THE RELATIONSHIP BETWEEN MATHEMATICS TEACHERS' PROBABILITY APPROACHES AND MISCONCEPTIONS

by

Mine Doğucu

B.A, Mathematics, Smith College, 2009

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

Graduate Program in Secondary School Science and Mathematics Education

Boğaziçi University

2013

ACKNOWLEDGEMENTS

First of all I would like to thank my advisor Gülseren Karagöz Akar, for her endless support and patience throughout this research. Without her help and smile this work would have been impossible. With her perspective on mathematics education, learning and teaching, I have grown academically.

I would also like to express my gratitude to Assist. Prof. Emine Adadan who has accepted to be on the committee in such a short notice and has taken the time to read my thesis. She has been very supportive throughout her course and has in fact helped me in identifying my thesis topic and set me on this journey.

I would also like to thank Assist. Prof. Engin Ader, for taking the time to read my work, providing me insightful feedback and asking me questions that have helped me shape my ideas.

I would also like to thank Prof. Ali Baykal for always being there to listen both for academic and personal stories , sharing his wisdom, inspiring me in the field of measurement and evaluation. I feel fortunate to have made it just in time to Boğaziçi to be able to be his student. I can not imagine my M.S. courses without his wit.

My sincere thanks go to Assist. Prof. Buket Yakmacı Güzel who has provided me invaluable feedback during her course and during the early stages of my thesis. She has always been ready to listen to me as the head of the department and help me adjust to Boğaziçi. Her positivity has always motivated me during my years at Boğaziçi.

I would like to thank Assist Prof. Sibel Kazak for her support during every stage of the research. Without her help on the instrument and the approaches, this thesis would not have been possible.

I am grateful to Binnaz Melin, who has not only helped me since my high school years but also helped me tremendously during the data collection process and has gotten me in touch with many principals.

I would like to thank Atakan Gürkan and my friends Şirin Çalışkan and Ümit Aslan for their endless technical support and Güneş Ertaş for her invaluable feedback during every stage of the research.

I also would like to thank Cumhur Aydın for making me feel lucky and happy.

My special thanks go to two of my best friends Nihan Pınar and Emre Şahin who have always been there for me since my childhood, who laugh and cry with me and who have always been there with their emotional support during this research.

I am sincerely thankful to Asuman Okman and Carol Caswell for always believing in me.

Last but not least, I am thankful to my parents Nadide and Reşit and my siblings Yılmaz, Arzu, Fatih and Abdullah for their unconditional love and support. I love each and every one of them dearly.

ABSTRACT

THE RELATIONSHIP BETWEEN MATHEMATICS TEACHERS' PROBABILITY APPROACHES AND MISCONCEPTIONS

The literature on probability offers three distinct definitions: theoretical, experimental and subjective. Frameworks suggest that these three approaches should be adopted in probability classrooms. The aim of this research was to examine which of the three approaches that mathematics teachers adopt, whether there is a relationship between the experience level of the teachers and the approach they adopt, whether there is a relationship between the experience level and their success level in answering questions related to misconceptions and whether there is a relationship between the approach and success level. Results of the study indicate that there is not a significant relationship between experience level and approach and also the relationship between experience level and success level is not significant. However, for five items the relationship between approach and success level was found to be significant. Overall results have shown that, participants who have adopted all of the three approaches were more likely to score higher than the ones who have solely used the classical approach.

ÖZET

MATEMATİK ÖĞRETMENLERİNİN OLASILIK YAKLAŞIMLARI İLE KAVRAM YANILGILARI ARASINDAKİ İLİŞKİ

Olasılık literatürü üç farklı tanım sunmaktadır: teorik, deneysel ve öznel. Eğitim çerçeveleri bu üç yaklaşımın da olasılık derslerinde ele alınmasını önermektedir. Bu araştırmanın amacı matematik öğretmenlerinin hangi yaklaşımları kullandığını, matematik öğretmeninin deneyim seviyesi ile kullanılan yaklaşım arasında bir ilişki olup olmadığını, deneyim seviyesi ile kavram yanılgılarıyla bağdaşmış sorulardaki başarı seviyesi arasında bir ilişki olup olmadığını ve kullanılan yaklaşım ile başarı seviyesi arasında ilişki olup olmadığını gözlemlemektir. Bulgular deneyim seviyesi ile yaklaşım arasında ve deneyim seviyesi ile başarı seviyesi arasında anlamlı bir ilişki olmadığını göstermektedir. Fakat, beş madde için yaklaşım ve başarı arasındaki ilişki anlamlıdır. Genel olarak bakıldığında, üç yaklaşımı da kullanan katılımcıların, sadece teorik yaklaşımı kullanan katılımcılardan daha yüksek skor elde etmesinin daha olası olduğu görülmüştür.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

C(n,r)	Combination n choose r
P(A)	Probability of Event A
φ_c	Cramer's Phi Correlation Coefficient

LIST OF ACRONYMS/ABBREVIATIONS

AEC	Australian Education Council
GPA	Grade Point Average
IEA	International Association for the Evaluation of Educational Achievement
MEB	Ministry of Education in Turkey
NCTM	National Council of Teachers of Mathematics
TIMSS	Trends in International Mathematics and Science Study

1. INTRODUCTION

1.1. Background

Probability is a science of making predictions and decisions under uncertainty (Garfield & Ahlgren, 1988; Shaughnessy, 1977; Konold, 1989; NCTM, 2002; MEB, 2009). It is a novel topic in mathematics compared to others such as algebra. Despite being new, it is a crucial field especially in the information age. In its history, it was considered to be important for gamblers and mathematicians, however in today's world it is important for anyone in public. The information that people read or hear in the news every day about weather predictions, economics, medical tests etc. mostly contain a probability fact. Understanding these facts requires a basic understanding of the probability theory. By scholars, understanding probability in the 21^{st} century was predicted to be as important as mastering arithmetic in the 20^{th} century (Falk & Konold, 1992).

As probability gained more importance in real life by general public, the education level expanded itself from universities to K-12 schools. Curriculum designers have started including probability along with statistics in K-12 curricula. NCTM's Principles and Standards for School Mathematics includes probability as part of data analysis and probability along with four other content areas (2005). Similarly the curriculum in Australia includes probability as part of chance and data (AEC, 1990).

The relationship between probability inclusion in mathematics curricula and international testing is bilateral. The inclusion of probability in curricula of many countries has caused international testing to adopt probability as a component. Data representation, analysis and probability was among the six domains that was tested in TIMSS 1995. 14% of the items were regarding this domain. On the other hand, inclusion of probability in international tests has caused some countries such as Egypt to include probability in the national curriculum (Aggarwala, 2004).

The ongoing curriculum reform in mathematics education in Turkey has also implemented changes in the probability content which was not part of the curriculum before. In the new curriculum probability education starts at fourth grade level with an objective of teaching vocabulary related to probability such as “possible, certain, impossible, equally likely”. The second change implemented in the curriculum is the variety of objectives adopted such as the inclusion of dependent and independent events at the eighth grade level. The new curriculum states the importance of probability and statistics education by intending a goal to raise students who are conscious citizens and consumers, and who are aware of the importance of the field for the individual, society, various branches of science and various jobs (MEB, 2009).

What sets current Turkish curriculum apart from the rest of the curricula around the world is that it mentions three different approaches to probability. Even though definitions of the approaches will be explained later in detail, these definitions are namely: theoretical, experimental and subjective. These approaches are introduced at the 8^{th} grade level. Most of the other curricula are based on the theoretical approach and some others suggest the inclusion of the experimental. Since the real life probability questions such as the weather prediction and financial analysis are not straight forward, it is suggested that probability education should be inclusive of experimental and subjective approaches (Albert, 2003). One of the objectives of the current Turkish curriculum is that students are able to explain classical, theoretical and subjective probability.

The curriculum reform in Turkey is an ongoing one. The newer curriculum that is planned to be implemented starting in September 2013 has changes in probability topics. One of the changes is the exclusion of subjective probability and the other change is the age the different approaches are taught. The newer curriculum limits the above mentioned objective to theoretical and experimental approaches at 12^{th} grade level. By focusing on the three approaches held by teachers, the results of this study might provide reasons as to why the re-inclusion of the subjective probability into the reformed curriculum is necessary.

2. LITERATURE REVIEW

The initial studies on probability were conducted by psychologists in 1970s. These studies mainly focused on the heuristics that people rely on when making judgments about uncertainty. An heuristic is a strategy used when solving a problem that is based on the person's own experience, learning and discovery. It can be considered as a mental shortcut when making a decision (Kahneman & Tversky, 1972). The research by psychologists offer an extensive array of misconceptions that arise when people rely on these heuristics.

The earlier research also focused on the effects of age on probabilistic thinking and examine the relationship between misconceptions and age. In the early 80's as the curriculum designers suggested probability to be included in the mathematics curriculum, the need for educational research had risen. However there was still limited research on the teaching of probability in the late 80's (Garfield & Ahlgren, 1988). Mathematics educators started research on probability education in early 90's and there is still limited research when compared to other fields of mathematics (Jones & Thornton, 2005).

The research on probability education using different approaches of probability including subjective probability is limited and does not appear in the literature before the 21st century. The framework for teacher knowledge in probability includes different approaches to probability (Kvatinsky & Even, 2002). The research on teachers' adaptation of these approaches will be discussed later in the chapter.

As stated above, the research on misconceptions is relatively old and an extensive one while the research on different approaches is novel and limited. This chapter aims to summarize the research on these subtopics while trying to synthesize the relationship between the two and provide a summary on research focusing on the factors affecting probabilistic thinking such as age, culture and formal education.

2.1. Definitions of Probability

Any research on probability would require a definition of probability; however, there is not a single agreed-upon definition of probability. This section attempts to provide the different definitions found in the literature and exemplify the different viewpoints and terms by using probability questions. These definitions are namely classical, experimental and subjective and are summarized in Table 2.1. The explanations of these terms are as follows.

The broadest definition of probability can be stated as the measure of how likely that the event will occur (NCTM, 2002). Theoretical or classical probability is the ratio of number of outcomes in the set of interest to the total number of outcomes in the sample space. This definition is based on the assumption that the outcomes are equally likely (Batanero, Henry, & Parzysz, 2005). Chernoff uses the term school probability in place of classical probability because of its wide use in schools in mathematics classrooms (2010).

The frequency or empirical viewpoint is based on the assumption that if a random experiment is repeated many times then the probability of an event would be the relative frequency of the event in the set of all experimental trials. Despite the fact that the frequency definition ameliorates the classical definition by extending events to the situations where outcomes are not likely, it omits the situations in which the random process cannot be replicated (Albert, 2003).

There is also a philosophical difference between the classical and experimental definitions of probability. Chernoff attributes the difference between these two definitions to the epistemological division between empiricism and rationalism. He states that the classical approach is the product of rationalist view in which people believe that knowledge is derived from pure reason, whereas in empiricism knowledge is based on sensory perception. On the one hand, the classical approach calculates probability deductively without the need to conduct an experiment (*a priori*). On the other hand, in the frequency approach, one relies on the experiment (*a posteriori*). In other words:

On their senses and sense perception (2008).

The dispute about the definition of probability has given rise to another viewpoint, the subjective view. Usually associated with De Finetti's "Probability does not exist" claim, the subjective view assumes that probability does not exist objectively. According to this view, probability is a person's own description of uncertainty about the world (Nau, 2001).

Different problems can be solved using different definitions. For instance the probability of obtaining a prime number when rolling a single die can be solved objectively using both the classical and frequency definitions. From the classical perspective the probability is $\frac{1}{2}$ since the set of interest $\{2, 3, 5\}$ has three elements, the sample space $\{1, 2, 3, 4, 5, 6\}$ has six elements and the ratio between the number of these sets is $\frac{1}{2}$. From the frequency perspective this ratio is based on the outcome of the experiment of rolling a die and it is also $\frac{1}{2}$. The frequency approach is based on the hypothesis that as the number of trials approaches to infinity the experimental probability approaches the theoretical probability. Since it is impossible to reach infinity, it is suggested that simulation tools (calculators, probability software) are adapted in mathematics classrooms in order to increase number of trials when teaching the frequency approach (Jones & Thornton, 2005).

As stated before, both of these approaches are inadequate in situations where the outcomes are not equally likely and the experiment is not repeatable. Consider the probability of a specific student getting an A on a final. Even though the sample space is a finite one containing all the grade letters that the student can get, the elements in the set are not equally likely and the act of taking the final can not be repeated. Even if the student is allowed to take the final again, the conditions would have changed thus it would not be a repetition of the same experiment. In situations like this, subjective view does not have any limitation as the probability of the event solely depends on the person's own view of the event (Albert, 2003). The answer to this problem may vary from impossible to certain depending on the person. The probability of getting an A would be highly likely if the student had an high GPA, and it would be unlikely if the student had a low GPA. There are other factors such as the student's health and motivation that has an influence on the subjective probability in this case.

The definitions of subjective probability found in the literature are multivalent (Chernoff, 2010). One point of view about the subjective probability in the mathematics education is as irrational as De Finetti's aforementioned statement that the probability of an event is a degree of person's belief (Chernoff, 2008). The other view defines the subjective probability as the degree of rational belief (Chernoff, 2010).

The distinction between these definitions is crucial. For instance, in an activity provided in Turkey's national curriculum provides a statement about the chance of rain according to different students in a classroom (MEB, 2009). However, these statements of students are not based on a reason and are provided as examples of subjective probability. It can be assumed that the Turkish curriculum is considering subjective probability to be a degree of person's own belief solely based on the person rather than any rational reason.

In another activity that merges subjective and frequency approaches, trying to find the probability of an Hershey's kiss landing on its base, students make predictions and continue to change their predictions as they gather data. In this activity, students first make a prediction based on the physical attributes of the candy. Then they spill the candies from a cup ten times and discuss their earlier predictions and make further predictions. The newer subjective probabilities that students assign are based on the frequency, mean or median of the data they have gathered or close to one of these numbers. For example one student finds the frequency as 37% but changes her subjective prediction to 40% since she likes even numbers more. (Richardson, 2002). It is clear that this activity that is presented as an example to subjective probability is different from the one provided in the Turkish curriculum as it is based on experiments and gathered data and can be considered to be more rational.

The distinction between these two viewpoints of subjective probability is crucial, however, the terminological difference has not yet found its place in mathematics education literature (Chernoff, 2010). In this study, an earlier and wider definition of subjective probability will be adopted. Subjective probability will be considered to be a person's own belief of likelihood of an event, whether this belief is rational or not is beyond the scope of this study.

Table 2.1. Different Views of Probability.

Theories of Probability	**Definition**
Classical **Theoretical** ***a priori***	Ratio of the set of interest to the sample space
Frequency **Experimental** ***a posteriori*** **Empirical**	The frequency of the event happening as experiment results
Subjective **Bayesian** **Intuitive**	Likelihood based on person's own views

2.2. Misconceptions of Probability

This section is intended to summarize the misconceptions that people have while thinking probabilistically. Understanding the different conceptions of probability provides a wide perspective on probabilistic thinking and gives an insight into what choices people make when faced with uncertain situations. These studies do not necessarily provide information on why people make choices that they make but rather on what choices they make. Understanding these choices is a vital step in understanding the decision making mechanism of people. The seven misconceptions found in the literature are *representativeness*, *availability*, *negative and positive recency effect*, *equiprobability bias*, *conjunction fallacy*, *time axis fallacy*, *insensitivity to sample size.* These misconceptions are explained in detail.

One of the first studies determining and defining the heuristics was conducted by the psychologists Kahneman and Tversky (1972). In their research they explore the *representativeness* heuristics according to which the probability of an event is determined by the degree that the sample is similar to the population, in other words, the probability of an event depends on how much the sample represents the population. One example of an error based on this heuristic is when people are asked to compare the probability of sequences of births in two families BBBBBB (B being a birth of a boy) and BGGBBG (G being birth of a girl), people tend to state that the second sequence BGGBBG is more likely because it is similar to the parent population which is all the births in the world, however the sequences are almost equally likely.

Looking at the birth problem from the classical perspective it is easy to see that both sequences are equally likely. The probability of a girl is $\frac{1}{2}$ which is same as the probability of a boy. Then the probability of the first sequence would be $(\frac{1}{2})^5$ which is equal to the probability of the second sequence. However the frequency approach to this question may be misleading. According to the frequency approach as the number of trials approach to infinity the probability of an event would approach to the theoretical probability. If the length of the sequence is to be increased from 6 to a larger number (theoretically up to infinity) then one would expect the number of girls and boys to be equal in the sequence. The frequency approach in this case might mislead to a belief that the second sequence is more likely since the number of girls and boys are equal. However, frequency of many trials for a single child would lead to the correct answer which is the same as the theoretical approach.

Another heuristic defined is the *availability* heuristics according to which the probability of an event is related to the ease of remembering relevant instances of the event. When people are asked to compare the number of English words beginning with r with the ones that have r as the third letter, they tend to state that there are more words beginning with r even though the latter is more frequent (Tversky & Kahneman, 1973). In this situation this heuristic may be related to subjective probability because it is based on the person's remembrance of words. If such a relation exists then people who adopt the subjective approach would likely be relying on this heuristic.

Another question found in the literature assessing this misconception which is also used in this study is the comparison of probabilities of choosing 2 people out of 10 people with choosing 8 people out of 10 people (Fischbein & Schnarch, 1997). People relying on *availability* heuristic tend to state that the probability of choosing 2 people out of 10 people is greater than the probability of choosing 8 people out of 10 people, however, using the combinatorics rule these probabilities are equal since C(10,2) = C(10,8).One can arrive at the combinations formula using both the classical and frequency approach. However, using both approaches is arduous since classical probability would require listing all possible outcomes in the sample space and frequency approach would require repetition of the selection process. People who use this formula may not be aware of the approach behind it.

Another misconception is *negative and positive recency effect* (Fischbein, 1975). After a coin flip of three heads one may believe that the fourth flip is more likely to be tails. This is named the *negative recency effect*, also known as the *gambler's fallacy* in the literature. If one believes the fourth flip is more likely to be heads then it is named as the *positive recency effect*. The probability of a head in the fourth flip is equal to the probability of a tail. Using the classical approach in this problem would lead to a correct answer since the P (HHHH) = $(\frac{1}{2})^4$= P (HHHT). The frequency approach in this case might be misleading as the number of trials (coin flips) increase the number of heads and tails would equal to each other, favoring T in the fourth flip, however for the frequency approach one has to focus on a single coin flip rather than the whole sequence. Focusing on the fourth flip the frequency approach will yield to equal number of heads and tails -as the number of trials increase- suggesting that H and T have an equal probability in the fourth flip.

Another alternative conception that people display is the *equiprobability bias* according to which people tend to believe that events are equally likely when they actually are not (Lecoutre, 1992). In Fischbein and Schnarch's the same bias is named as the *misconception of compound and simple events* (1997). A compound event is two or more simple events happening at the same time. For instance, if the simple event of obtaining a six in a roll die and the simple event of obtaining a five in a roll die are to happen simultaneously then the compound event of obtaining a five and a six would occur. When comparing two compound events of obtaining ‘a double six’ and ‘a five and a six’ people tend to state that these probabilities are equally likely, however the probability of obtaining ’a five and a six’ is higher.

Looking at the two dice problem from the classical perspective may be beneficial or not depending on the person’s ability to recognize the sample space. It would be misleading to adopt the classical approach if the person thinks of a compound event as an intersection and calculates the probability as P(6) x P(5) $= \frac{1}{6}$ x $\frac{1}{6} =$ P(6) x P(6). However if a person is able to recognize the sample space of rolling a two dice experiment that has 36 elements including $\{5, 6\}$, $\{6,5\}$ and $\{6,6\}$ may lead to a conclusion that P (5 and 6) $= \frac{2}{36} > \frac{1}{36}$. Approaching the same problem from the frequency perspective would require a simulation tool.

The *conjunction fallacy* is another misconception that people have (Tversky & Kahneman, 1983). Probability of a conjunction P $(A \cap B)$, two events occurring at the same time, is always less than the probability of the event A, P(A), and is also less than the P(B). When people are confronted with the following description and are asked to choose between the two choices such as in the following situation:

> Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.
> a) Linda is a bank teller
> b) Linda is a bank teller and is active in the feminist movement

people tend to choose the latter choice as more probable even though the first choice is a more general statement.

The subjective approach in this situation may be misleading since Linda’s history on social justice may lead to a belief that she is active in the feminist movement and thus ignore the fact that the likelihood of Linda being a bank teller is higher than both events combined. Both the classical and frequency approach may lead to the correct thinking that $P(A) \geq P(A \cap B)$. People who possess this sort of thinking may not state the approach they use to arrive at this judgment.

The *time axis fallacy* also known as the *Falk phenomenon* is the understanding that people hold that is related to cause and effect (Falk, 1979). In situations of conditional probability, for example drawing a ball from a bag without replacement, people have an understanding that events A and B (in the chronological order) are related in a way that P(B) is related to the P(A) but not vice versa. The outcome of the event A has an effect on the outcome of the event B but not the other way around. In other words the outcome of the event A can be used to predict the outcome of the event B but not vice versa.

The question shown in Figure 2.1 is used in the literature to assess *time axis fallacy* (Fischbein & Schnarch, 1997). It is also adopted in this study.

Yoav and Galit each receive a box containing two white marbles and two black marbles.
(A.) Yoav extracts a marble from his box and finds out that it is a white one. Without replacing the first marble, he extracts a second marble. Is the likelihood that this second marble is also white smaller than, equal to, or greater than the likelihood that it is a black marble?
(B.) Galit extracts a marble from her box and puts it aside without looking at it. She then extracts a second marble and sees that it is white. Is the likelihood that the first marble she extracted is white smaller than, equal to, or greater than the likelihood that it is black?

Figure 2.1. Item Assessing Time Axis Fallacy.

In both i and ii the probability of obtaining a white marble is smaller than obtaining a black one. In both of the situations P(white) $= \frac{1}{3}$ and P (black) $= \frac{2}{3}$. The misconception occurs when people tend to believe that probability of obtaining a white marble is smaller than obtaining a black marble in the first situation and is greater in the second situation. In the second situation when the first draw is unknown the sample space becomes misleading. It can be considered as composed of 2 white and 2 black marbles however that is not the case since the second draw is white for certain. Then the sample space in the first draw must be consisting of 1 white and 2 black marbles. Approaching this situation from a classical perspective can lead to a correct answer given that the sample space is clear. The frequency approach in this case would require a simulation tool.

Insensitivity to sample size is a heuristic people rely on when calculating probabilities of events without paying any attention to the sample size (Tversky & Kahneman, 1982). People with this insensitivity tend to think that getting at least two heads out of three coin flips is as likely as getting at least two hundred heads out of three hundred coin flips. However, the latter has a smaller probability. Such misconception may be a result of a classical approach $\frac{2}{3} = \frac{200}{300}$ which does not lead to a correct thinking. The effect of sample size can be seen easier using the frequency approach. When considering coin flips one would expect half of the outcomes to have equal number of heads and tails. As the number of trials increase this ratio would approach $\frac{1}{2}$. In this situation where the number of trials is 300, one would expect to see sequences mostly consisting of 150 heads and 150 tails. There will be less sequences with 200 or more heads.

This section has examined the misconceptions found in the literature and stated their possible relation to different approaches of probability. Through these relationships it can be seen that different approaches can be used when solving problems assessing these misconceptions. None of the misconceptions were suggested to be solved using subjective probability. As stated earlier in the chapter, subjective probability does not give an absolute single answer to any probability problem. The problems assessing misconceptions require a single answer that is either right or wrong. Using subjective probability when solving questions, such as the ones above, may be misleading whereas using classical and frequency approaches depending on the situation may be rewarding.

2.3. Factors Affecting Probability

2.3.1. Age

Age is one of the factors considered to have an effect on probabilistic thinking. The research on the effect of age can be divided into two: the ones focusing on the initial development of probabilistic thinking focusing on the age when human beings first start to make probabilistic judgments and the second type focusing on how the probabilistic thinking develops across ages.

One of the first studies exploring the effect of age on the development was of Piaget's. In his book co-written with Inhelder he details all the experiments conducted with children (1975). They worked with children aged 5-12. One of the experiments in the book is the well-known ball drawing from a bag which is the scenario of some of the probability questions found in almost every mathematics textbooks today. Participants in the study were asked the color of the token that will be drawn from an opaque bag. They were also provided with the same number of colored tokens on the table.

Based on the experiment, researchers concluded that younger children (aged 5-7) do not make predictions based on ratio but rather on different criteria. Some of these different criteria are their color preference, physical characters of tokens and the order of tokens on the table. Participants at this specific age level were said not to have an understanding of ratio of elements and randomizing. Children aged (8 to 10) were able to base their choices on ratio however they failed to do so in situations when tokens were not replaced. Older children had a grasp of understanding the situation with replacement. The findings of the study can be interpreted as younger children having the tendency to use subjective approach when they are faced with uncertainty. As children grow up their understanding of ratio helps them to a tendency to use the classical approach.

Perhaps one of the research methods to examine the effect of age would be a longitudinal study. However, such studies are hard to conduct with difficulty of access to such a sample. One longitudinal study found in the literature is of Green's which compared the probability concepts of 7-10 year olds with 11-14 year olds (1990). The study measured two concepts: participants' understanding of randomness and their abilities to compare odds in a ball drawing situation.

Participants were asked to generate a sequence of heads and tails of 50 trials but not actually flip a coin. The study concluded that the students had an understanding of randomness as having equal frequencies of the elements in the sample space at both age levels, however, the understanding of independence of trials and variation were not part of their understanding and these concepts were not acquired as age increased for these groups. For comparison of odds, the study states that the age has a positive effect on comparing probabilities especially for those with high ability. The results of this study overlap with Piaget's. As age increases and students have a better understanding of ratio and they are more able to compare odds.

Another study in the literature examining the effects of age on the development of probabilistic thinking was conducted by Fischbein and Schnarch (1997). The study compared grade 5 (ages 10-11), grade 7 (ages 12-13), grade 9 (14-15), grade 11 (ages 16-17) and college students who had never received formal instruction in probability. In the study it was seen that the age had a positive effect on overcoming representativeness and negative recency effect misconceptions. The conjunction fallacy was constant among grades 5, 7, and 9 and diminished for grade 11 and college students. The only stable misconception found in the study was compound and simple events (equiprobability bias). There were three misconceptions identified on which the age had negative effect: availability, effect of sample size, and the effect of time-axis.

In both Green's and Fischbein and Schnarch's studies it can be seen that the relationship between the age and the success of overcoming misconceptions is not linear. For certain misconceptions the success increases and for some it decreases. Overall all the studies focusing on age can be interpreted as follows: Up to a certain age (12) children develop the concept of ratio and as they are learning the concept they start to rely less on subjective probability. However, lack of reliance on subjective probability as a result of increase in age does not lead to reliance on classical or frequency approaches that would lead to a correct decision making mechanism under uncertainty in different situations.

2.3.2. Culture

There is limited research on the impact of culture on probabilistic thinking. Shaughnessy states the need for cross-cultural studies in order to provide evidence on whether probabilistic thinking and misconceptions vary across cultures or are universal (1992). Impact of culture in probabilistic thinking may be related to subjective probability since it is based on person's own belief of the world.

The study conducted by Amir and Williams tries to establish the link between culture and probabilistic thinking (1999). The researchers define culture as a construct comprised of language, beliefs, and experiences. The research uses both quantitative and qualitative data. The quantitative data shows that the lack of probabilistic vocabulary is an obstacle in attaining success in probabilistic thinking. In terms of beliefs students mostly display belief in religion, God, superstition, however such beliefs do not have any effect on probabilistic thinking. In terms of experience, there were students who had experience with coin flipping, die rolling, playing cards and board games, however such experience did not lead to any success in probabilistic thinking.

Overall in terms of religion and experience there was not any relation found between these aspects and success in probability, however students still display subjective approaches in the interviews. For instance, students expressed that the devices used in classrooms such as coins, dice, wheel of fortune are not random and the result displayed on these devices are dependent on the person operating the device. The findings of the study can be interpreted as students having subjective view of probability but displaying objective approaches when given tests.

Despite the importance of the relation between culture and probabilistic thinking, the literature does not provide adequate research on the effects of culture (Shaughnessy, 1992; Greer & Mukhopadhyay, 2005). It is suggested that further studies are needed in the cultural analysis of probability along with historical and epistemological. Studies focusing on these topics would increase awareness about the importance of subjectivity in probability and would take the field beyond determinism (Greer & Mukhopadhyay, 2005).

2.3.3. Formal Education

Misconceptions in the literature have been deeply researched by psychologists and mathematics educators who have been focusing on delivery mode of probability in classrooms in order to overcome these misconceptions. The inclusion of probability in mathematics curricula has led to activity and material development. Despite the rise in material and course development in the late 80's, the lack of research left the teaching part of the field still unstructured (Garfield & Ahlgren, 1988). The research conducted in the last twenty years enlightens us about teaching, however there is still limited research compared to teaching of other branches of mathematics education (Greer & Mukhopadhyay, 2005).

The research on teaching of probability relies on Fischbein's classification of intuitions (1975). In his study intuitions were classified into two as primary and secondary. Primary intuitions are cognitive beliefs that are derived from a person's own experience and that are not a result of systematic instruction. For example, in a ball drawing situation a student might believe that a red ball is more likely to be drawn based on the fact that red is his favorite color. Primary intuitions can be considered as the basis of subjective probability.

Secondary intuitions are the beliefs based on instruction which constitute the objective probability. A student working on the task above may replace his intuition to "red is more likely because there are more red balls". In a situation like this, according to Fischbein, the student is not losing his primary intuition -in fact it may reappear in a different task- but rather associating secondary intuitions with a certain task. The goal of probability education is to help students prioritize the secondary intuition (Fischbein, 1975). Contrary to the belief that helping students raises a cognitive conflict between two intuitions by providing rules and formulae, practice or new experience can cause a change in intuition (Fischbein, 1987). It is important to provide with students opportunities through which they can experience different probabilistic situations.

In order to provide students with different experiences, there has been a variety of research on computer use as a source for simulation and data generation in probability classrooms. Computers offer many simulation tools for data generation. The teaching activities designed by Kazak and Konold (2010) integrate the use of simulation tools into teaching of probability. During the implementation students were asked about their perception and made predictions on tasks followed by a collection of data physically. The data collection process helped to raise a conflict between students' prior thinking and formal probabilistic thinking. This conflict led to a need for more data collection thus to a need for using a simulation tool. Authors concluded that simulation tools provide opportunities for students to put their informal theories of probability to test. Simulation tools are especially important in cases where it is impossible to collect a large set of data in classrooms. Such tools serve the purpose of teaching the frequency view of probability.

In a similar study using a different computer program as a simulation tool, the researcher examined students' thinking process of the total of two dice (Pratt, 2000). The study began with identifying students' prior, informal theories (local resources) of probability and tried to raise a conflict between students' thinking model and the formal theory. The tasks using simulation helped students forge formal theories of probability, however when faced with a different situation rather than these new theories, the informal theories were used. This study sets a very crucial point about the formal education of probability. As seen in this study, raising a cognitive conflict is not sufficient to internalize the formal probability theories and make use of them in different situations. In other words, even when students are instructed using objective approaches and rely on these approaches in a certain task, does not necessarily guarantee their reliance on the same approaches in a different task.

Probability tasks involving computer use focus on theoretical and frequency views of probability. Most curricula depend solely on these two definitions. However, in Turkish curriculum there is an objective stating that the student knows the difference between theoretical, experimental and subjective probabilities (MEB, 2009). Subjective probability is taught as a definition along with a couple of examples. One of the examples given under subjective probability is shown in Figure 2.2.

<u>Öznel olasılık:</u> Bugün yağmur yağma olasılığı Melike'ye göre %60, Yavuz'a göre %80'dir. Cümlede, yağmurun yağma olasılık değerlerinin neden farklı olduğu açıklatılır.

Figure 2.2. Subjective Probability Example in MEB Textbook.

The activity states that the chance of rain according to Melike is 60% and according to Yavuz it is 80%. The difference between probabilities of rain is to be explained by students. However, in further exercises that are not listed under this objective, students are not given the option of using a subjective approach and teachers are not directed to encourage different approach usage.

Research involving all three different approaches in an educational setting is limited. One of the studies involving the three approaches is of Albert's research conducted with 75 college students enrolled in an introductory statistics course (2003). In the study students are given a test which contains three classical type problems, three frequency type problems and three subjective type problems. The classical problems were found to be the easiest ones to solve. Students hesitated to use personal belief on questions and some of them even tried to compute probabilities of situations which did not even make sense.

In his article Albert compares the usual probability tasks with the ones that are found in real life such as weather prediction that does not involve a specific, regular method. Since the real life probability questions are not as straight forward as the typical problems he suggests that probability education should be more focused on real life and involve more of frequency and subjective approaches.

Rast's doctoral dissertation involves an implementation of teaching model that is inclusive of subjective probability tasks (2005). In her mixed research design the control group receives instruction solely based on the traditional (theoretical and frequency) view of probability whereas the experimental group receives instruction based on both traditional and subjective views of probability. The study is conducted with fourth, fifth and sixth grade students and at all levels there was no significant difference between the means of probability knowledge. The qualitative data shows that misconceptions labeled in the research are subjective judgments. This study shows that adopting all three approaches was not enough in overcoming misconceptions.

Despite the lack of statistical evidence in her dissertation, Rast provides suggestions for inclusion of different approaches, especially the subjective approach, in the mathematics curriculum. One of her suggestions is that rather than solely teaching the three approaches, students should also be taught when to use the appropriate approach. Students should be taught to validate their answers with a rational reasoning even if they choose the subjective approach. In other words, adopting the subjective approach in a classroom does not give the student the right to answer whatever he believes in. Another suggestion that Rast makes is that teachers should internalize the subjective approach and should be able to direct students towards the different approaches depending on appropriateness. The last suggestion she makes is regarding teacher preparation. She suggests that pre-service and in-service teachers should be trained in different approaches since the traditional curriculum does not include these approaches.

2.3.3.1. Teachers. One of the main actors in education of probability is the teachers. Students' probabilistic thinking depends heavily on teacher's understanding of probability (Stohl, 2005; Memnun, 2008). According to Greer and Mukhopadhya the inclusion of probability in curricula in many countries were due to the political forces and educational fashion and it was done in a way that did not pay sufficient attention to teacher preparation (2005). Compared to research focusing on students' understanding of probability there has been less research focusing on teachers (Stohl, 2005).

Many teachers do not like teaching probability (Kvatinsky & Even, 2002). Teachers' negative attitudes towards probability causes a less efficient learning environment and cause students to have negative attitudes towards probability (Bulut, 2001). The reluctance of teachers to teach probability may be related to their lack of subject-matter knowledge (Shaughnessy, 1992). One of the studies focusing on teachers' subject-matter knowledge was conducted by Bulut, Kazak and Yetkin (1999). Despite the fact that some of the participants had taken a probability course at college level, pre-service mathematics teachers were found to be insufficient in basic probability concepts. Authors suggest that pre-service teachers who lack information on basic probability subjects would have difficulty teaching the subject efficiently in their teaching careers.

Apart from the lack of subject-content knowledge, another difficulty that teachers face is the nature of probability (Stohl, 2005). Unlike other fields in mathematics the rules in probability are not "certain". Stohl gives an example of rolling a die in which we are interested in getting a "four". The theoretical probability 1/6 is an estimate. The actual probability which also includes the complexity of physics (e.g. friction). The conflict between theory and the "real" experience brings a difficulty in teaching. She suggests that using both the classical and frequency approach would lead to eliminating this conflict thus better learning for students.

Apart from the literature on probability, as part of the literature on teacher preparation, years of experience of teachers is considered to be one of the factors predicting teachers' competence. Studies show that there is a relationship between teachers' effectiveness and their years of experience (Darling-Hammond, 1999). In his study, Rockoff showed that students of teachers with 10 or more years of experience score higher in tests than the students of those with less experience (Rockoff, 2004).

In summary, the literature provides adequate research on misconceptions. It is also clear that age (up to a certain point), culture and teachers are important factors in education of probability. The suggestions in the literature emphasize inclusion of different approaches. Teachers' content knowledge is crucial and the approaches adopted by the teachers may be related to subject-matter knowledge. Given the effects of teaching experience in other domains, experience level of the teacher may also be a factor in the field of probability. In this regard, this study investigated the relationship between experience level, approaches and success level in solving probability problems.

2.4. Significance of the Study

The importance of probability is stated in different curricula of different cultures (NCTM, 2002; MEB, 2009). However, the literature states a need for research on teachers (Shaughnessy, 1992; Garfield & Ahlgren, 1988; Stohl, 2005). There is research focusing on teachers' subject-matter knowledge through which we conclude that the subject-matter knowledge of teachers is inadequate.

There are studies focusing on different approaches of probability and how students learn through these approaches (Albert, 2003; Rast, 2005). However there are not any studies combining both the teacher aspect and the probability definitions. In this research study, teachers' approaches were examined which may be related to their subject-matter knowledge.

When teaching probability, the aim is to avoid misconceptions and to teach secondary intuitions. The experts of the field are those who hold correct modes of decision making under uncertainty. The aim of probability education is to teach students to think like experts who can predict on certain outcomes of events. The literature on experts state that age and necessary preparations as factors of superior performance on different domains (Ericsson & Lehmann, 1996). Considering that the literature on probability provides adequate research on the effects of age there should be more research focusing on necessary preparations required to master probability.

The researcher hypothesized that one aspect of the necessary preparation in probability is the different approaches. In this study, teachers' probability approaches were examined. Teachers are thought to be experts who are teaching the topic. Understanding their approaches to probability would enlighten the mathematics education fields on whether different definitions of probability are part of the necessary preparation required for mastery.

3. STATEMENT OF THE PROBLEM

The aim of the study was to examine teachers' approaches to probability questions identifying misconceptions while focusing on their experience level.

3.1. Research Questions

(i) What types of probability approaches do pre-service and in-service teachers adopt while reasoning on different probability questions that are affiliated with misconceptions?
(ii) What is the relationship between the experience level and the adopted probability approach?
(iii) What is the relationship between the experience level and the success level in answering questions correctly?
(iv) What is the relationship between the adopted probability approach and the success level?

3.2. Variables

There are three variables in this study.

(i) Approach: This variable is a categorical (nominal) variable defining the approach or approaches adopted by the teacher when solving probability questions. Despite the fact that there are three main categories for this variable; classical, frequency, subjective, there are seven possible combinations of these variables; classical, frequency, subjective, classical and frequency, frequency and subjective, classical and subjective and the last one including all of the three approaches. There were more categories found for modes of thinking that may be related to these approaches that are less broader than the approaches themselves and more case specific. These categories will be introduced and explained in Chapter 5.
(ii) Experience level: This variable is an ordinal variable based on teachers' years

of experience in teaching mathematics. There are three categories as pre-service teachers with no formal teaching experience (Experience Level 0), teachers with a teaching experience up to 10 years but not including those with 10 years of experience (Experience Level 1), teachers with teaching experience of 10 or more years (Experience Level 2). Experts attain their highest level of performance after a decade of necessary preparation (Ericsson & Lehmann, 1996). The 10-year rule was obliged in this study when categorizing participants based on their years of experience.

(iii) Success level: This variable has two different definitions at different levels of the study. When analyzing a response to a specific misconception question success level was a dichotomous ordinal variable with only two categories (correct and incorrect) which shows whether the participant was successful in answering the question or not. In the analysis combining all the questions, success level was a score out of 7 depending on the success on items identifying seven different misconceptions taken into account. The total success score includes success on all items except for item 3 and 9 assessing subjectivity. The success on remaining 8 questions is used in determining the total success level. However, it should be noted that items 5 and 10 identify the same misconception and success on both items was analyzed as success on having overcome the misconception hence the total score of 7 rather than 8.

In the second research question which aims to examine the relationship between experience level and the probability approaches, experience level is the predictor variable and probability approach is the criterion variable.

In the third research question which aims to examine the relationship between experience level and the success level, experience level is the predictor variable and success level is the criterion variable.

In the fourth research question which aims to examine the relationship between probability approach and the success level, probability approach is the predictor variable and success level is the criterion variable.

3.3. Research Hypotheses

First research question is a qualitative one and does not have a hypothesis that can be computed. Based on the literature, especially Albert's study (2003) the researcher hypothesized that teachers will mostly adopt the classical approach with very few adopting the frequency approach and very few adopting the subjective approach.

The hypotheses for the rest of the research questions are as follows.

- $H_0 2$: There is not a significant relationship between the experience level and the adopted probability definition.
- $H_A 2$: There is a significant relationship between the experience level and the adopted probability definition.
- $H_0 3$: There is not a significant relationship between the experience level and the success level of answering questions correctly.
- $H_A 3$: There is a significant relationship between the experience level and the success level of answering questions correctly.
- $H_0 4$: There is not a significant relationship between the adopted probability definition and the success level.
- $H_A 4$: There is a significant relationship between the adopted probability definition and the success level.

4. METHODOLOGY

4.1. Sampling Method

The sampling method used in this study was purposive sampling. In purposive sampling, the researcher selects a sample based on her purpose and experience (Gay, Mills, & Airasian, 2009). In order to examine the different approaches, the researcher selected samples from private schools where language of instruction is English. Such schools are more likely to use technological devices in mathematics classrooms compared to public schools due to financial limitations in public schools. As stated earlier, frequency approach is related with simulation thus requires technological devices in classrooms. The researcher hypothesized that different approaches may be adopted by teachers teaching in such schools.

4.2. Participants

The total number of participants was 72. The intended minimum number of participants was 15 for each experience level, however since the researcher had access to schools, she continued to collect data. The number of experienced teachers was higher. 17 participants were pre-service mathematics teachers. 22 participants had less than ten years of teaching experience. 33 participants had ten or more years of teaching experience.

4.3. Research Design

The nature of the research question has shaped the research design of the study. In seeking an answer to the first research question, identifying the probability approaches used by the mathematics teachers, qualitative data was collected through the open ended parts of the questions.

The responses to the questions have also been quantitative, measuring the variable success level. This variable, assessing whether the question was right or wrong was a result of a multiple choice thus quantitative.

For the rest of the research questions examining a relationship between variables; experience level, approach and success level, data were quantified and quantitative statistics were used in seeking the relationship.

The study merged both the quantitative and qualitative data and used qualitative analysis categorization along with quantitative correlation techniques thus the research design of the study is a mixed methods design (Pinto, 2010).

4.4. Method

The researcher examined the literature depending on the research questions. Through her research of the literature she developed in the instrument. Based on the suggestions of the commitee members, in order not to lose the reliability of the items, the questionnaire was developed in English.

The items were analyzed by an expert of measurement and evaluation in terms of the construct of the test, number of items and the order of items. Then second round of review was conducted by a probability expert who has examined the items in terms of content validity. The expert has also reviewed the demographic information that was intended to be collected in this study. After suggestions and critique, the expert and the researcher came to an agreement on the items.

The researcher used her personal network to contact the private schools, mainly contacted the principals. She sent an e-mail to potential participants asking whether they would volunteer or not. Those who responded were contacted for an appointment. The volunteers picked the time of the day depending on their teaching schedule. They were given unlimited amount of time to finish all of the items. Most of the participants took around 40 minutes to an hour, with an exception of one who took 2 hours to complete all of the items.

After the data collection process, the researcher examined the responses and put the data into categories and some items into subcategories. The categories and subcategories were discussed between the researcher and her advisor. The definitions of these categories and subcategories along with the data were given to another graduate student who used these categories to analyze the data. This step was conducted for plausability of categories. The researcher and the graduate student came to an agreement on categorization of all items for every participant. These categorizations were later used for statistical analysis.

4.5. Instrument

Appendix A shows the questionnaire used in the study. There are two parts found in the questionnaire. The first part starts with information about the research followed by 10 probability questions. The second part consists of demographic information on participant's years of teaching experience, the levels of schools the participant has taught, whether the participant has taken statistics and probability courses at undergraduate and graduate levels, whether they use a graphing calculator and simulation programs in their classrooms as these factors were predicted to have a possible relationship with the success level.

Part I of the questionnaire consists of questions that are mostly adopted from the literature. The names of people and places have been changed with Turkish names in order to avoid cultural effects. Part II was given to participants after they have finished Part I so that they would not be directed by questions in Part II such as whether they simulate data or not. Since one of the factors that Part I was assessing was whether participants use the frequency approach or not, by asking about simulation before hand might have directed the participants to use the frequency approach. For this reason, Part II was given after Part I was completed.

In this section all the items will be explained in detail. Table 4.1 summarizes the items and lists the misconception the item is assessing and the approach that can be used when solving the items. As stated earlier, the frequency approach requires simulation tools such as calculators and computer software. Since the participants did not have access to simulation tools, they were provided with a graphing calculator. Simulation on the calculators is possible, however it is not guaranteed that the participant knows how to simulate data on a calculator.

All the items except for item 3 have been used in other research studies. The items were found to be reliable. Especially the items measuring misconceptions have been used several times in the literature. For this study, two experts have reviewed the instrument. A professor of measurement and evaluation has reviewed the instrument in terms of test design. An expert on probability education who has publications in international journals has reviewed the instrument in terms of content validity. After discussions the researcher and the expert have arrived at a consensus.

First item assessing the representativeness adapted from Fischbein and Schnarch (1997) is as follows:

> In a lotto (Sayısal Loto) game, one has to choose 6 numbers from a total of 49. Banu has chosen 1, 2, 3, 4, 5, 6. Mehmet has chosen 39, 1, 17, 33, 8, 27. Who has a greater chance of winning? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

Without the simulation tools this question can not be solved with the frequency approach however it can be solved using the classical approach. The sample space here is 49 and for each drawing there is $\frac{1}{49}$ chance of guessing correctly. The probability of Banu winning and Mehmet winning are both equal to $\frac{1}{49} \cdot \frac{1}{48} \cdot \frac{1}{47} \cdot \frac{1}{46} \cdot \frac{1}{45} \cdot \frac{1}{44}$.

Second item assessing negative and positive recency effect adapted from Fischbein and Schnarch (1997) is as follows:

> When tossing a coin, there are two possible outcomes: either heads or tails. Nurcan flipped a coin three times and in all cases heads came up. Nurcan intends to flip the coin again. What is the chance of getting heads the fourth time? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

Using the classical approach would lead to a correct answer since the P (HHHH) = $(\frac{1}{2})^4$= P (HHHT). The frequency approach in this case might be misleading as the number of trials (coin flips) increasing the number of heads and tails would equal to each other. The correct mode of thinking for the frequency approach would be focusing on the experimental outcomes of the fourth flip only. This would require data simulation for the fourth flip thus a simulation tool.

Item 3 has been developed by the researcher based on Albert's suggestion that situations such as lottery winning would derive a subjective approach (2003). Item 3 is as follows:

> Above is an example of a Turkish lottery ticket (Milli Piyango) consisting of 6 digits. Turkish National Lottery Administration (Milli Piyango İdaresi) has a lottery drawing on 9th, 19th and 29th of every month. What is your chance of winning the lottery in 2013? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

The question is asked along with an image and in a way to give an opportunity for classical, frequency and subjective approaches. The best approach to this question is the subjective approach since it is actually asking for the participant's own chance of winning the lottery. The researcher hypothesized that the perfect solution might include statements such as the probability is zero since I never buy lottery tickets, it is low since I do not buy lottery tickets often, it is extremely high as I buy many tickets every week and I am a really lucky person. This item will not be used to measure success level since it is not assessing a misconception. Participants were expected to be tricked by this question and use classical approach since it contains numeric data.

A similar question similar to Item 4 has been discussed earlier in Section 2.2. The item assessing negative and positive recency effect from Fischbein and Schnarch (1997) is as follows:

> Tayfun dreams of becoming a doctor. He likes to help people. When he was in high school, he volunteered for the Turkish Red Crescent (Kızılay) organization. He accomplished his studies with high performance and served in the army as a medical attendant. After ending his army service, Tayfun registered at the university. Which seems to you to be more likely? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.
>
> A) Tayfun is a student of the medical school.
>
> B) Tayfun is a student.

The subjective approach in this situation may be misleading since Tayfun's history with medical practices may lead to a belief that he is a student of the medical school thus ignore the fact that the likelihood of Tayfun being a student is higher than both events combined. Both the classical and frequency approach may lead to the correct thinking that $P(A) \geq P(A \cap B)$, event A representing Tayfun as a student, event B representing Tayfun in a medical school. People who possess this thinking may not state the approach they use to arrive at this judgment. One of the methods of solution to this problem is through drawing a Venn diagram such as the one in Figure 4.1. Through this figure it can be seen that if Tayfun is a student of the medical school, then he is a student for sure.

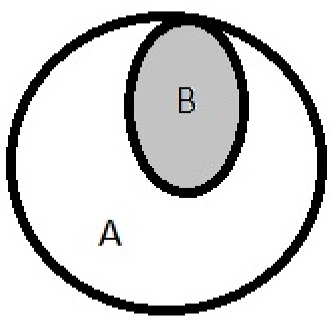

Figure 4.1. Sets A and B.

Item 5 has been discussed earlier in Section 2.2. The item assessing insensitivity to sample size from Fischbein and Schnarch (1997) is as follows:

> K: The likelihood of getting heads at least twice when tossing three coins
> L: The likelihood of getting heads at least 200 times out of 300 times
> A) K is smaller than L
> B) K is equal to L
> C) K is bigger than L
> Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

As stated earlier, this misconception may be a result of a classical approach $\frac{2}{3} = \frac{200}{300}$ which does not lead to a correct thinking. The effect of sample size can be seen easier using the frequency approach. When considering coin flips one would expect half of the outcomes to have equal number of heads and tails. As the number of trials increase this ratio would approach $\frac{1}{2}$. In this situation, where the number of trials is 300, one would expect to see sequences mostly consisting of 150 heads and 150 tails. There will be less sequences with 200 or more heads. The law of large numbers with the frequency approach must be used in order to solve this problem.

Item 6 has also been discussed earlier in Section 2.2. The item assessing availability from Fischbein and Schnarch (1997) is as follows:

> K The number of possibilities when choosing 2 members from among 10 candidates
> L The number of possibilities when choosing 8 members from among 10 candidates
> A) K is smaller than L
> B) K is equal to L
> C) K is bigger than L
> Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

Using the combinatorics rule these probabilities are equal since C(10,2) = C(10,8). One can arrive at the combinations formula using both the classical and frequency approach. However, using both approaches is arduous since classical probability would require listing all possible outcomes in the sample space and frequency approach would require repetition of the selection process. People who use this formula may not be aware of the approach behind it. Thus the researcher calls this a formula.

Item 7 has also been discussed earlier in Section 2.2. The item assessing equiprobability bias from Fischbein and Schnarch (1997) is as follows:

> Suppose one rolls two dice simultaneously. Which of the following has a greater chance of happening? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.
> A) Getting the pair 5-6
> B) Getting the pair 6-6
> C) Both have the same chance.

Approaching this item from a classical approach may not be adequate as the misconception occurs when people think that event in A and event in B are equally likely. For those who can not see that $\{6,5\}$ is also a possible outcome these two event would be equally likely.

Item 8 has also been discussed earlier in Section 2.2. The item assessing time axis fallacy is from Fischbein and Schnarch (1997) is as follows:

> Emre and Nihan each receive a box containing two white marbles and two black marbles.
> i) Emre extracts a marble from his box and finds out that it is a white one. Without replacing the first marble, he extracts a second marble. Is the likelihood that this second marble is also white smaller than, equal to, or greater than the likelihood that it is a black marble? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.
> ii) Nihan extracts a marble from her box and puts it aside without looking at it. She then extracts a second marble and sees that it is white. Is the likelihood that the first marble she extracted is white smaller than, equal to, or greater than the likelihood that it is black? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

The misconception occurs when people tend to believe that probability of obtaining a white marble is smaller than obtaining a black marble in the first situation and is greater in the second situation. In the second situation when the first draw is unknown the sample space becomes misleading. It can be considered as composed of 2 white and 2 black marbles, however, that is not the case, since the second draw is white for certain. Then the sample space in the first draw must be consisting of 1 white and 2 black marbles. Approaching this situation from a classical perspective can lead to a correct answer given that the sample space is clear. The frequency approach in this case would require a simulation tool.

Item 9 is adopted from Albert's study (2003) in order to assess subjective probability. In his study he was only working with college students thus could ask the probability of the participant getting married in the following five years. For this study, the participants were likely to be married thus they were asked the probability of their getting divorced in the following five years. Those with an objective approach could respond as the question being wrong or as the answer being impossible to be calculated. Those with a subjective approach could respond based on experience, personal life etc.. This question will not be used to assess success level since it does not have a single answer and is not assessing a misconception.

> If you are single what is the probability that you will get married in the next five years? If you are married what is the probability that you will get divorced in the next five years? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

Since it is more difficult to use the frequency approach without a simulation tool, an extra question assessing insensitivity sample size has been added to the questionnaire. This misconception is the only one that is easier to be solved using the frequency approach and not the classical. Item 10 has been adopted from Garfield's Statistical Reasoning Assessment (1988) and is as follows:

> 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? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.
> 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.

Those with a classical approach to this problem may use the $\frac{40}{50} = \frac{8}{10}$ which would lead to a wrong answer. The researcher hypothesized that this question can only be solved using the frequency approach objectively. Since the number of trials (births) are higher in Hospital B, according to law of large numbers, the ratio of girls to boys is more likely to be close to 50%.

Table 4.1. Summary of Items.

Item Number	**Misconception**	**Approach**
1	Representativeness	Classical
2	Negative and positive recency effect	Classical
3	-	Subjective
4	Conjunction fallacy	Formula
5	Insensitivity to sample size	Frequency
6	Availability	Formula
7	Equiprobability bias	Classical
8	Time axis fallacy	Classical
9	-	Subjective
10	Insensitivity to Sample Size	Frequency

As seen from Table 4.1, items 3 and 9 which provide situations for subjective approaches do not assess misconceptions. These items were not used towards calculating the success level of the participant. There are a total of four questions that can be solved correctly using the classical approach and there are a total of two questions that can be solved using the frequency approach.

4.6. Data Analysis

The responses to all items were analyzed by the researcher and categorized into different approaches. Another graduate student in the department has analyzed all of the items for inter-rater reliability. The inter-rater reliability was found to be 100 % for all the items after discussions.

The first part of the analysis consisted of looking at the choices that participants made. These choices are usually stated as A, B, C in the questionnaire. For instance in the first item the choices were not explicitly given, however the participant had three choices; the probability of Banu winning the lottery is higher, the probability of Mehmet winning the lottery is higher or both of the probabilities are equal to each other.

When analyzing the choices that participants made, the responses were coded as correct or incorrect. However there were some responses that were left blank. In all of the items if the participant has not stated a choice the answer was coded as no answer. On the other hand, there have been some participants who have stated that any of the choices is possible. Such responses were coded as no preference as they do not state a preference of a choice over another.

Each item required the participants to state the reasons behind their answers. The analysis of the responses given to the why question make up the second part of the analysis. In this part of the analysis main categories have been classical, frequency and subjective or different combinations of these three since the aim of the research is to analyze the different approaches in probability. However, in some cases the answers were not necessarily representative of these approaches or where possible these approaches were divided into subcategories. In such cases, different categories for reasonings have emerged. These different categories will be discussed in Chapter 5 in detail.

In some cases, some of the participants failed to state the reason behind their choice. Such responses have been coded as no reason. Participants who were in the no reason category provided responses that could not be interpreted as a reason. These responses were usually in the form that A has a higher probability because it is more likely. Such statements do not answer why A is more likely. In some cases, there was evidence of misunderstanding the question or the participant answering a question different than the one asked. Such responses have been coded as irrelevant.

The third part consists of the statistical analysis of the results found after the data was analyzed. The statistics are provided with cross tables that show the frequency of responses. Along with the cross tables, Cramer's phi (φ_c) was used in order to analyze the relationship between experience level and approach. Cramer's phi (φ_c) is a coefficient used when analyzing the relationship between two nominal variables (Agresti, 2002). Despite the fact that experience level is an ordinal variable, in other words, at a higher level of scaling, based on Richards' statement that any measure that is associated with lower levels of scaling can be used with the higher levels of scaling (1998). Cramer's phi (φ_c) was also used when analyzing the relationship between approach and success level since the variables in this relationship are also nominal and ordinal respectively.

When analyzing the relationship between experience level and success level Gamma was used since both of the variables are ordinal and Gamma is the correlation coefficient measuring the relationship between two ordinal variables (Agresti, 2002).

When $\varphi_c = 0$ it is said that there is not a relationship between the variables. When $\varphi_c < 0.09$ the relationship is said to be very weak. When $0.10 < \varphi_c < 0.19$ the relationship is said to be weak. When $0.20 < \varphi_c < 0.29$ the relationship is said to be moderate. When $\varphi_c > 0.30$ the relationship is said to be strong. The same intervals state similar strength for Gamma coefficient however, the range of Gamma coefficient is between -1 and 1 (Agresti, 2002). For example a Gamma coefficient of -0.40 means that there is a strong relationship between two variables but the relationship is inverse, as one variable increases, the other decreases.

The researcher would like to note that the gender pronouns used in this study such as he or she does not represent the gender of the participant. Data regarding gender was not collected in this study.

5. RESULTS

This chapter is divided into three sections: Results for each item, summary of analysis of each item and overall results for all items. In the first section, each item will be examined individually and information on how data was analyzed and results for each item will be provided in detail. In the second section the questionnaire will be examined as a whole. In this chapter data analysis will be discussed and statistical results will be given.

5.1. Results for Each Item

In this section results for each item will be provided. Categories for each item will be explained and examples of responses in certain categories will be provided. The categories of variables will be shown in Cross Tables which were used in calculating the statistical figures.

The statistical relationships will be provided in the order of the research questions. The order is as follows:

(i) Relationship between experience level and approach
(ii) Relationship between experience level and success level
(iii) Relationship between approach and success level

5.1.1. Item 1

First item of the instrument is as follows:

> In a lotto (Sayısal Loto) game, one has to choose 6 numbers from a total of 49. Banu has chosen 1, 2, 3, 4, 5, 6. Mehmet has chosen 39, 1, 17, 33, 8, 27. Who has a greater chance of winning? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

The responses were analyzed at two levels; success level and probability approach. For success level, responses that state only that Mehmet has a higher chance of winning were labeled as incorrect. There were 6 incorrect responses. The responses that state that they have an equal chance of winning were labeled as correct. There were 65 correct responses. 1 participant calculated the probability of winning, however he did not state whether the calculation is for Banu's or Mehmet's chance of winning. This response was coded as no preference. 17 responses were coded as no reason.

In terms of explanations to this item; the responses were divided into six categories. The first category was the classical approach in which the participants either stated the sample space or the set of interest or both. 42 responses were categorized as classical. Figure 5.1 is an example of a response of an in-service teacher with 12 years of experience which was categorized as classical. In this response it can been seen that the participant takes both the number of elements in the sample space, 49, and the set of interest, 6, into account.

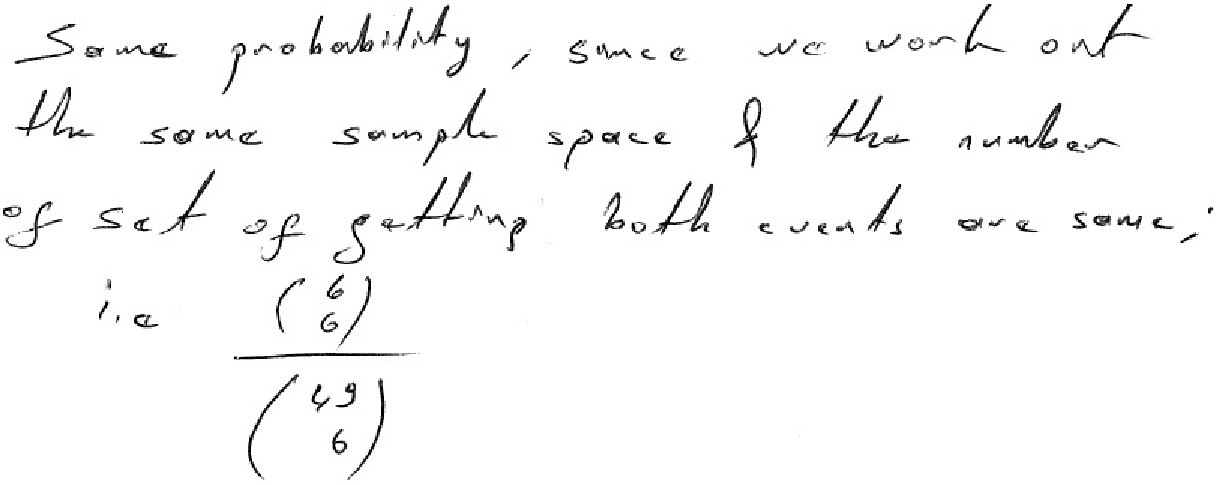

Figure 5.1. Classical Response to the First Item.

The second category for this item was the subjective category in which participants either stated that their intuition that Mehmet is more likely to win the lottery because Banu's sequence consists of consecutive numbers and such sequence does not win in real life or some stated that if they were to play lottery they would choose

Mehmet's sequence because it looks more random. There were 4 subjective responses.

The frequency or experimental approach was found only in 1 response of a participant who stated on the experiment of the selecting 6 consecutive integers. The participant has also stated on the theoretical probability as being equal. In other words, the participant is aware that the theoretical probabilities are equal, however he also states that experimentally if you repeat the trials it is unlikely to get 6 consecutive integers. This response was coded as classical and frequency since the participant focuses both on the theoretical and experimental approaches.

7 responses included both the classical and subjective approach such participants stated that theoretically Mehmet and Banu have an equal chance of winning and subjectively they chose Mehmet's sequence over Banu's for various different reasons such as their own experience of watching lottery results. Responses of these participants were labeled as correct since they showed some knowledge of the theoretical probability but simultaneously showed subjective reasoning as well. Figure 5.2 is an example of a response of a pre-service teacher which was categorized as classical and subjective.

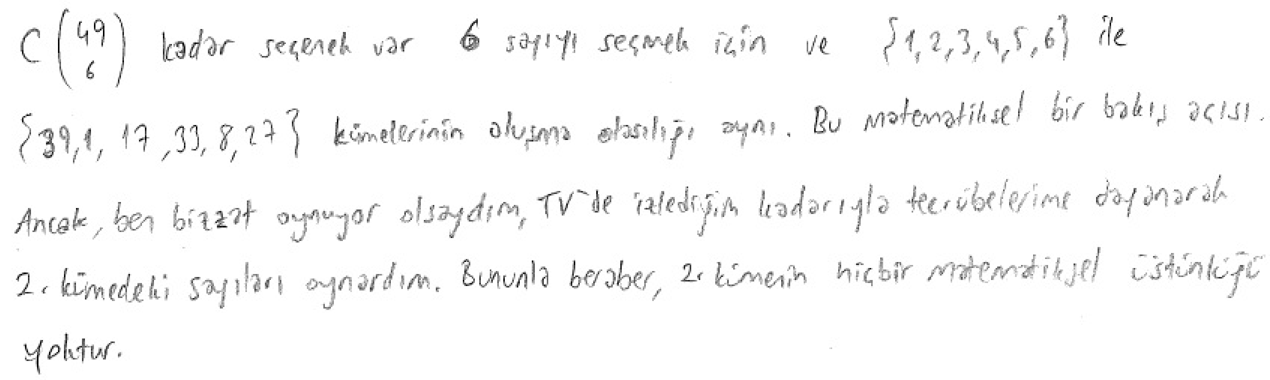
$C\binom{49}{6}$ kadar seçenek var 6 sayıyı seçmek için ve {1, 2, 3, 4, 5, 6} ile {39, 1, 17, 33, 8, 27} kümelerinin oluşma olasılığı aynı. Bu matematiksel bir bakış açısı. Ancak, ben bizzat oynuyor olsaydım, TV'de izlediğim kadarıyla tecrübelerime dayanarak 2. kümedeki sayıları oynardım. Bununla beraber, 2. kümenin hiçbir matematiksel üstünlüğü yoktur.

Figure 5.2. Classical and Subjective Approach to the First Item.

From the participant's response it can be seen that he differentiates the mathematical knowledge from his own behavior. He states that if he were to play he would have chosen Mehmet's sequence even though it is not mathematically superior.

There has been 1 participant who has stated that the chances are equal but intuitively it feels like consecutive numbers are less likely. She also stated that it also depends where the drawing takes place, the size of the ball etc.. This response was categorized as no reason and subjective. The first part of her response where she states that the probabilities are equal was categorized as no reason since she does not explain why the probabilities are equal and the second part where she comments based on her intuition was categorized as subjective thus leading to a new category of no reason and subjective.

The distribution of different approaches among different experience levels can be seen in Table B.1. For the relationship between experience level and approach level Cramer's phi (φ_c) was found to be 0.243 indicating that the relationship is moderate. However the approximate significance was found to be 0.583 which indicates that the relationship is not significant.

The distribution of different responses among different experience levels can be seen in Table B.2. It can been seen in the table that a participant at Experience Level 1 (less than 10 years of experience) has not stated a preference between Banu or Mehmet's chance of winning the lottery. This participant's data has been removed from the data set when examining the relationship between experience level and success level. For the relationship between experience level and success level Gamma was found to be 0.110 indicating that the relationship is weak.

The distribution of different responses among different approaches can be seen in Table B.3. The participant who has not stated a preference over Banu or Mehmet's chance of winning the lottery has been removed from the data set when examining the relationship between approach and success Level. For the relationship between approach and success level Cramer's phi (φ_c) was found to be 0.643 indicating that the relationship is strong and is statistically significant (0.000).

When the no response item is removed in Table B.3, it can be seen that the least successful approach has been the subjective approach. Given that the participant has used the subjective approach, the probability of a correct response was 0.25. All the combinations of approaches, classical, classical and frequency, classical and subjective had a success rate of 1. Participants who have failed to provide a reason had a success rate of 0.82. This rate can be considered high in terms of having overcome the misconception, however it is important to note that it is still lower than the success rate of the participants who were able to state a reason.

5.1.2. Item 2

The second item of the instrument is as follows:

> When tossing a coin, there are two possible outcomes: either heads or tails. Nurcan flipped a coin three times and in all cases heads came up. Nurcan intends to flip the coin again. What is the chance of getting heads the fourth time? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

The responses that stated that the probability of Nurcan getting heads the fourth time as $\frac{1}{2}$ was marked as correct. There were 64 correct responses. There were 8 responses of participants who failed to state the probability as $\frac{1}{2}$ thus were coded as incorrect.

One category in the analysis of responses to this question was classical. There were 64 participants who solely used the classical approach. The participants with a classical approach stated that there are two possible outcomes in the sample space: heads or tails. Some of the participants examined the situation by looking at the four flips as H, H, H, H and H, H, H, T and concluded that there are still 2 possible outcomes focusing on the last flip. There were 7 participants who showed evidence of using the classical approach but failed to calculate the theoretical probability of getting heads in the fourth flip. Some of such responses stated that P (HHHH) $=(\frac{1}{2})^4$. These participants failed to see that the P(HHHH)$=1.1.1.\frac{1}{2}$ since the first three heads had already happened and their probability is 1.

Any response that contained the word experimental probability or explained the situation with infinitely many or large number of flips were classified as a frequency approach. There was only 1 participant who solely relied on the frequency approach.

Any response that questioned the fairness of the coin after getting three heads consecutively was classified as a subjective approach because whether the coin is fair or not is not given in the question and these participants show a personal belief on the fairness. These participants have talked about the situation when the coin is unfair and stated that P(H)=1. Participants who have stated that they solved the question assuming the coin being fair were not classified as subjective. These participants have not commented on the situation when the coin is unfair. They have just solved the question assuming that the coin is fair.

There were no responses that were classified solely as subjective, however there were subjective responses along with classical. There were 4 responses that were classified as classical and subjective. Figure 5.3 shows an example of a response that was classified as classical and subjective since the participant is aware of the fact that P(H)=$\frac{1}{2}$ however he states that it feels like the fourth flip will be the same (heads).

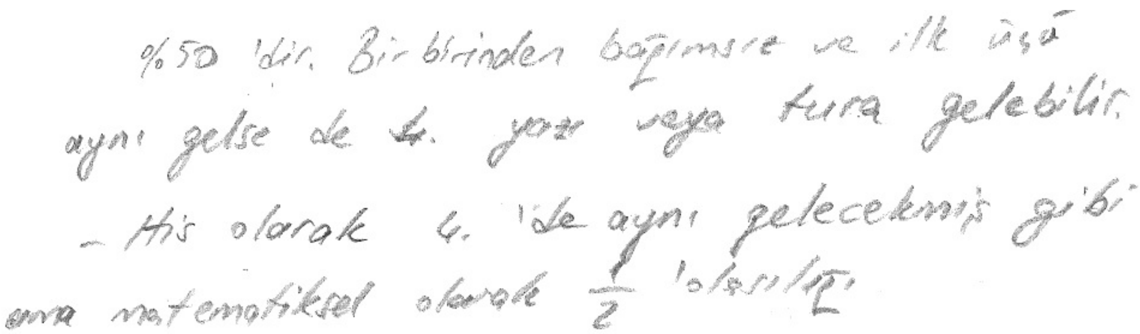

Figure 5.3. Classical and Subjective Approach to the Second Item.

There were two participants who has responded using all of the classical, frequency and subjective approaches. For instance, one of these participants has stated that the probability is always $\frac{1}{2}$ for heads and tails because there are two outcomes. This statement has been labeled as the classical approach since it contains the desired outcome and the sample space. She also stated that this is based on the fact that if we flip the coin 10 quintillion times, heads and tails both will show up around 5 quintillion times. This has been labeled as the frequency approach since it talks about a large number, in other words approaching infinitely many trials. She also stated that Nurcan's flipping the coin 3 times and getting heads and the fourth flip are independent events. She also stated that the coin might be unfair since not enough information is given in the question. She stated that if she knows that the probability of heads is increased to for example 60% heads and 40% tails then she would choose heads. This has been labeled as the subjective approach since the participant is putting her own belief into the solution that the coin might be unfair.

1 participant showed that the probability of heads is $\frac{1}{2}$ but have not stated any reason. This response has been coded in the no reason category.

The distribution of different approaches among different experience levels can be seen in Table B.4. For the relationship between experience level and approach Cramer's phi (φ_c) was found to be 0.198 at a 0.687 significance level indicating that the relationship is moderate but not significant.

The distribution of responses among different experience levels can be seen in Table B.5. For the relationship between experience level and success level Gamma was found to be -0.342 at a 0.275 significance level indicating that the relationship is negatively strong meaning that as experience level increases, success level decreases however this relationship is not statistically significant.

The distribution of responses among different approaches can be seen in Table B.6. For the relationship between approach and success level Cramer's phi (φ_c) was found to be 0.351 indicating that the relationship is a strong one, however the approximate significance 0.064 indicates that the relationship is not significant.

5.1.3. Item 3

The third item of the instrument is as follows:

> Above is an example of a Turkish lottery ticket (Milli Piyango) consisting of 6 digits. Turkish National Lottery Administration (Milli Piyango İdaresi) has a lottery drawing on 9th, 19th and 29th of every month. What is your chance of winning the lottery in 2013? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

This question was asked to assess whether participants include their own behavior as part of their probability calculation. The question directly asks the participant's own chance of winning the lottery rather than the theoretical probability for the whole year. The answer to this question varied between classical approaches which include numerical calculations some including the total number of tickets printed, number of drawings in a year. There were 62 participants who used the classical approach when answering this question.

Subjective approaches on the other hand include participant's own habit of buying tickets and/or their thought on theoretical probability being different than real life. 1 participant solely answered this question using the subjective approach focusing on the fact of him buying a ticket. 6 participants used the subjective approach along with the classical.

There have been 3 irrelevant responses to this question. These responses compared the chance of winning the lottery in 2013 to other years and stated that the probability does not change over the years and has not provided any calculations or further information on their reasoning. These responses have been coded as irrelevant.

The distribution of different approaches among different experience levels can be seen in Table B.7. For the relationship between experience level and approach Cramer's phi (φ_c) was found to be 0.139 indicating that the relationship is a weak one.

5.1.4. Item 4

The fourth item of the instrument is as follows:

> Tayfun dreams of becoming a doctor. He likes to help people. When he was in high school, he volunteered for the Turkish Red Crescent (Kızılay) organization. He accomplished his studies with high performance and served in the army as a medical attendant. After ending his army service, Tayfun registered at the university. Which seems to you to be more likely? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.
> A) Tayfun is a student of the medical school.
> B) Tayfun is a student.

The correct answer to this question is the choice B. 45 responses with the choice B has been coded as correct and 13 responses with the choice A as incorrect. There have been 13 responses without a preference, stating that both of the choices could happen and 1 participant has not stated anything and has been coded as no answer.

When analyzing the reasoning behind the responses to item 4, more categories within the classical approach have emerged. In order to provide detailed subcategories, certain event, subset, outcome approach were used.

Frequency approach was not adopted in this question. 1 response has been coded as classical and was not put into any of the subcategories. One of the subcategories within the classical approach was the certain event. 18 responses were coded solely in this subcategory. These participants have made mathematical statements such as P(B)=1 or they have stated verbally about the fact that B is a certain event, B happening for sure or the fact that it is a given fact that Tayfun is a university student.

Another subcategory within the classical approach was called the subset. 6 participants have either stated using mathematical symbols $A \subset B$ or verbally that A is a subset of B. These responses have been coded in the subset category.

The last subcategory within the classical approach was named the outcome approach. In this subcategory participants have stated on the events either happening or not thus assigning the probability of $\frac{1}{2}$ to every event. In this case, the statements were in the form of Tayfun either is at medical school or not, the probability is $\frac{1}{2}$. The naming of this category was adopted from Konold (1989). In his study Konold states that people focus on outcomes of trials thus assign probabilities to events at a basic level of either happening or not. The responses in this item adopt Konold's outcome approach, however, Konold's model of outcome approach is broader and has more elements than assigning half and half probabilities.

There were 19 participants who have used only the subjective approach. The subjective responses have been verbal answers discussing the university entrance exam and Tayfun's interests. The subjective approaches have differed between correct and incorrect responses. For instance one participant stated that B is more likely because it is really difficult to get in to medical school and he may not have done well in the exam. Despite the fact the answer is right the reasoning behind it is not, based on the fact that B is a more general statement than A.

The categories described above are main categories, however there have been many responses which contained more than one reasoning thus they were coded as combinations of categories. 1 response was coded as certain event and outcome approach. 8 responses were coded as certain event and subjective. 1 responses was coded as certain event and subset. 2 were coded as subjective and outcome approach. 1 was coded as subset and outcome approach. 2 were coded as subset and subjective.

9 participants have not provided any reason in their responses. 1 response was coded as no preference.

The distribution of different approaches with subcategories among different experience levels can be seen in Table B.8. In order to analyze the relationship between experience level and approach the response that was classified as no preference has been removed from the data set. For the relationship between experience level and approach, Cramer's phi (φ_c) was found to be 0.485 at a 0.057 significance level indicating that the relationship is strong but not significant. The same relationship was analyzed by excluding the subcategories. In other words the subcategories were coded under the classical approach. Cramer's phi (φ_c) was found to be 0.229 indicating that the relationship is moderate however not statistically significant (0.284).

The distribution of responses among different experience levels can be seen in Table B.9. In order to analyze the relationship between experience level and responses the response that was classified as no preference has been removed from the data set. For the relationship between experience level and response Gamma was found to be 0.224 at a 0.408 significance level indicating that the relationship is moderate however not statistically significant.

The distribution of responses among different approaches can be seen in Table B.10. In order to analyze the relationship between approaches and success level the responses that were classified as no preference or no answer have been removed from the data set. For the relationship between approach and success level Cramer's phi (φ_c) was found to be 0.866 at a 0.000 significance level indicating that the relationship is strong and significant. From Table B.10, it can be seen that the least successful categories were the outcome approach and subjective and outcome approach with a success rate of 0. The second least successful approach was the subjective approach with a success rate of 0.17 followed by the no reason category with a success rate of 0.85. All the other categories had a success rate of 1.

Further statistical analysis was conducted considering emerging the subcategories as a single classical approach category. Table B.11 represents the cross tabulation of this relationship. The relationship is once again found to be strong with a Cramer's phi (φ_c) of 0.752 at 0.000 significance level. From Table B.11, it can be seen that classical approach had the highest success rate 0.97 followed by classical and subjective with a success rate of 0.31. Participants who have not provided a reason had a success rate of 0.21. Participants who have solely used the subjective approach had a success rate of 0.07.

5.1.5. Item 5

The fifth item of the instrument is as follows:

> K: The likelihood of getting heads at least twice when tossing three coins
> L: The likelihood of getting heads at least 200 times out of 300 times
> A) K is smaller than L
> B) K is equal to L
> C) K is bigger than L
> Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

The correct response to this question was choice C. There were 39 correct responses. 7 participants have not chosen an answer.

The three approaches continued to come up in the fifth item as well. However classical approach has been the most widely used. This approach has been divided into three subcategories as ratio, calculation and induction. 2 of the classical responses were not in any of these subcategories.

10 participants responded to the question focusing on the ratio $\frac{2}{3}$ either verbally or numerically. This ratio was sometimes used to set up the proportion $\frac{2}{3} = \frac{200}{300}$. These responses have been coded as ratio.

Some of the participants have used the classical approach to calculate the probability for the event K and L. Since the number of coin flips is an high number in the event L, they chose to find a pattern as the number of coin flips increase. These responses usually contained multiple periods as a symbol to show the pattern. Such responses were categorized as calculation. There were 27 participants who have only used calculation. Figure 5.4 is an example of a response of a teacher with 12 years of experience which was categorized as calculation. The periods can be seen in the figure.

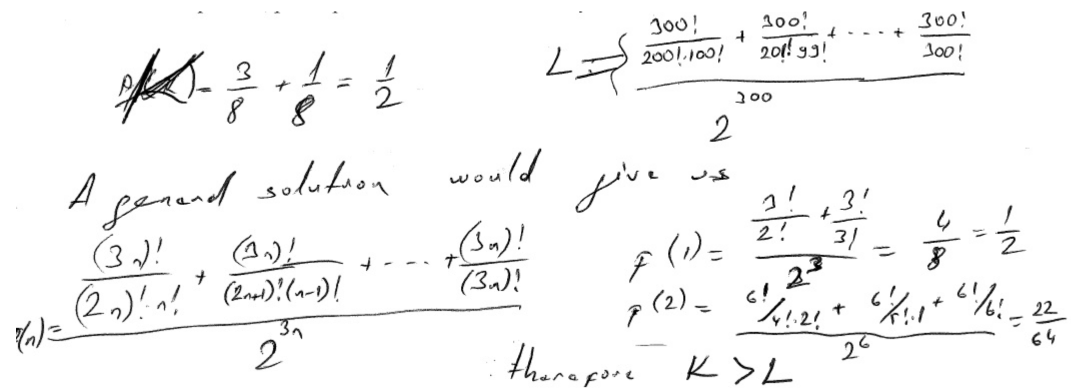

Figure 5.4. Response that is Coded as Calculation.

Figure 5.5 is an example of a response of a teacher with 29 years of experience which was categorized as induction. In this category, rather than calculating the probability of getting at least 200 hundred heads out of 300 coin flips, participants have created a new sample space that is proportional to 300. These participants tended to create a new sample space consisting of 6 -or some other proportional number- coin flips and calculated the probability of getting at least 4 heads -or some other proportional number-. 12 responses were coded as induction.

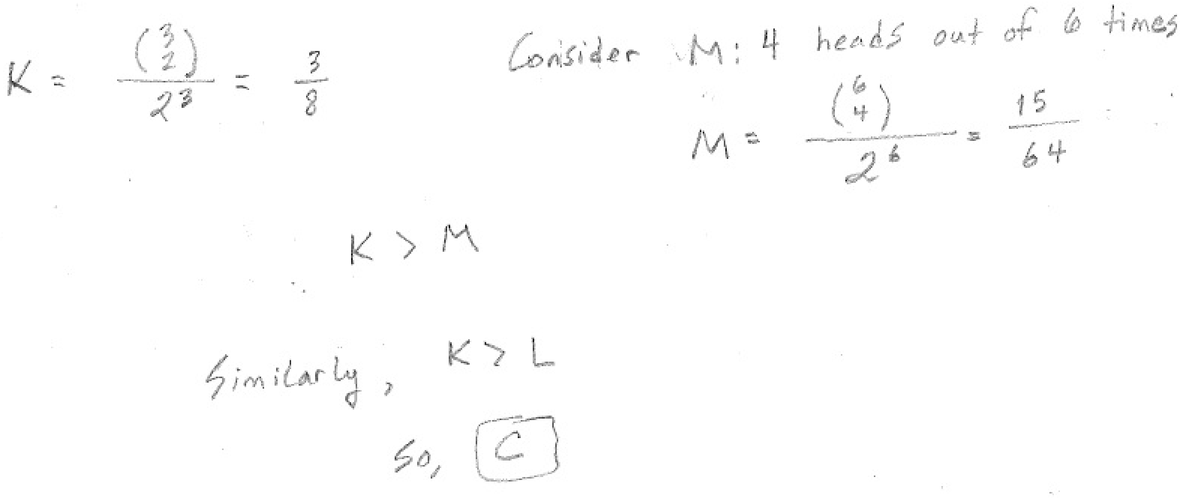

Figure 5.5. Response that is Coded as Induction.

Participants with the frequency approach have stated similar responses that L is less likely because more trials would make the probability equal to $\frac{1}{2}$. They chose K since the number of trials is less. 8 participants relied only on the frequency approach.

The responses in the subjective approach included participant's feelings or that L looking more difficult than K. There were not any participants who relied solely on the subjective approach.

Similar to Item 4, responses in this item also contained different combinations of the reasonings mentioned above. 1 participant adopted both the frequency and subjective approaches. 3 participants used calculation and adopted the frequency approach. 4 participants used calculation and induction. 2 participants used ratio and frequency. 1 participant used ratio and calculation. 1 participant used ratio and relied on the subjective approach.

The distribution of different approaches among different experience levels can be seen in Table B.12. In order to analyze the relationship between experience level and approach the responses that were classified as no answer has been removed from the data set. Cramer's phi (φ_c) was found to be 0.393 indicating that the relationship is strong however it is not significant (0.447). When induction, calculation and ratio are listed under classical approach Cramer's phi (φ_c) was found to be 0.393 indicating that the relationship is still strong but still not significant (0.166).

The distribution of responses among different experience levels can be seen in Table B.13. In order to analyze the relationship between experience level and responses the response that was classified as no answer has been removed from the data set. For the relationship between experience level and response Gamma was found to be -0.042 indicating that the relationship is weak.

The distribution of responses among different approaches can be seen in Table B.14. In order to analyze the relationship between approach and success level the response that was classified as no answer has been removed from the data set.

Cramer's phi (φ_c) of 0.558 indicated that there is a strong relationship and it is significant (0.000). The highest success rates as 1 was found for the following categories and subcategories: frequency and subjective, no reason, induction and frequency. The second highest success rate as 0.83 belonged to the inductive responses, followed by solely frequency users and induction and calculation group with a success rate of 0.75. Despite the trend of success of those who have adopted the frequency approach along with others, those who have combined it with ratio were not as successful and had a success rate of 0. Ratio in general was not a successful category. Those who have solely used ratio had a success rate of 0.1 and those who have brought it together with the subjective had a success rate of 0.

When calculation, induction and ratio responses were categorized as classical, Cramer's phi (φ_c) was found to be 0.243 indicating that there is a moderate relationship but the relationship is not significant (0.574).

5.1.6. Item 6

The sixth item of the instrument is as follows:

> K The number of possibilities when choosing 2 members from among 10 candidates L The number of possibilities when choosing 8 members from among 10 candidates
>
> A) K is smaller than L
> B) K is equal to L
> C) K is bigger than L
>
> Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

51 participants chose the correct response B whereas 12 participants chose A and 6 participants chose C. 3 participants have not provided an answer. 1 participant has not provided a reason.

Compared to other questions, this item does not ask a probability of an event but rather the size of a sample space. Participants have showed different ways of calculating the sample space and some of these solutions are related to the approaches.

The solutions were divided into five categories. The two of the categories were combinations and permutations. In the combinations category participants have stated that C(10,2)=C(10,8). Half of the participants have solely used combinations when calculating the number of possibilities.

In the permutations category, participants have focused on selecting two specific people so when selecting two people there are 10 choices for the first person and 9 choices for the second one, so there would be 90 ways of choosing two people. Similarly there are 10.9.8.7.6.5.4.3 ways of choosing 8 people among 10 people according to their reasoning. 5 participants have used permutations. Figure 5.6 shows a response of a teacher with 4 years experience that was categorized as permutations.

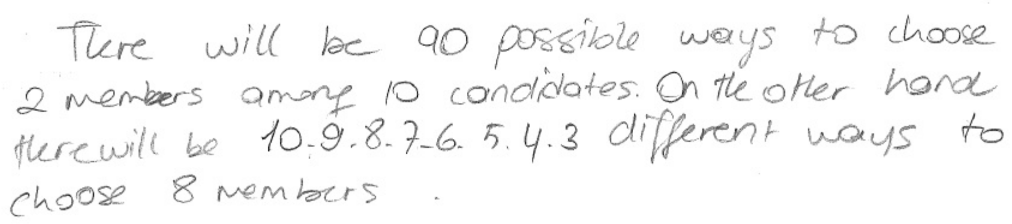
There will be 90 possible ways to choose 2 members among 10 candidates. On the other hand there will be 10.9.8.7.6.5.4.3 different ways to choose 8 members .

Figure 5.6. Response that is Coded as Permutations.

16 participants have stated verbally that there is no difference between choosing 2 people and 8 people since when choosing 2 people the act of not choosing the remaining 8 is an act of choosing the 8 people. This category was labeled as not choosing. 14 participants have used the not choosing reasoning along with the classical and 2 participants have used not choosing along with the classical.

All of the combinations, permutations and not choosing are different ways of counting the number of possibilities and are not directly related to the definitions of probability.

Another category was the ratio category in which participants compared $\frac{2}{10}$ and $\frac{8}{10}$. 12 people used this reasoning. Despite the fact that the question is not asking for a probability calculation, these participants were comparing a probability of events with set sizes of 2 and 8 in a sample space of size 10. While making this calculation they use the classical approach.

Another verbal category, other than not choosing category, that did not require any calculations was the quantity category. Responses that fit in this category state that since 2 people is fewer in number it would be easier to select 2 people. In fact this kind of reasoning is closely related to misconception. As stated in the definition of the availability misconception, people tend to focus on what is more available to them. In this case, 2 people are more available than 8 people. 2 participants used the quantity reasoning. 1 of them used it along with the combinations. Figure 5.7 shows part of a response that was categorized as quantity.

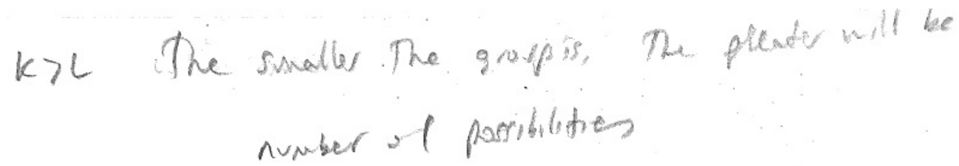

Figure 5.7. Response that is Coded as Quantity.

The distribution of different approaches among different experience levels can be seen in Table B.15. In order to analyze the relationship between experience level and approach Cramer's phi (φ_c) was calculated and found to be 0.383 and at a significance level of 0.173 indicating that the relationship is strong but not significant. It should be noted that for this item, this analysis can not be done using the three approaches as it was not directly asking the probability of an event.

The distribution of responses among different experience levels can be seen in Table B.16. In order to analyze the relationship between experience level and responses the responses that were classified as no answer or no preference have been removed from the data set. For the relationship between experience level and response Gamma was found to be 0.333 indicating a strong relationship however the approximate significance of 0.136 indicates that the relationship is not significant.

The distribution of responses among different approaches can be seen in Table B.17. In order to analyze the relationship between approach and success level the responses that were classified as no answer has been removed from the data set. Cramer's phi (φ_c) was found to be 0.900 and at a significance level of 0.000 indicating that the relationship is strong and significant. However, it is important to state that these approaches do not necessarily represent the probability definitions and includes approaches to a problem rather than probability theory. Nonetheless, the question was not eliminated from the questionnaire since the data regarding success level gathered from this item will be used in calculating total success.

From Table B.17 it can be seen that those who were in permutations, quantity or ratio category had a success rate of 0. Those who were in not choosing, not choosing and combinations categories had a success rate of 1. Those who have solely used combinations had a success rate of 0.97.

5.1.7. Item 7

The seventh item of the instrument is as follows:

> Suppose one rolls two dice simultaneously. Which of the following has a greater chance of happening? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.
> A) Getting the pair 5-6
> B) Getting the pair 6-6
> C) Both have the same chance.

In this item choice A was the correct response. There have been 3 responses without a selection of the choices. These responses have been coded as no answer. 38 of the participants have answered this item correctly choosing the choice A. Only 1 participant chose B and 30 participants chose C.

Participants who have chosen A were the ones who have stated that getting 6-5 is as same as 5-6 or they have stated the numerical calculation of P(5-6) as $\frac{1}{18}$and P(6-6) as $\frac{1}{36}$. On the other hand, the ones who have chosen C tend to state that $\frac{1}{36}x\frac{1}{36} = \frac{1}{36}x\frac{1}{36}$ or state that probability of getting a 5 is $\frac{1}{6}$ and the probability of getting a 6 is also $\frac{1}{18}$ so they have to be equal. All of these responses were categorized as classical. The difference between those who chose A and C was not the approach they use but the sample space that they use. There were 58 participants who have solely used the classical approach.

There have been some participants who have focused on whether the dice look the same or not since any difference in appearance of the dice would change the probability. These responses were categorized as different dice. There were also participants who have stated that the order of the dice might make a difference depending on what the question is asking or they have directly asked whether the order makes a difference in their responses. These responses were also categorized as different dice since the order would matter in a situation where the dice are different. There were 6 participants who have focused on the difference of the dice.

There have been some subjective responses stating that the die might be unfair and getting 6-6 might be more likely or getting 6-6 more unlikely in real life but theoretically both of them are equally likely. The emphasis on real life experiences are coded as subjective. There were 3 participants who have used the subjective approach along with the classical.

Responses without any reason have been coded as no reason. There were 6 participants who have not provided a reason.

The distribution of different approaches among different experience levels can be seen in Table B.18. From the table it can be seen that different dice reasoning was always used along with a probability approach. When analyzing the relationship, different dice reasoning was eliminated from the data set. For example a response that was categorized as classical and different dice was considered to be classical only, since the focus is on the probability approach rather than scenario of the question. Cramer's phi (φ_c) was calculated and found to be 0.052 indicating that the relationship is weak.

The distribution of responses among different experience levels can be seen in Table B.19. In order to analyze the relationship between experience level and responses, the responses that was classified as no answer have been removed from the data set. For the relationship between experience level and success level Gamma was found to be 0.246 indicating that the relationship is moderate but not significant (0.212).

The distribution of responses among different approaches can be seen in Table B.20. In order to analyze the relationship between approach and success level, the responses that were classified as no answer has been removed from the data set. Cramer's phi (φ_c) was found to be 0.227 and at a significance level of 0.129 indicating that the relationship is moderate but not significant.

5.1.8. Item 8

The eighth item of the instrument is as follows:

> Emre and Nihan each receive a box containing two white marbles and two black marbles.
>
> i) Emre extracts a marble from his box and finds out that it is a white one. Without replacing the first marble, he extracts a second marble. Is the likelihood that this second marble is also white smaller than, equal to, or greater than the likelihood that it is a black marble? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.
>
> ii) Nihan extracts a marble from her box and puts it aside without looking at it. She then extracts a second marble and sees that it is white. Is the likelihood that the first marble she extracted is white smaller than, equal to, or greater than the likelihood that it is black? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

The two parts to this item has been analyzed separately. First, the analysis of both sections will be explained in detail and later the two parts will be combined in order to measure success level. In both parts classical approach was widely used. No other approach was adopted in the first part and in the second part subjective approach was the only other approach adopted and was adopted by only 1 participant.

This item can not be examined in terms of different approaches to probability theory since almost all of the responses contain the classical approach. However different categories within the classical approach will be analyzed. For example both the conditional probability answers and answers that focused on both draws are actually classical responses since they indicate the desired outcome and sample space in every condition. These subcategories will be explained in detail.

<u>5.1.8.1. Item 8i.</u> In the first part, responses that have stated that the probability of getting a white marble is smaller than a black marble have been coded as correct. There were 65 correct responses.

In this part, there have been participants who have focused solely on the second draw that have been coded as second draw. Among these responses the participants use the classical approach and state that there are 2 black marbles and 1 white marble in the second draw and $P(B) > P(W)$. Figure 5.8 shows a response of a pre-service teacher that illustrates the sample space of the second draw. There were 56 participants in total who have solely focused on the second draw in the item.

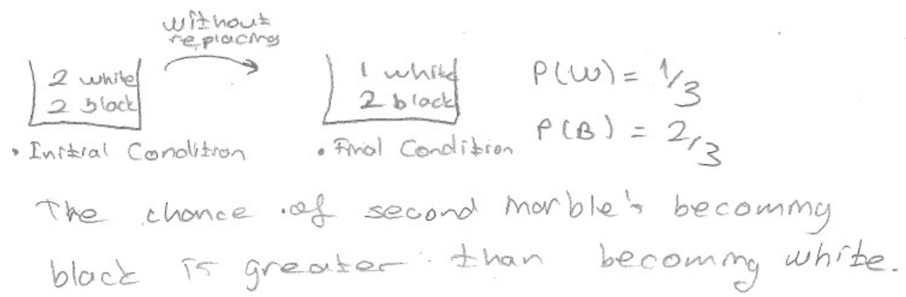

Figure 5.8. Response that is Coded as Second Draw.

The responses that focus on the first and second draw have been coded as two draws. Responses fitting into this category state the probability of white on first draw and white on the second draw and compare it with white on first draw and black on second draw. Similarly some responses contain the whole sample space of possible outcomes such as { (black, black), (black,white), (white,white), (white,black)}.

Figure 5.9 shows a response of a teacher with 10 years experience that calculates the probabilities for two draws. There have been 9 participants who have used the two draws reasoning.

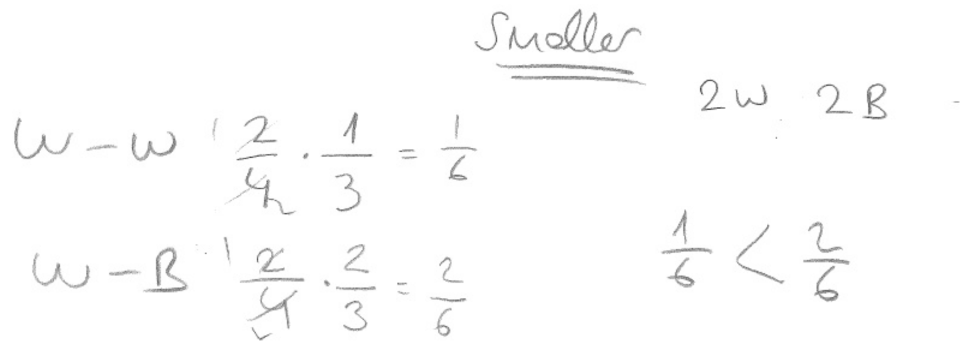

Figure 5.9. Response that is Coded as Two Draws.

Another category in the analysis of the first part of the item was the conditional probability. Responses with conditional probability trees or using the formula $P(A\backslash B) = \frac{P(A \cap B}{P(A)}$ were coded as conditional probability. There were 3 participants who have used conditional probability, 1 of these participants used it along with the second draw reasoning.

There have been two responses that compared the probability of drawing a white marble on the first draw and the probability of drawing a white marble on the second draw. These responses have been coded as irrelevant. There were 2 irrelevant responses and 2 responses that did not provide a reason.

The distribution of different approaches among different experience levels can be seen in Table B.21. In order to analyze the relationship between experience level and approach Cramer's phi (φ_c) was calculated and found to be 0.290 and at a significance level of 0.277 indicating that the relationship is moderate but not significant.

The distribution of responses among different experience levels can be seen in Table B.22. In order to analyze the relationship between experience level and success level the responses that were classified as no answer and irrelevant have been removed from the data set. For the relationship between experience level and success level Gamma was found to be 0.185 indicating that the relationship is weak.

The distribution of responses among different approaches can be seen in Table B.23. In order to analyze the relationship between approach and success level the responses that were classified as no answer and irrelevant have been removed from the data set. Cramer's phi (φ_c) was found to be 0.550 and at a significance level of 0.000 indicating that the relationship is significant.

From Table B.23 it can be seen that all the categories except for two draws category had a success rate of 1. For two draws category, success rate was found to be 0.67.

5.1.8.2. Item 8ii. Responses that stated that probability of drawing a white marble is smaller than black marble have been coded as correct. There were 30 correct responses.

Similar to the first part, some participants have solely focused on the first draw. These responses have been coded as first draw. There were 5 responses that have focused on the first draw. These responses also include the ones that state that the situation is the same as the first part, the sample space of the first draw consisting of 2 black marbles and 1 white marble. Figure 5.10 shows an example of a response of a pre-service teacher that was coded as first draw.

For the second part some participants have focused on both of the draws and these responses have also been coded as two draws. These responses examine the situation of getting black on first draw and white on second draw and white on first draw and white on second draw. Figure 5.11 shows an example of a response of a teacher with 15 years of experience that was coded as two draws. There were 20 participants who have focused on both of the draws.

Similar to the first part, for the second part some participants used the conditional

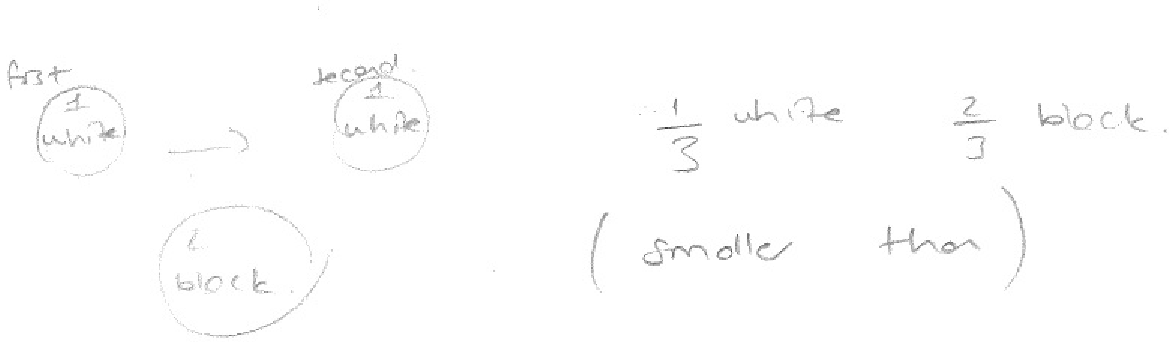

Figure 5.10. Response that is Coded as First Draw.

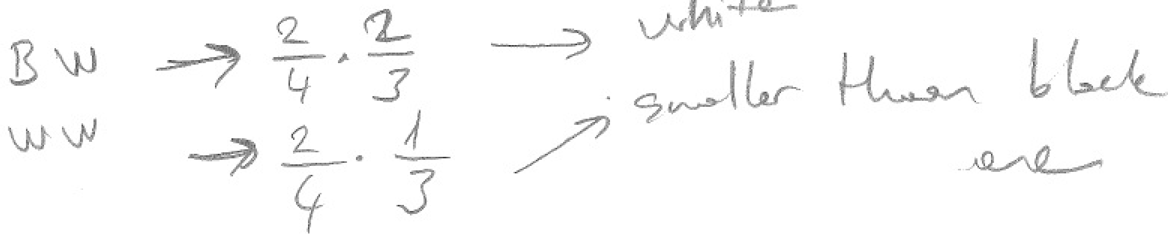

Figure 5.11. Response that is Coded as Two Draws.

probability trees or the formula. These responses have been coded as conditional probability. There were 10 participants who have used the conditional probability.

Lastly, in this part some participants have stated that the probability of getting black is equal to getting white on the first draw since there are 2 black and 2 white marbles at the beginning. Their reasoning statements included that the first draw is independent of the second draw. These responses were coded as independent draw. There were 22 participants who used the independent reasoning. 1 of these participants used it along with the subjective approach.

There were 13 participants who have misunderstood the question as Nihan putting the first marble back into the bag. Any response stating that Nihan putting the marble back into the bag has been coded as irrelevant.There were 2 responses without a reason.

There has been one participant who stated that he answered based on his intuition. This response has been coded as subjective.

The distribution of different approaches among different experience levels can be seen in Table B.24. In order to analyze the relationship between experience level and approach Cramer's phi (φ_c) was calculated and found to be 0.339 and at a significance level of 0.168 indicating that the relationship is strong but not significant.

The distribution of responses among different experience levels can be seen in Table B.25. In order to analyze the relationship between experience level and success level the responses that were classified as no answer and irrelevant have been removed from the data set. For the relationship between experience level and success level Gamma was found to be 0.142 indicating that the relationship is weak.

The distribution of responses among different approaches can be seen in Table B.26. In order to analyze the relationship between approach and success level the responses that were classified as no answer and irrelevant have been removed from the data set. Cramer's phi (φ_c) was found to be 0.869 and at a significance level of 0.000

indicating that the relationship is strong and significant.

From Table B.26 it can be seen that the least successful categories were independent, independent and subjective with success rates of 0. The most successful categories were no reason and first draw with success rates of 1. Conditional probability category had a success rate of 0.89 and two draws category had a success rate of 0.84.

As stated in Section 2.2, the misconception occurs when a person thinks that probability of white is bigger in the first situation but does not see the same in the second situation. In order to analyze this the responses in both parts were brought together.

When analyzing the relationship between experience level and approach, the second draw category in the first part and the first draw category in the second part were merged into a new category called single draw. All the responses that were irrelevant were removed from the data set. When analyzed Cramer's phi (φ_c) was found to be 0.476 at a 0.171 significance level indicating that the relationship is strong but not significant.

When analyzing the relationship between experience and success level, a participant was defined to be successful if he could answer both parts of the item correctly. Responses that did not provide an answer have been removed from the data set. When analyzed Gamma was found to be 0.086 indicating that the relationship is weak.

When analyzing the relationship between approach and success level Cramer's phi (φ_c) was found to be 0.071 indicating that the relationship is weak.

5.1.9. Item 9

The ninth item of the instrument is as follows:

> If you are single what is the probability that you will get married in the next five years? If you are married what is the probability that you will get divorced in the next five years? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

Since the question was assessing whether the participants provide subjective reasons or not most of the responses were verbal explanations. One of the categories for analysis of this item consists of responses that state that: The answer can not be computed; it is not a mathematical question; or it is not related to probability theory. These responses have been coded as not mathematical. 13 responses. were coded as not mathematical.

There were 16 answers that were irrelevant to the question such as answers comparing probability of marriage and divorce. This might be due to a misunderstanding of the question. These answers have been coded as irrelevant.

8 participants stated that they need more data in order to answer this question. They were seeking for marriage, divorce, age etc. statistics in order to solve the question. These responses have been coded as more data.

9 participants have stated a numeric answer along with an explanation. For example a participant was stating that his chances of getting a divorce is very unlikely and it is 10%. These responses have been coded as numerical.

30 of the responses that contain numeric calculations have specifically used the probability $\frac{1}{2}$. They have also stated that events either happen or not thus probability is $\frac{1}{2}$. These responses have been coded as outcome approach.

11 participants stated that the answer depends on other factors such as culture, family, age etc. These responses have been coded as other factors.

Responses that contained any information on participant's own experiences, behavior and perceptions have been coded as subjective. 12 responses were found to be subjective.

1 participant has not answered the question.

The distribution of different approaches among different experience levels can be seen in Table B.27. The table contains a long list of responses to this item as most of the responses were verbal and different modes of reasonings were held by the participants.

Irrelevant responses have been removed from the data when analyzing. For the relationship between experience level and approach Cramer's phi (φ_c) was found to be 0.534 at a 0.469 significance level indicating that the relationship is strong but not significant. It is impossible to find a significant relationship in the detailed data provided in Table B.27 as the sample size is relatively small.

The purpose of this item was to identify those who adopt subjective approach and to provide a basis for subjective probability by posing a scenario that is less numerical than Item 3. Therefore, participants were divided into two wider categories as subjective and not subjective. It is impossible to state that those whose responses are coded as not subjective, do not adopt subjective probability in general. This category is only limited to this item and those who are categorized as not subjective may display knowledge of subjective probability in other fields of life.

Table B.28 displays the distribution of subjective and not subjective groups among different experience level. For this relationship Cramer's phi (φ_c) was found to be 0.269 indicating a moderate but an insignificant (0.80) relationship.

5.1.10. Item 10

The tenth item of the instrument is as follows:

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? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.
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.

The correct response to this item was B. There were 34 correct and 33 incorrect responses. 1 participant failed to provide an answer. 4 participants have not stated a preference between the two hospitals.

This item is similar to question 5. Both of them measure the insensitivity sample size misconception. 4 participants have recognized the fact that both of these questions were similar and stated it in their responses. These responses have been coded as question 5 for those who have stated the similarity. All of these participants had answered item 5 with using the classical approach.

Similar to question 5, some have tried finding a pattern for 50 births and calculated or estimated the probability using the classical approach. These responses were coded as calculation. There were 11 participants who have calculated the probabilities.

3 participants have focused on what constitutes 80% in both hospitals and calculated that in Hospital A 40 births, and in hospital B 8 births make up the 80%in both hospitals. They have also stated that $\frac{40}{50} = \frac{8}{10}$. These responses have been coded as ratio.

6 participants have chosen Hospital A because it has higher number of births. These responses have been categorized as higher number. This kind of thinking is contradicts with the law of large numbers. This item is assessing insensitivity to sample size, and correct responses should focus on the fact that as sample size increases experimental probability should reach to theoretical probability. In hospital A the sample size is larger so it is more likely that half of the newborn will be girls. Participants

who focus on higher number of births tend to believe that higher the number of births, higher the probability of girls would be.

Some participants focused on the fact that the probability of getting a girl is always a half in both hospitals. These responses have been coded as the outcome approach. Some of these participants state that because the probability is a half there is no way that the hospitals would reach the 80% probability.

There was only 1 participant who used the subjective approach who stated that he feels like the two situations seem equal.

There were 10 participants who have used an inductive reasoning in this item. These participants created a new sample space with fewer births and calculated the probability using the classical approach. 2 of the induction responses were provided along with ratio.

The distribution of different approaches among different experience levels can be seen in Table B.29. Cramer's phi (φ_c) was found to be 0.405 indicating that the relationship but not significant (0.365).

For further analysis some of the subcategories of classical approach was coded back as classical. Question 5 category consisting of those who have stated that Item 10 is same as Item 5 have been coded as Classical as all of these participants had used the classical approach in Item 5. Those who have said that Item 10 is same as Item 5 and used the frequency approach when solving Item 10 have been coded as Classical and Frequency. Those who have used ratio, induction and calculation have also been coded as classical. Higher Number and Outcome Approach categories were eliminated from the data set as it is not related to any of the approaches. In this case, Cramer's phi (φ_c) was found to be 0.272 at a significance level of 0.514 indicating that the relationship is moderate but not significant.

The distribution of responses among different experience levels can be seen in Table B.30. In order to analyze the relationship between experience level and responses, the responses that were classified as no answer and no preference have been removed from the data set. For the relationship between experience level and response Gamma was found to be -0.044 indicating that the relationship is weak.

The distribution of responses among different approaches can be seen in Table B.31. In order to analyze the relationship between approach and success level, the responses that were classified as no answer and no preference have been removed from the data set. Cramer's phi (φ_c) of .860 indicated that there is a strong relationship and it is significant (0.000).

When calculation, induction, Question 5 and ratio responses were categorized as classical, Cramer's phi (φ_c) was found to be 0.553 indicating that there is a strong relationship and the relationship is (0.002). It can been seen from Table B.32 that when the approaches are listed from the highest success rate to lowest the list is as follows: classical and frequency (1.00), frequency (0.94), classical (0.46), subjective (0).

5.2. Summary of Analyses of Each Item

To provide a summary of all the relationships tested statistically Table 5.1 has been provided. The cells written in bold represent a strong and significant relationship. It is clear that there have not been a significant relationship between experience level and approach or experience level and success in any of the items, however in Items 1, 4, 6, 8i and 8ii the relationship between approach and success level was found to be strong and significant. It should be noted, though, for items 6 and 8 this relationship is not attributed to the definitions of probability but rather to the solution method.

Table 5.2 provides a summary of approaches used in all items along with the no reason and irrelevant categories. From this table it can be seen that classical approaches has been adopted in all of the items. Frequency approach was the least used approach among all others. It was used in Items 1, 2, 5 and 10. On the other hand subjective

Table 5.1. Strength of Relationship Between Variables.

	Experience Level and Approach	Experience Level and Success Level	Approach and Success Level
Item 1	moderate	weak	**strong**
Item 2	moderate	negatively strong	strong
Item 3	weak	N/A	N/A
Item 4	moderate	moderate	**strong**
Item 5	strong	weak	moderate
Item 6	strong	strong	**strong**
Item 7	weak	moderate	moderate
Item 8i	moderate	weak	**strong**
Item 8ii	strong	weak	**strong**
Item 8 Total	strong	weak	weak
Item 9	strong	N/A	N/A
Item 10	moderate	weak	**strong**

Table 5.2. Different Approaches Used in All Items.

	Classical	Frequency	Subjective	No Reason	Irrelevant
Item 1	✓	✓	✓	✓	
Item 2	✓	✓	✓	✓	
Item 3	✓		✓		✓
Item 4	✓		✓	✓	
Item 5	✓	✓	✓	✓	
Item 6	N/A				
Item 7	✓		✓	✓	
Item 8i	✓			✓	✓
Item 8ii	✓		✓	✓	✓
Item 9	✓		✓		✓
Item 10	✓	✓	✓	✓	

approach was more widely used than frequency approach. It was used in all items except for 3, 6 and 8i.

The responses which were coded as no reason were found in all items except for Items 3 and 9, in other words the items examining subjective probability. It can be said that there were participants who were not able to provide a reason for their answer in questions assessing misconceptions. Contrary to the no reason category, irrelevant responses were found in Items 3 and 9. It can be said that there were participants who have answered irrelevantly in items assessing subjective probability. Irrelevant responses were also found in both parts of Item 8.

5.3. Results for all items

In order to analyze results for all items, first, the approaches used in different item were combined. For instance if a participant who responded to some items with the classical approach and at least one item with the frequency approach then the participant has been labeled as classical and frequency. The participant can mainly use the classical approach and may only use the frequency approach once but there is still evidence of the participant having an idea of the frequency approach even if not using it in every case.

All of the participants used the classical approach. 18 participants have only used the classical approach. There have been 31 participants who have used the frequency approach. 10 of them used the frequency approach along with the classical approach. 21 of them used it along with the classical and subjective approaches. 44 participants have used the subjective approach. 23 of them used it along with the classical approaches. 21 of them used it along with the classical and frequency approaches.

Item 5 and Item 10 measured the same misconception: Insensitivity to sample size. When calculating a total score for the success level in all items, those who have given a correct response both in item 5 and item 10 have been considered successful. Those who were able to give a correct response in either one of the items were not considered successful. It is impossible to state that these participants have overcome their misconception since they can not show evidence of their knowledge in both cases.

The distribution of approaches among different experience levels can be seen in Table B.33. For the relationship between experience level and success level, Cramer's phi (φ_c) was found to be 0.144 indicating that the relationship is weak.

The distribution of responses among different experience levels can be seen in Table B.34. For the relationship between experience level and success level, Gamma was found to be 0.182 indicating that the relationship is weak.

The distribution of responses among different approaches can be seen in Table B.35. Cramer's phi (φ_c) was found to be 0.360 and at a significance level of 0.043 indicating that the relationship is strong and significant.

None of the participants who have answered all the items, related to misconceptions, correctly, adopted the three approaches at the same time. Given that the participant has answered all the items correctly, it is more likely that she has adopted classical and frequency approach.

Highest percentage (33.33%) of participants who have adopted the classical approach have answered 5 items correctly Highest percentage of participants who have adopted the classical and subjective approaches (30.43%) have answered 3 items correctly. Highest percentage of both the classical and frequency approach adopters (40.00%) and those who have adopted all the three approaches (33.33%) have answered 6 items correctly.

6. CONCLUSION AND DISCUSSION

In this chapter, firstly, answers to research questions will be concluded based on the results. Secondly implications will be discussed given the conclusions of this study and the studies provided in the literature. Lastly, limitations of the study will be discussed and suggestions for further research will be provided given the limitations.

6.1. Research Question 1

Research Question 1 was as follows:

> What types of probability approaches do pre-service and in-service teachers adopt while reasoning on different probability questions that are affiliated with misconceptions?

The results show that classical approach was the most widely used approach and frequency approach was the least used approach. Items 3 and 9 assessing subjective probability were the only two questions where no reason category was not found, also these two items were among the three items where irrelevant responses were found. Possible explanations to these findings will be discussed in this section.

In seeking an answer to the first research question it was revealed that teachers use all of the three approaches while reasoning on different probability questions that are affiliated with misconceptions. All the teachers have used the classical approach throughout the study and classical approach was among the approaches used in every item. It was the most widely used approach. It is also the mostly taught approach in school curricula. In his paper defining probability terminologies, Chernoff calls classical probability as “the school probability” (2010) which justifies the findings of this study.

Frequency approach was the least used approach in this study. One reason might be due to the lack of technological devices provided along with the questionnaire. However, it was impossible to provide a computer software for a wide range of participants as there are many different software used to simulate data.

Demographic information gathered from the participants in Part II of the instrument revealed that 18 participants use a graphing calculator in the classroom, however none of the participants have used the graphing calculator to simulate data or at least have not stated that they have used it for such a purpose. One reason for this could be the fact that teachers may be unaware of the probability tools provided in graphing calculators.

Another information from Part II reveals that only 21 participants uses simulation in probability classrooms. Considering the suggestions found in the literature, probability classrooms should integrate technology use in order to integrate the frequency approach as data simulation is faster and less arduous and more data can be generated (Kvatinsky & Even, 2002; Kazak & Konold, 2010).

The importance of simulation for frequency approach has also been found in this study as well. The relationship between whether the teachers used the frequency approach and whether they use simulation in classrooms was found to be strong and significant with Cramer's phi (φ_c) value of 0.306 at 0.06 significance. In total only 5 teachers stated that they use both the graphing calculator and simulation tools in their probability classrooms.

More interestingly, there was also a relationship between experience level and simulation use. For this relationship Gamma was found to be -0.508 at at 0.005 significance level indicating that the relationship is negatively strong and significant. In other words, less experienced teachers are more likely to use a simulation tool.

Items 5 and 10 were the only two questions assessing the same misconception. The reason for repeating a similar scenario was that as the researcher hypothesized, this item was related to the frequency approach. The findings of these two items reveal two important things. First, teachers who show evidence of knowing the frequency approach in at least one case may not necessarily adopt it in the another case. There were only 4 participants who have adopted the frequency approach both in Item 5 and 10. This may be due to the fact that teachers may not have internalized the frequency approach.

Secondly, frequency approach was more widely used. In item 10, 20 participants used it whereas in Item 5, 14 participants used it. The difference in these numbers may be due to the fact that item 10 is on birth rates and item 5 has coin flips, and coin flip questions are more popular thus people might be more reliant on the classical approach. This misconception needs to be assessed at a further level with different scenarios including mainstream and less mainstream probability scenarios in a further research.

Table 5.2 showed earlier that in Items 3 and 9 which were assessing subjective probability were the only two times that did not have a no reason category. In other words, in subjective questions participants were able to state a reason.

There is not any evidence to make a claim for the reason behind this finding, however, the researcher hypothesizes that when participants think using the classical approach, their thinking is faster for two reasons: 1) they have internalized the classical approach as it is widely used; 2) items assessing the misconceptions (items excluding 3 and 9) had more clear sample spaces. If the thinking was faster as the researcher hypothesizes then the participants might have failed to express the thinking process and might have only expressed the answer. Such a hypothesis can not be clarified without further research that involves interviews.

It is also important to note that these two items were among the three items which had a category of irrelevant answers. In the irrelevant responses, participants tried to answer a different question rather the one being asked. This might be due to the fact that subjective questions are less mainstream in probability and when confronted with a question that is not similar to the questions that participants have seen before, they may have misunderstood the question or changed the question to a similar question that they already know of. However, there is not enough evidence for this claim and further research is needed.

6.2. Research Question 2

Research Question 2 was as follows:

> What is the relationship between the experience level and the adopted probability approach?

Experience level of the teachers did not have any significant relationship with the approach they adopted. In other words, teachers in specific experience levels do not necessarily adopt a specific approach.

As seen earlier in Table 5.1, the relationship between experience level and approach was not found to be significant in any of the items. For items 5, 6, 8 and 9 this relationship was found to be strong, however not significant. One reason behind the insignificance of this relationship could be the relatively small sample size of the study. This relationship may have been found to be strong if sample size were to be increased.

6.3. Research Question 3

Research Question 3 was as follows:

> What is the relationship between the experience level and the success level of answering questions correctly?

Experience level of the teachers did not have any significant relationship with the success level. From the results, it can be concluded that there is not enough statistical evidence to state that experienced teachers are more likely to have overcome the misconceptions in probability.

As seen earlier in Table 5.1, the relationship between experience level and success level have been found to be strong in item 6 and negatively strong in item 2 however, these relationships were not found to be signifcant. One reason behind the insignificance of this relationship could be the relatively small sample size of the study. This relationship may have been found to be strong if sample size were to be increased.

It is important to make a distinction between the amount of experience and the quality of it. The literature agrees that children should be taught probability starting at an early age in order to build intuitions on actual experience but mere experience is not sufficient (Greer & Mukhopadhyay, 2005). In other words, what matters is the quality of experience rather than the amount of experience that students are exposed to. This statement might also be a ground for explaining the lack of relationship between teachers' experience level and success level. It might not be how many years teachers have actual experience in teaching but rather the quality of teaching they were subjected to and approach.

6.4. Research Question 4

Research Question 4 was as follows:

> What is the relationship between the adopted probability approach and the success level?

As seen earlier in Table 5.1, the relationship between experience level and success level have been found to be strong and significant for items 1, 4, 6, 8i, 8ii, and 10. Possible explanations for the difference in success levels for these items will be explained in this section.

In Items 1 and 4, it was seen that participants who have used the subjective approach solely have been among the least successful. It can be concluded that subjective probability when used on its own may lead to misconceptions. In addition, results have suggested that when subjective probability is combined with the classical approach, the success level rises. Chernoff's call for the need to make a clear distinction between subjective probabilities based on person's own belief and rational belief is also validated in this study (2010). When subjective probability is solely used, the success level decreases. In order to increase learning in probability, subjective reasoning found in learners has to be combined with rational reasoning (Pratt, 2000).

Item 1 having a scenario of a lottery draw can be considered to be taking part in everday life relatively more, when compared to the scenarios of other items that consist of coin flips and extracting marbles. Participants who have solely stated subjective responses to this item may have not connected it with the classical approach due to the relation of this item with real life behavior although they have shown evidences of knowing the classical approach in other items.

The relation of conjuction fallacy identified in Item 4 and real life is even more inevitable. The first part of the item gives information on Tayfun's real life experiences. Participants who have used the subjective approach have focused on Tayfun's real life experiences rather than seeing the probabilities based on the conjunction principle as $P(A) \geq P(A \cap B)$. For this item, it can be concluded that when faced with real life situation, participants have hard time connecting the probability question with a mathematical formula.

Despite the fact that classical approach is the most widely used, in both parts of item 8 which were found to be significant, it has been shown that classical probability on its own is not adequate. In both parts, the way participants used the classical approach made a difference determining their success. It can be concluded that it is not only the approach that makes the difference but also how the approach is used.

Similarly in item 10 identifying insensitivity to sample size misconception, it was also seen that the way classical approach is used makes a difference in determining success. More than half (53.57 %) of the participants who have used the classical approach in item 10 have failed to provide a correct answer. In item 10 it was more advantageous to use the frequency approach. Almost all (94.73%) of the participants who have used the frequency approach have answered this questions correctly. In situations, similar to item 10 where identifying sample space is ardous, classical approach is not as beneficial as the frequency approach.

Analysis of all items showed that among classical approach adopting teachers, highest percentage of them were likely to get 5 items correctly whereas among those who have adopted all of the three approaches, highest percentage was likely to get 6 items correctly. The difference in success levels of these different categories might be due to the fact that insensitivity to sample size was related to the frequency approach. There are situations, in which classical approach itself is not adequate or ardous like the case in item 10. It can be concluded that depending on the nature of the problem, different approaches may be adopted.

6.5. Implications

Despite the fact that the sample of the study consisted of mathematics teachers and not the students, there are still some implications of this study about the approaches that can be considered for probability classrooms. These implications will be discussed in this section.

As seen in this study, even teachers have subjective reasoning about uncertain situations. It is hard to ignore the intuition or the subjective reasoning one has when thinking probabilistically. It is clear from the findings that subjective probability, on its own, may not be useful. However, this does not mean that subjective probability should be taken out of the probability classroom. In fact the literature states that people can hold both the rational and subjective reasoning at the same time and the goal of probability education should be to prioritize the rational (Pratt, 2000).

In Section 2.1 it was explained that classical and frequency approaches can not answer all the probability questions in real life. Classical approach is limited to the situations where the sample space is clear and frequency approach is limited to the cases where the experiments can be repeated. In real life, probability questions do not necessarily have a clear sample space or the repeatability of the experiments may not be possible (Albert, 2003). If the aim of the probability classrooms is to help students find answers to real life questions then the subjective approach should be embraced. Results of this study indicate that adopting subjective approach is not disadvantageous. In fact, given that subjective approach can help answer more questions, such adoption could be more advantageous.

It should also be noted that adopting subjective probability solely can lead to misconceptions. As the findings of this study indicate, when subjective probability is used as the only approach, the success level decreases. The activity described in Section 2.1 found in the Turkish curriculum appears to be a weak one according to the findings of this study because the activity does not combine subjective probability with the classical or the frequency approach.

Many activities in the literature dedicate time to students' use of subjective probability and try to build on that (Pratt, 2000; Richardson, 2002). Ignoring the subjectivity of people may not be constructive when teaching. In fact the findings of the study also suggest that teacher preparation in probability can also build on teachers' subjectivity and combine it with the other approaches.

The frameworks and curricula suggest adoption of the frequency approach in probability classrooms and teachers should be capable of using simulation tools (Kvatinsky & Even, 2002; NCTM, 2002). The significant relationship between frequency approach adoption and simulation use concluded in this study implies that teacher training programs should also embrace simulation activities especially for more experienced teachers.

6.6. Limitations of the Study

One of the limitations of the study was during the sample selection process. For quantitative research, it is important to select participants randomly (Gay et al., 2009). However, for practical reasons, it is impossible to do random selection. In this study the schools or the participants were not selected randomly. In order to be able to examine different approaches the researcher used the purposive sampling method.

During the sample selection process, school administrators were contacted and teachers were kindly asked whether they would volunteer or not. Teachers who have chosen to volunteer might already have an interest in probability and thus results may not be indicative of the whole population, excluding teachers who are not interested in probability.

The second limitation of the study was the mode of the assessment. Apart from assessing what choices participants make, it was also measuring why they make those choices. However, some participants have failed to provide a reason behind their choices. This might be due to the culture of multiple choice tests in Turkey, where test takers rarely express reasons behind their choices.

Another limitation of the test was the number of items. Each misconception -except for insensitivity to sample size- was assessed only once. The researcher did not increase the number of items, in order to keep the duration of the test durable for participants. The results could differ if the same misconceptions were to be assessed with the same participants on different items.

Last limitation of the test was the lack of technological devices provided during the research. Participants were only provided a graphing calculator which can simulate data. Only 18 (25%) participants stated that they use a graphing calculator while teaching. If participants were to be provided with more tools such as computer software they could have been more reliant on frequency approach.

6.7. Suggestions for Further Research

Results gained from this study and limitations of the study indicate possible suggestions for further research. First suggestion can be having multiple number of research focusing on each misconception separately. Measuring all misconceptions simultaneously is time consuming and tedious for participants. However if a scale is to be developed for each misconception separately, teachers' understanding and reasoning can be better examined.

The second suggestion may be using interviewing methods. With interviewing methods, missing data can be more easily eliminated and participants may be able to provide the reasons behind their choices. The second advantage of interviews would be to ask if the participant can solve the question with different approaches. With verbal prompts, participants might come up with different approaches to the problems.

APPENDIX A: MAKING DECISIONS OF PROBABILITY QUESTIONNAIRE

INFORMATION ABOUT THE RESEARCH

I am an M.S. candidate studying Secondary School Science and Mathematics Education at Bogazici University. For my thesis, I am working on how teachers make decisions about probability. I would like to thank you for volunteering to participate in the study. If you feel under pressure while solving the test or before the test, you may choose to resign out of the study **at any point**.

This instrument consists of two parts. The first part consists of 10 questions on probability. While solving these questions, please make sure that you show all your work. In this research, I am interested in your way of thinking. If you think this test is limiting you in any way please state the limitations you face. For example if you would like to draw a circle and in need of a compass and cannot draw without one, please state that you needed a compass in order to solve the question and make sure to include how you would have used the compass. Also if you cannot think of anything for some problems, you can state that.

The second part consists of questions on your academic and teaching background. It also has questions on your teaching of probability. **The second part will be given to you after you finish the first part.**

The questions are in English. Depending on which language you choose to express yourself in, you can use Turkish and/or English when answering the questions. If you think you are unable to understand the instructions and thus unable to answer the questions, please inform the researcher.

Sorular İngilizcedir. Hangi dilde kendinizi daha iyi ifade ettiğinize bağlı olarak soruları Türkçe ve/veya İngilizce kullanarak cevaplandırabilirsiniz. Yönergeleri anlamadığınızı düşünüyorsanız ve soruları yanıtlayamayacağınıza inanıyorsanız, araştırmacıyı bilgilendirin. Bu paragraf her iki dilin de kullanabilineceğini örneklendirmek için verilmiştir.

You will be provided a TI-84 calculator along with the test. You can use any feature of the calculator while solving the questions. **Please state the features you have used or thought of using when solving the questions.**

The purpose of this questionnaire is to understand how you make decisions. So, when answering the questions, please provide a detailed explanation of your reasoning. <u>If you think there is more than one possible explanation to the questions, please provide all the solutions that you can think of.</u>

Your name is not asked and please do not write it on the form. Your identity will be kept **anonymous** and c**onfidential**.

Once again, thank you for volunteering to participate,

Mine Dogucu

MAKING DECISIONS ABOUT PROBABILITY QUESTIONNAIRE
PART I

1) In a lotto (Sayısal Loto) game, one has to choose 6 numbers from a total of 49. Banu has chosen 1, 2, 3, 4, 5, 6. Mehmet has chosen 39, 1, 17, 33, 8, 27. Who has a greater chance of winning? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

2) When tossing a coin, there are two possible outcomes: either heads or tails. Nurcan flipped a coin three times and in all cases heads came up. Nurcan intends to flip the coin again. What is the chance of getting heads the fourth time? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

3)

Above is an example of a Turkish lottery ticket (Milli Piyango) consisting of 6 digits. Turkish National Lottery Administration (Milli Piyango İdaresi) has a lottery drawing on 9th, 19th and 29th of every month. What is your chance of winning the big prize in 2013? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

4) Tayfun dreams of becoming a doctor. He likes to help people. When he was in high school, he volunteered for the Turkish Red Crescent (Kızılay) organization. He accomplished his studies with high performance and served in the army as a medical attendant. After ending his army service, Tayfun registered at the university. Which seems to you to be more likely? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

A) Tayfun is a student of the medical school.
B) Tayfun is a student.

5) K: The likelihood of getting heads at least twice when tossing three coins | L: The likelihood of getting heads at least 200 times out of 300 times

A) K is smaller than L
B) K is equal to L
C) K is bigger than L

Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

6) K: The number of possibilities when choosing 2 members from among 10 candidates | L: The number of possibilities when choosing 8 members from among 10 candidates

A) K is smaller than L
B) K is equal to L
C) K is bigger than L

Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

7) Suppose one rolls two dice simultaneously. Which of the following has a greater chance of happening? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of

A) Getting the pair 5-6
B) Getting the pair 6-6
C) Both have the same chance

8) Emre and Nihan each receive a box containing two white marbles and two black marbles.

i) Emre extracts a marble from his box and finds out that it is a white one. Without replacing the first marble, he extracts a second marble. Is the likelihood that this second marble is also white smaller than, equal to, or greater than the likelihood that it is a black marble? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

ii) Nihan extracts a marble from her box and puts it aside without looking at it. She then extracts a second marble and sees that it is white. Is the likelihood that the first marble she extracted is white smaller than, equal to, or greater than the likelihood that it is black? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

9) If you are single what is the probability that you will get married in the next five years? If you are married what is the probability that you will get divorced in the next five years? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

10) 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? Why do you think so? If you think there is more than one possible explanation to the question, please provide all the solutions that you can think of.

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.

MAKING DECISIONS ABOUT PROBABILITY QUESTIONNAIRE
PART II

For how many years have you been teaching mathematics formally? __________________________

Please select the level of schools that you have taught mathematics. If you are a pre-service teacher please select your department.
___Primary
___Secondary
___Both primary and secondary

Have you ever taken a Probability or Statistics course at a university level? If yes, what courses did you take? Could you state the topics you have covered in those courses?

____

____

____

Have you ever taken a Probability or Statistics course at the graduate level? If yes, what courses did you take? Could you state the topics you have covered in those courses?

____

____

____

Do you use a graphing calculator when teaching? If yes, please provide examples briefly. That is, explain how you use them.

____
____
____
____
____

Do you conduct simulation activities (with calculator and/or computer) in your probability or statistics classrooms? If yes, please provide examples briefly. That is, explain how you use them.

____
____
____
____
____

APPENDIX B: CROSS TABLES OF RESULTS

Table B.1. Item 1 Experience Level and Approach Cross Table.

		Experience Level			**Total**
		0	1	2	
Approach	Classical	13	12	17	42
	Classical and Frequency	0	0	1	1
	Classical and Subjective	2	1	4	7
	No Reason	1	7	9	17
	No Reason and Subjective	0	0	1	1
	Subjective	1	2	1	4
Total		17	22	33	72

Table B.2. Item 1 Experience Level and Success Level Cross Table.

		Experience Level			**Total**
		0	1	2	
Response	Correct	16	18	31	65
	Incorrect	1	3	2	6
	No Preference	0	1	0	1
Total		17	22	33	72

Table B.3. Item 1 Approach and Success Level Cross Table.

		Approach						
		Classical	Classical and Frequency	Classical and Subjective	No Reason	No Reason and Subjective	Subjective	**Total**
Success Level	Correct	41	1	7	14	1	1	65
	Incorrect	0	0	0	3	0	3	6
	No Preference	1	0	0	0	0	0	1
Total		42	1	7	17	1	4	72

Table B.4. Item 2 Experience Level and Approach Cross Table.

		Experience Level			Total
		0	1	2	
Approach	Classical	16	21	27	64
	Classical, Frequency and Subjective	1	0	1	2
	Classical and Subjective	0	1	3	4
	Frequency	0	0	1	1
	No Reason	0	0	1	1
Total		17	22	33	72

Table B.5. Item 2 Experience Level and Success Level Cross Table.

		Experience Level			Total
		0	1	2	
Response	Correct	16	20	28	64
	Incorrect	1	2	5	8
Total		17	22	33	72

Table B.6. Item 2 Approach and Success Level Cross Table.

		Approach					**Total**
		Classical	Classical, Frequency and Subjective	Classical and Subjective	Frequency	No Reason	
Success Level	Correct	57	2	4	0	1	64
	Incorrect	7	0	0	1	0	8
Total		64	2	4	1	1	72

Table B.7. Item 3 Experience Level and Approach Cross Table.

		Experience Level			**Total**
		0	1	2	
Approach	Classical	15	20	27	62
	Classical and Subjective	1	2	3	6
	Subjective	0	0	1	1
	Irrelevant	1	0	2	3
Total		17	22	33	72

Table B.8. Item 4 Experience Level and Approach (with subcategories) Cross Table.

		Experience Level			**Total**
		0	1	2	
Approach	Classical	0	0	1	1
	Certain Event	3	6	9	18
	Certain Event and Outcome Approach	0	1	0	1
	Certain Event and Subjective	3	5	0	8
	Certain Event and Subset	1	0	0	1
	Outcome Approach	2	0	1	3
	No Preference	0	0	1	1
	No Reason	1	2	6	9
	Subjective	3	5	11	19
	Subjective and Outcome Approach	2	0	0	2
	Subset	1	3	2	6
	Subset and Outcome Approach	1	0	0	1
	Subset and Subjective	0	0	2	2
Total		17	22	33	72

Table B.9. Item 4 Experience Level and Success Level Cross Table.

		Experience Level			**Total**
		0	1	2	
Response	Correct	9	16	20	45
	Incorrect	5	3	5	13
	No Preference	3	2	8	13
	No Answer	0	1	0	1
Total		17	22	33	72

Table B.10. Item 4 Approach (with subcategories) and Success Level Cross Table.

		Approach													**Total**
		Class.	Cert. Event	Cert. Event and Outc. App.	Cert. Event and Subj.	Cert. Event and Subset	Outc. App.	No Pref.	No Reason	Subj.	Subj. and Outc. App.	Subset	Subset and Outc. App.	Subset and Subj.	
Success Level	Correct	1	18	1	8	1	0	0	6	2	0	6	1	1	45
	Incorrect	0	0	0	0	0	1	0	1	10	1	0	0	0	13
	No Pref.	0	0	0	0	0	2	1	1	7	1	0	0	1	13
	No Answer	0	0	0	0	0	0	0	1	0	0	0	0	0	1
Total		1	18	1	8	1	3	1	9	19	2	6	1	2	72

Table B.11. Item 4 Approach and Success Level Cross Table.

		Approach				**Total**
		Classical	Classical and Subjective	No Reason	Subjective	
Success Level	Correct	28	9	6	2	45
	Incorrect	1	1	1	10	13
Total		29	10	7	12	58

Table B.12. Item 5 Experience Level and Approach (with subcategories) Cross Table.

		Experience Level			**Total**
		0	1	2	
Approach	Classical	0	1	1	2
	Frequency	3	3	2	8
	Frequency and Subjective	0	1	0	1
	No Reason	1	0	0	1
	Calculation	6	5	16	27
	Calculation and Frequency	0	2	1	3
	Calculation and Induction	1	3	0	4
	Ratio	3	2	5	10
	Ratio and Frequency	1	0	1	2
	Ratio and Calculation	0	1	0	1
	Ratio and Subjective	0	0	1	1
	Induction	2	4	6	12
Total		17	22	33	72

Table B.13. Item 5 Experience Level and Success Level Cross Table.

		Experience Level			**Total**
		0	1	2	
Response	A	5	3	6	14
	B	3	2	7	12
	C	9	16	14	39
	No Answer	0	1	6	7
Total		17	22	33	72

Table B.14. Item 5 Approach (with subcategories) and Success Level Cross Table.

		Approach												Total
		Class.	Freq.	Freq. and Subj.	No Reason	Calc.	Calc. and Freq.	Calc. and Induct.	Ratio	Ratio and Freq.	Ratio and Calc.	Ratio and Subj.	Induct.	
Response	A	1	2	0	0	7	0	1	1	1	0	0	1	14
	B	0	0	0	0	1	0	0	8	1	0	1	1	12
	C	1	6	1	1	14	2	3	1	0	0	0	10	39
	NoAnswer	0	0	0	0	5	1	0	0	0	1	0	0	7
Total		2	8	1	1	27	3	4	10	2	1	1	12	72

Table B.15. Item 6 Experience and Approach (with subcategories) Cross Table.

		Experience Level			**Total**
		0	1	2	
Approach	Combination	6	9	21	36
	Combination and Not Choosing	3	5	6	14
	Combination and Quantity	0	0	1	1
	Not Choosing	0	2	0	2
	No Reason	1	0	0	1
	Permutation	1	3	1	5
	Quantity	0	0	1	1
	Ratio	5	3	3	11
	Ratio and Combination	1	0	0	1
Total		17	22	33	72

Table B.16. Item 6 Experience and Success Level Cross Table.

		Experience Level			**Total**
		0	1	2	
Response	A	5	3	4	12
	B	9	16	26	51
	C	1	3	2	6
	No Answer	2	0	1	3
Total		17	22	33	72

Table B.17. Item 6 Approach (with subcategories) and Success Level Cross Table.

		Approach									Total
		Comb.	Comb. and Not Choosing	Comb. and Quant.	Not Choosing	No Reason	Permut.	Quant.	Ratio	Ratio and Comb.	
Response	A	0	0	0	0	0	1	0	11	0	12
	B	35	14	0	2	0	0	0	0	0	51
	C	1	0	0	0	0	4	1	0	0	6
	No Answer	0	0	1	0	1	0	0	0	1	3
Total		36	14	1	2	1	5	1	11	1	72

Table B.18. Item 7 Experience Level and Approach (with Different Dice) Cross Table.

		Experience Level			**Total**
		0	1	2	
Approach	Classical	15	18	25	58
	Classical and Different Dice	0	1	4	5
	Classical, Different Dice and Subjective	0	0	1	1
	Classical and Subjective	1	1	0	2
	No Reason	1	2	3	6
Total		17	22	33	72

Table B.19. Item 7 Experience Level and Success Level Cross Table.

		Experience Level			**Total**
		0	1	2	
Response	A	7	12	19	38
	B	0	0	1	1
	C	10	9	11	30
	No Answer	0	1	2	3
Total		17	22	33	72

Table B.20. Item 7 Approach (with Different Dice) and Success Level Cross Table.

		Approach					**Total**
		Classical	Classical and Different Dice	Classical, Different Dice and Subjective	Classical and Subjective	No Reason	
Response	A	36	1	0	1	0	38
	B	1	0	0	0	0	1
	C	21	3	0	1	5	30
	No Answer	0	1	1	0	1	3
Total		58	5	1	2	6	72

Table B.21. Item 8i Experience Level and Approach (with subcategories) Cross Table.

		Experience Level			Total
		0	1	2	
Approach	Conditional Probability	0	2	0	2
	Irrelevant	0	0	2	2
	No Reason	0	0	2	2
	Second Draw	16	17	23	56
	Second Draw and Conditional Probability	0	0	1	1
	Two Draws	1	3	5	9
Total		17	22	33	72

Table B.22. Item 8i Experience Level and Success Level Cross Table.

		Experience Level			Total
		0	1	2	
Response	$P(W) < P(B)$	16	21	28	65
	$P(W) = P(B)$	1	0	1	2
	$P(W) > P(B)$	0	1	0	1
	Irrelevant	0	0	2	2
	No Answer	0	0	2	2
Total		17	22	33	72

Table B.23. Item 8i Approach (with subcategories) and Success Level Cross Table.

		Approach						**Total**
		Conditional Probability	Irrelevant	No Reason	Second Draw	Second Draw and Conditional Probability	Two Draws	
Response	$P(W) < P(B)$	2	0	1	55	1	6	65
	$P(W) = P(B)$	0	0	0	0	0	2	2
	$P(W) > P(B)$	0	0	0	0	0	1	1
	Irrelevant	0	2	0	0	0	2	2
	No Answer	0	0	1	1	0	0	2
Total		2	2	2	56	1	9	72

Table B.24. Item 8ii Experience Level and Approach(with subcategories) Cross Table.

		Experience Level			Total
		0	1	2	
Approach	Conditional Probability	0	2	8	10
	First Draw	3	0	2	5
	Irrelevant	3	4	6	13
	Independent	5	9	7	21
	Independent and Subjective	0	1	0	1
	No Reason	0	0	2	2
	Two Draws	6	6	8	20
Total		17	22	33	72

Table B.25. Item 8ii Experience Level and Success Level Cross Table.

		Experience Level			Total
		0	1	2	
Response	$P(W) < P(B)$	8	7	15	30
	$P(W) = P(B)$	6	11	9	26
	Irrelevant	3	4	6	13
	No Answer	0	0	3	3
Total		17	22	33	72

Table B.26. Item 8ii Approach(with subcategories) and Success Level Cross Table.

		Approach							Total
		Conditional Probability	First Draw	Irrelevant	Independent	Independent and Subjective	No Reason	Two Draws	
Response	$P(W) < P(B)$	8	5	0	0	0	1	16	30
	$P(W) = P(B)$	1	0	0	21	1	0	3	26
	Irrelevant	1	0	12	0	0	0	0	13
	No Answer	0	0	1	0	0	1	1	3
Total		10	5	13	21	1	2	20	72

Table B.27. Item 9 Experience Level and Approach (with subcategories) Cross Table.

		Experience Level			**Total**
		0	1	2	
Approach	Data	1	2	2	5
	Data and Numerical	0	1	0	1
	Outcome Approach	2	6	4	12
	Outcome Approach and Other Factors	0	0	1	1
	Outcome Approach and Subjective	0	1	1	2
	Irrelevant	0	0	1	1
	Irrelevant and Outcome Approach	8	2	3	13
	Irrelevant and Numerical	0	0	1	1
	Irrelevant, Numerical and Ratio	0	0	1	1
	Not Mathematical	2	3	3	8
	Not Mathematical and Data	1	0	2	3
	Not Mathematical and Outcome Approach	0	1	0	1
	Not Mathematical and Other Factors	0	0	1	1
	Numerical	1	1	0	2
	No Answer	1	0	0	1
	Other Factors	0	2	3	5
	Other Factors and Data	0	2	2	4
	Other Factors, Subjective and Outcome Approach	0	0	1	1
	Subjective	0	1	2	3
	Subjective and Outcome Approach	1	0	0	1
	Subjective and Numerical	0	0	5	5
Total		17	22	33	72

Table B.28. Item 9 Experience Level and Approach (Subjective Comparison) Cross Table.

		Experience Level			Total
		0	1	2	
Approach	Subjective	1	2	9	12
	Not Subjective	15	20	23	58
Total		17	22	33	72

Table B.29. Item 10 Experience Level and Approach (with subcategories) Cross Table.

		Experience Level			Total
		0	1	2	
Approach	Classical	0	2	0	2
	Frequency	4	6	8	18
	Frequency and Quest. 5	0	1	1	2
	Outcome Approach	2	0	3	5
	Higher Number	2	1	3	6
	No Reason	2	2	8	12
	Calculation	4	5	2	11
	Quest. 5	0	1	1	2
	Ratio	0	0	3	3
	Subjective	1	0	0	1
	Induction	1	3	4	8
	Induction & Ratio	1	1	0	2
Total		17	22	33	72

Table B.30. Item 10 Experience Level and Success Level Cross Table.

		Experience Level			**Total**
		0	1	2	
Response	A	4	2	3	9
	B	6	14	14	34
	C	5	5	14	24
	No Answer	1	0	0	1
	No Preference	1	1	2	4
Total		17	22	33	72

Table B.31. Item 10 Approach (with subcategories) and Success Level Cross Table.

		Approach												**Total**
		Class.	Freq.	Freq. and Quest. 5	Outc. App	Higher No.	No. Reason	Calc.	Quest 5.	Ratio	Subj.	Induct.	Induct. and Ratio	
Response	A	0	1	0	0	6	0	1	1	0	0	0	0	9
	B	2	16	2	0	0	3	10	1	0	0	0	0	34
	C	0	0	0	4	0	6	0	0	3	1	8	2	24
	No Answer	0	0	0	0	0	1	0	0	0	0	0	0	1
	No Preference	0	1	0	1	0	2	0	0	0	0	0	0	4
Total		2	18	2	5	6	12	11	2	3	1	8	2	72

Table B.32. Item 10 Approach and Success Level Cross Table.

		Approach					**Total**
		Classical	Frequency	Classical and Frequency	No Reason	Subjective	
Success Level	Correct	13	16	2	2	0	33
	Incorrect	15	1	0	6	1	23
Total		28	17	2	8	1	56

Table B.33. All Items Experience Level and Approach Cross Table.

		Experience Level			Total
		0	1	2	
Approach	Classical	5	6	7	18
	Classical and Frequency	3	4	3	10
	Classical, Frequency and Subjective	3	6	12	21
	Classical and Subjective	6	6	11	23
Total		17	22	33	72

Table B.34. All Items Experience Level and Success Level Cross Table.

		Experience Level			Total
		0	1	2	
Success Level	1	0	0	1	1
	2	3	2	3	8
	3	4	5	6	15
	4	3	3	6	12
	5	4	5	8	17
	6	3	6	6	15
	7	0	1	3	4
Total		17	22	33	72

Table B.35. All Items Approach and Success Level Cross Table.

		Approach				Total
		Classical	Classical and Frequency	Classical, Frequency and Subjective	Classical and Subjective	
Success Level	1	0	0	1	0	1
	2	0	3	1	4	8
	3	4	1	3	7	15
	4	4	0	4	4	12
	5	6	0	5	6	17
	6	3	4	7	1	15
	7	1	2	0	1	4
Total		18	10	21	23	72

REFERENCES

Aggarwala, N. K., 2004, *Quality Assessment of Primary and Middle Education in Mathematics and Science(TIMSS)*, Eaton-Batson International Management & Consulting Services, Tahoma Park, MD, USA.

Agresti, A., 2002, *Categorical Data Analysis*, Wiley-Interscience, Hoboken, NJ, USA.

Albert, J. H., 2003, "College Students' Misconceptions of Probability", *The American Statistician*, Vol. 57, pp. 37-45.

Amir, G. S., & J. S. Williams, 1999, "Cultural Influences on Children's Probabilistic Thinking", *Journal of Mathematical Behavior*, Vol. 18, pp. 85-107.

Australian Educational Council, 1990, *A National Statement on Mathematics for Australian Schools: A Joint Project of the States, Territories and the Commonwealth of Australia Initiated by the Australian Education Council*, Curriculum Corporation for the Australian Education Council, Carlton, Australia.

Batanero, C., M. Henry, & B. Parzysz, 2005, "The Nature of Chance and Probability" in G. A. Jones (ed), *Exploring Probability in School*, pp. 15-37, Springer.

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

Bulut, S., S. Kazak, & E. Yetkin, 1999, "Matematik Öğretmen Adaylarının Olasılık Kavramları ile İlgili Yeterliliklerinin İncelenmesi", *D.E.Ü. Buca Eğitim Fakültesi Dergisi* , Vol. 11 , pp. 384-394.

Chernoff, E. J., 2008, "The State of Probability Measurement in Mathematics Education", *Philosophy of Mathematics Education Journal*, Vol. 23.

Chernoff, E. J., 2010, "Coming to Terms with Probability Terminology", *Vector: The*

Official Journal of the BC Association of Mathematics Teachers, Vol. 51 , pp. 13-16.

Darling-Hammond, L., 1999, *Teacher Quality and Student Achievement: A Review of State Policy Evidence*, University of Washington: Center for the Study of Teaching and Policy.

Ericsson, K. A., & A. Lehmann, 1996, "Expert and Exceptional Performance: Evidence of Maximal Adaptation to Task Constraints", *Annual Review of Psychology*, Vol. 47, pp. 273-305.

Falk, R., 1979, Revision of Probabilities and the Time Axis. In Proceedings of the Third International Conference for the Psychology of Mathematics Education , pp. 64-66, Warwick, UK: University of Warwick.

Falk, R., & C. Konold, 1992, "The Psychology of Learning Probability", In F. Gordon & S. Gordon (eds), *Statistics for the 21st century*, pp. 151-164, The Mathematical Assocation of America.

Fischbein, E., 1975, *The Intuitive Sources of Probabilistic Thinking in Children*, Reidel, Dordrecht, The Netherlands.

Fischbein, E., 1987, *Intitution in Math and Science*, Reidel, Dordrecht, The Netherlands.

Fischbein, E., & D. Schnarch, 1997, "The Evolution with Age of Probabilistic, Intuitively Based Misconceptions", *Journal for Research in Mathematics Education*, Vol. 28 .

Garfield, J., & A. Ahlgren, 1988, "Difficulties in Learning Basic Concepts in Probability and Statistics: Implications for Research", *Journal for Research in Mathematics Education*, Vol. 19 , pp. 44-63.

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

Green, D., 1990, "A Longitudinal Study of Pupils' Probability Concepts", Paper presented at International Conference on Statistics Teaching, Dunedin, New Zealand.

Greer, B., & S. Mukhopadhyay, 2005, "Teaching and Learning the Mathematization of Uncertainity: Historical, Cultural, Social and Political Contexts", In G. A. Jones (ed), *Exploring Probability in School*, pp. 297-324, Springer.

Jones, G. A., & C. A. Thornton, 2005, "An Overview of Research into the Teaching and Learning of Probability". In G. A. Jones (ed), *Exploring Probability in School*, pp. 64-92, Springer.

Kahneman, D., & A. Tversky, 1972, "Subjective Probability: Representativeness", *Cognitive Psychology*, Vol. 3 , pp. 430-454.

Kazak, S., & C. Konold, 2010, "Development of Ideas in Data and Chance Through the Use of Tools Provided by Computer-Based Technology", Paper presented at *International Conference on Teaching Statistics*, Ljubljana, Slovenia.

Konold, C., 1989, "Informal Conceptions of Probability", *Cognition and Instruction*, pp. 59-98.

Kvatinsky, T., & R. Even, 2002, "Framework for Teacher Knowledge and Understanding About Probability", Paper presented at *International Conference on Teaching Statistics*, Cape Town, South Africa.

Lecoutre, M. P., 1992, "Cognitive Models and Problem Spaces in 'Purely Random' Situations", *Educational Studies in Mathematics*, Vol. 23 , pp. 557-568.

Memnun, D. S., 2008, "Olasılık Kavramlarının Öğrenilmesinde Karşılaşılan Zorluklar, Bu Kavramlarının Öğrenilememe Nedenleri ve Çözüm Önerileri", *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, Vol. 9, pp. 89-101.

Milli Eğitim Bakanlığı, 2009, *İlköğretim Matematik Dersi 6-8. Sınıflar Öğretim Programı ve Kılavuzu,* Ankara.

National Council of Teachers of Mathematics, 2002, *Navigating Through Data Analysis and Probability in Grades 3-5*, Reston,VA, USA: National Council of Teachers of Mathematics.

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

Nau, R. F., 2001, "De Finetti Was Right: Probability Does not Exist". *Theory and Decision*, Vol. 51 , pp. 89-124.

Piaget, J., & B. Inhelder, 1975, *The Origin of the Idea of Chance in Children*, Norton, New York, USA.

Pinto, R. M., 2010, "Mixed Methods Design", In N. J. Salkind (ed), *Encyclopedia of Research Design*, pp. 64-92, SAGE.

Pratt, D., 2000, "Making Sense of the Total of Two Dice", *Journal for Research in Mathematics Education*, Vol. 31 , pp. 602-625.

Rast, J. D., 2005, *A Comparison of Learning Subjective and Traditional Probability in Middle Grades*, Unpublished doctoral dissertation, Georgia State University.

Richards, W. D., 1998, *The Zen of Emprical Research*, Empirical Press, Vancouver, BC, CA.

Richardson, M., 2002, "What is the Probability of a Kiss? (It's not What You Think)", *Journal of Statistics Education*, Vol. 10 .

Rockoff, J. E., 2004, "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data". In Papers and Proceedings of the *One Hundred Sixteenth*

Annual Meeting of the American Economic Association, pp. 247-252, The American Economic Review, San Diego, CA.

Shaughnessy, J. M., 1977, "Misconceptions of Probability: An Experiment with a Small-Group, Activity-Based, Model Building Approach to Introductory Probability at the College Level", *Educational Studies in Mathematics*, Vol. 8 , pp. 295-316.

Shaughnessy, J. M., 1992, "Research in Probability and Statistics: Reflections and Directions", In D. A. Grouws (ed), *Handbook of Research on Mathematics Teaching and Learning*, NCTM, Reston, VA.

Stohl, H., 2005, "Probability in Teacher Education and Development", In G. A. Jones (ed), *Exploring Probability in School*, pp. 345-366, Springer.

Tversky, A., & D. Kahneman, 1973, "Availability: A Heuristic for Judging Frequency and Probability", *Cognitive Psychology*, Vol. 5 , pp. 297-232.

Tversky, A., & D. Kahneman, 1982, "Judgments of and by representativeness", In D. Kahneman, P. Slovic, & A. Tversky (eds), *Judgment Under Uncertainity: Heuristics and Biases*, pp. 84-100, Cambridge University Press.

Tversky, A., & D. Kahneman, 1983, "Extensional Versus Intuitive Reasoning: The Conjuction Fallacy in Probability Judgment", *Psychological Review*, Vol. 90 .