

Table 5.35. Descriptive statistics for misconceptions involving averages based on the taking SCED354.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
SCED354 taken	19	.434	.500	.201	.000	.75	.250	-.170
SCED354 not taken	10	.425	.250	.237	.250	.750	.500	.742

As in the table, descriptive statistics for preservice mathematics teachers who took SCED354 (N=19) were found $\bar{x}$ = .434 and .201 standard deviation. For preservice mathematics teachers who did not take SCED354 (N=10), $\bar{x}$ = .425 and .237 standard deviation were found. According to the skewness values, the distribution of the scale scores for both groups was assumed to be normal.

Table 5.36. Independent samples t-test for misconceptions involving averages based on the taking SCED354.

Groups	*M*	*SD*	*t*	*df*	*p*
SCED354 taken	.434	.201	.110	27	.252
SCED354 not taken	.425	.237			

Table 5.36. shows that preservice mathematics teachers who took the SCED354 course were not statistically significant (p = .252) from preservice mathematics teachers who did not take the course on this scale.

Outcome Orientation Misconception: The scale includes responses to alternative e of 2nd, alternatives a and b of 3rd, alternatives a, b, and d of 11th, alternative c of 12th, and alternative b of 13th questions. Results of this scale were calculated over 1.60. Based on data analysis, scores changed between 0 and .40, and there were some participants who did not have an outcome orientation misconception.

Table 5.37. Descriptive statistics for outcome orientation misconception based on departments.

Groups	*N*	Mean	Median	Std. Deviation	Min	Max	IQR	Skewness
Primary School Level	20	.180	.200	.144	.000	.400	.200	.152
Secondary School Level	9	.244	.200	.167	.000	.400	.300	-.501

In the table, descriptive statistics for primary school preservice mathematics teachers (N=20) were found $\bar{x}$ = .180 and .144 standard deviation. For secondary school mathematics teachers (N=9), $\bar{x}$ = .244 and .167 standard deviation were found. According to the skewness values, distributions of the scale scores for both primary school preservice mathematics teachers and secondary school preservice mathematics teachers were assumed to be relatively normal.

Table 5.38. Independent samples t-test for outcome orientation misconception based on departments.

Groups	***M***	***SD***	***t***	***df***	***p***
Primary School Level	.180	.144	1.06	27	.297
Secondary School Level	.244	.167			

Table 5.38. shows that primary school preservice mathematics teachers were not statistically different (p = .297) from secondary school preservice mathematics teachers for outcome orientation misconception.

Table 5.39. Descriptive statistics for outcome orientation misconception based on the taking SCED354.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
SCED354 taken	19	.211	.200	.156	.000	.400	.400	-.096
SCED354 not taken	10	.180	.200	.148	.000	.400	.250	.166

In the table, descriptive statistics for preservice mathematics teachers who took the SCED354 (N=19) were found $\bar{x}$ = .211 and .156 standard deviation. For preservice mathematics teachers who did not take the SCED354 (N=10), $\bar{x}$ = .180 and .148 standard deviation were found. According to the skewness values, the distribution of the scale scores for both groups was assumed to be normal.

Table 5.40. Independent samples t-test for outcome orientation misconception based on the taking SCED354.

Groups	***M***	***SD***	***t***	***df***	***p***
SCED354 taken	.211	.156	.510	27	.756
SCED354 not taken	.180	.148			

Table 5.40. shows that preservice mathematics teachers who took the SCED354 course were not significantly different (p = .756) from preservice mathematics teachers who did not take the course on this scale.

Good Samples Have to Represent a High Percentage of the Population: The scale includes responses to alternatives b and c of the 7^{th} and alternatives a and d of the 16^{th} questions. Results of this scale were calculated over 2. Based on data analysis, scores changed between 0 and 1. So, there were participants who did not have misconceptions about this scale.

Table 5.41. Descriptive statistics for good samples have to represent a high percentage of the population based on departments.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
Primary School Level	20	.288	.250	.272	.000	1.00	.500	1.03
Secondary School Level	9	.361	.250	.309	.000	1.00	.380	.929

In the table, descriptive statistics for primary school preservice mathematics teachers (N=20) were found $\bar{x}$ = .288 and .272 standard deviation. For secondary school mathematics teachers (N=9), $\bar{x}$ = .361 and .309 standard deviation were found. According to the skewness values, the distribution of the scale scores for secondary school preservice mathematics teachers assumed normal, but the normality test for the primary school preservice mathematics teachers failed.

Table 5.42. Mann-Whitney U Test for good samples have to represent a high percentage of the population based on departments.

Groups	***N***	**Mean Rank**	**Sum of Ranks**	***U***	***Z***	***p***
Primary School Level	20	14.3	287	76.5	-.667	.532

Table 5.42. Mann-Whitney U Test for good samples have to represent a high percentage of the population based on departments. (cont.)

Groups	***N***	**Mean Rank**	**Sum of Ranks**	***U***	***Z***	***p***
Secondary School Level	9	16.5	149			

The scores of the Mann-Whitney U test are shown in Table 5.42. Primary school and secondary school preservice mathematics teachers did not differ on this scale. A statistically significant difference between the groups was not found ($U = 76.5$; $p > 0.05$). According to the mean ranks, the scores of primary school preservice mathematics teachers are lower than secondary school preservice mathematics teachers. However, a statistically significant difference requires more than just this difference.

Table 5.43. Descriptive statistics for good samples have to represent a high percentage of the population based on the taking SCED354.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
SCED354 taken	19	.302	.250	.329	.000	1.00	.500	1.04
SCED354 not taken	10	.325	.250	.169	.000	.500	.250	-.434

Descriptive statistics for preservice mathematics teachers who took the SCED354 (N=19) were found $\bar{x}$ = .302 and .329 standard deviation as in the Table 5.43. For preservice mathematics teachers who did not take the SCED354 (N=10), $\bar{x}$ = .325 and .169 standard deviation were found. According to the skewness values, the distribution of the scale scores for preservice mathematics teachers who did not take the course was relatively normal.

Table 5.44. Mann-Whitney U Test for good samples have to represent a high percentage of the population based on the taking SCED354.

Groups	***N***	**Mean Rank**	**Sum of Ranks**	***U***	***Z***	***p***
SCED354 taken	19	14.1	269	78.5	-.793	.428
SCED354 not taken	10	16.7	167			

The scores of the Mann-Whitney U test are shown in Table 5.44. Preservice mathematics teachers did not differ in considering that good samples have to represent a high percentage of the population. A statistically significant difference between the groups was not found ($U = 78.5$; $p > 0.05$). If the mean ranks are regarded, the scores of the participants who took the SCED354 course are higher than those participants who did not take the course, but this difference is not sufficient for a statistically significant difference.

Law of Small Numbers: This scale includes answers to alternative a of the 12th and alternative c of the 14th questions. Scores of this scale were calculated over 1. According to data analysis, scores changed between 0 and 1. So, some participants strongly had misconceptions and did not have them at all.

Table 5.45. Descriptive statistics for law of small numbers based on departments.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
Primary School Level	20	.400	.500	.308	.000	1.00	.500	.120
Secondary School Level	9	.278	.500	.220	.000	.500	.500	-.271

In the table, descriptive statistics for primary school preservice mathematics teachers (N=20) were found $\bar{x}$ = .400 and .308 standard deviation; for secondary school mathematics teachers (N=9), $\bar{x}$ = .278 and .264 standard deviation were found. According to the skewness values, distributions of the scale scores for both primary school preservice mathematics teachers and secondary school preservice mathematics teachers were assumed to be relatively normal.

Table 5.46. Independent samples t-test for law of small numbers based on departments.

Groups	***M***	***SD***	***t***	***df***	***p***
Primary School Level	.400	.308	-1.03	27	.312
Secondary School Level	.278	.264			

Table 5.46. shows that primary school preservice mathematics teachers were not statistically significant (p = .312) from secondary school preservice mathematics teachers for law of small numbers.

Table 5.47. Descriptive statistics for law of small numbers based on the taking SCED354.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
SCED354 taken	19	.395	.500	.315	.000	1.00	.500	.173
SCED354 not taken	10	.300	.500	.258	.000	.500	.500	-.484

Table 5.47 shows descriptive statistics for preservice mathematics teachers who took the SCED354 (N=19) were found $\bar{x}$ = .395 and .315 standard deviation. For preservice mathematics teachers who did not take the SCED354 (N=10), $\bar{x}$ = .300 and .258 standard deviation were found. According to the skewness values, the distribution of the scale scores for both groups was assumed to be normal.

Table 5.48. Independent samples t-test for law of small numbers based on the taking SCED354.

Groups	***M***	***SD***	***t***	***df***	***p***
SCED354 taken	.395	.315	.815	27	.877
SCED354 not taken	.300	.258			

Table 5.48. shows that preservice mathematics teachers who took the SCED354 course were not significantly different (p = .877) from preservice mathematics teachers who did not take the course on this scale.

Representativeness Misconception: The scale includes responses to alternatives a, b, and d of the 9th, alternative e of the 10th, and alternative c of the 11th questions. Scores were calculated over 1.67. Based on data analysis, scores changed between 0 and .2. Some participants did not have representativeness misconceptions.

Table 5.49. Descriptive statistics for representativeness misconception based on departments.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
Primary School Level	20	.100	.000	.045	.000	.200	.000	4.47
Secondary School Level	9	.000	.000	.000	.000	.000	.000	-

In the table, descriptive statistics for primary school preservice mathematics teachers (N=20) were found $\bar{x}$ = .100 and .045 standard deviation. For secondary school mathematics teachers (N=9), $\bar{x}$ = .000 and .000 standard deviation were found since students had no misconception. According to the skewness values, the distribution of the scale scores for primary school preservice mathematics teachers failed, and the normality for secondary school preservice mathematics teachers was ignored.

Table 5.50. Mann-Whitney U Test for representativeness misconception based on departments.

Groups	***N***	**Mean Rank**	**Sum of Ranks**	***U***	***Z***	***p***
Primary School Level	20	15.3	305	85.5	-.671	.835
Secondary School Level	9	14.5	131			

The scores of the Mann-Whitney U test are presented in Table 5.50. Primary and secondary school preservice mathematics teachers did not differ on this scale. A statistically significant difference between the groups was not found (U = 85.5; $p > 0.05$). Based on the mean ranks, scores of secondary school preservice mathematics teachers for representativeness misconception are lower than primary school preservice mathematics teachers. Nevertheless, this difference is insufficient to consider as statistically significant.

Table 5.51. Descriptive statistics for representativeness misconception based on the taking SCED354.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
SCED354 taken	19	.011	.000	.046	.000	.200	.000	4.36
SCED354 not taken	10	.000	.000	.000	.000	.000	.000	-

In the table, descriptive statistics preservice mathematics teachers who took the SCED354 (N=19) were found $\bar{x}$ = .011 and .046 standard deviation. For participants who did not take the course (N=10), $\bar{x}$ = .000 and .000 standard deviation were found since students did not have any misconceptions. According to the skewness values, the distribution of the scale scores for preservice mathematics teachers taking the course failed, and the normality for secondary school preservice mathematics teachers was ignored.

Table 5.52. Mann-Whitney U Test for representativeness misconception based on the taking SCED354.

Groups	***N***	**Mean Rank**	**Sum of Ranks**	***U***	***Z***	***p***
SCED354 taken	19	15.3	290	90.0	-.725	.468
SCED354 not taken	10	14.5	145			

The scores of the Mann-Whitney U test are shown in Table 5.52. Preservice mathematics teachers did not differ in their misconceptions of representativeness. A statistically significant difference between the groups was not found (U = 90.0; $p > 0.05$). According to the mean ranks, the scores of the participants who took the SCED354 course are higher than those who did not take the course, but this difference is not sufficient for a statistically significant difference.

Correlation Implies Causation: This scale includes answers to alternatives b and e of the 16th question. Scores of this scale were calculated over 2. According to the data analysis, scores changed between 0 and .2. Some participants had strong misconceptions and did not have them at all.

Table 5.53. Descriptive statistics for correlation implies causation based on departments.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
Primary School Level	20	.700	.500	.801	.000	2.00	1.00	.627
Secondary School Level	9	.444	.000	.726	.000	2.00	1.00	1.50

In the table, descriptive statistics for primary school preservice mathematics teachers (N=20) were found $\bar{x}$ = .700 and .801 standard deviation. For secondary school mathematics teachers (N=9), $\bar{x}$ = .444 and .726 standard deviation. According to the skewness values, the distribution of the scale scores for primary school preservice mathematics teachers was considered relatively normal, but normality for secondary school preservice mathematics teachers failed.

Table 5.54. Mann-Whitney U Test for correlation implies causation based on departments.

Groups	***N***	**Mean Rank**	**Sum of Ranks**	***U***	***Z***	***p***
Primary School Level	20	15.8	316	74.0	-.840	.472

Table 5.54. Mann-Whitney U Test for correlation implies causation based on departments. (cont.)

Groups	***N***	**Mean Rank**	**Sum of Ranks**	***U***	***Z***	***p***
Secondary School Level	9	13.22	119			

The scores of the Mann-Whitney U test are presented in Table 5.54. Primary and secondary school preservice mathematics teachers did not differ on this scale. A statistically significant difference between the groups was not found ($U = 74.0$; $p > 0.05$). Based on the mean ranks, scores of secondary school preservice mathematics teachers for correlation imply causation is lower than primary school preservice mathematics teachers, but a statistically significant difference requires more than just this difference.

Table 5.55. Descriptive statistics for correlation implies causation based on the taking SCED354.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
SCED354 taken	19	.684	1.00	.749	.000	2.00	1.00	.616
SCED354 not taken	10	.500	.000	.850	.000	2.00	1.25	1.36

As in the table, descriptive statistics for preservice mathematics teachers who took the SCED354 (N=19) were found $\bar{x}$ = .684 and .749 standard deviation. For preservice mathematics teachers who did not take the SCED354 (N=10), $\bar{x}$ = .500 and .850 standard deviation were found. According to the skewness values, the distribution of the scale scores for preservice mathematics teachers who have taken the course was relatively normal.

Table 5.56. Mann-Whitney U Test for correlation implies causation based on the taking SCED354.

Groups	***N***	**Mean Rank**	**Sum of Ranks**	***U***	***Z***	***p***
SCED354 taken	19	15.8	301	79.0	-.817	.484
SCED354 not taken	10	13.4	134			

The socres of the Mann-Whitney U test are stated in Table 5.56. Preservice mathematics teachers did not differ in consideration of correlation implies causation. A statistically significant difference between the groups was not found (U = 79.0; $p > 0.05$). When mean ranks are considered, the scores of the participants who took the SCED354 course are higher than those who did not. However, this difference is insufficient to be counting as statistically significant difference.

Equiprobability Bias: The scale includes responses to alternative c of the 13th, alternative a of the 18th, alternative d of the 19th, and alternative d of the 20th questions. Results of this scale were calculated over 1. Based on data analysis, scores changed between 0 and 1 and there are some individuals strongly had misconceptions and did not have them.

Table 5.57. Descriptive statistics for equiprobability bias based on departments.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
Primary School Level	20	.662	.750	.327	.000	1.00	.000	-1.30
Secondary School Level	9	.667	.750	.306	.000	1.00	.380	-1.28

In the table, descriptive statistics for primary school preservice mathematics teachers (N=20) were found $\bar{x}$ = .662 and .327 standard deviation. For secondary school mathematics teachers (N=9), $\bar{x}$ = .667 and .306 standard deviation. According to the skewness values, assumptions for normality of the scale scores for primary and secondary school preservice mathematics teachers failed.

Table 5.58. Mann-Whitney U Test for equiprobability bias based on departments.

Groups	***N***	**Mean Rank**	**Sum of Ranks**	***U***	***Z***	***p***
Primary School Level	20	15.2	305	85.4	-.234	.835
Secondary School Level	9	14.5	131			

The scores of the Mann-Whitney U test are given in Table 5.58. Primary and secondary school preservice mathematics teachers did not differ on this scale. A statistically significant difference between the groups was not found (U = 85.4; $p > 0.05$). In the consideration of mean ranks, results of secondary school preservice mathematics teachers for equiprobability bias are lower than primary school preservice mathematics teachers. On the other hand, a statistically significant difference requires more than just this difference.

Table 5.59. Descriptive statistics for equiprobability bias based on the taking SCED354.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
SCED354 taken	19	.658	.750	.314	.000	1.00	.000	-1.46
SCED354 not taken	10	.675	.750	.334	.000	1.00	.560	-1.06

As in the table, descriptive statistics for preservice mathematics teachers who took the SCED354 (N=19) were found $\bar{x}$ = .658 and .314 standard deviation. For preservice mathematics teachers who did not take the SCED354 (N=10), $\bar{x}$ = .675 and .334 standard deviation were found. According to the skewness values, distributions of the scale scores for groups failed.

Table 5.60. Mann-Whitney U Test for equiprobability bias based on the taking SCED354.

Groups	***N***	**Mean Rank**	**Sum of Ranks**	***U***	***Z***	***p***
SCED354 taken	19	14.7	278	89.5	-.278	.804
SCED354 not taken	10	15.6	156			

The scores of the Mann-Whitney U test are shown in Table 5.60. Preservice mathematics teachers did not differ in equiprobability bias. A statistically significant difference between the groups was not found (U = 89.5; $p > 0.05$). When mean ranks are considered, the scores of the participants who did not take the SCED354 course are higher than participants taking the course. On the other hand, this difference is insufficient to be counted as statistically significant difference.

Groups Can Only Be Compared If They Are the Same Size: This scale includes answers to alternative a of the 16th question. Scores of this scale were calculated over 1. According to the data analysis, scores changed between 0 and .1. Some participants had this misconception, and others did not.

Table 5.61. Descriptive statistics for groups can only be compared if they are the same size based on departments.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
Primary School Level	20	.400	.000	.503	.000	1.00	1.00	.442
Secondary School Level	9	.222	.000	.441	.000	1.00	.500	1.62

In the table, descriptive statistics for primary school preservice mathematics teachers (N=20) were found $\bar{x}$ = .400 and .503 standard deviation. For secondary school mathematics teachers (N=9), $\bar{x}$ = .222 and .441 standard deviation. According to the skewness values, the distribution of the scale scores for primary school preservice mathematics teachers was considered relatively normal, but assumptions of normality for secondary school preservice mathematics teachers were failed.

Table 5.62. Mann-Whitney U Test for groups can be only compared if they are the same size based on departments.

Groups	***N***	**Mean Rank**	**Sum of Ranks**	***U***	***Z***	***p***
Primary School Level	20	15.8	316	74.0	-.916	.472
Secondary School Level	9	13.2	119			

The scores of the Mann-Whitney U test are stated in Table 5.62. Primary and secondary school preservice mathematics teachers did not differ on this scale. A statistically significant difference between the groups was not found (U = 74.0; $p > 0.05$). According to the mean ranks, the scores of secondary school preservice mathematics teachers are lower than priamary school preservice mathematics teachers, but a statistically significant difference requires more than just this difference.

Table 5.63. Descriptive statistics for groups can only be compared if they are the same size based on the taking SCED354.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
SCED354 taken	19	.421	.000	.507	.000	1.00	1.00	.348
SCED354 not taken	10	.200	.000	.422	.000	1.00	.250	1.78

In the table, descriptive statistics for preservice mathematics teachers who took the SCED354 (N=19) were found $\bar{x}$ = .421 and .507 standard deviation. For preservice mathematics teachers who did not take the SCED354 (N=10), $\bar{x}$ = .200 and .422 standard deviation were found. According to the skewness values, the distribution of the scale scores for preservice mathematics teachers who took the course was relatively normal.

Table 5.64. Mann-Whitney U Test for groups can be only compared if they are the same size based on the taking SCED354.

Groups	***N***	**Mean Rank**	**Sum of Ranks**	***U***	***Z***	***p***
SCED354 taken	19	16.1	306	74.0	-1.17	.242
SCED354 not taken	10	12.9	129			

The scores of the Mann-Whitney U test are shown in Table 5.64. Preservice mathematics teachers did not differ in this scale. A statistically significant difference between the groups was not found ($U = 74.0$; $p > 0.05$). If the mean ranks are regarded, the scores of the participants who took the SCED354 course are higher than those who did not take the course, but this a statistically significant difference requires more than just this difference.

5.1.3. Total Scores

In this part, descriptive statistics of total scores of correct reasoning skills and misconceptions are given for both categorical variables separately as department and SCED354 course. Then, the groups' results are compared, and statistical analysis of both groups is given.

Firstly, to examine the difference between the total score of primary school preservice mathematics teachers (N=20) and secondary school preservice mathematics teachers (N=9), skewness of the distributions was analyzed for correct reasoning skills score and misconceptions scores. According to the skewness values, the distributions for both groups in the total scores were assumed to be normal. Therefore, the results of the groups were compared with an independent sample t-test for the groups. Then, to examine the difference between total scores of correct reasoning skills and misconceptions of preservice mathematics teachers who have taken SCED354: Teaching Probability and Statistics course and who have not, the same procedures were repeated.

<u>5.1.3.1. Correct Reasoning Skills Score.</u> The scores for each scale item assessing correct reasoning skills are summed to determine the overall correct reasoning skills score. This score is computed on a scale of 10. If an individual possesses all the correct reasoning skills, then s/he could achieve a maximum of 10 points in total for their correct reasoning skills score. Findings indicate that scores ranged from 3.03 to 7.17. Therefore, there was no participant who answered every question either completely correctly or not.

Table 5.65. Descriptive statistics for correct reasoning skills score based on departments.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
Primary School Level	20	4.95	4.87	1.06	3.03	7.17	1.37	.069
Secondary School Level	9	5.22	5.34	.954	3.73	6.87	1.46	.045

As in the Table 5.65., descriptive statistics for primary school preservice mathematics teachers (N=20) were found $\bar{x}$ = 4.95 and 1.06 standard deviation; for secondary school mathematics teachers (N=9), $\bar{x}$ = 5.22 and .948 standard deviation were found. According to the skewness values, the distributions of the scale scores for primary and secondary school preservice mathematics teachers were assumed to be relatively normal.

Table 5.66. Independent samples t-test for correct reasoning skills score based on departments.

Groups	***M***	***SD***	***t***	***df***	***p***
Primary School Level	5.09	1.17	.535	27	.597
Secondary School Level	5.32	.874			

Table 5.66. shows that primary school preservice mathematics teachers were not statistically different (p = .597) from secondary school preservice mathematics teachers on this scale.

Table 5.67. Descriptive statistics for correct reasoning skills score based on the taking SCED354.

Groups	***N***	**Mean**	**Median**	**Std. Deviation**	**Min**	**Max**	**IQR**	**Skewness**
SCED354 taken	19	4.94	4.87	1.06	3.03	7.17	1.54	.042
SCED354 not taken	10	5.21	5.34	.943	3.73	6.87	1.47	.168

In the table, descriptive statistics for preservice mathematics teachers who took the SCED354 (N=19) were found $\bar{x}$ = 4.94 and 1.06 standard deviation. For preservice mathematics teachers who did not take the SCED354 (N=10), $\bar{x}$ = 5.21 and .943 standard deviation were found. According to the skewness values, the distribution of the scale scores for both groups was assumed as normal.

Table 5.68. Independent samples t-test for correct reasoning skills score based on the taking SCED354.

Groups	***M***	***SD***	***t***	***df***	***p***
SCED354 taken	4.94	1.06	-.663	27	.658
SCED354 not taken	5.21	.943			

Table 5.68. shows that preservice mathematics teachers who took the SCED354 course were not statistically different (p = .658) from preservice mathematics teachers who did not take the course on this scale.

<u>5.1.3.2. Misconceptions Score.</u> Score are calculated by summing the individual scale item scores that contribute to misconceptions. The total of this summation is computed to be 12.27. If and individual possesses all these misconceptions, then s/he could achieve a maximum of 12.27 points in their overall misconception score. The findings indicate that the scores varied between 1.25 and 4.75. Therefore, no one possessed all the misconceptions while also being free of any.

Table 5.69. Descriptive statistics for misconceptions score based on departments.

Groups	*N*	Mean	Median	Std. Deviation	Min	Max	IQR	Skewness
Primary School Level	20	3.07	2.70	.975	1.65	4.75	1.78	.328
Secondary School Level	9	2.62	2.50	.817	1.25	3.65	1.40	-.352

Descriptive statistics for primary school preservice mathematics teachers (N=20) were found $\bar{x}$ = 3.07 and .975 standard deviation; for secondary school mathematics teachers (N=9), $\bar{x}$ = 2.62 and .817 standard deviation were found as in the Table 5.69. According to the skewness values, distributions of the scale scores for primary and secondary school preservice mathematics teachers were assumed to be relatively normal.

Table 5.70. Independent samples t-test for misconceptions score based on departments.

Groups	*M*	*SD*	*t*	*df*	*p*
Primary School Level	3.07	.975	-1.20	27	.241
Secondary School Level	2.61	.817			

Table 5.70. shows that primary school preservice mathematics teachers were not statistically significant (p = .241) from secondary school preservice mathematics teachers on this scale.

Table 5.71. Descriptive statistics for misconceptions score based on the taking SCED354.

Groups	*N*	Mean	Median	Std. Deviation	Min	Max	IQR	Skewness
SCED354 taken	19	3.16	3.40	1.01	1.65	4.75	1.85	.075
SCED354 not taken	10	2.57	2.48	.688	1.25	3.65	.610	-.174

In the table, descriptive statistics for preservice mathematics teachers who took the SCED354 (N=19) were found $\bar{x}$ = 3.16 and 1.01 standard deviation. For preservice mathematics teachers who did not take the SCED354 (N=10), $\bar{x}$ = 2.57and .688 standard deviation were found. According to the skewness values, the distribution of the scale scores for both groups was assumed to be normal.

Table 5.72. Independent samples t-test for misconceptions score based on the taking SCED354.

Groups	***M***	***SD***	***t***	***df***	***p***
SCED354 taken	3.16	1.01	1.54	27	.017
SCED354 not taken	2.57	.688			

Table 5.72. shows that preservice mathematics teachers who took the SCED354 course were statistically significant (*p* =.017) than preservice mathematics teachers who did not take the course on this scale.

5.2. Preservice Mathematics Teachers Reflections

This section includes two parts. In the first part, explanation about the reflections of preservice mathematics teachers studying at Bogazici University regarding their statistical knowledge to teach statistics and their experiences in learning statistics are stated.

5.2.1. Statistical Knowledge to Teach Statistics

Primary and secondary school preservice mathematics teachers' statistical knowledge to teach statistics was examined based on their explanations. In order to examine their understanding, specifically conceptual understanding and their pedagogical approach, the 1st, 15th, and 17th questions of SRA (Garfield, 2003) asked them such that if your student selects the choice a, b, c, and d, how do you interpret his/ her response and what is your position to explain the question. With these what-if cases, participants 'responses were analyzed. Then, some of the responses to the remaining nine questions were also considered to examine the pedagogical approach of preservice mathematics teachers.

<u>5.2.1.1. Statistical Concepts.</u> In this part, preservice mathematics teachers' knowledge about statistical concepts such as measures of central tendency, measures of dispersion, difference between mean and mode, finding the average, data representation, normal distribution and skewness, outliers, theoretical probability and experimental probability, omnipresence of variability, and probability and statistics were examined with the codes created based on the participants' explanation. Categories and codes were described according to how many times they were repeated in Table 5.73.

Table 5.73. Categories and codes to explain preservice mathematics teachers' knowledge of statistical concepts.

Categories	**Codes**
Measures of Central Tendency	Considering the mean as accurate results (15)
	Considering the decimals as more accurate results (7)
Measures of Dispersion	Discussion and consideration of standard deviation (18)
Difference Between Mean and Mode	Distinction between mode, median, and average (26)
Finding the Average	Adding and division algorithm (29)
	Meaning of average (14)
Data Representation	Considering the given graphs (10)
	Discussing the given information (21)
Normal Distribution and Skewness	Discussion of distributions as normal and not normal (3)
Outliers	Considering the outliers in data analysis (20)
Theoretical Probability and Experimental Probability	Theoretical probability vs experimental probability (6)
Omnipresence of Variability	Discussion of other possible cases (8)
	Uncertainty (1)
Probability and Statistics	Nature of mathematics vs statistics (3)
	Discussion of the philosophy of probability (2)
	Probability (9)
	Procedural knowledge in probability (2)
	Probability and statistics connection (3)
	Emphasis on making predictions and interpretation (10)

According to the number of repetitions of the codes, it was determined that preservice mathematics teachers mostly mentioned adding and division algorithms; in other words, they mostly talked about their procedural knowledge to find the mean and average. For instance, the following quotas give insight into how preservice mathematics teachers explain the average regarding their procedural knowledge.

> "Students who choose the choice of d in the first question understand the mean since s/ he uses adding numbers and dividing with the number of elements. If the n numbers are given to us, the sum of these numbers and divided by n is the mean" (Secondary school preservice mathematics teacher, correct reasoning skills score is 6.87, misconception score is 2.0, no experience in SCED354 course)
>
> "The average of the number of children is 2.2; when you multiply it by 50, that is the number of houses in the village, then you find how many children there are." (Primary school preservice mathematics teacher, correct reasoning skills score is 5.34, misconception score is 3.4, experience in SCED354 course)

Based on the findings, only one of the participants emphasized the concept of uncertainty. Other participants mentioned the possibility of cases when they discussed the alternatives to the questions, but only one participant directly used the uncertainty of the cases.

> "I tried to say there are inferences, interpretations. In other words, some situations might vary from person to person, even from experiment to experiment. Therefore, there is no certainty in here" (Secondary school preservice mathematics teacher, correct reasoning skills score is 7.54, misconception score is 2.0, experience in SCED354 course)

<u>5.2.1.2. Pedagogical Approach.</u> In this part, preservice mathematics teachers' pedagogical approaches in instruction design, objectives, real-life examples, the role of students, technology integration, and additional sources were examined with the codes created based on the participants' explanations. Categories and codes were described according to how many times they were repeated in Table 5.74.

Table 5.74. Categories and codes to explain preservice mathematics teachers' pedagogical approaches.

Categories	**Codes**
Instruction Design	Lesson plan (7,3)
Objectives	Primary school level objectives (1)
	Secondary school level objectives (1)
Real-Life Examples	Giving real-life examples (4, 10, 3)
	Conducting research (3)
The Role of Students	Students Engagement (8)
Technology Integration	Use of technology (4)
Additional Sources	Need a book (2)

According to the number of repetitions of the codes, it was determined that preservice mathematics teachers mostly mentioned real-life examples for probability and statistics; in other words, they mostly talked about teaching probability and statistics by using real-life examples and from real data. For example, the following quota helps to understand which examples the preservice mathematics teachers use when teaching probability and statistics.

> "I would use a weather report as an introductory example. For instance, I would ask students, "What is the weather forecast for tomorrow?". Then, I would encourage them to research by looking at the previous year's data. I would encourage them to make predictions about tomorrow. I would ask what the weather forecast means. I would ask them to discuss it." (Secondary school preservice mathematics teacher, correct reasoning skills score is 7.54, misconception score is 2.0, experience in SCED354 course)

Based on the findings, preservice mathematics talked about the learning outcomes of the curriculum the least. They did not comment on or explain their field's objectives and curriculum content. In other words, thier knowledge about the curriculum and its content could not be recognized, as stated by one participant, as they do not have enough information on the curriculum.

5.2.2. Experiences in Learning Statistics

Primary and secondary school preservice mathematics teachers' experiences were examined based on their explanations, and they were categorized as learners and preservice teachers. Nine open-ended questions were asked to them, such as "Please describe the things you have learned throughout your experience as a learner of how to teach probability and statistics courses." and "Think about your experiences…Could you please list three positive/ negative areas that you observed?". With this part, participants' reflections on their learning and teaching approach were tried to be explained.

<u>5.2.2.1. As A Learner.</u> In this part, preservice mathematics teachers' experience as learners was discussed. Whether they have already taken probability and/ or statistics courses, whether they have already taken teaching probability and statistics courses, and their strengths and weaknesses in probability and/ or statistics were asked to participants. Responses were coded regarding the learners of the SCED354 course and those of other course(s).

Some participants explained that their understanding of probability and statistics improved after the SCED354: Teaching Probability and Statistics course. Following quotas helps us understand the change in their understanding.

> "In my eyes, mathematics was a very stressful thing. I mean, it was very clear before this course. Well, the square of i equals -1; it could only exist in an imaginary universe and not in real numbers. This is true and exact in mathematics. In statistics, there can be at least one black-swan. Statistics allowed me to look at a broader perspective. Actually, I cannot say anything for sure anymore." (Secondary school preservice mathematics teacher, correct reasoning skills score is 7.54, misconception score is 2.0, experience in SCED354 course)

> "I observed that my language was changed after this course. Before the course, everything was clear and precise for me. However, after this course, I realized there is a possibility for a thing and that there are gray areas in life, like mathematics."

Primary school preservice mathematics teacher, correct reasoning skills score is 6.07, misconception score is 1.65, experience in SCED354 course)

5.2.2.2. As A Preservice Teacher. This section discussed preservice mathematics teachers' experience as preservice teachers. Participants could develop minimal experience in such experience until they entered the classroom (Stuart & Thurlow, 2000). Therefore, the primary focus was their willingness to teach probability and statistics.

All participants confirmed that they need to improve how they teach statistics. For instance, Participant D mentioned that I can explain the concept of mean rather than its procedural part, but I might have difficulty explaining it to my students. In other words, she emphasized that she was not confident about teaching statistics and had doubts about teaching. In addition to that, participants preferred teaching functions, algebra, and geometry to teaching probability and statistics. Participant A directly said, "I would like to teach algebra or geometry instead of teaching probability and statistics." Also, Participant B directly said, "I can feel myself in danger since I have to teach probability and statistics. I feel safest when I teach algebra."

6. CONCLUSION AND DISCUSSION

There are four sections in this chapter. Firstly, the findings are examined to respond the research questions. Secondly, the study's limitations are outlined. Following that, the implications for teacher education programs and suggestions for further research are addressed.

The first and second questions are as follows: i. What are the total correct reasoning skills scores and total misconceptions scores of preservice mathematics teachers studying at a high-ranked public university in Turkey, depending on whether they have taken a formal course on teaching probability and statistics or not? ii. Is there a statistically significant difference in correct reasoning skills scores and misconceptions scores between primary and secondary school preservice mathematics teachers studying at Bogazici University? and ii. a. Is there a statistically significant difference in correct reasoning skills scores and misconceptions scores between primary and secondary school preservice mathematics teachers at Bogazici University who have taken the SCED 354: Teaching Probability and Statistics course and those who have not? These questions are analyzed based on the participants' scores using the SRA (Garfield, 2003).

The conclusions related to the third research question, "What are the reflections of primary and secondary school preservice mathematics teachers at Bogazici University regarding their statistical knowledge to teach statistics and their experiences in learning statistics?" are discussed based on the participants' explanations.

6.1. Statistical Reasoning Skills

This section discusses the results of the correct reasoning skills scores and misconceptions scores, which are based on the Statistical Reasoning Assessment (Garfield, 2003). Scales for correct reasoning skills and misconceptions were given in Table 2.2 and Table 2.3. (See Statistical Reasoning Assessment Section).

6.1.1. Scores of Correct Reasoning Skills

Correctly Interprets Probabilities: Responses to the second and third questions contributed to this scale score. The average score would have been 1 if every participant had provided correct answers to these questions. The findings revealed that preservice mathematics teachers from primary ($\bar{x}$ = .800) and secondary ($\bar{x}$ = .611) schools achieved relatively high scores in accurately interpreting probabilities. Furthermore, preservice mathematics teachers who completed SCED354 ($\bar{x}$ = .790) performed better than those who did not take the course ($\bar{x}$ = .650). While both groups answered the second question correctly 100% of the time, their success in the third question was not as high. Consequently, the variation in the group means may stem from the third question, which also involves outcome orientation. This aspect was elaborated on under the outcome orientation misconception scale. The interpretation required participants to comprehend a message that included a quantitative statement to use as a basis for forming an assumption or judgment (Gal, 2005). For example, in our scenario, participants were expected to interpret a 15% likelihood of developing a rash in the second question. They should understand the probabilistic statement of 15% as representing 15 individuals out of 100 who experienced a rash while using the medication. Based on the results, both primary and secondary school mathematics teachers successfully comprehended the statements relating probability.

Understand How to Select an Appropriate Average: The scale score was derived from the responses to the 1st, 4th, and 17th questions. The average score would have been 1 if every participant had provided correct answers to these questions. However, the actual means were $\bar{x}$ = .583 for preservice mathematics teachers in primary schools and $\bar{x}$ = .630 for those in secondary schools. Thus, the data indicates that the understanding of how to select an appropriate average among both groups of preservice mathematics teachers is relatively high. Additionally, preservice mathematics teachers who completed SCED354 scored an average of ($\bar{x}$ = .580) while those who did not take the course had a higher average $\bar{x}$ = .630. An analysis of the response frequencies revealed that the percentage of correct answers for the 4th question was low in both groups. Conversely, a majority of preservice mathematics teachers from primary (80%) and secondary (77%) schools were able to correctly respond to the 17th question. In the 1st and 4th questions, participants were tasked with calculating the appropriate average for a given data set, while the 17th question required them to interpret the data set's mean. It seems that determining how to calculate an average from a data set may pose greater challenges for both group of preservice school mathematics teachers. Similar to this, Groth and Bergner (2006) state that primary school preservice mathematics teachers faced difficulties in identifying suitable center measures for provided data. Therefore, in light of these points, it is crucial to emphasize preservice teachers' understanding of selecting the appropriate average.

Correctly Computes Probability: Responses to this scale were based on the 8^{th}, 13^{th}, 18^{th}, 19^{th}, and 20^{th} questions. The average score would have been 1 if every participant had provided correct answers to these questions. Conversely, the data indicated that the reasoning abilities of primary ($\bar{x}$ = .370) and secondary ($\bar{x}$ = .356) school preservice mathematics teachers in calculating probability were quite low. Additionally, preservice mathematics teachers who completed SCED354 ($\bar{x}$ = .347) and those who did not take the course ($\bar{x}$ = .400) achieved relatively higher scores. While the 8th question was correctly responded by 100% of primary and 89% of secondary school preservice mathematics teachers, only a few participants from both groups provided correct responses to the other questions. The high scores on the 8th question may be attributed to the nature of the question, which included drawing marbles from two boxes, a familiar probability scenario often found in textbooks. However, questions that tested combinatorial reasoning asked the participants to differentiate between the sample spaces of rolling two 5s, and one 6 and one 6. The participants' relatively low performance in combinatorial reasoning had a negative impact on their overall mean score. Although there was no statistically significant difference between the scores of primary and secondary school preservice mathematics teachers on this scale, the findings related to each question highlight the need to focus on combinatorial reasoning. In other words, the results suggest that both groups of preservice teachers struggle with combinatorial reasoning, prompting a recommendation for teacher training programs to incorporate the enhancement of this skill.

Understand Independence: The responses to the 9^{th}, 11^{th}, and options c, d, and f of the 10^{th} question formed this scale score. If every participant had responded correctly to all these questions, the average would have been 1.67. The results indicated that preservice mathematics teachers in both primary ($\bar{x}$ = 1.30) and secondary ($\bar{x}$ = 1.30) schools scored well in grasping the concept of independence. Furthermore, those preservice mathematics teachers who completed SCED354 ($\bar{x}$ = 1.33) performed better than those who did not take the course ($\bar{x}$ = 1.23). This scale evaluates the understanding that multiple outcomes from flipping a fair coin are independent of one another and do not influence subsequent outcomes. The findings revealed that preservice mathematics teachers in both primary and secondary schools have a solid comprehension of this concept. The most frequently chosen alternatives for these questions were correct, with the proportion of the most popular answers exceeding 65%, except for alternative d of the 10th question. 50% of primary and 33% of secondary school preservice mathematics teachers chose option d of the10th question. The average scores remained unaffected by the lower percentages of correct answers for alternative d in the 10th question, as the other questions were answered correctly at high rates. In conclusion, the results demonstrated no statistically significant difference in the understanding of independence between primary and secondary school mathematics teachers studying at Bogazici University.

Understand Sampling Variability: Responses to the 14th and 15th questions formed the basis for this scale score. The average score would have been 1 if every participant had provided correct answers to these questions. In contrast, their actual means were $\bar{x}$ = .225 for primary and $\bar{x}$ = .278 for secondary school preservice mathematics teachers. Additionally, preservice mathematics teachers who completed SCED354 ($\bar{x}$ = .225) had different scores compared to those who did not take the course ($\bar{x}$ = .30). The findings revealed that preservice mathematics teachers' grasp of sampling variability was significantly low, regardless of their department or the courses they had taken. The proportion of both groups selecting the correct answer was minimal. Moreover, the most frequently selected alternatives by participants indicated the presence of misconceptions. The misconception identified in the 14th question was related to the law of small numbers, as elaborated in 6.1.2. For the 15th question, the findings further suggested that participants failed to acknowledge variation and made comparisons of the groups based solely on averages. This aligned with Leavy's (2006) findings that preservice mathematics teachers tend to emphasize averages without considering variation when analyzing data. In the 15th question, the most prevalent response suggested that primary and secondary school preservice mathematics teachers may have compared the data sets based on their distribution shapes. Thus, there is an imperative to enhance preservice mathematics teachers' analytical skills in comparing data sets and comprehending sampling variability.

Distinguishes Between Correlation and Causation: Responses to the 16th question formed the basis of this scale score. The average score would have been 1 if every participant had provided correct answers to this question. However, the group means were determined to be $\bar{x}$ = .350 for primary school preservice mathematics teachers and $\bar{x}$ = .556 for secondary school preservice mathematics teachers. It was observed that 35% of primary and 56% of secondary school preservice mathematics teachers selected the correct response. These findings suggest that the ability of primary school preservice mathematics teachers to differentiate correlation and causation was lower than that of their secondary school counterparts. In relation to the question, only a small number of participants understood that a relationship between variables does not imply the cause-and-effect relation. Additionally, among both groups, the prevalent response indicated a misunderstanding that correlation signifies causation.

Correctly Interprets Two-way Tables: Responses to the fifth question formed this scale score. This question comprised two components: the first evaluated the ability to read the two-way table, while the second assessed the capability to interpret it. The findings indicated that preservice mathematics teachers in primary ($\bar{x}$ = .550) and secondary ($\bar{x}$ = .667) schools successfully interpreted two-way tables. Additionally, preservice mathematics teachers who completed the SCED354 course ($\bar{x}$ = .580) and those who did not ($\bar{x}$ = .600) achieved relatively high scores. Both groups frequently answered correctly, although a higher percentage of secondary school preservice mathematics teachers (66%) answered correctly compared to primary school preservice mathematics teachers (50%). There was no statistically significant difference in their scores on this scale.

Understand Importance of Large Samples: The scale score was derived from the responses to the 6th and 12th questions. The average score would have been 1 if every participant had provided correct answers to these questions. Findings indicated that preservice mathematics teachers in primary ($\bar{x}$ = .866) and secondary ($\bar{x}$ = .833) schools demonstrated a solid understanding of the significance of large sample sizes. The proportion of correct responses to these questions was notably high. For example, 85% of primary and 77% of secondary preservice mathematics teachers from Bogazici University accurately responded to the 6th question. This suggests that both groups are able to prioritize sample size when drawing inferences. Additionally, the percentages of preservice mathematics teachers in primary (75%) and secondary (77%) schools who based their decisions on scientific findings rather than anecdotal evidence were also relatively strong. Furthermore, results indicated that participants possess the essential concepts necessary for generalizing results to the wider population.

In summary, the reasoning skills of preservice mathematics teachers in primary and secondary education were discussed. They were able to accurately interpret probability, grasp the concept of independence, and recognize the significance of large sample sizes. However, there remains a lack of understanding when it comes to selecting appropriate averages, accurately calculating probability, understanding sampling variability, differentiating between correlation and causation, and properly interpreting two-way tables. Consequently, it is essential to prioritize the development of these reasoning skills.

6.1.2. Scores of Misconceptions

Misconceptions Involving Averages: This scale included responses to alternatives a and c from the 1st question, alternatives a and b from the 4th, alternatives b and f from the 15th, and alternatives a and e from the 17th questions. The average score would have been 1 if every participant had provided correct answers to these questions. The findings indicated that primary ($\bar{x}$ = .425) and secondary ($\bar{x}$ = .444) school preservice mathematics teachers harbor misunderstandings concerning averages. Despite the relatively low mean scores, these issues need addressing, as some preservice mathematics teachers struggle to identify outliers and mistakenly consider averages to be the most frequent number. The 1st and 4th questions were particularly significant since they evaluated the ability to determine the appropriate averages, with some participants demonstrating greater success in this area. Additionally, the results aligned with understanding the correct selection of an average, as previously discussed. Both primary and secondary school mathematics teachers require a deep understanding of how to choose appropriate averages and necessitate clarification on the topic. As noted in the work of Groth and Bergner (2006), these misconceptions may stem from a lack of conceptual understanding of averages. When enhancing statistical reasoning in primary and secondary school mathematics teachers, it is crucial to take this misconception into account.

Outcome Orientation Misconception: Answers to the 2nd, 12th, and 13th, alternatives a and b of 3rd, and alternatives a, b, and d of 11th questions constituted this scale score. The average score would have been 1.60 if every participant had provided correct answers to these questions. Results showed that primary ($\bar{x}$ = .180) and secondary ($\bar{x}$ = .244) school preservice mathematics teachers studying at Bogazici University could not have outcome orientation misconceptions. The percentages of the alternatives selected by both groups are very low. On the other hand, in contrast to other questions, alternatives of the 3rd and 13th questions, which highlighted outcome orientations, were relatively high. For example, in the 13th question, the percentages for the groups were around 45%, but it was around 1% in the 2nd question. It could be more beneficial to investigate the outcome orientation thoroughly among primary and secondary due to the variations in percentages.

Good Samples Have to Represent a High Percentage of the Population: Responses to alternatives b and c of the 7th and alternatives a and d of the 16th questions constituted scores of this scale. The average score would have been 2 if every participant had provided correct answers to these questions. Based on the results, primary ($\bar{x}$ = .288) and secondary ($\bar{x}$ = .361) school preservice mathematics teachers could not have a misconception that good samples have to represent a high percentage of the population. However, some participants had this misconception. The percentages of the alternatives chosen by both groups were different. For example, 30% of preservice mathematics teachers in primary schools and 29% in secondary schools selected option b for the 7th question. Regarding option c of that same question, the percentages were 45% and 66%, respectively. Consequently, the beliefs of primary and secondary preservice mathematics teachers about good samples needing to reflect a large portion of the population may require further investigation.

Law of Small Numbers: Answers to the 12th and 14th questions constituted the scores of this scale. The average score would have been 1 if every participant had provided correct answers to these questions. Results showed that primary ($\bar{x}$ = .400) and secondary ($\bar{x}$ = .278) school preservice mathematics teachers had a misunderstanding related with the law of small numbers. The percentages of both groups who selected the alternative a of the 12th question were consistently low. On the other hand, 30% of primary and 55% of secondary school preservice mathematics teachers studying at Bogazici University chose the alternative c of the 14th question. In the 14th question, Hospital A has a larger sample compared to Hospital

B. Based on the law of large numbers, as the sample size increases, the results tend to converge with the population statistics, while smaller samples tend to show greater deviations in results. In the scenario, the ratio of girls varied significantly in the smaller hospital, indicating that the likelihood of observing a particular event was not the same for both hospitals. Additionally, the choice of alternative c suggests that size was not a factor for both preservice teachers who selected it. Furthermore, the concept of sampling variability was addressed earlier. There was a consistent logic regarding sampling variability and a misunderstanding related to sample size.

Representativeness Misconception: Responses to alternatives a, b, and d of the 9th, alternative e of the 10th, and alternative c of the 11th questions constituted the score of this scale. If all participants chose every one of these options, the average would be 1.67. The findings indicated that preservice mathematics teachers in primary ($\bar{x}$ = .100) and secondary ($\bar{x}$ = .000) schools do not exhibit misconceptions about representativeness. Moreover, the percentages for these options selected by both groups could be disregarded, except for 5% of primary school preservice mathematics teachers who opted for alternative c in the 11th question. Individuals with this misconception failed to grasp the concept of independence. For instance, they believed that the sequence HTHTHT was more probable than THTTTH when flipping a fair coin six times, as they identified a pattern in the sequence of HTHTHT. Conversely, each outcome of heads and tails was independent, meaning the probabilities of the different outcomes remained the same. Preservice mathematics teachers in primary and secondary school level demonstrated a strong understanding of independence as it relates to this topic. Although the results of this scale reveal that neither group had a representativeness misconception, this may be due to their previous exposure to similar types of questions. Consequently, it may be beneficial to further explore the representativeness misconception among preservice mathematics teachers and how their reasoning influences their decisions.

Correlation Implies Causation: This scale score included the alternatives b and e of the 16th question. The average score would have been 2 if every participant had provided correct answers to this question. Based on the results, primary ($\bar{x}$ = .700) and secondary ($\bar{x}$ = .444) school preservice mathematics teachers could not have misunderstanding about correlation and causation. Conversely, both alternatives were the most frequently selected response among preservice mathematics teachers for primary and secondary schools. When analyzing the frequencies, it was noted that the percentage of preservice mathematics teachers from primary (40%) and secondary (29%) schools who opted for either of these two choices was significant. Furthermore, the findings from the correlation and causation scale reinforced the notion that neither group demonstrated sufficient ability in differentiating between correlation and causation. Consequently, it is essential to address the misconception held by preservice mathematics teachers that correlation indicates causation.

Equiprobability Bias: Selection of the alternative c of the 13th, alternative a of the 18th, alternative d of the 19th, and alternative d of the 20th questions constituted the scores of this scale. The average score would have been 1 if every participant had provided correct answers to these questions. Primary ($\bar{x}$ = .662) and secondary ($\bar{x}$ = .667) school preservice mathematics teachers got high scores for this misconception. In the analysis of accurately calculating probability, it was noted that preservice mathematics teachers struggled with probability problems that involved combinatorial reasoning. For example, the likelihood of obtaining two 5s and that of obtaining one 5 and one 6 were perceived as equally probable when two dice were thrown at the same time. These outcomes were viewed as random occurrences, leading to the conclusion that the probabilities of both events are the same. Conversely, the occurrence of one 5 and one 6 can be represented by the combinations of two distinct events $\{56, 65\}$. High performance on the equiprobability bias scale and low performance on accurately computed probability were seen to correlate with one another. It may be necessary for programs in primary and secondary mathematics education to place greater focus on equiprobability bias.

Groups Can Only Be Compared If They Are the Same Size: Alternative a of the 6th question constituted the scores of this scale. The average score would have been 1 if every participant had provided correct answers to this question. Results showed that primary ($\bar{x}$ = .400) and secondary ($\bar{x}$ = .222) school preservice mathematics teachers do not have such misconceptions. Garfield (2003) stated that measures of center and distribution should be considered when comparing groups so that data can be seen as accumulated. In understanding sampling variability, primary and secondary school preservice mathematics teachers' reasoning was very low. Even if they did not have such a misconception, their reasoning to compare the data sets should be investigated.

In conclusion, the misconceptions held by primary and secondary school mathematics teachers were analyzed. Aside from equiprobability bias, they did not display significant misconceptions. Nonetheless, the existing misconceptions still warrant careful attention. Furthermore, preservice mathematics teachers at Bogazici University, both at the primary and secondary levels, showed a need for additional education regarding statistical measures. They demonstrated proficiency in interpreting averages but struggled with choosing the right averages. Their capability to interpret averages may stem from their procedural knowledge of calculating the mean via addition and division methods. This aligns with existing literature indicating that they predominantly possess procedural knowledge (Groth & Bergner, 2006). Additionally, their challenges in selecting suitable averages may arise from insufficient reasoning about data, since the type of data influences the choice of averages. Even though their misconceptions were not pronounced, they still encountered difficulties in selecting appropriate averages. Regarding measures of spread, findings indicated that preservice teachers did not consider variation when comparing data distributions and relied on averages for their comparisons. Preservice mathematics teachers still need to improve their reasoning when it comes to measures of spread, even they did not hold the misunderstanding that groups can only be compared when they are of equal size. Consequently, it is essential to advance preservice mathematics teachers' reasoning about data and statistical measures prior to their instruction in statistics.

Preservice mathematics teachers' notions about sampling should also be taken into account. They recognized the importance of large samples and that a representative sample does not necessarily have to represent a high portion of the population. However, they demonstrated a limited understanding of variation in sampling. Their lack of reasoning ability about measures of spread might be the cause of this. Moreover, they did not grasp the concept of randomness, which is crucial for sampling. As a result, it may be necessary to place greater emphasis on randomness in mathematics teacher education. Furthermore, their limited reasoning regarding measures of spread impacts their understanding of samples. Therefore, these measures should also be prioritized in mathematics teacher education programs.

Results also showed a statistically significant difference between the total misconception scores of participants who took the SCED354 course and those who did not ($p = .017$). Even if the participants took the course, they still had misconceptions relating to statistical reasoning. Therefore, the course's content might be improved in order to overcome the preservice mathematics teacher's misconceptions and misunderstandings. Specifically, preservice primary school mathematics teachers had more misconceptions than secondary school preservice mathematics teachers. This might be because of the other mathematics elective courses in their programs. In other words, even though there is no obligation to take the SCED354 course in the secondary school mathematics education program, there are more compulsory mathematics elective courses such that students have to take the courses from the mathematics department.

6.2. Preservice Mathematics Teachers Reflection

This study aimed to examine the reflections of primary and secondary school preservice mathematics teachers studying at Bogazici University regarding their statistical knowledge to teach statistics and their experiences. This current situation was elucidated based on the preservice mathematics teachers' knowledge of some statistical concepts, pedagogical approaches, and their experiences.

6.2.1. Statistical Knowledge to Teach Statistics

The analysis of preservice mathematics teachers' statistical understanding focused on their understanding of various statistical concepts, including measures of central tendency, measures of dispersion, and the distinction of mode and mean. Findings showed that preservice mathematics teachers predominantly rely on procedural knowledge, particularly computation of mean. For instance, participants explained averages through straightforward algorithms that indicated their procedural knowledge rather than a conceptual understanding of the topics. While the participants demonstrated awareness of basic statistical concepts, such as standard deviation, outliers, and normal distributions, only one participant emphasized the significance of variability and uncertainty in statistical analysis. This lack of emphasis suggests a gap in their understanding of the nature of statistics, which mainly considers variability and involves interpretations rather than certainties.

When discussing the pedagogical approaches of preservice mathematics teachers, they frequently cited the importance and usage of real-life examples in teaching statistics. Their emphasis on using practical and research-based scenarios, such as weather forecasts, to engage students illustrates a positive approach toward statistics, making statistics more understandable. However, there is a noticeable lack of discussion regarding objectives and curriculum, indicating that preservice mathematics teachers feel unprepared, and this gap points to a need for more emphasis on objectives and curriculum content in the courses, as well as how to teach probability and statistics.

6.2.2. Experience in Learning Statistics

Preservice mathematics teachers stated that their understanding of probability and statistics significantly improved after the SCED354 course. They expressed a transformation in their perspectives, moving from a precise area of mathematics to a mode of unpredictable understanding that includes uncertainty and variability. They also improved their conceptual understanding of statistical concepts. These improvements are essential for developing a deep comprehension of statistical concepts and obtaining a key developmental understanding of statistics.

In terms of their experiences as preservice teachers, participants expressed a need for further development in teaching statistics. Many of them voiced concerns about their ability to teach statistical concepts confidently. Their preference to teach subjects like algebra and geometry over probability and statistics reflects their discomfort and lack of self-efficacy. This highlights the need for professional development that builds confidence and competence in teaching statistics.

6.3. Limitations and Implications for Further Research

The research aimed to examine the statistical reasoning skills of preservice primary and secondary school mathematics teachers studying at Bogazici University, as well as their current competencies in teaching statistics.

Preservice mathematics teachers have difficulties with sampling concepts. This may hinder their ability to pay more attention to the impact of sample size on results. This research highlights the need to improve preservice teachers' understanding of sampling concepts, particularly random sampling. In other words, it is recommended that primary and secondary school education programs place a great emphasis on this subject prior to beginning the classroom practices.

For preservice mathematics teachers, measures of central tendency are the most commonly discussed topic in statistical measures (Leavy, 2002; Canada, 2008). They struggle with accounting for outliers when calculating the mean and distinguishing it from the mode. Their performance regarding measures of spread could also be enhanced. The findings of this study indicate a necessity to enhance preservice teachers' conceptual understanding of statistical measures, particularly in relation to variation.

Research indicates that the knowledge held by teachers influences the knowledge acquired by students (Heaton & Mickelson; Yolcu, 2012). When preservice mathematics teachers struggle to grasp the concepts related to statistical measures and sampling, or if they carry misconceptions such as equiprobability bias, their students are likely to also misunderstand these concepts or adopt these misconceptions. It is essential to enhance the statistical reasoning abilities of preservice mathematics teachers. To achieve better reasoning skills, a course focusing on the teaching of probability and statistics is necessary. In this research, it was found that 85% of primary school mathematics teachers and 22% of secondary school mathematics teachers participated in a course concerning the instruction of probability and statistics. However, the results of this study revealed that both groups exhibited deficiencies in their reasoning skills and held misconceptions. Thus, there is a need for a deeper examination of these reasoning skills and misconceptions. The findings support the idea that probability and statistics teaching courses should be available for both types of programs, with course content tailored to the needs of each. Additionally, this study revealed that despite preservice mathematics teachers having taken a methods course, they still faced challenges in their reasoning skills and held misconceptions. Consequently, mathematics education programs might need to increase the number of courses focused on statistics or separate these from probability courses to effectively enhance statistical reasoning in preservice mathematics teachers.

The individuals involved in this research were preservice mathematics teachers from primary and secondary schools at Bogazici University. The initial intention was to include preservice mathematics teachers from 13 universities across Turkey, covering both primary and secondary mathematics education programs. However, the number of university participants who completed the assessment during the collection of data phase was inconsistent. In order to avoid the effects of this variance on the results, preservice mathematics teachers from Bogazici University were chosen as the primary focus, and the study was conducted solely with this group. Consequently, the findings of this study are not applicable to preservice mathematics teachers from other institutions. Additional research involving diverse participants, particularly from different universities in Turkey, could be undertaken to gain a better understanding of the statistical reasoning framework of Turkish preservice mathematics teachers.

REFERENCES

Alkas Ulusoy, C., and Kayhan Altay, M, 2017, "Analyzing the Statistical Reasoning Levels of Pre-Service Elementary School Teachers in the Context of a Model Eliciting Activity", *International Journal of Research in Education and Science (IJRES),* Vol. 3, No. 1, pp. 20-30.

Akoglu, K., 2017, Blending Online Coursework and Small Learning Communities to Examine Professional Growth in Teaching Statistics: A Phenomenological Case Study (Unpublished Doctoral Dissertation), North Caralina State University, North Carolina

Ball, D. L., and Bass, H., 2000, *Interweaving Content and Pedagogy in Teaching and Learning to Teach: Knowing and Using Mathematics*, CT: Ablex Publishing, Westport.

Ball, D. L., Lubienski, S. T., and Mewborn, D. S., 2001, *Research on Teaching Mathematics: The Unsolved Problems of Teachers' Mathematical Knowledge*, Fourth Edition, Macmillan, New York.

Ball, D. L., Thames, M. H., and Phelps, G., 2008, "Content Knowledge for Teaching: What Makes It Special?", *Journal of Teacher Education*, Vol. 59, No. 5, pp. 389-407.

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

Burgess, T., 2006, "A Framework for Examining Teacher Knowledge as Used in Action While Teaching Statistics", *Proceedings of Seventh International Conference on Teaching Statistics,* edited by Rossman & B. Chance, International Statistical Institute, Brazil.

Britz, G., Emerling, D., Hare, L., Hoerl, R., and Shade, J., 1997, "How to Teach Others to Apply Statistical Thinking", *Quality Progress*, Vol. June 1997 issue, pp. 67-79.

Canada, D. L., 2008, “Conceptions of Distribution Held by Middle School Students and Preservice Teachers”, *Proceedings of ICMI Study 18 and 2008 IASE Round Table of Conference,* edited by C. Batanere, G. Burrill, C. Reading, and A. Rossman.

Chance, B., 2002, “Components of Statistical Thinking and Implications for Instruction and Assessment”, *Journal of Statistics Education*, Vol. 10, No. 3, pp. 1-14.

Cobb, G. W., and Moore, D. S., 1997, “Mathematics, Statistics, and Teaching”, *The American Mathematical Monthly*, Vol. 104, No. 9, pp. 801-823.

Creswell, J. W, 2015, *A Concise Introduction to Mixed Methods Research*, Second Edition.CA: SAGE, Thousand Oaks.

Creswell, J. W., and Planco Clark, V. L., 2011, *Designing and Conducting Mixed Methods Research,* Second Edition, CA: SAGE, Thousand Oaks.

Curcio, F. R., 1987, “Comprehension of Mathematical Relationships Expressed in Graphs”, *Journal of Research in Mathematics Education*, Vol. 18, No. 5, pp. 382-393.

Delmas, R. C., 2002, “Statistical Literacy, Reasoning, And Learning: A Commentary”, *Journal of Statistics Education*, Vol. 10. No. 3, pp. 1-11.

Delmas, R. C., 2004, *A Comparison of Mathematical and Statistical Reasoning*”, Kluwer Academic Publishers, Netherlands.

Dransfield, S. B., Fisher, N., and Vogel, N. J., 1999, “Using Statistics and Statistical Thinking to Improve Organizational Performance”, *International Statistical Review*, Vol. 67. No. 2, pp. 99-150.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., and Schaeffer, R., 2007, *Guidelines for assessment and instruction in statistics education (GAISE) Report: A Pre-K-12 curriculum framework*, Alexandria, VA: American Statistical Association.

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,* Springer.

Gal, I., and Garfield, J., 1997, *Curricular Goals And Assessment Challenges In Statistics Education*, IOS Press, Amsterdam.

Galotti, K. M., 1989, "Approaches to Studying Formal and Everyday Reasoning", *Psychological Bulletin*, Vol. 105, No. 3, pp. 331-351.

Garfield, J., 1991, "Evaluating Students' Understanding of Statistics: Development of the Statistical Reasoning Assessment", *Proceedings of the Thirteenth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education,* Vol. 2, Blacksburg, VA, pp. 1-7.

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

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

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

Garfield, J., and Chance, B., 2000, "Assessment in Statistics Education: Issues and Challenges", *Mathematics Thinking and Learning*, Vol. 2, No. 1, pp. 99-125.

Garfield, J., DelMas, R., and Chance, B., 2003, “Web-Based Assessment Resource Tools for Improving Statistical Thinking”*, The American Educational Research Association,* Chicago.

Garfield, J. B., and Gal, I., 1999, *Teaching and Assessing Statistical Reasoning*, Reston: National Council Teachers of Mathematics.

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

Groth, R. E., 2007, “Toward a Conceptualization of Statistical Knowledge for Teaching”, *Journal for Research in Mathematics Education*, Vol. 38, No. 5, pp. 427-437.

Groth, R. E., 2013, “Characterizing Key Developmental Understandings and Pedagogically Powerful Ideas Within a Statistical Knowledge for Teaching Framework”, *Mathematical Thinking and Learning*. Vol. 15, No. 2, pp. 121-145.

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

Heaton, R. M., and Mickelson, W. T., 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.

Hill, H. C., Rowan, B., and Ball, D. L., 2005, “Effects of Teachers’ Mathematical Knowledge for Teaching on Student Achievement”, *American Educational Research Journal*, Vol. 42, No. 2, pp. 371-406.

Hill, H. C., Schilling, S. G., and Ball, D. L., 2004, “Developing Measures of Teachers’ Mathematical Knowledge for Teaching”, *The Elementary School Journal*, Vol. 105, No. 1, pp.11-30.

IBM Corp. Released 2023. IBM SPSS Statistics for Macintosh, Version 29.0.0.0 (241). Armonk, NY: IBM Corp

Jacobbe, T., Foti, S., Case, C., and Whitaker, D., 2014, "High School (Ages 14-18) Students' Understanding of Statistics", *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July 2014) Flagstaff, Arizona, USA,* edited by K. Makar, B. de Sousa, & R. Gould, The Netherlands: International Statistical Institute, Voorburg.

Kahneman, D., Slovic, P., and Tversky, A., 1982, *Judgement Under Uncertainty: Heuristics and Biases*, Cambridge University Press, Cambridge.

Karatoprak, R., 2014, *Assessing Preservice Mathematics Teachers' Statistical Reasoning*, M.S. Thesis, Boğaziçi University, İstanbul.

Kazak, S., 2009, *Olasılık Konusu Öğrencilere Neden Zor Gelmektedir?*Pegem A Yayıncılık, Ankara.

Kazak, S., 2008, *Öğrencilerin Olasılık Konularındaki Kavram Yanılgıları İle Öğrenme Zorlukları*, Pegem A Yayıncılık, Ankara.

Konold, C., 1989, "Informal Conceptions of Probability", *Cognition and Instruction*, Vol. 6, No. 1, pp. 59-98.

Konold, C., 1990, *Chance Plus: A Computer-Based Curriculum for Probability and Statistics*, Final Report to the National Science Foundation, Scientific Reasoning Research Institute, University of Massachusetts, Amherst.

Lavigne, N., and Lajoie, S., 2007, "Statistical Reasoning of Middle School Children Engaged in Survey Inquiry", *Contemporary Educational Psychology,* Vol. 32, No. 4, pp. 630-666.

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

Lovett, J. N., 2016, *The Preparedness of Preservice Secondary Mathematics Teachers to Teach Statistics:A Cross-Institutional Mixed Methods Study,* Ph.D. Thesis, North Carolina State University.

MacKay, R. J., and Oldford, W., 1994, *Stat 231 Course Notes Fall 1994*, Waterloo: University of Waterloo.

Mallows, C., 1998, “The Zeroth Problem”, *The American Statistician*, Vol. 52, No. 1, pp. 1-9.

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

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

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

Moore, D. S., 1988, “Should Mathematicians Teach Statistics?”, *The College Mathematics Journal*, Vol. 19, No. 1, pp. 3-7.

Moore, D. S., 1990, *Uncertainty*, National Academies Press, Washington.

Moore, S. D., 1997, “New Pedagogy and New Content: The Case of Statistics”, *International Statistical Review*, Vol. 65, No. 2, pp. 123-165.

National Council of Teachers of Mathematics, 2000, *Principles And Standards For School Mathematics*, Restom.

Oxford University Press, “Experience”. In *Oxford English Dictionary*, 2024.

Pereira-Mendoza, L., 2002, “Would You Allow Your Accountant to Perform Surgery? Implications for Education of Primary Teachers”, *Proceedings of The Sixth International Conference on Teaching of Statistics*, edited by B. Phillips, The Netherlands: International Statistical Institute, Cape Town.

Privitera, G. J., 2012, *Statistics for the Behavioral Sciences*. Sage Publications, Inc.

Pfannkuch, M., and Ben-Zvi, D., 2011, *Developing Teachers' Statistical Thinking*, The Netherlands: Springer, Voorburg.

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

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

Silverman, J., and Thompson, P. W., 2008, “Toward a Framework for the Developments of Mathematical Knowledge for Teaching”, *Journal of Mathematics Teacher Education*, Vol. 11, No. 6, pp. 499-511.

Simon, M. A., 2006, “Key Developmental Understandings in Mathematics: A Direction for Investigating and Establishing Learning Goals”, *Mathematical Thinking and Learning*, Vol. 8, No. 4, pp. 359-371.

Shaughnessy, J. M., 2007, *Research on Statistics Learning And Reasoning*, NC: IAP, Charlotte.

Shulman, L. S., 1986, “Those Who Understand: Knowledge Growth in Teaching”, *Educational researcher*, Vol. 15, No. 2, pp. 4-14.

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.

Snee, R., 1990, “Statistical Thinking and Its Contribution to Quality”, *The American Statistician,* Vol. 44, No. 2, pp. 116-121.

Stuart, C., and Thurlow, D., 2000, “Making in Their Own: Preservice Teachers’ Experiences, Beliefs, and Classroom Practices”, *Journal of Teacher Education*, Vol. 51, No. 2, pp. 113-121.

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

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

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

Wallman, K. K., 1993, “Enhancing Statistical Literacy: Enriching Our Society”, *Journal of the American Statistical Association*, Vol. 88, No. 421, pp. 1-8.

Watson, J. M., 1997, *Assessing Statistical Thinking Using the Media*, I. Gal, J. Garfield, and Y. Gal Edition, IOS Press.

Watson, J., and Callingham, R., 2003, “Statistical Literacy: A Complex Hierarchical Construct”, *Statistics Education Research Journal*, Vol. 2, No. 2, pp. 3-46.

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

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

Yükseköğretim Kurulu (YÖK)., 2018, "Programların Güncellenmesi Gerekçeleri, Getirdiği Yenilikler Ve Uygulama Esasları".

APPENDIX A: STATISTICAL REASONING ASSESSMENT – SRA

Purpose: The purpose of this survey is to indicate how you use statistical information in everyday life.

Take your time: The questions require you to read and think carefully about various situations. If you are unsure of what you are being asked to do, please sent an email to researcher.

The following pages consists of multiple-choice questions about probability and statistics. Read the question carefully before selecting an answer.

1. A small object was weighed on the same scale separately by nine students in a science class. The weights (in grams) recorded by each student are shown below.
6.2. 6.0. 6.0. 15.3 6.1. 6.3. 6.2. 6.15. 6.2
The students want to determine as accurately as they can the actual weight of this object. Of the following methods, which would you recommend they use?
a. Use the most common number, which is 6.2.
b. Use the 6.15 since it is the most accurate weighing.
c. Add up the 9 numbers and divide by 9.
d. Throw out the 15.3, add up the other 8 numbers and divide by 8.

2. The following message is printed on a bottle of prescription medication:
WARNING: For applications to skin areas there is a 15% chance of developing a rash. If a rash develops, consult your physician.
Which of the following is the best interpretation of this warning?
a. Don't use the medication on your skin, there's a good chance of developing a rash.
b. For application to the skin, apply only 15% of the recommended dose.
c. If a rash develops, it will probably involve only 15% of the skin.
d. About 15 of 100 people who use this medication develop a rash.
e. There is hardly a chance of getting a rash using this medication.

3. The Springfield Meteorological Center wanted to determine the accuracy of their weather forecasts. They searched their records for those days when the forecaster had reported a 70% chance of rain. They compared these forecasts to records of whether or not it actually rained on those particular days.

The forecast of 70% chance of rain can be considered very accurate if it rained on:

a. 95% - 100% of those days.

b. 85% - 94% of those days.

c. 75% - 84% of those days.

d. 65% - 74% of those days.

e. 55% - 64% of those days.

4. A teacher wants to change the seating arrangement in her class in the hope that it will increase the number of comments her students make. She first decides to see how many comments students make with the current seating arrangement. A record of the number of comments made by her 8 students during one class period is shown below.

Student Initials								
	A.A.	R.F.	A.G.	J.G.	C.K.	N.K.	J.L.	A.W.
Number of comments	0	5	2	22	3	2	1	2

She wants to summarize this data by computing the typical number of comments made that day. Of the following methods, which would you recommend she use?

a. Use the most common number, which is 2.

b. Add up the 8 numbers and divide by 8.

c. Throw out the 22, add up the other 7 numbers and divide by 7.

d. Throw out the 0, add up the other 7 numbers and divide by 7.

5. A new medication is being tested to determine its effectiveness in the treatment of eczema, an inflammatory condition of the skin. Thirty patients with eczema were selected to participate in the study. The patients were randomly divided into two groups. Twenty patients in an experimental group received the medication, while ten patients in a control group received no medication.

The results after two months are shown below.

	Experimental group (Medication)	Control group (No Medication)
Improved	8	2
No Improvement	12	8

Based on the data, I think the medication was:

1. somewhat effective 2. basically ineffective

If you choose option 1 select the one of the explanations below that best describes your reasoning.

a. 40% of the people (8/20) in the experimental group improved.

b. 8 people improved in the experimental group while only 2 improved in the control group.

c. In the experimental group, the number of people who improved is only 4 less than the number of who didn’t improve (12-8), while in the control group the difference is 6 (8-2).

d. 40% of the patients in the experimental group improved (8/20), while only 20% improved in the control group (2/10).

If you choose option 2 select the one of the explanations below that best describes your reasoning.

a. In the control group, 2 people improved even without the medication.

b. In the experimental group, more people didn’t get better than did (12 vs 8).

c. The difference between the numbers who improved and didn’t improve is about the same in each group (4 vs 6).

d. In the experimental group, only 40% of the patients improved.

6. Listed below are several possible reasons one might question the results of the experiment described above. Place a check by every reason you agree with.

a. It’s not legitimated to compare the two groups because there are different numbers of patients in each group.

b. The sample of 30 is too small to permit drawing conclusions.

c. 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.

d. I'm not given enough information about how doctors decided whether or not patients improved. Doctors may have been biased in their judgments.
e. I don't agree with any of these statements.

7. A marketing research company was asked to determine how much money teenagers (ages 13 - 19) spend on recorded music (cassette tapes, CDs and records). The company randomly selected 80 malls located around the country. A field researcher stood in a central location in the mall and asked passers-by who appeared to be the appropriate age to fill out a questionnaire. A total of 2,050 questionnaires were completed by teenagers. On the basis of this survey, the research company reported that the average teenager in this country spends $155 each year on recorded music.
Listed below are several statements concerning this survey. Place a check by every statement that you agree with.
a. The average is based on teenagers' estimates of what they spend and therefore could be quite different from what teenagers actually spend.
b. They should have done the survey at more than 80 malls if they wanted an average based on teenagers throughout the country.
c. The sample of 2,050 teenagers is too small to permit drawing conclusions about the entire country.
d. They should have asked teenagers coming out of music stores.
e. The average could be a poor estimate of the spending of all teenagers given that teenagers were not randomly chosen to fill out the questionnaire.
f. The average could be a poor estimate of the spending of all teenagers given that only teenagers in malls were sampled.
g. Calculating an average in this case is inappropriate since there is a lot of variation in how much teenagers spend.
h. I don't agree with any of these statements.

8. Two containers, labeled A and B, are filled with red and blue marbles in the following quantities:

Container	Red	Blue
A	6	4
B	60	40

Each container is shaken vigorously. After choosing one of the containers, you will reach in and, without looking, draw out a marble. If the marble is blue, you win $50. Which container gives you the best chance of drawing a blue marble?

a. Container A (with 6 red and 4 blue)

b. Container B (with 60 red and 40 blue)

c. Equal chances from each container.

9. Which of the following sequences is most likely to result from flipping a fair coin 5 times?

a. H H H T T

b. T H H T H

c. T H T T T

d. H T H T H

e. All four sequences are equally likely.

10. Select one or more explanations for the answer you gave for the item above.

a. Since the coin is fair, you ought to get roughly equal numbers of heads and tails.

b. Since coin flipping is random, the coin ought to alternate frequently between landing heads and tails.

c. Any of the sequences could occur.

d. If you repeatedly flipped a coin five times, each of these sequences would occur about as often as any other sequence.

e. If you get a couple of heads in a row, the probability of a tails on the next flip increases.

f. Every sequence of five flips has exactly the same probability of occurring.

11. Listed below are the same sequences of Hs and Ts that were listed in Item 8. Which of the sequences is <u>least</u> likely to result from flipping a fair coin 5 times?
a. H H H T T
b. T H H T H
c. T H T T T
d. H T H T H
e. All four sequences are equally unlike.

12. The Caldwells want to buy a new car, and they have narrowed their choices to a Buick or a Oldsmobile. They first consulted an issue of Consumer Reports, which compared rates of repairs for various cars. Records of repairs done on 400 cars of each type showed somewhat fewer mechanical problems with the Buick than with the Oldsmobile.
The Caldwells then talked to three friends, two Oldsmobile owners, and one former Buick owner. Both Oldsmobile owners reported having a few mechanical problems, but nothing major. The Buick owner, however, exploded when asked how he liked his car:
First, the fuel injection went out — $250 bucks. Next, I started having trouble with the rear end and had to replace it. I finally decided to sell it after the transmission went. I'd never buy another Buick.
The Caldwells want to buy the car that is less likely to require major repair work. Given what they currently know, which car would you recommend that they buy?
a. I would recommend that they buy the Oldsmobile, primarily because of all the trouble their friend had with his Buick. Since they haven't heard similar horror stories about the Oldsmobile, they should go with it.
b. I would recommend that they buy the Buick in spite of their friend's bad experience. That is just one case, while the information reported in Consumer Reports is based on many cases. And according to that data, the Buick is somewhat less likely to require repairs.
c. I would tell them that it didn't matter which car they bought. Even though one of the models might be more likely than the other to require repairs, they could still, just by chance, get stuck with a particular car that would need a lot of repairs. They may as well toss a coin to decide.

13. Five faces of a fair die are painted black, and one face is painted white. The die is rolled six times. Which of the following results is more likely?
a. Black side up on five of the rolls; white side up on the other roll
b. Black side up on all six rolls
c. a and b are equally likely

14. 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.

15. Forty college students participated in a study of the effect of sleep on test scores. Twenty of the students volunteered to stay up all night studying the night before the test (no-sleep group). The other 20 students (the control group) went to bed by 11:00 p.m. on the evening before the test. The test scores for each group are shown in the graphs below. Each dot on the graph represents a particular student's score. For example, the two dots above the 80 in the bottom graph indicate that two students in the sleep group scored 80 on the test.

```
.   .   .   .   .
.   :   :   :   :   :   .   .   .       .
------------------------------------------------------------
30                  40    50    60    70    80    90    100
```
Test Scores: No- Sleep Group

```
.   .   .
.   .   .   .   :   :   :   :   :   :   .
------------------------------------------------------------
30     40     50     60    70    80    90    100
```
Test Scores: Sleep Group

Examine the two graphs carefully. Then choose from the 6 possible conclusions listed below the one you most agree with.

a. The no-sleep group did better because none of these students scored below 40 and the highest score was achieved by a student in this group.
b. The no-sleep group did better because its average appears to be a little higher than the average of the sleep group.
c. There is no difference between the two groups because there is considerable overlap in the scores of the two groups.
d. There is no difference between the two groups because the difference between their averages is small compared to the amount of variation in the scores.
e. The sleep group did better because more students in this group scored 80 or above.
f. The sleep group did better because its average appears to be a little higher than the average of the no-sleep group.

16. For one month, 500 elementary students kept a daily record of the hours they spent watching television. The average number of hours per week spent watching television was 28. The researchers conducting the study also obtained report cards for each of the students. They found that the students who did well in school spent less time watching television than those students who did poorly. Listed below are several possible statements concerning the results of this research. Place a check by every statement that you agree with.
a. The sample of 500 is too small to permit drawing conclusions.
b. If a student decreased the amount of time spent watching television, his or her performance in school would improve.
c. Even though students who did well watched less television, this doesn't necessarily mean that watching television hurts school performance.
d. One month is not a long enough period of time to estimate how many hours the students really spend watching television.
e. The research demonstrates that watching television causes poorer performance in school.
f. I don't agree with any of these statements.

17. The school committee of a small town wanted to determine the average number of children per household in their town. They divided the total number of children in the town by 50, the total number of households. Which of the following statements must be true if the average children per household is 2.2?

a. Half the households in the town have more than 2 children.

b. More households in the town have 3 children than have 2 children.

c. There are a total of 110 children in the town.

d. There are 2.2 children in the town for every adult.

e. The most common number of children in a household is 2.

f. None of the above.

18. When two dice are simultaneously thrown it is possible that one of the following two results occurs: Result 1: A 5 and a 6 are obtained. Result 2: A 5 is obtained twice.

Select the response that you agree with the most:

a. The chances of obtaining each of these results is equal

b. There is more chance of obtaining result 1.

c. There is more chance of obtaining result 2.

d. It is impossible to give an answer. (Please explain why)

19. When three dice are simultaneously thrown, which of the following results is most likely to be obtained?

a. Result 1: "A 5, a 3 and a 6"

b. Result 2: "A 5 three times"

c. Result 3: "A 5 twice and a 3"

d. All three results are equally likely.

20. When three dice are simultaneously thrown, which of these three results is least likely to be obtained?

a. Result 1: "A 5, a 3 and a 6"

b. Result 2: "A 5 three times"

c. Result 3: "A 5 twice and a 3"

d. All three results are equally unlikely.

APPENDIX B: CORRECT REASONING SKILLS AND MISCONCEPTIONS

1. A small object was weighed on the same scale separately by nine students in a science class. The weights (in grams) recorded by each student are shown below.
6.2 6.0 6.0 15.3 6.1 6.3 6.2 6.15 6.2
The students want to determine as accurately as they can the actual weight of this object. Of the following methods, which would you recommend they use?
a. Use the most common number, which is 6.2.
(Misconceptions: Involving Averages-averages are the most common number)
b. Use the 6.15 since it is the most accurate weighing.
c. Add up the 9 numbers and divide by 9.
(Misconceptions: Involving Averages-fails to take outlier into consideration when computing the mean)
d. Throw out the 15.3, add up the other 8 numbers and divide by 8.
(Correct Reasoning Skills: Understand how to select an appropriate average)

2. The following message is printed on a bottle of prescription medication:
WARNING: For applications to skin areas there is a 15% chance of developing a rash. If a rash develops, consult your physician.
Which of the following is the best interpretation of this warning?
a. Don't use the medication on your skin, there's a good chance of developing a rash.
b. For application to the skin, apply only 15% of the recommended dose.
c. If a rash develops, it will probably involve only 15% of the skin.
d. About 15 of 100 people who use this medication develop a rash.
(Correct Reasoning Skills: Correctly interpret probabilities)
e. There is hardly a chance of getting a rash using this medication.
(Misconceptions: Outcome orientation)

3. The Springfield Meteorological Center wanted to determine the accuracy of their weather forecasts. They searched their records for those days when the forecaster had reported a 70% chance of rain. They compared these forecasts to records of whether or not it actually rained on those particular days.

The forecast of 70% chance of rain can be considered very accurate if it rained on:
a. 95% - 100% of those days.
(Misconceptions: Outcome orientation)
b. 85% - 94% of those days.
(Misconceptions: Outcome orientation)
c. 75% - 84% of those days.
d. 65% - 74% of those days.
(Correct Reasoning Skills: Correctly interpret probabilities)
e. 55% - 64% of those days.

4. A teacher wants to change the seating arrangement in her class in the hope that it will increase the number of comments her students make. She first decides to see how many comments students make with the current seating arrangement. A record of the number of comments made by her 8 students during one class period is shown below.

Student Initials								
	A.A.	R.F.	A.G.	J.G.	C.K.	N.K.	J.L.	A.W.
Number of comments	0	5	2	22	3	2	1	2

She wants to summarize this data by computing the typical number of comments made that day. Of the following methods, which would you recommend she use?
a. Use the most common number, which is 2.
(Misconceptions: Involving Averages-averages are the most common number)
b. Add up the 8 numbers and divide by 8.
(Misconceptions: Involving Averages-fails to take outlier into consideration when computing the mean)
c. Throw out the 22, add up the other 7 numbers and divide by 7.
(Correct Reasoning Skills: Understand how to select an appropriate average)
d. Throw out the 0, add up the other 7 numbers and divide by 7.

5. A new medication is being tested to determine its effectiveness in the treatment of eczema, an inflammatory condition of the skin. Thirty patients with eczema were selected to participate in the study. The patients were randomly divided into two

groups. Twenty patients in an experimental group received the medication, while ten patients in a control group received no medication. The results after two months are shown below.

	Experimental group (Medication)	Control group (No Medication)
Improved	8	2
No Improvement	12	8

Based on the data, I think the medication was:

1. somewhat effective 2. basically ineffective

If you choose option 1 select the one of the explanations below that best describes your reasoning.

a. 40% of the people (8/20) in the experimental group improved.

b. 8 people improved in the experimental group while only 2 improved in the control group.

c. In the experimental group, the number of people who improved is only 4 less than the number of who didn't improve (12-8), while in the control group the difference is 6 (8-2).

d. 40% of the patients in the experimental group improved (8/20), while only 20% improved in the control group (2/10).

(Correct Reasoning Skills: Correctly interpret two-way tables)

If you choose option 2 select the one of the explanations below that best describes your reasoning.

a. In the control group, 2 people improved even without the medication.

b. In the experimental group, more people didn't get better than did (12 vs 8).

c. The difference between the numbers who improved and didn't improve is about the same in each group (4 vs 6).

d. In the experimental group, only 40% of the patients improved.

6. Listed below are several possible reasons one might question the results of the experiment described above. Place a check by every reason you agree with.

a. It's not legitimated to compare the two groups because there are different numbers of patients in each group.

(Misconceptions: Groups can only be compared if they are the same size)

b. The sample of 30 is too small to permit drawing conclusions.
(Correct Reasoning Skills: Understand the importance of large samples)
c. 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.
d. I'm not given enough information about how doctors decided whether or not patients improved. Doctors may have been biased in their judgments.
e. I don't agree with any of these statements.

7. A marketing research company was asked to determine how much money teenagers (ages 13 - 19) spend on recorded music (cassette tapes, CDs and records). The company randomly selected 80 malls located around the country. A field researcher stood in a central location in the mall and asked passers-by who appeared to be the appropriate age to fill out a questionnaire. A total of 2,050 questionnaires were completed by teenagers. On the basis of this survey, the research company reported that the average teenager in this country spends $155 each year on recorded music.
Listed below are several statements concerning this survey. Place a check by every statement that you agree with.
a. The average is based on teenagers' estimates of what they spend and therefore could be quite different from what teenagers actually spend.
b. They should have done the survey at more than 80 malls if they wanted an average based on teenagers throughout the country.
(Misconceptions: Good samples have to represent a high percentage of the population)
c. The sample of 2,050 teenagers is too small to permit drawing conclusions about the entire country.
(Misconceptions: Good samples have to represent a high percentage of the population)
d. They should have asked teenagers coming out of music stores.
e. The average could be a poor estimate of the spending of all teenagers given that teenagers were not randomly chosen to fill out the questionnaire.
f. The average could be a poor estimate of the spending of all teenagers given that only teenagers in malls were sampled.

g. Calculating an average in this case is inappropriate since there is a lot of variation in how much teenagers spend.
h. I don't agree with any of these statements.

8. Two containers, labeled A and B, are filled with red and blue marbles in the following quantities:

Container	Red	Blue
A	6	4
B	60	40

Each container is shaken vigorously. After choosing one of the containers, you will reach in and, without looking, draw out a marble. If the marble is blue, you win $50. Which container gives you the best chance of drawing a blue marble?
a. Container A (with 6 red and 4 blue)
b. Container B (with 60 red and 40 blue)
c. Equal chances from each container.
(Correct Reasoning Skills: Correctly computes probability-understands probabilities as ratios)

9. Which of the following sequences is most likely to result from flipping a fair coin 5 times?
a. H H H T T
(Misconceptions: Representativeness)
b. T H H T H
(Misconceptions: Representativeness)
c. T H T T T
d. H T H T H
(Misconceptions: Representativeness)
e. All four sequences are equally likely.
(Correct Reasoning Skills: Understand independence)

10. Select one or more explanations for the answer you gave for the item above.
a. Since the coin is fair, you ought to get roughly equal numbers of heads and tails.
b. Since coin flipping is random, the coin ought to alternate frequently between landing heads and tails.
c. Any of the sequences could occur.
(Correct Reasoning Skills: Understand independence)
d. If you repeatedly flipped a coin five times, each of these sequences would occur about as often as any other sequence.
(Correct Reasoning Skills: Understand independence)
e. If you get a couple of heads in a row, the probability of a tails on the next flip increases.
f. Every sequence of five flips has exactly the same probability of occurring.
(Correct Reasoning Skills: Understand independence)

11.Listed below are the same sequences of Hs and Ts that were listed in Item 8. Which of the sequences is <u>least</u> likely to result from flipping a fair coin 5 times?
a. H H H T T
(Misconceptions: Outcome orientation)
b. T H H T H
(Misconceptions: Outcome orientation)
c. T H T T T
(Misconceptions: Representativeness)
d. H T H T H
(Misconceptions: Outcome orientation)
e. All four sequences are equally unlike.
(Correct Reasoning Skills: Understand independence)

12. The Caldwells want to buy a new car, and they have narrowed their choices to a Buick or a Oldsmobile. They first consulted an issue of Consumer Reports, which compared rates of repairs for various cars. Records of repairs done on 400 cars of each type showed somewhat fewer mechanical problems with the Buick than with the Oldsmobile.
The Caldwells then talked to three friends, two Oldsmobile owners, and one former Buick owner. Both Oldsmobile owners reported having a few mechanical problems,

but nothing major. The Buick owner, however, exploded when asked how he liked his car:

First, the fuel injection went out — $250 bucks. Next, I started having trouble with the rear end and had to replace it. I finally decided to sell it after the transmission went. I'd never buy another Buick.

The Caldwells want to buy the car that is less likely to require major repair work. Given what they currently know, which car would you recommend that they buy?

a. I would recommend that they buy the Oldsmobile, primarily because of all the trouble their friend had with his Buick. Since they haven't heard similar horror stories about the Oldsmobile, they should go with it.

(Misconceptions: Law of small numbers)

b. I would recommend that they buy the Buick in spite of their friend's bad experience. That is just one case, while the information reported in Consumer Reports is based on many cases. And according to that data, the Buick is somewhat less likely to require repairs.

(Correct Reasoning Skills: Understand the importance of large samples)

c. I would tell them that it didn't matter which car they bought. Even though one of the models might be more likely than the other to require repairs, they could still, just by chance, get stuck with a particular car that would need a lot of repairs. They may as well toss a coin to decide.

(Misconceptions: Outcome orientation)

13. Five faces of a fair die are painted black, and one face is painted white. The die is rolled six times. Which of the following results is more likely?

a. Black side up on five of the rolls; white side up on the other roll

(Correct Reasoning Skills: Correctly computes probability-uses combinatorial reasoning)

b. Black side up on all six rolls

(Misconceptions: Outcome orientation)

c. a and b are equally likely

(Misconceptions: Equiprobability bias)

14. 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)

(Correct Reasoning Skills: Understand sampling variability)

c. The two hospitals are equally likely to record such an event.

(Misconceptions: Law of small numbers)

15. Forty college students participated in a study of the effect of sleep on test scores. Twenty of the students volunteered to stay up all night studying the night before the test (no-sleep group). The other 20 students (the control group) went to bed by 11:00 p.m. on the evening before the test. The test scores for each group are shown in the graphs below. Each dot on the graph represents a particular student's score. For example, the two dots above the 80 in the bottom graph indicate that two students in the sleep group scored 80 on the test.

```
.   .   .   .   .
.   :   :   :   :   :   .   .   .      .
-----------------------------------------------------------
30                    40    50    60    70    80    90    100
```

Test Scores: No- Sleep Group

```
.   .   .
.   .   .   .   :   :   :   :   :   :   .
-----------------------------------------------------------
30    40    50    60    70    80    90    100
```

Test Scores: Sleep Group

Examine the two graphs carefully. Then choose from the 6 possible conclusions listed below the one you most agree with.

a. The no-sleep group did better because none of these students scored below 40 and the highest score was achieved by a student in this group.

b. The no-sleep group did better because its average appears to be a little higher than the average of the sleep group.

(Misconceptions: Involving Averages-compares groups based on their averages)

c. There is no difference between the two groups because there is considerable overlap in the scores of the two groups.
d. There is no difference between the two groups because the difference between their averages is small compared to the amount of variation in the scores.
(Correct Reasoning Skills: Understand sampling variability)
e. The sleep group did better because more students in this group scored 80 or above.
f. The sleep group did better because its average appears to be a little higher than the average of the no-sleep group.
(Misconceptions: Involving Averages-compares groups based on their averages)

16. For one month, 500 elementary students kept a daily record of the hours they spent watching television. The average number of hours per week spent watching television was 28. The researchers conducting the study also obtained report cards for each of the students. They found that the students who did well in school spent less time watching television than those students who did poorly. Listed below are several possible statements concerning the results of this research. Place a check by every statement that you agree with.
a. The sample of 500 is too small to permit drawing conclusions.
(Misconceptions: Good samples have to represent a high percentage of the population)
b. If a student decreased the amount of time spent watching television, his or her performance in school would improve.
(Misconceptions: Correlation implies causation)
c. Even though students who did well watched less television, this doesn't necessarily mean that watching television hurts school performance.
(Correct Reasoning Skills: Distinguish between correlation and causation)
d. One month is not a long enough period of time to estimate how many hours the students really spend watching television.
(Misconceptions: Good samples have to represent a high percentage of the population)
e. The research demonstrates that watching television causes poorer performance in school.
(Misconceptions: Correlation implies causation)
f. I don't agree with any of these statements.

17. The school committee of a small town wanted to determine the average number of children per household in their town. They divided the total number of children in the town by 50, the total number of households. Which of the following statements must be true if the average children per household is 2.2?

a. Half the households in the town have more than 2 children.

(Misconceptions: Involving Averages-confuses means with median

b. More households in the town have 3 children than have 2 children.

c. There are a total of 110 children in the town.

(Correct Reasoning Skills: Understand how to select an appropriate average)

d. There are 2.2 children in the town for every adult.

e. The most common number of children in a household is 2.

(Misconceptions: Involving Average-averages are the most common number)

f. None of the above.

18. When two dice are simultaneously thrown it is possible that one of the following two results occurs: Result 1: A 5 and a 6 are obtained. Result 2: A 5 is obtained twice.

Select the response that you agree with the most:

a. The chances of obtaining each of these results is equal.

(Misconceptions: Equiprobability bias)

b. There is more chance of obtaining result 1.

(Correct Reasoning Skills: Correctly computes probability-uses combinatorial reasoning)

c. There is more chance of obtaining result 2.

d. It is impossible to give an answer. (Please explain why)

19. When three dice are simultaneously thrown, which of the following results is most likely to be obtained?

a. Result 1: “A 5, a 3 and a 6”

(Correct Reasoning Skills: Correctly computes probability-uses combinatorial reasoning)

b. Result 2: “A 5 three times”

c. Result 3: “A 5 twice and a 3”

d. All three results are equally likely.
(Misconceptions: Equiprobability bias)

20. When three dice are simultaneously thrown, which of these three results is least likely to be obtained?
a. Result 1: “A 5, a 3 and a 6”
b. Result 2: “A 5 three times”
(Correct Reasoning Skills: Correctly computes probability-uses combinatorial reasoning)
c. Result 3: “A 5 twice and a 3”
d. All three results are equally unlikely
(Misconceptions: Equiprobability bias)

APPENDIX C: INTERVIEW PROTOCOL FOR PRESERVICE MATHEMATICS TEACHER

Estimated duration: 45 minutes

Thanks for your participation. Please note that you are free to answer or not answer any questions, and you are free to stop and cancel this interview anytime you want. I will ask 12 questions about your current situation in teaching statistics. Thanks again… Please let me know when you are ready.

Prompts / Questions:

Part A: Knowledge: Results of Statistical Reasoning Assessment showed that preservice mathematics teachers studying at Bogazici University have lacked understanding in selecting appropriate averages and understanding sampling variability relating to reasoning about statistical measures and reasoning about samples.

I am very interested to learn more about your position to improve these reasoning skills or to overcome the related misconceptions.

1. If your student selects choice a, b, c and d, how do you interpret her/ his response and what is your position to explain the question. Please tell me about your ideas for each choice.

A small object was weighed on the same scale separately by nine students in a science class. The weights (in grams) recorded by each student are shown below.

6.2 6.0 6.0 15.3 6.1 6.3 6.2 6.15 6.2

The students want to determine as accurately as they can the actual weight of this object. Of the following methods, which would you recommend they use?

a. Use the most common number, which is 6.2.

b. Use the 6.15 since it is the most accurate weighing.

c. Add up the 9 numbers and divide by 9.

d. Throw out the 15.3, add up the other 8 numbers and divide by 8.

2. If your student selects statements a, b, c, d, e, and f, how do you interpret her/ his response and what is your position to explain the question. Please tell me about your ideas for each item.

Forty college students participated in a study of the effect of sleep on test scores. Twenty of the students volunteered to stay up all night studying the night before the test (no-sleep group). The other 20 students (the control group) went to bed by 11:00 p.m. on the evening before the test. The test scores for each group are shown in the graphs below. Each dot on the graph represents a particular student's score. For example, the two dots above the 80 in the bottom graph indicate that two students in the sleep group scored 80 on the test.

```
.  .  .  .  .
.  :   :  :  :  :  .  .   .     .
--------------------------------------------------------------
30 40 50 60 70 80 90 100
```

Test Scores: No- Sleep Group

```
.   .  .
.  .  .   .  :  :  :  :  :  :  .
--------------------------------------------------------------
30     40    50    60   70   80   90   100
```

Test Scores: Sleep Group

Examine the two graphs carefully. Then choose from the 6 possible conclusions listed below the one you most agree with.

a. The no-sleep group did better because none of these students scored below 40 and the highest score was achieved by a student in this group.

b. The no-sleep group did better because its average appears to be a little higher than the average of the sleep group.

c. There is no difference between the two groups because there is considerable overlap in the scores of the two groups.

d. There is no difference between the two groups because the difference between their averages is small compared to the amount of variation in the scores.

e. The sleep group did better because more students in this group scored 80 or above.

f. The sleep group did better because its average appears to be a little higher than the average of the no-sleep group.

3. If your student selects statements a, b, c, d, e and f, how do you interpret her/ his response and what is your position to explain the question. Please tell me about your ideas for each choice.

The school committee of a small town wanted to determine the average number of children per household in their town. They divided the total number of children in the town by 50, the total number of households. Which of the following statements must be true if the average children per household is 2.2?

a. Half the households in the town have more than 2 children.

b. More households in the town have 3 children than have 2 children.

c. There are a total of 110 children in the town.

d. There are 2.2 children in the town for every adult.

e. The most common number of children in a household is 2.

f. None of the above.

Part B: Lived Experience:

1. Please tell me about your experience in teaching probability and statistics course (SCED 354: Teaching Probability and Statistics) before you participated in these professional developments?

a. Have you ever taken any probability and/ or statistics course?

b. Have you ever been a part of teaching probability and/ or statistics?

c. Have you ever been a part of preparing a lesson plan to teach probability and/ or statistics?

2. Please describe the things you have learned throughout your experience as a learner of how to teach probability and statistics courses.

3. Please describe your position in using digital tools such as TinkerPlot, CODAP, and how you interacted in those?

a. Throughout the lectures, how have your perspectives about teaching statistics changed or not?

4. Please describe the things you learned about teaching statistics throughout your experiences as a preservice mathematics teacher to teach statistics.

I am very interested to learn more about how your participation in the course may or may not have impacted your teaching practices.

5. If we walked into a class where you were teaching a statistics lesson, describe what we would see?
a. Specifically, thinking about the aim of teaching probability and statistics courses, how would it be different from before? Please consider the course elements as the definition of concepts, the use of technology, the use of statistical investigative cycle, and the use of real and messy data.

6. Think about your experiences… Can you please tell me at least three things that were positive for you in the areas you think went well throughout the course? Could you please list three positive areas you observed? Tell me your opinion about these.

7. Think about the areas you think did not go well throughout the project. Could you please list three negative areas (or areas you were concerned) you observed? Tell me what your opinion is about these.

8. The goal of this course was to positively impact preservice mathematics teachers' practices in statistics. How do you think the course would be improved to impact these practices more positively?

9. What comments or questions do you have for the researcher.