COLLEGE STUDENTS' ACHIEVEMENT AND UNDERSTANDING OF EXPERIMENTAL AND THEORETICAL PROBABILITY: THE ROLE OF TASKS

By

Irini Papaieronymou

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Mathematics Education

2012

ABSTRACT

COLLEGE STUDENTS' ACHIEVEMENT AND UNDERSTANDING OF EXPERIMENTAL AND THEORETICAL PROBABILITY: THE ROLE OF TASKS

By

Irini Papaieronymou

This study examined the role of particular tasks implemented through two instructional methods on college students' achievement and understanding of probability. A mixed methods design that utilized a pre-test and post-test was used. This included treatment and control groups, each comprised of students in three sections of an introductory statistics course taught by the researcher at a college in Cyprus. During the study, students in the treatment group worked in small groups on four in-class activities about experimental and theoretical probability (Instructional Method B), and students in the control group worked in small groups on solutions to four sets of probability problems from the course textbook (Instructional Method A).

An initial analysis of pre-test scores indicated that the students in the control group had comparable initial probability knowledge to the students in the treatment group. Quantitative as well as qualitative analysis were then carried out to address the research questions. With regards to students' *achievement* on the multiple-choice items on probability, the results of the Wilcoxon Signed-Ranks test which was carried out to analyze gain scores indicated that the multiple-choice scores of students in the control group were significantly lower on the post-test compared to the pre-test. In the case of the treatment group, student scores on the multiple-choice items did not increase significantly from the pre-test to the post-test. Possible explanations to this phenomenon are provided in the last chapter of this dissertation. In addition, an analysis of

n*ormalized* gain scores was carried out. Positive as well as negative normalized gains existed in both groups. The Mann-Whitney test resulted in a p-value of 0.001 (< 0.05) indicating that the normalized gain scores of the treatment group were significantly different from the normalized gain scores of the control group. Relative to students' achievement on the open-ended items included on the post-test, the Mann-Whitney test resulted in a p-value of 0.001 indicating that the scores of the treatment group on these items were significantly higher than the scores of the control group. Therefore, Instructional Method B was successful in producing significantly better *achievement* scores than Instructional Method A.

In this dissertation, students' *understanding* of probability was measured through a distractor analysis and a qualitative analysis of audio-taped student conversations. Specific to the distractor analysis relative to probabilistic heuristics defined in the literature, the mean percentage of students who applied these mostly increased in the case of the control group whereas it mostly decreased in the case of the treatment group. These results are in line with past research that indicated that the use of activity-based instruction may help students with respect to probabilistic misconceptions (e.g. Shaughnessy, 1977, 1981).

For the purposes of qualitative analysis, the framework by Jones, Thornton, Langrall and Tarr (1999) was used. Based on the results of the qualitative analysis of students' levels of reasoning relative to the constructs presented on the framework i) Instructional Method B (treatment) produced better results than Instructional Method A (control) relative to experimental probability and the concept of sample space and ii) it was difficult to identify which instructional method had a better effect on students' understanding of theoretical probability, conditional probability and independence, discrete probability distributions and the binomial distribution.

To my family, for their support throughout the years
And most of all
To my beloved husband George, for all his love, patience and understanding

ACKNOWLEDGMENTS

In my pursuit to carry out this dissertation in my home country Cyprus, while being registered as a PhD student at Michigan State University, USA I faced many challenges which were overcome with the help of many people.

First, I would like to express my gratitude to my wonderful advisor, Dr Sharon Senk who supported and guided me throughout my graduate studies in mathematics education. Dr Senk's thoughtful feedback, invaluable insights, challenging questions, and time devoted to emailing and holding skype meetings with me, helped tremendously in completing each step of this dissertation and keeping me on track to graduation. She has been instrumental in helping me become a researcher and a more reflective mathematics instructor, and for these I will forever be indebted!

Special thanks go to the rest of my committee members. Dr Jack Smith's thought-provoking emails and personal conversations pointed me to useful sources of literature and helped me in the formulation of a definition of the term 'understanding' in this study. The statistical expertise of Dr Jennifer Kaplan and Dr Vincent Melfi, along with their feedback, helped me with the methodological and quantitative analyses aspects of this study. Thank you all for the time you devoted to this dissertation!

This study certainly would not have been possible without the students who volunteered to participate. Thank you for your cooperation and willingness to be part of my research! Moreover, special thanks go to Lisa Keller, Margaret Iding and Jean Beland for their prompt

emails whenever I needed them and for their help in overcoming any administrative issues that arose during the PhD process!

While working on this dissertation, I was lucky to meet a wonderful and kind-hearted man, George, who I am happy and proud to be able to call my husband! His constant caring, love, patience, sense of humor, and … wedding proposal, were sources of inspiration during the writing process of the dissertation. Thank you for standing by me, for being my friend and lifetime companion, and for traveling to the USA with me!

A deep thank you goes to my parents who supported me at every step of the way during my studies and made my stay in the USA possible. My gratitude also goes to my uncle Petros, for his guidance throughout my life, and my aunt Pitsa for her support and constant care. Thank you to my wonderful brother, Michael whose love and support have always accompanied me in life! I would also like to thank my grandparents, Christos and Irini, for instilling in me the importance of education and a love for learning. Had they been alive I know that they would have been proud of my accomplishments.

Special thanks go to my best friend Georgia who always stood by me. Thank you for the countless hours you spent listening to me, for the myriad times you made me laugh, and for your reassurance and belief in me whenever I reached an obstacle in my way! My gratitude also goes to my dear friend Beste who supported me in my times of need and helped me not feel lonely while at Michigan State University. You will always be in my heart! My sincerest thanks also go to my friend Electra and her parents, uncle Nicos and auntie Anna, who had been my family while in Michigan! Moreover, I would like to thank my friends Elena, Tina, Costas, Neelambari, Carole, Nicole, Vicky, and Roselyn for making my life more enjoyable with their friendship!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

'Tis all a checker-board of nights and days
Where destiny with men for pieces plays
Hither and thither moves, and mates, and slays,
And one by one back in the closet lays.

(Edward Fitzgerald's translation of *The Rubáiyát of Omar Khayyám*, 1879; as cited in Woolfson, 2008, p. 191).

1.1 Importance of Probability

As the *Rubáiyát* notes, in life we are moved "hither and thither"; that is, life is unpredictable, consisting of "a series of chance happenings that toss you this way and that, sometimes for good and sometimes for ill" (Woolfson, 2008, p. 191). In 1998 people in Great Britain were astounded by the news of the death of the son of a couple in Cheshire. The boy, Harry, was aged eight weeks at the time of his death. The astonishment came about due to the fact that a little more than a year prior to Harry's death, the same couple lost their son Christopher, aged eleven weeks at that time. In response to this sequence of events, the couple was arrested and after a trial that hit headline news, the mother, Sally Clark, was charged with the murder of her two sons.

What was the evidence that led to the mother's arrest? Sally Clark insisted that the cause of her sons' deaths was SIDS (Sudden Infant Death Syndrome). Recent research has revealed that the cause of SIDS is the existence of a brain abnormality in infant victims. This abnormality prevents infants from sensing high carbon dioxide and low oxygen levels thus increasing their risk of inhaling their own exhaled breath (Hawkes, 2006). However, these research results were not known at the time of Sally Clark's trial.

Pediatric consultant Sir Roy Meadow acted as an expert witness for the prosecution in the Sally Clark trial and although he was principally supposed to provide medical evidence, he also made a statistical statement that led the court to its decision to charge the mother with murder (Batt, 2004). The statement by Meadow indicated that the probability of two children in the same family dying of SIDS was 1 in 73 million. This probability sent Sally Clark to jail to serve two life sentences.

The court's decision caused an incorrect belief to arise: that the probability of a very rare event occurring is the same as the probability that the defendant is innocent, also known as the *Prosecutor's Fallacy* (Kaplan & Kaplan, 2006, p. 191). Later on, this flaw in reasoning was recognized, with the Royal Statistical Society protesting against Meadow's claim through a public announcement. Along with medical records - that were not available at the first trial - that the first son had indeed died from a respiratory infection, as well as records that Sally Clark was not an abusive or uncaring mother, the woman was eventually released from prison in 2003 and Sir Roy Meadow was taken off the medical rolls.

The above unpleasant event indicates the role of probability in courts. Yet, probability is encountered in many other aspects of everyday life. Probabilistic claims are made by weather forecasters, physicians, journalists, and election polls. In order to be a well-informed citizen one needs to understand the language and basic ideas of probability (Gal, 2004; Scheaffer, Watkins, & Landwehr, 1998; Utts, 2003) and how probability can be used to model our world (Moore, 1997; Penas, 1987).

Given these assertions about the role of probability in everyday life, it is not surprising that since the late 1950s there has been a strong call for an increase in the inclusion of probability in mathematics curricula. In Europe, 1958 was marked by a study of mathematics education of

the member countries of the Organization for European Economic Cooperation (OEEC), which in turn led to an international conference the following year in France. One of the conference conclusions was that statistics and probability should be introduced in the curriculum (Exarchakos, 1988).

Following this, the 1960s were also marked by suggestions for the inclusion of probability in school curricula. In Europe, and in particular in Great Britain, a reform project was carried out during this decade, namely the *School Mathematics Project*, which was aimed at middle and high school students (Lordou-Kaspari, 2003). The project writers were concerned with the gap between school-level and university-level mathematics as well as with the absence of applications in school mathematics; so, they included statistics and probability in school syllabi. In the US, a group of mathematicians and National Science Foundation (NSF) representatives published *Goals for School Mathematics* in 1963, in which the importance of "some 'feeling' for probability" for all students was indicated (Jones, 1970, p. 291).

Subsequently, in the US, the National Council of Supervisors of Mathematics (NCSM) identified probability as one of the basic skills that students should acquire (1977). Furthermore, in 1983 the National Commission for Excellence in Education (NCEE) published *A Nation at Risk*, a report aimed at pointing out the immediate need for reform in education. In the report, the NCEE suggested that high school graduates should understand and be able to apply elementary probability and statistics. In the context of this new reform movement in the USA, the University of Chicago School Mathematics Project (UCSMP) initiated the development of a comprehensive set of curriculum materials for grades K-12 starting in 1983. Among the goals of UCSMP was "to give increased emphasis to "newer" mathematics, especially statistics and probability" (Lordou-Kaspari, 2003, p. 117).

It should be noted that in the early 1980s, the results of the Second International Mathematics Study (SIMS) revealed that in grade 8, probability was not considered important or was not taught at all, both in US schools as well as in schools internationally (Travers and Westbury, 1989). However, this has changed since then. On the Third International Mathematics and Science Study 1999 (TIMSS), 67% of US 8th grade mathematics teachers indicated that they spent at least one period teaching simple probabilities (TIMSS International Study Center, 2000). A possible factor contributing to this increased attention on probability in US classrooms is the set of recommendations set forth by US national professional organizations at the secondary (NCTM, 1989, 2000) and post-secondary levels (MAA, 1998). In particular, the National Council for Teachers of Mathematics (NCTM, 2000) specified in the *Data Analysis and Probability Standard* that emphasis should be placed on the understanding and application of basic probability concepts. According to the NCTM probability concepts should become increasingly sophisticated as students move through the grades. Moreover, at the post-secondary level, the Mathematical Association of America (MAA, 1998) and the American Mathematical Association of Two-Year Colleges (Cohen, 1995) have also supported an increase in the importance placed on probability.

With regards to Cyprus – where this study took place - the results of TIMSS 2007 revealed that at the 8th grade level only 3% of class time is devoted to data and chance (Mullis, Martin and Foy, 2008). Topics relative to this domain are considered to be for the more able students and as such, only 3% of Cypriot 8th graders receive formal instruction in this domain; that is, this 3% of Cypriot students receive the 3% of instructional time devoted to data and chance.

Given the importance of probability in everyday life and its recommended increased focus in school curricula, it is encouraging that the Ministry of Education and Culture in Cyprus has recently initiated educational reform efforts which include revisions to the national curriculum in grades K-12 (The Ministry of Education and Culture, Republic of Cyprus, 2008). In 2010, reports were published by the Ministry of Education and Culture of Cyprus regarding the intended curriculum for each school subject, including mathematics. This report specifies the content, procedures, applications and experiences that students are expected to acquire while in school. Probability is introduced at the elementary school level with increased emphasis as students move through the grades. The details of the mathematics report relative to the domain of probability are provided in Chapter 3 of this dissertation.

1.2 Teaching of Probability

Over the past couple of decades, there have been various reform initiatives concerning the content and means of instruction in mathematics classrooms at the college level (MAA, 1998). In the US, the MAA (1998) recommends that lectures be replaced by more active teaching methods in which group work is used and emphasis is placed on "making sense of inherently quantitative situations, problem formulation and heuristics at the expense of mechanics" (Retrieved May 10, 2009 from http://www.maa.org/past/ ql/ql_toc.html).

In addition to initiatives taken in the mathematics education community, undergraduate education research in the STEM fields (science, technology, engineering, and mathematics) has raised concerns regarding the teaching and learning in these areas. Undergraduate education in many STEM fields relies heavily on lectures which do not promote "genuine understanding" needed for further studies and work in these fields (National Research Council, 2003; Baldwin, 2009, p. 10). Instead, teaching environments should be designed that promote the involvement of

all students through the use of instructional procedures such as cooperative learning, "which affects the head and hand while simultaneously affecting the heart, thereby potentially reversing the negative trends noted in higher education" (Smith, Douglas, and Cox, 2009, p. 21).

Specific to the area of statistics and probability, many researchers have recommended that there be a change in the way statistics courses are taught (Chance, 1997; Garfield 1994; Shaughnessy, 1981; Watts, 1991; see also Keeler & Steinhorst, 2001). "One area that has received less focus in this literature is the teaching of probability" (Keeler & Steinhorst, 2001, Retrieved April 24, 2009 from http://www.amstat.org/publications/jse/v9n3/keeler.html). Much has been written about people's misconceptions and use of heuristics regarding judgment under uncertainty but there is a lack of research on solutions to this phenomenon (Keeler & Steinhorst, 2001). People hold preconceptions about probability so instruction needs to be changed and be carried out in such a way as to facilitate learning. In 1988, Garfield and Ahlgren recommended that more research be carried out on teaching probability in ways that aid people overcome their misconceptions whereas in his 1992 review of research on probability, Shaughnessy called for research to be carried out on instruction related to students' conceptual knowledge of probability.

Relative to these issues, the NCTM (2000) promotes the use of manipulatives and emphasizes the study of real-world problems and their connections to data analysis and probability. Moreover, several researchers suggest that instruction allows students to build models and develop their thinking capability (Pollak, 1968; Klamkin, 1968, Fitzgerald, 1975; see also Shaughnessy, 1977). To this end, it is recommended that statistics instruction relies less on lecturing and more on active learning that uses group problem-solving, activities and discussions (Cobb, 2000). Given that curriculum reform has brought data handling to the forefront, less emphasis should be placed on "formal" probability and "an empirical frequency-based approach

to probability that is also an important foundation for later work in theoretical probability" should be used (Watson, 2006, p.127).

Since Shaughnessy's 1992 review of research on probability and statistics, a considerable number of studies have been carried out on the teaching and learning of probability (Jones, Langrall & Mooney, 2007). However, most of this research relates to students' thinking of probability with a minimal amount having been carried out relative to instructional methods (Hirsch & O'Donnell, 2001). In addition, Shaughnessy (1992) indicated a lack of research in probability learning and teaching outside of western countries. Since then, very few studies have been found outside the western culture and only one was cited in a recent review by Jones, Langrall & Mooney (2007) which was carried out in Cyprus that examined elementary school students' probabilistic thinking. Furthermore, Jones, Langrall & Mooney (2007) indicated that "there has been little research on students' thinking about experimental probability and even less on students' understanding of the connections between theoretical and experimental probability" (p. 946).

During the three academic years spanning 2008-2011, the researcher of this dissertation study taught mathematics and statistics at a private college in Cyprus. In the courses that students take at this college, emphasis is placed on application problems in the area of business since the college is a specialized business school. During the introductory statistics course offered at the college and which the researcher first taught in spring 2009, these application problems were the most challenging for students. Students at this college tend to be mostly native Cypriots whose first language is Modern Greek. However, the medium of teaching is English which poses difficulties for several students. Moreover, many of these students were weak in probability and found it difficult to think critically and reason through probability problems. During

conversations with another mathematics instructor at the college, the researcher was informed that students in the introductory statistics course in previous years carried similar weaknesses and attitudes towards probability.

With regards to probability, it was not surprising to the researcher to come across student weaknesses. The national curriculum in public secondary schools in Cyprus includes minimal exposure to probability which occurs towards the end of grade 12 (Ignatiou and Zotos, 2003). Given that this introductory statistics course (Statistics I) provides the first substantial amount of exposure that these students have to probability, and that along with Statistics II, may be the only formal statistics courses that the students at this college take that include material on probability, it is important that these courses be taught in the most effective way.

Since these students are a vital element of the future workforce of Cyprus, it is important to understand their thinking capabilities and improve their critical thinking skills. For these to occur, instruction needs to provide students with opportunities to express their thinking and reasoning through ways that motivate them to do so. Studies have shown that college students "do not think critically and reflectively about important societal issues" but "courses in statistical thinking have the potential to improve students' general reasoning capabilities" (Derry et al., 2000, p. 748). Students "need opportunities on a regular basis to engage with tasks that lead to deeper, more generative understandings about the nature of mathematical concepts, processes, and relationships" (Stein, Smith, Henningsen, & Silver, 2000, p.15; as cited in Jones, 2004, p.3). Therefore, the statistics courses that the students at this college are required to take could be taught in such a way as to foster their critical thinking capabilities. Perhaps the use of in-class group activities that promote students' reasoning about probability might have a positive effect on students' achievement and understanding of probability.

1.3 Purpose of the Study

With the above issues under consideration, the purpose of this study was to examine the role of particular tasks on college students' achievement and understanding of probability. In particular, the study aimed to examine the potential benefits of performing probability experiments that generate real data and completing probability activities on college students' achievement and understanding of experimental and theoretical probability. The study took place in a college-level introductory statistics course and examined the effects of:

i. an instructional method that combined lectures and small-group cooperative learning sessions during which students solved probability problems (control group) and
ii. an instructional method that combined lectures and small-group cooperative learning sessions during which students completed activities involving probability experiments that generate real data (treatment group).

1.4 Research Questions

The research questions addressed in this study involve the use of two instructional methods:

Instructional Method A: Using lectures and small-group cooperative learning sessions during which students solve probability problems and

Instructional Method B: Using lectures and small-group cooperative learning sessions during which students use activities involving probability experiments that generate real data to make connections between experimental and theoretical probability.

Given Instructional Method A and Instructional Method B

1) What are the effects of using each of these instructional methods on college students' achievement on probability and on their understanding of experimental and theoretical probability?
2) Does Instructional Method B have a better effect on college students' achievement on probability and on their understanding of experimental and theoretical probability than Instructional Method A?

1.5 Definitions

1.5.1 Experimental Probability

Experimental or empirical probability involves first the collection of data through experiments or simulations (Jones, Thornton, Langrall & Tarr, 1999) and then the use of relative frequency to provide *a posteriori* approximation to the probability of an event (Mojica, 2006). According to this approach, probability is "the hypothetical number towards which the relative frequency tends when stabilizing (Batanero, Henry, & Parzysz, 2005, p. 23; see also von Mises, 1928/1952). The modern generalization of this stabilization of the relative frequency after a large number of trials is called the *Law of Large Numbers*.

1.5.2 Theoretical Probability

Theoretical probability involves the determination of the probability of an event through *a priori* approach (Levine, Krehbiel, & Berenson, 2010); that is, before conducting any experimental trials. A theoretical probability may be computed mathematically using numerical or geometrical methods (Burdzy, 2009; Jones, 2004) and it is based on the idea of equally likely outcomes (Aczel & Sounderpandian, 2002; Burdzy, 2009).

Laplace was the first who attempted to define probability using a mathematical interpretation at the beginning of the 19^{th} century. His interpretation was based on equally likely

events and so, excluded the possibility of unequally likely events. Laplace (1814/1995; see also Burdzy, 2009, p. 16) provided a definition of probability as follows:

> The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability, which is thus simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all the cases possible.

1.5.3 Probability Experiments with Real Data

The statistics education community recommends that students in statistics classrooms have access to and experiences collecting, analyzing and using real data (Hall and Rowell, 2008). The American Statistical Association (ASA) (Aliaga et al., 2005; Franklin & Garfield, 2006) indicates that real data comes in various forms: archival data, data generated in the classroom, and data generated through simulations. According to Franklin and Garfield (2006) teachers may use textbooks, journals, data repositories found on the web, data from a practicing research scientist, as well as data generated through surveys or activities completed in class as sources of real data.

Real data in this study was in the form of data generated through experiments (e.g. rolling dice) carried out by students (treatment group) or in the form of archival data (treatment and control groups). The course textbook (Levine, Krehbiel, and Berenson, 2010), which was used in all sections of the introductory statistics course in which the study took place, provided real data sets. Since the students in all sections of the course were given opportunities to work with textbook problems and were assigned homework from the textbook, all participants were provided with opportunities to work with real data. However, the students in the control group only dealt with real data provided through archived data sets in the course textbook whereas the

students in the treatment group, in addition and prior to working with archived data sets, performed experiments in which they generated real data themselves in small groups.

1.5.4 Cooperative Learning

Cooperative learning is defined as a "structured, systematic instructional strategy" (Cooper & Mueck, 1990, p. 68) in which students are assigned specific roles and work towards a common goal while being responsible for their own learning. Cooperative learning makes use of small groups of two to ten students (Springer, Stanne, & Donovan, 1999) who work on assignments, in-class activities or problems until all group members understand the material (Gunawardena, 1998).

Research identified five elements of cooperative learning groups. First, such situations are characterized by *positive interdependence*; students perceive that they must make a joint effort and that each member has a unique contribution they can make to the group (Johnson & Johnson, n.d.; Johnson & Johnson, 1985; Effandi and Zanaton, 2007). Positive interdependence may be structured by asking group members to fulfill specific assigned roles, to agree on an answer for the group or by giving the group a shared grade (Smith, Douglas and Cox, 2009).

A second element is that of *individual and group accountability*; the group is accountable for achieving the set goal and each member is accountable for making a contribution to the group (Johnson & Johnson, 1985). Individual accountability may be structured by giving individual exams (Effandi and Zanaton, 2007). In order to achieve group accountability, the instructor may provide feedback to the group regarding their performance on the activity or problem.

Third, cooperative learning promotes *face-to-face interactions* among students. The instruction must ensure that students interact with one another in order to help each other

complete the task, explain orally to one another how to solve problems and discuss the strategies used (Smith, Douglas and Cox, 2009).

The fourth component of cooperative learning groups is that of *teamwork skills*. Such skills may be introduced by assigning students different roles in their groups (Smith, Douglas and Cox, 2009). The last element of cooperative learning groups is that of *group processing*. In structuring group processing, instructors should provide students with a specific (rather than vague) task which is complex enough to warrant a group, and also provide sufficient time for them to work in groups.

In this study cooperative learning was used as part of instruction in all course sections (control and treatment). The small student groups were set up by the instructor based on criteria that are described in Chapter 4: Methods of this dissertation. In the treatment group cooperative learning was promoted through the use of small-group in-class activities. In the control group cooperative learning was promoted through the use of small-group sessions during which students solved probability problems. In all sections of the course, students were assigned specific roles and received a shared grade for work completed in groups during class (*positive interdependence*). Assigned roles aimed at promoting *face-to-face interactions* among group members as they completed an activity or solved a problem. Groups received feedback on their performance and a shared score for the group work submitted to the instructor (*group accountability*) but were examined individually on the course final exam (*individual accountability*). In addition, students were assigned different roles and these were rotated between activities/problem sessions (*teamwork skills*). Moreover, activities were well-structured (treatment group), problem sets were specifically selected from the course textbook (control group) and students were provided with sufficient time to work on these (*group processing*).

1.5.5 Achievement and Understanding

Although an extensive literature can be found relative to *mathematical understanding*, a definition of *probabilistic understanding* has not been explicitly established. Researchers defined terms such as *probabilistic reasoning* and *probabilistic thinking* both of which entail the idea of *understanding*. A detailed review of the mathematical meaning of understanding along with definitions of probabilistic reasoning and thinking, as well as the available models for examining students' probabilistic reasoning are provided in Chapter 2 of this dissertation.

As Dewey (1933) indicated, "to understand is to grasp meaning" (p. 132) and this *meaning* may be acquired by seeing something "in its *relations* to other things" (p. 137). In a similar manner, Thompson and Saldanha (2003) defined understanding as "assigning meanings according to a web of connections the person builds over time" (p. 99). However, Thompson and Saldanha indicate that these connections are constructed through interactions that the person has with "with his or her own interpretations of settings and through interactions with other people" (p. 99). According to Skemp (1971), "to understand something means to assimilate it into an appropriate schema" where a schema is "a conceptual structure" (as cited in Resnik & Ford, 1981, p. 167). In this study, the extended meaning of understanding given by Thompson and Saldanha (2003) of "assimilation to a scheme" is used; this allows for correct as well as incorrect or inappropriate understandings (i.e. misconceptions) people may have.

In statistics education the term probabilistic *reasoning* is more prevalent as compared to the term *understanding*. In particular, Jolliffe (2005) defined *probabilistic reasoning* as "understanding and being able to explain and justify probabilistic processes" (p. 326). Garfield (2003) defined *reasoning about uncertainty* as "[U]nderstanding and using ideas of randomness, chance, likelihood to make judgments about uncertain events; knowing that not all outcomes are

equally likely; knowing how to determine the likelihood of different events using an appropriate method" (p. 25).

The definition of understanding used in this study combines i) the definitions of understanding provided by Dewey (1933) and Thompson and Saldanha (2003) and ii) the definitions of probabilistic reasoning expressed by Garfield (2003) and Jolliffe (2005). Therefore, *students' understanding of probability* in this study means to grasp the meaning of probability concepts by viewing them in relation to other concepts and to be able to explain the processes used in the solution of probability problems or in the completion of probability activities as students interact with one another in small groups. Such an understanding includes, as Garfield stated, knowledge that not all outcomes have an equal probability of occurring and the ability to use an appropriate method to determine the probability of an event. Students' understanding of probability includes both appropriate as well as inappropriate understandings. Therefore, two things needed to be specified for each item on the pre-test and post-test used in this study with regards to probabilistic understanding:

i) What is the goal of the item? That is, what probability content or concept is the item addressing and the student trying to understand?

and

ii) What has the student actually understood or what misconception does he/she hold relative to the particular item?

In this dissertation, achievement is measured quantitatively by considering students' responses to a probability pre-test and post-test. Understanding is measured in two ways: i) a distractor analysis of student responses to the multiple-choice items on the pre-test and post-test and ii) a qualitative analysis of audio-taped conversations as students worked in groups on

activities or problem sets. A distractor analysis, by itself, provides useful, but not sufficient, information on students' understanding since it fails to offer a complete view of what understanding is. This is so because multiple-choice items do not provide access to a responder's reasoning and subsequently, to a responder's explanation as to why he/she chose a particular response. Without access to a responder's reasoning, an incorrect response to a multiple-choice item does not directly indicate the use of a particular misconception; it simply indicates the "application" of a given misconception (J. P. Smith, personal communication, December 14, 2011). Therefore, both a distractor analysis and a qualitative analysis of student conversations were carried out in the attempt to gain insight on students' understanding of probability in this study.

1.6 Overview of Methods

This dissertation took place in an introductory statistics course taught by the researcher in spring 2010 at a college in Cyprus. The study aimed to examine the effects of two instructional methods (as described in section 1.3) on college students' achievement and understanding of experimental and theoretical probability. These two methods were used during instruction on the probability component of the course.

Participants included 44 college students split among three sections of the course. The majority of these students used Modern Greek as their first language and English as their second language. The medium of instruction in the course was English. Two of the sections served as the treatment group whereas one section served as the control group.

In order to address the research questions posed in this study, a mixed-methods design was used. A pre-test and a post-test were given to participants. Students in the treatment group worked in small groups on four in-class activities about experimental and theoretical probability,

and students in the control group worked in small groups on solutions to four sets of probability problems. In each group student conversations were audio-recorded. These data sources allowed for both quantitative and qualitative analyses of data. All of the activities used in the treatment group had an experimental and a theoretical component and all included parts which required students to compare experimental and theoretical probabilities (see Appendix C). This provided the means for emphasizing the connection between theoretical and experimental probability for this relation is not directly apparent to students (Jones, Thornton, Langrall and Tarr, 1999). Yet, "[I]t is that reciprocal dynamic of theoretically computed probabilities and observed relative frequencies that may best contribute to the development of efficient probabilistic intuition" (Fischbein and Gazit, 1984, p. 3).

Specifically in this dissertation, *achievement* is measured quantitatively through the performance of students on a probability pre-test and post-test, using statistical analysis methods as described in Chapter 4. *Understanding* is measured through a distractor analysis of multiple-choice student responses to the pre-test and post-test items as well as through a qualitative analysis of audio-taped student conversations as students worked in small groups on activities or problem sets. The framework for the qualitative analysis of the audio tapes is provided in Chapter 2 and the way in which it was used is presented in Chapter 6.

1.7 Overview of Chapters

The study is organized into seven chapters. *Chapter 1: Introduction* includes the rationale for conducting this research study along with the research questions and definitions of terms used. In *Chapter 2: Literature Review* a thorough review of the available literature on issues related to this dissertation is given. Specifically, the following are reviewed: i) probability misconceptions held by secondary school students, college students and adults; ii) the use of the

experimental approach and real data in teaching and learning probability; iii) the effects of cooperative learning and iv) the literature on mathematical and probabilistic understanding. Also, the framework used in this study is presented. *Chapter 3: Cyprus Educational System* provides an overview of the Cyprus educational system since the study was conducted at a college in Cyprus. Moreover, a detailed account is provided on educational reform efforts currently taking place in Cyprus specific to the area of mathematics and to the teaching of probability. *Chapter 4: Methods* provides a description of the research design, participants, procedures and instruments used in the control and treatment groups as well as the types of analysis carried out. *Chapter 5: Students' Achievement and Understanding of Probability – Results of Quantitative Analysis* is a presentation of the results obtained through the quantitative analysis of the data whereas *Chapter 6: Students' Understanding of Probability – Results of Qualitative Analysis* is a presentation of the results obtained through the qualitative analysis of the data. Last, *Chapter 7: Discussion and Conclusions* presents a discussion of the results along with conclusions drawn from the study. This last chapter includes the strengths and limitations of this dissertation as well as recommendations for further research.

CHAPTER 2

LITERATURE REVIEW

This study is informed by various bodies of literature. First, an examination of the available research on students' probabilistic reasoning is included in this chapter. This section focuses on the use of heuristics as well as the difficulties and misconceptions held by secondary school students and adults. Since the participants in this study were beginning college students, it was expected that they would use such heuristics, encounter such difficulties and carry probability misconceptions as those listed in the literature. Second, a review of research concerning the use of experiments that generate real data in mathematics and statistics classrooms at the upper secondary and undergraduate levels is provided. This second section is related to one of the instructional methods implemented in this dissertation. Third, the effects of using cooperative learning are discussed since this learning approach was employed in this study. Next, the literature on the mathematical meaning of *understanding* is reviewed along with related terms defined in the probability literature since this study examined students' *understanding* of probability. Last, the theoretical framework used in this study is described.

2.1 Probabilistic Reasoning and Understanding

Although *understanding* is a term widely used and extensively defined in mathematics education, in statistics education the term *probabilistic reasoning* is more prevalent. Some research on probabilistic reasoning focused on school children (Green, 1982a; Shaughnessy, 1981, 1992, 2003; Watson & Moritz, 2002) whereas another line of research focused on the reasoning of college students and adults (Kahneman, Slovic, & Tversky, 1982; Konold, 1989, 1993, 1995). This review focuses on the probabilistic reasoning – difficulties, misconceptions, and use of heuristics - of secondary school students and adults.

People often fail to use the rules and methods learned in statistics courses when making decisions under conditions of uncertainty thus resulting in misconceptions. "A misconception is a student's erroneous concept that produces a systematic pattern of errors" (Khazanov, 2008, p. 180). Such misconceptions act as obstacles in students' ability to master probability concepts. The literature reviewed below suggests that probability misconceptions are resistant to change (Konold, 1995). The issue is that, when faced with a correct concept, one does not immediately perform a "complete replacement" of the erroneous one with the correct one (Vosniadou and Verschaffel, 2004; as cited in Khazanov, 2008, p. 182). The likely scenario is that the misconception is attached to the correct concept, leading to student inconsistencies in probabilistic reasoning (Konold, 1995).

At other times, people use "intuitive strategies (often referred to as heuristics) for dealing informally with situations that might otherwise be formally analyzed in terms of probability" (Pratt, 2000, p. 603). These heuristics may lead to quick and reasonable conclusions but may also result in judgment that comes at odds with probability theory. Commonly used heuristics relative to probability are discussed in detail in the literature review that follows.

2.1.1 Research on Heuristics

Research on probabilistic reasoning has been carried out by educators as well as by psychologists such as Kahneman, Slovic, and Tversky (1982) who pursued an area of inquiry labeled as *judgment under uncertainty*. Participants in the series of studies that these psychologists completed were 1500 college and college-preparatory students. Under the line of inquiry pursued by Kahneman, Slovic, and Tversky (1982), humans make decisions using heuristics such as *representativeness* and *availability* in their effort to estimate the probability of an event or to compare the probabilities of outcomes. For Kahneman and Tversky (1972),

heuristics are "strategies that statistically naïve people use to make probability estimates" (see also Jones and Thornton, 2005, p. 74). These strategies can be helpful but may also lead to misconceptions.

According to the *representativeness* heuristic, an event is probable to the extent that it is representative of the population from which it is drawn or to the extent that it is characteristic of the process that generates it. For example, people tend to think that the sequence of coin tosses HTHTTH is more likely than HHHTTT because the latter does not appear to be random, and that HTHTTH is also more likely than the sequence HHHTHH because the latter is not characteristic of the fairness of coins (Kahneman et al., 1982). In particular, the results of Kahneman, Slovic and Tversky's (1982) experiment revealed that 82% of students indicated that the birth order BGBBBB was less likely than GBGBBG. Using a situation involving coin tossing, Konold et al. (1993) found that 38% of 7^{th} graders and 33% of 11^{th} graders used representativeness. Moreover, research carried out by Batanero, Serrano and Garfield (1996) with 137 fourteen-year-old students who had not previously studied probability and with 130 eighteen-year-old students who had received formal instruction on probability indicated that representativeness was widely used among students in both groups. Furthermore, Fast (1997; see Carter, 2005) used a similar question to that of Kahneman et al. involving coin tosses and found that one third of the participating pre-service teachers did not give the correct answer. This was in agreement with results of a study by Ulep (1990) with pre-service secondary mathematics teachers.

It has been noted though that as students age, the use of the representativeness "misconception" decreases (Fischbein and Schnarch, 1997, p. 101). Moreover, research by Hirsch and O'Donnell (2001; see Shay, 2008) with 263 college students provided evidence that the use of representativeness decreased based on the number of statistics courses taken. In

particular, 37.5% of participants who had taken two or more statistics courses used this heuristic (Shay, 2008).

The representativeness heuristic manifests itself in the *gambler's fallacy* or *negative recency* in which one believes that a particular outcome is due to occur because it has not done so for a while. Under the negative recency effect the person believes "intuitively that the alternating outcomes seem to better represent a random sequence", just like "the gambler believes the events will balance at the end" (Bamberger, 2003, p. 28). On the other hand, under the *positive recency* a person predicts that an outcome that has occurred repeatedly in the past will keep occurring (Jones and Thornton, 2005). Fischbein and Schnarch (1997) found that negative recency decreased with age whereas positive recency remained stable with age.

The second judgmental heuristic, namely *availability*, states that people think the probability of an event is based on how easily occurrences of similar events can be brought to mind. For example, when people using this heuristic are asked to estimate the probability of heart-attack among middle-aged people, they tend to base their estimate on instances of heart-attack among their middle-aged acquaintances (Kahneman et al., 1982). So, a person who has more acquaintances who are middle-aged and had a heart-attack would give a higher probability estimate than someone who does not have as many acquaintances with these characteristics. In a study carried out by Fischbein and Schnarch (1997) with 20 students in each of grades 5, 7, 9, and 11 as well as with 18 undergraduate pre-service mathematics teachers, it was found that the frequency of the availability "misconception" grew stronger with age (p. 102). Moreover, Ulep (1990) found that the availability heuristic was widely employed by pre-service secondary mathematics teachers.

A third heuristic accounted for in the work of Kahneman and Tversky was that of *adjustment and anchoring*. Under this heuristic, people start from an initial value (an anchor), adjust it according to information given in the problem and make an inadequate probability estimate (Tversky and Kahneman, 1974; see Jones and Thornton, 2005). There are two instances to adjustment and anchoring: the *conjunction fallacy* and the *disjunction fallacy*. For example, when asked which has the bigger chance, two consecutive sixes on a die or a single six on the next roll of a die, someone using the conjunction fallacy would say that the first one has the bigger chance (Jones and Thornton, 2005). That is, people using the conjunction fallacy think that a compound probability can be higher than the probability of each single event whereas the probability law states that $P(A \cap B) \leq \min(P(A), P(B))$. In contrast, someone using the disjunction fallacy would say that getting a single six on the next roll of a die has a bigger chance than getting at least one six in three rolls. Fischbein and Schnarch (1997) found that the conjunction fallacy was very strong among students up to grade 9 and tended to be less strong among high school and college students. In a study carried out by Diaz and de la Fuente (2007) with 414 students in a college introductory statistics course, the conjunction fallacy was observed in 71% of the responses (p. 136).

Although the use of heuristics may result in misconceptions, it can also lead to reasonable estimates that are correct enough for practical purposes (Hirsch & O'Donnell, 2001; Shaughnessy 1992). For example, availability is helpful in matters of decision-making whereas representativeness is central to statistics where the aim is for a random sample to be representative of the parent population so that the results of a study may be generalizable (Borovcnik, 1986; see Shaughnessy, 1992).

> Our task as mathematics educators is to point out circumstances in which judgmental heuristics can adversely affect people's decisions, and to distinguish these from situations in which such heuristics are helpful. We are obliged to point out this difference to our students; it is not that there is something wrong with the way our students think, just that they – and we – can carry the usefulness of heuristics too far (Shaughnessy, 1992, p. 479).

The above statements as well as arguments brought forward in the last decade by Stanovich (1999) and Gigerenzer (1991a, 1991b; see Stanovich 1999) point to alternative interpretations of what Kahneman, Slovic, and Tversky (1982) demonstrated as "systematic irrationalities" in human cognition. In particular, Gigerenzer (1991b) criticized the use of heuristics to account for errors because i) they are errors "only from a particular narrow and challengeable interpretation of probability"; ii) the explanations the heuristics attempt to provide are a "little more than redescriptions of the phenomena they are intended to explain"; and iii) the heuristics "are largely undefined concepts and can post hoc be used to explain almost everything" (p. 102).

Stanovich and West (2000; as cited in Stanovich and West, 2003) argue that "research in the heuristics and biases tradition has not demonstrated human irrationality at all and that a Panglosian position … which assumes perfect human rationality is the proper default position to take". According to the Panglosian position, human irrationality is impossible although human behavior may, in some instances, deviate from a normative model. Such deviations in human behavior may be explained by: i) *performance errors* which may be attributed to the subject's lack of attention, memory lapses, or other minor psychological malfunctions; ii) *incorrect norm application*; the experimenter has implemented the wrong normative model so, the problem lies with the experimenter rather than with the subject; and iii) *alternative task construal*; the subject

interprets the task differently than the experimenter and thus responds to a different problem (Stanovich, 1999).

As Nickerson (2004) points out the use of heuristics has gains as well as pitfalls. As long as the heuristics work well in the situation they are applied, they may require simpler computations and only a modest effort in addressing a problem, however, precision and consistency of results are lost.

2.1.2 Research on Probabilistic Difficulties and Misconceptions

2.1.2.1 Law of Large Numbers

Piaget and Inhelder (1975) support that by age 12 children are fully aware of the effect of sample size and the role of the Law of Large Numbers. In the experiments used by these researchers, children were shown a box made up of two equal-sized compartments, with a funnel at the top middle part. Balls were dropped in the box and children were asked whether a large sample or a small sample of balls would result in a uniform distribution. Children indicated that the larger sample was more likely to generate such a distribution (see Sedlmeier and Gigerenzer, 1997). Piaget and Inhelder concluded that children possessed an intuitive understanding of the Law of Large Numbers. In agreement with this, Evans and Pollard (1985) concluded that

> Overall subjects did quite well as intuitive statisticians in that their judgments tended, over the experiments as a whole, to move in the direction required by statistical theory as the levels of mean difference, sample size and variability varied (p. 68-69).

An analysis of the data collected by Zimmerman (2002) from 23 students in a high school Advanced Placement Statistics class indicated that participants were able to recognize the effect of many trials on empirical probability. “They recognized that with enough trials, the probability would fluctuate less and settle towards a theoretical value” (p. 167). In addition, use of a graphing calculator had a considerable effect on students’ ability to reason about probability

simulations during instruction. The graphing calculator enabled students to perform many trials quickly "which in turn led to a deeper understanding of the long-run effects to empirical probability" (p. 168). These results were in agreement with those of Aspinwall and Tarr (2001) who examined the effects of an instructional program that made use of manipulatives in carrying out simulations. Similar to the Zimmerman (2002) study, subjects were able to recognize the effect of repeated trials on the probability of an event and developed a conceptual understanding of the Law of Large Numbers.

However, another group of studies supports a different view than the one presented above. In a study carried out by Fischbein and Schnarch (1997) with students in grades 5, 7, 9, 11 and prospective teachers, participants were asked to respond to two problems relating to the effect of sample size. The researchers indicated that the misconception that sample size is irrelevant was prominent among responses and that this misconception grew with age.

Jones and Harris (1982) investigated university undergraduate students' understanding of the Law of Large Numbers by means of three different tasks. The first task was similar to that used by Piaget and Inhelder (1975) and involved the distribution of marbles between two compartments; the second was a set of written tasks similar to those used by Kahneman and Tversky (1972); and the third asked students to compare samples of counters drawn from a parent population (see Jones and Harris, 1982). Participants performed well on the Piagetian task but tended to ignore the relevance of the Law of Large Numbers when faced with tasks similar to those used by Kahneman and Tversky, thus confirming the results of the latter. However, participants performed significantly better on the tasks similar to those of Kahneman and Tversky if they had first been tested on the Piagetian task. "Therefore, failure on the Kahneman

and Tversky task is typically due, not to a complete lack of insight, but to a failure to apply that insight to a relevant instance" (Jones and Harris, 1982, p. 487).

Little is known about the explanations behind these inconsistencies in research results relative to people's attendance to sample size. For instance, Bar-Hiller (1979) suggested that a factor influencing the use of sample size is that people tend to attend to "relative sample size (relative to the population size)" and not "absolute" sample size (see Sedlmeier and Gigerenzer, 1997, p. 13). However, the results of her study were inconsistent and therefore, did not entirely support this claim.

Sedlmeier and Gigerenzer (1997) carried out a review of studies relative to the use of sample size which led them to distinguish between two types of tasks which had rarely been distinguished in the literature: *frequency distribution tasks* and *sampling distribution tasks*. "A frequency distribution is a distribution of values from *one* sample" whereas a sampling distribution is "a distribution of means from independent samples of fixed size, drawn from the same population" (p. 4). In frequency distribution tasks people judge how well the sample mean estimates the population mean; in sampling distribution tasks people deal with the variance of sampling distributions. This distinction helped explain the inconsistencies in the results of these studies. Sedlmeier and Gigerenzer (1997) noted that "the distinction between frequency and sampling distribution tasks was shown to be a strong predictor of participants' use of sample-size information" (p. 8). Moreover, when asked to construct sampling distributions, people tend to construct frequency distributions.

2.1.2.2 Other Basic Probability Concepts

Fischbein, Nello and Marino (1991) undertook a study in which 618 students aged 9-14 were asked to identify various types of events. The results indicated that students tended to

confuse ‘certain’ and ‘possible’ events because, as the researchers explained, “usually, one tends to relate the notion of ‘certain’ to that of ‘uniqueness’” (p. 527). Similarly, Chan (1997; see Li, 2000) found that 35% of the 425 participants aged 16-18 confused ‘certain’ and ‘possible’ events. Based on questionnaire responses and student interviews he concluded that this confusion might be due to students’ misconception that i) a ‘certain’ event is the same as a ‘possible’ event since both types of events are indicative that something will happen or ii) each type of event consists of a single outcome but the student was unable to think of an outcome that would happen for sure in a given situation such as in rolling a fair die (Li, 2000).

The linguistic difficulties associated with the terminology used for probability events are also highlighted in a study by Green (1982a) in which participants included 2,930 students aged 11-16. The results indicated that students tended to think of i) ‘very likely’ to mean ‘certain’ or ‘always happens’ and ii) ‘not very likely’ and ‘unlikely’ to mean the same as ‘impossible’ or ‘cannot happen’. Students who tended to think that ‘low probability’ means ‘impossibility’ were surprised when an outcome with a low probability still occured in an experiment (Konold, 1988; see Ulep, 1990).

A study carried out by Green (1979, 1982b) with 3,000 students aged 11-16, examined students’ thinking on probability concepts such as randomness, sample space, most likely event, compound events and independence. Students at all levels exhibited difficulties distinguishing between random and nonrandom distributions as well as with items relating to sample space. Also, students had problems with the multiplication principle. In addition, Green (1983) asked participants to decide which one of Bag A – containing 3 black and 1 white counters – or Bag B – containing 6 black and 2 white counters – gives a bigger chance of picking a black counter. More than 50% of the students chose Bag B because the bag had more black counters. Green

concluded that students relied on the *absolute* size instead of the *relative* size of outcomes when comparing the likelihood of events. Alternatively one could argue that the students did not understand proportional reasoning.

Furthermore, in a study by LeCoutre (1992) with 600 students, participants were asked to decide which of two colors of marbles is more likely to be picked up from a bag containing two red and one black marble. Many indicated that both colors had the same chance of being picked up. LeCoutre labeled the students' belief that all outcomes are equally likely as the "equiprobability" bias; a bias that in the particular study persisted among children of all ages regardless of their exposure to probability. The results of a study carried out by Albert (2003) also revealed that college students in an introductory statistics course made use of the equiprobability bias. The study took place before the participants received any formal instruction in probability so, the results were indicative of their prior knowledge. The equiprobability bias has also been exhibited among students across various cultures (Batanero, Serrano, & Garfield, 1996; Jones, Langrall and Mooney, 2007). This bias could be explained by students' commonly held belief that in order for an experiment to be "fair", all outcomes need to be equally likely (Jacobs, 1999; Shaughnessy, 2003).

Yet, although students hold the "fairness" belief, they sometimes tend to believe that although all numbers on a die have an equal chance of occurring, the number 6 is the least likely to occur when rolling a die (Konold et al., 1993). When students in a study carried out by Green (1983) were asked to choose a number on a fair die that would give them the bigger chance of winning a prize, the majority chose middle numbers such as 3 or 4 and avoided numbers 1 and 6.

Green (1982b) found that students often consider experimental outcomes to be equally probable even though relative frequencies obtained from actually carrying out the experiments

show differently. In support of this issue, Shaughnessy (1977) found that students carry out such an assignment of probabilities even when they perform the experiment themselves. In his study, students performed many trials of an experiment and concluded that outcomes cannot be equally likely. However, when they were requested to provide a probability model for the experiment they assigned an equal probability to each outcome, claiming that theoretically the outcomes are equally likely and pointing towards the belief that "every probability model is a uniform model" (Shaughnessy, 1977, p. 303). In general, Green (1982b) found that students assign equal probabilities to outcomes when they should not, and at times they assign unequal probabilities to outcomes that are indeed equally likely.

A substantial amount of research on probabilistic thinking among college students has been carried out by Konold (1989, 1993, 1995). An interview study with undergraduate students indicated that people reason according to the *outcome approach* in which the goal seems to be predicting the outcome of the next single trial (Konold, 1989). For example, when students were asked to interpret the meaning of "70% chance of rain today" they predicted that "[I]t's going to rain today" (p. 68). Individuals who use the outcome approach tend to interpret probability values as follows: a 50% chance means a total lack of knowledge about the outcome thus leading to a "don't know" decision; a probability value sufficiently higher than 50% (close to 75%) leads to a "yes" decision; and a probability value sufficiently lower than 50% leads to a "no" decision (Konold 1989, 1995). If the probability is close to 0% students view the event as impossible and if the probability is close to 100% they view the event as certain. Only when the probability is close to 50% they view an event as random. Ulep (1990) also found that pre-service secondary mathematics teachers made use of the outcome approach.

Related to the outcome approach and equiprobability bias is a misconception identified by Rubel (2006), namely the *50/50 approach*. One of the questions that the 173 male participants in grades 5, 7, 9, and 11 were asked to respond to involved the probability of getting one head and one tail when two coins are tossed. More than half of the students provided the correct answer of ½. However, when asked to justify their response they used the outcome approach, predicting the probability of getting heads or tails on a single toss of a coin. In a follow-up interview with 33 of the participants, one of the questions asked for the probability of getting all tails when three coins were tossed. One of the students who said 50% gave the justification that "unless something affects the way the quarters come down, it's still going to be equal" (p. 52). Overall, 40% of participants used the 50/50 approach on at least two questions.

Students also tend to find the concepts of independence and mutual exclusivity difficult. Based on the results of a survey of 219 participants enrolled in a college-level introductory probability and statistics course, "students tend to impose an artificial separation between the concepts of mutual exclusivity and independence" (Manage and Scariano, 2010, p. 16). The survey was administered at the end of the semester, after students had received formal instruction on the two concepts. A substantial percentage (68.3%) of them believed that if two events are mutually exclusive then they are independent; that is, they held the misconception that "independence" means "separation" (Manage and Scariano, 2010, p. 16). Also, 36.1% of the students thought that if two events are independent then they are mutually exclusive. According to the researchers, the reason for the decrease in the percentage (68.3% to 36.1%) is that when students come across the term "mutually exclusive" they immediately think of "separation" as "independence" whereas when faced with the term "independence" no such immediate intuition comes to mind.

Antoine (2000) also found that the 66 liberal arts students in a finite mathematics course tended to confuse independence with mutual exclusivity and vice versa, and had difficulty providing examples illustrating each of these types of events. When students were asked to indicate whether events were independent based on given probability values, they tended to apply previously learned definitions incorrectly into formulas. For example, some students considered the union and intersection of events to be the same. Moreover, students focused on key words which had a non-mathematical meaning. In one of the items two events were defined as: "Event A: Getting at least one tail and Event B: Exactly three tails" (p. 61). A typical response to whether these were independent was: "Events are dependent because the word "tails" can be found in the intersection (overlap) of events A and B" (p. 61). An issue that came up was that arithmetic and algebra skills played a role in students' ability to solve problems on independence since many of them incorrectly used principles of multiplying with exponents and checking the equality of equations.

2.1.2.3 Conditional Probability

Research was also carried out with regards to students' reasoning about conditional probability (Pollatsek, Well, Konold & Hardiman, 1987; Watson & Moritz, 2002). Watson and Moritz (2002) investigated probability or frequency estimates that students in grades 5-11 give for a conditional event and its inverse. For example, their research included items such as the following:

> Please estimate:
> (a) Out of 100 men, how many are left-handed?
> (b) Out of 100 left-handed adults, how many are men?
> Please estimate:
> (a) The probability that a woman is a school teacher.
> (b) The probability that a school teacher is a woman. (p. 66)

Student performance on the conditional frequency item was higher than on the probability item. Watson and Moritz claimed that the grammatical structure of the items affected performance with the statement *out of* in the frequency item leading students to the relevant conditions. This claim was in agreement with the findings by Pollatsek et al. (1987) who investigated college students' reasoning about conditional probability and found that they confused a conditional event and its inverse.

Relative to research on biases associated with conditional probability, the work of Falk (1979, 1989; see Batanero and Sanchez, 2005) is of significance. In particular, Falk demonstrated the *fallacy of the time axis* using the responses of 88 university students to the following problem:

> An urn contains two white balls and two red balls. We pick up two balls at random, one after the other without replacement. (a) What is the probability that the second ball is red, given that the first ball is also red? (b) What is the probability that the first ball is red, given that the second ball is also red?

Although students performed well in part (a), they were confused with part (b) because "they thought that an event couldn't condition another event that occurs before it" (Batanero and Sanchez, 2005, p. 251). These students confused conditional with causal thinking. In particular, they claimed that the answer to part (b) was 0.5 because the second event does not affect the first since the second ball had not been drawn at the time of drawing the first ball (Falk, 1979, 1989; see Batanero and Sanchez, 2005).

Falk (1986, 1988) pointed out a number of reasons as to why students find situations that involve conditional probabilities difficult. One of the reasons was related to the grammatical structure of items as also indicated by Pollatsek et al. (1987) and Watson and Moritz (2002), whereas other reasons included: i) students' difficulty recognizing the conditioning event; and ii)

students' confusion of conditionality versus causality leading to a confusion of an event and its inverse. That is, people often confuse P(A|B) and P(B|A) which Falk (1986; see also Diaz and de la Fuente, 2007) termed the *fallacy of the transposed conditional*.

Another study related to misconceptions about conditional probability was carried out by Gras and Totohasina (1995; see also Batanero and Sanchez, 2005). Participants included seventy-five 17 to 18-year-old secondary school students. Responses were indicative of three misconceptions relative to conditional probability: i) the *chronological conception* in which students view that in P(A|B), the conditioning event B should always precede event A; ii) the *causal conception* in which students view that the conditioning event B is the cause and event A is the consequence; and iii) the *cardinal conception* in which students view that $P(A \mid B) = \frac{Card(A \cap B)}{Card(B)}$. It should be noted that the cardinal conception is correct in the case of an equiprobable sample space. Gras and Totohasina suggest that the chronological and causal conceptions have a cognitive origin whereas the cardinal conception is brought on through instruction (see Batanero and Sanchez, 2005).

Research carried out by Diaz and de la Fuente (2007) with 414 freshmen in a college-level introductory statistics course revealed that receiving formal instruction on probability was unrelated to some of the biases relative to conditional probability. The participants had studied conditional probability in secondary school and received two weeks of formal instruction on conditional probability prior to completing the study questionnaire. Based on the results, even after formal instruction, students persistently confused independence with mutual exclusiveness and used the chronological conception of independence. Moreover, 31% of the students confused conditional with joint probability and 59% employed the confusion of the transposed conditional

(p. 136). Although some biases remained persistent, some others improved. For example, after instruction participants were able to easily compute "without-replacement" conditional probabilities.

2.1.2.4 Binomial Distribution

Several researchers carried out studies pertaining to people's conceptualizations of the binomial distribution. Stokes (1972; see Krause, 2001) taught the two components of the binomial distribution formula separately with reference to a fictitious substance named "quirnon", without teaching anything else about probability. Students were asked to respond to a problem which was then revised into a second version as follows:

> "$\frac{N!}{r!(n-r)!}$ = The number of different sequences of alpha and beta particles
>
> $p^r \cdot (1-p)^{n-r}$ = The probability of a sequence"
>
> Version 1: "If a quantity of quirnon emits eight particles what is the probability that four of them will be alpha particles and four of them will be beta particles?"
>
> Version 2: "If a quantity of quirnon emits eight particles, what is the probability that any one sequence out of the set of all sequences that contains four alpha particles and four beta particles will occur?" (as cited in Krause, 2001, p. 31)

When the first version of the problem was used, most students were unable to provide a solution whereas when the second version was used most of them solved it. Krause criticized the study indicating that i) it provided "no useful information about the students"; ii) since nothing else, apart from the particular procedure, was taught about probability, there was "no reason for Ss to realize that in the first problem they should take into consideration all of the possible permutations"; and iii) the use of the term "sequences" in the second version "influences the interpretation of the problem by cueing the procedure for counting the number of different permutations" and thus, Stokes "communicates the "deep structure" of the problem" to the students whereas students should bring the "deep structure" to the problem themselves (p. 32).

In a later study, Fischbein (1975) used the following problem with 13, 15, and 17 year-olds: "A warehouse contains 90% ripe apples, and 10% green apples. If three apples are taken at random, what is the probability that two of them will be ripe?" Although, as Fischbein indicated, participants had the necessary knowledge to solve the problem (that is, they were aware of the addition law and the multiplication law), only 60% provided a solution. The common response was $\left(\frac{9}{10}\right)^2 \cdot \frac{1}{10}$ instead of $3 \cdot \left(\frac{9}{10}\right)^2 \cdot \frac{1}{10}$ leading Fischbein to conclude that participants lacked an intuition of a compound event. Since Fischbein did not clearly state the subjects' level of prior knowledge and math experience, an alternative explanation for this result may be that students had 'translation' issues and misinterpreted the problem as asking for the probability of a "single outcome" (Konold, 1989; Krause, 2001). Fischbein simply stated that students possessed the necessary knowledge to tackle the problem but "what he probably meant was that they knew the procedures needed to calculate the correct answer" which does not necessarily mean that subjects understood the problem and its relation to these procedures (Krause, 2001, p. 30).

2.2 Probability Experiments with Real Data

In the attempt to aid students in overcoming any probabilistic misconceptions and difficulties they may have, mathematics and statistics educators recommend that instruction on probability be carried out in a way that is different from traditional (lecture-based) instruction (Franklin and Garfield, 2006; MAA, 1998; Shaughnessy, 1977). The NCTM (2000) indicates that "[D]oing mathematics involves discovery" and that students "should learn to investigate their conjectures using concrete materials" (p. 57). Given this statement and considering the increased demand by employers for university graduates who are able to work cooperatively to solve problems in their work environments (Mackisack, 1994), it is essential that college

students are provided with opportunities that help them develop their problem-solving skills as well as their verbal and written capabilities of presenting the results of their solution methods. In order to help students develop such skills, instruction in introductory statistics courses should include the use of experiments and real data collected and analyzed by the students themselves.

The American Statistical Association supports the use of real data sets which are of interest to students in statistics classrooms (Aliaga et al., 2005; Franklin & Garfield, 2006). Among the benefits of using such data sets is that they allow students to appreciate the difference between empirical and theoretical approaches to explaining and predicting phenomena under conditions of uncertainty (Batanero & Sanchez, 2005). Also, working with real data, points towards the need for various probability distributions (Batanero & Sanchez, 2005).

In a study carried out by Amit and Jan (2006) with high-achieving students in grades 6-9 in Israel, game tasks involving coin tossing were used in the acquisition of probability concepts. Participants had no formal background in probability and no formal teaching intervention was implemented during the study. Instead, students built their knowledge of probability through interactions with one another as they worked in triads. As a result of actively participating in games, the students invented their own probability terminology and understanding of probability concepts. In particular, students gained insight that there is a difference between theoretical and experimental probability and that a link exists between probability and sample size.

Research by Shaughnessy (1977) with college undergraduate students in an elementary probability and statistics course at Michigan State University, revealed that these students had a positive attitude towards the course when it was taught in a small-group, experimental activity-based manner. In addition, in comparison to two lecture-based sections of the course, the two experimental sections were more successful at helping students overcome probability

misconceptions. Day-to-day observations by the researcher in one of the experimental sections indicated that "college students can learn to discover some elementary probability models and formulas while working on probability experiments in small groups" (p. 313-314). The author developed and used nine activities in probability, combinatorics, game theory, expected value, and elementary statistics. The students in the experimental activity-based section worked on these activities in groups of four or five students. The activities required students to perform experiments, collect and analyze data, and reach conclusions based on their data. The approach used in the experimental group was that of "problem-solving, model-building" with the instructor circulating among the groups and using "the technique of 'answering' a question with another question" so as to encourage students to think problems through for themselves (p. 299). Shaughnessy (1981) recommends that an activity-based, experimental approach should be used for the study of introductory probability and statistics since, according to evidence he gathered through his research work, "an initial formalistic approach to probability is unlikely to help students overcome misconceptions" (p. 95). Moreover, he suggests that instruction should emphasize the use of simulations as a model-building and problem-solving tool.

In general, Garfield (1995) suggests that students learn better when they are involved in hands-on activities which may be completed in cooperative small groups allowing them to apply what they learn in new situations. Research carried out by Mackisack (1994) with sophomore undergraduate students majoring in mathematics points towards the benefits suggested by Garfield. The students in the Mackisack study were enrolled in a course that included design of experiments and which was taught by the researcher herself. One of the course requirements was that students work in small groups on a class project that involved experimentation in the form of collecting and analyzing data, and presenting results to the whole class. The outcomes of the

study indicated that the group project helped students learn concepts relating to design of experiments that would have otherwise been difficult to learn through standard textbook exercises. At the same time, being directly involved in data collection allowed students to be better able to critique experimental designs and question data interpretations.

In his work, Steinbring (1984; 1991; see Jones and Thornton, 2005) examined probability both from its empirical and theoretical forms and emphasized the connection between the two. Due to this connection, he supported the "simultaneous and mutual development" of the two views of probability (p. 79). According to Steinbring, probability instruction should be carried out as follows:

> … learning begins with personal judgments about a random situation; comparisons are made between the empirical situation and conjectured theoretical models, and finally these comparisons lead to generalizations and more precise characterizations of the random situation (Jones and Thornton, 2005, p. 78-79).

2.3 Cooperative Learning

Research carried out during the last few decades points towards the benefits of instructional methods that use cooperation among multiple students along with student engagement in explaining and reasoning with concepts (Aquilonius, 2005; Forman, 1996; Franklin & Garfield, 2006; Zieffler et al., 2008). This section provides a review of i) the characteristics of cooperative learning as an instructional method; ii) recommendations proposed by professional organizations and researchers that relate to the use of cooperative learning in the classroom; and iii) research outcomes (such as the effects on student attitudes and achievement) of studies carried out in which cooperative learning was used.

As Dewey (1969) stated, the fact is that "all human experience is ultimately social: that it involves contact and communication" (p. 38). An experience is based on the "transaction" taking

place between an individual and his environment; that is, the interactions one may have with other people as they discuss an idea or with the materials used to perform an experiment (p. 43-44). In the area of psychology, Vygotsky put forward the view of learning as a social process and emphasized the role of dialogue in instruction. "The mere exposure of students to new materials through oral lectures neither allows for adult guidance nor for collaboration with peers" (Cole, John-Steiner, Scribner and Souberman, 1978, p. 131).

An instructional approach which is based on Dewey's and Vygotsky's ideas of educative experience as a social process is that of cooperative learning. In such an approach, the instructor is no longer the "boss or dictator" but the "leader" of group activities (Dewey, 1969, p. 59). It should be noted though that assessing student learning under cooperative learning conditions is more challenging (Smith, Douglas and Cox, 2009). To this end, it is recommended that students' group work and group learning be encouraged and supported but that individual learning and performance be assessed (Johnson and Johnson, 2004).

Researchers suggest that when designing tasks which involve the use of small groups the following should be considered: i) the task should challenge all group members but the group, as a whole, should be able to attend to it and complete it; ii) authority should be turned over to the groups; iii) students should be assigned roles but these roles should not be static; and iv) groups should consist of three to four students with different skills (Zawojeski, Lesh and English, 2003; see also Kahveci and Imamoglu, 2007).

The importance of students communicating ideas with their classmates through tasks that require cooperative group work is emphasized by the NCTM (1991, 2000). Students should be provided with opportunities to work with classmates, to make and test conjectures and to listen to explanations given by other students (NCTM, 2000). In this manner, they can learn to reason

through class discussions with fellow students; this may allow them to compare their ideas with those of classmates and help them to modify or strengthen their own reasoning and clarify their understanding thus developing their critical thinking skills (Garfield, 1993, 1995; Giraud, 1997; NCTM, 2000). Since one of the content domains discussed by the NCTM is that of data analysis and probability, opportunities for communication in cooperative groups should be provided to students during instruction on probability.

In a study carried out by Leikin and Zaslavsky (1997) with four low-level ninth-grade mathematics classes, the findings showed that students in the small-group cooperative learning sections tended to be more active and remained more on-task while also having more opportunities to receive help, than students who received instruction in a conventional way (whole-class setting). In most cases that students in the cooperative learning sections asked for help, this was provided in the form of explanations. In this manner, the cooperative learning method allowed low-level students to engage in constructing explanations of the principles needed for the solution of problems. Furthermore, the results showed that through the exchange of knowledge, low-level students may develop their mathematical communications skills. In support of this, explainers in Dansereau's study (1988) learned more than students who received explanations who, in turn, learned more than those working individually.

Relative to these results, a study by Webb, Tropper, and Fall (1995) with 7^{th} graders working in small groups on units on operations with decimal numbers and fractions, revealed that the level of help received by a student from another student in the small group was a significant predictor of constructive behavior (e.g. reworking the problem after receiving help) which in turn was a strong predictor of post-test achievement. In general, student interactions and the promotion of mathematical communication are essential in achieving high quality

mathematical learning (Bishop, 1985; Brown and Campione, 1986; see Leikin and Zaslavsky, 1997).

Results from a study by Dees (1991) carried out in a college remedial mathematics course that included high school algebra and geometry indicated that cooperative learning helped students increase their problem-solving skills. The course consisted of a large lecture section and four laboratory sections to which students were randomly assigned. Two of the laboratory sections served as the treatment group and the other two as the control group. In the treatment group, students were organized with respect to cooperation whereas in the control group cooperation was not actively encouraged (nor discouraged). Students who used cooperative learning demonstrated significantly better results than students in the control group when it came to solving word problems in algebra and writing proofs in geometry.

Chance and Garfield (2002) found that instructional methods that promote active learning, such as cooperative learning, aid students in developing statistical reasoning. Additionally, student misconceptions in probability may be overcome through working in small groups (Gnanadesikan, Scheaffer, Watkins and Witmer, 1997; see also Pfaff and Weinberg, 2009; Shaughnessy, 1977, 1981; Webb, Tropper & Fall, 1995) on activities that confront such misconceptions (Sáenz, 1998). Proponents of the use of small groups indicate that students may learn better from their peers rather than from the instructor and that they may let go of a misconception if they hear a convincing argument from a peer (Shaughnessy, 1978; see Khazanov, 2005; Garfield 1995).

Another reason for advocating cooperative group work is that student achievement may be improved through cognitively demanding tasks set by teachers (Giraud, 1997; Hiebert et al., 2005; Potthast, 1999) which lead to multi-person communication among students and

simultaneously help maintain students' interest (Silver, Mesa, Morris, Star, & Benken, 2009). Dees (1991) found that students in a college remedial mathematics course performed the same under the cooperative learning and traditional methods of instruction on algebra tasks that did not involve complex thinking. However, students using the cooperative learning approach performed better on measures that tested higher cognitive skills. A meta-analysis of research carried out by Webb (1991) on the relationship between verbal interactions and learning in small groups in mathematics classrooms indicated that giving content-related elaborate explanations to teammates had a positive effect on achievement. In 50 out of 64 studies reviewed by Slavin (1995) where the use of cooperative learning in which group rewards were provided, significant positive effects on achievement were found. Meta-analyses which compared students' achievement indicated that "the average student taught by cooperative learning performs better than the average student taught with competitive and individualistic methods" (Johnson & Johnson, 1989; as cited in Potthast, 1999; Johnson, Maruyama, Johnson, Nelson, & Skon, 1981). Furthermore, the results of a study by Giraud (1997) support the hypothesis that "cooperative learning … is especially beneficial for those least prepared for statistics" (paragraph 36).

The relation between cooperative learning and students' attitudes has also been studied. Results from research carried out by Zainun (2001) and Mazlan (2002) indicate that when cooperative learning was used students held positive attitudes towards mathematics (see Effandi and Zanaton, 2007). A meta-analysis of studies relating to small-group learning revealed positive effects on achievement but also on persistence and attitudes among undergraduates in mathematics classrooms (Springer, Stanne, & Donovan, 1999).

Through teaching an undergraduate introductory statistics course, Gunawardena (1998) noticed that traditional instruction (lecturing) was not well suited for the students taking the

particular course at the particular research site. As a result, instead of using a traditional lecture-based instructional format, cooperative learning was used in the course during the subsequent semester. In this study, the first half of each class period was devoted to the introduction of new concepts whereas students worked in groups of three during the second half of the period. Student groups performed such activities as reading the text, understanding the concepts introduced at the beginning of class and solving problems. Students were also expected to work in their groups outside of class to complete homework problems. The results revealed that the new instructional method i) improved attendance; ii) increased the willingness of students to participate in class discussions; iii) helped students learn statistics with less anxiety; and iv) increased student-teacher interactions and student usage of office hours.

In another study carried out by Knypstra (2009) in an undergraduate econometrics course, students were placed in small groups and were encouraged to work actively on specific tasks. Cooperative learning and peer instruction were implemented with students explaining theory and problem solutions to other groups. Course evaluations revealed that students felt that they were more involved and more satisfied in this instructional format as compared to the traditional lecture-based format.

In addition to the above, a study by Keeler and Steinhorst (1995) revealed that when cooperative learning was used, more students completed the introductory statistics course. Research in STEM education suggests that an essential element of success in college is positive peer relationships whereas isolation is one of the best predictors for failure (Smith, Douglas, and Cox, 2009). Among the major reasons for students dropping out of college is the inability to i) form a network of classmates with whom they may work together, solve problems and discuss the material taught and ii) be involved in the classroom (Tinto, 1993). Overall, Shaughnessy and

Bergman (1993; see Benko, 2006) suggest that instruction on stochastics would be improved with the use of small groups in which students are actively involved in the exploration of probability concepts.

2.4 Understanding

The term *understanding* is widely used in the mathematics education community. In statistics education though, the term *probabilistic understanding* has not been explicitly defined. Instead, researchers have established definitions of *probabilistic reasoning* and *probabilistic thinking* both of which entail the idea of understanding. Since this study involves students' understanding of probability, this section presents a review of the literature on mathematical understanding, probabilistic reasoning and probabilistic thinking along with available models for examining students' probabilistic reasoning. Furthermore, the framework used in this study is provided.

The literature includes various forms of understanding such as *instrumental*, *relational*, *conceptual*, *procedural*, *formal*, *algorithmic* and *intuitive*. Some researchers advocate a specific type of understanding (e.g. Skemp, 1978a, 1978b) whereas others claim that "it is the combination and integration of the different types that is empowering" (Kvatinsky and Even, 2002, p. 4). This is especially true for the domain of probability where intuition may be misleading and so, using other types of understanding "could serve as control" (Kvatinsky and Even, 2002, p. 5).

2.4.1 Mathematical Understanding

According to Dewey (1933) "to understand is to grasp meaning" (p. 132) and this *meaning* may be acquired by seeing something "in its *relations* to other things" (p. 137). In his seminal work Skemp (1978a, 1978b) distinguished between three types of understanding in

learning mathematics: i*nstrumental*, *relational* and *logical*. Instrumental understanding involves "rules without reasons" (Skemp, 1978b, p. 9); that is, it reflects students' ability to use a rule when solving a problem without knowing why the rule works. Relational understanding though entails knowing *what* to do as well as *why* the rule works; that is, it entails an understanding between the task at hand and the mathematical content, and requires an ability to explain why the approach works. The third type of understanding, i.e. logical understanding, goes beyond instrumental and relational understanding. It involves both knowledge of what to do and why but in addition, it includes the ability to express the solution of the task correctly in written or symbolic form; that is, it entails knowledge of the conventions of school mathematics (Skemp, 1978a, 1978b).

In his earlier work Skemp (1971) had stated that "to understand something means to assimilate it into an appropriate schema" (as cited in Resnik and Ford, 1981, p. 167) where a 'schema' is "a conceptual structure" (Even and Tirosh, 2008, p. 206). Instrumental mathematics involves the formation of short-term schemas through which students may get to the correct answer more quickly. This type of mathematics makes use of an increasing number of rules, which may be regarded as "degenerate schemas", and which may be used successively in order to get from one step to the next when working on a problem (Resnik and Ford, 1981, p. 168). However, students learning instrumentally are not aware of the relationship between successive steps. On the other hand, schemas formed by relational mathematics are more adaptable to new situations which require conceptual connections (Resnik and Ford, 1981, p. 169).

Skemp advocated the use of relational mathematics and opposed to instrumental mathematics (Even and Tirosh, 2008). However, several mathematics education researchers raised concerns regarding the instrumental-relational dichotomy. For example, Resnik and Ford

(1981) argued that "automaticity of response", which can be obtained by memorizing certain procedures, can be beneficial because it frees up space in the working memory which can be used for more complex procedures. Moreover, Hiebert and his colleagues (Hiebert and Carpenter, 1992; Hiebert and Lefevre, 1986) as well as Star (2005) suggest that both procedural and conceptual knowledge are important components of understanding.

The widespread use of *conceptual knowledge* and *procedural knowledge* can be credited to the influential work of Hiebert and Lefevre (1986). In their book, conceptual knowledge is defined as "knowledge that is rich in relationships … a connected web of knowledge, a network in which the linking relationships are as prominent as the discrete pieces of information" (p. 3-4). Procedural knowledge includes i) knowledge of the symbols and syntax and ii) knowledge of the rules or procedures for manipulating symbols to arrive at a solution to a problem. Silver (1986) argues that educators need to consider the *relationships* among these two types of knowledge and not the distinctions between them because "distinctions are static, yet when one uses knowledge to perform a non-trivial task, the knowledge is used dynamically" (Kaplan, 2006, p. 19).

Although the definitions provided by Hiebert and Lefevre (1986) for conceptual and procedural knowledge have been influential, their use is problematic since they "suffer from a entanglement of knowledge type and knowledge *quality*" (De Jong & Ferguson-Hessler, 1996; Star, 2000; see Star, 2005, p. 408). As Star (2005) indicates,

> The term *conceptual knowledge* has come to encompass not only what is known (knowledge of concepts) but also one way that concepts can be known (e.g. deeply and with rich connections). Similarly, the term *procedural knowledge* indicates not only what is known (knowledge of procedures) but also one way that procedures (algorithms) can be known (e.g. superficially and without rich connections) (p. 408).

Note that in his reference to *procedures,* Star specifies that these are *algorithms*. In his work, he points out that there are various kinds of procedures and the type of connections associated with

each varies (see also Anderson, 1982). Algorithms constitute a set of steps that when followed in a particular order they lead to the solution to a problem. "Algorithms are apparently what Hiebert and Lefevre had in mind when they crafted their definition of procedural knowledge" (Star, 2005, p. 407). In such a case, Star agrees with Hiebert and Lefevre that (algorithmic) knowledge is superficial and not rich in connections.

However, other procedures are *heuristics* which can be "tremendously powerful assets in problem solving" (Star, 2005, p. 407). In particular, the use of heuristics requires that one makes a choice and "wise choices can indicate quite sophisticated and deep knowledge" (p. 407). In the case of probability, a substantial amount of research has been carried out relative to people's use of heuristics which has been reviewed earlier in this chapter (see section 2.1.1). Hiebert and Lefevre's definition of procedural knowledge though did not account for heuristics. Star argues that if *type* and *quality* (two "independent" characteristics of knowledge) are separated then this "allows for the reconceptualization of procedural knowledge as potentially deep" (p. 408). In his words,

> Deep procedural knowledge would be knowledge of procedures that is associated with comprehension, flexibility, and critical judgment and that is distinct from (but possibly related to) knowledge of concepts. (p. 408)

Recognizing that no single term captures entirely all that is entailed in mathematical knowledge and understanding, the National Research Council (2001) referred to *mathematical proficiency* which includes five "interwoven and interdependent" strands: *conceptual understanding*, *procedural fluency*, *strategic competence*, *adaptive reasoning* and *productive disposition* (p. 116). Conceptual understanding means the "comprehension of mathematical concepts, operations, and relations" (p. 116). Students with this type of understanding know more than facts and methods. They are actually able to form connections between new ideas and

their prior knowledge which allow them to reconstruct knowledge when forgotten. An important indicator of conceptual understanding is being able to use various representations for mathematical situations. Procedural fluency involves "carrying out procedures flexibly, accurately, efficiently, and appropriately" (p. 116). The study of algorithms as "general procedures" allows students to see that classes of problems can be solved using the same procedure, pointing towards the fact that mathematics is well-structured (p. 121). For this reason, "some algorithms are important as concepts in their own right" since they provide a link between procedural fluency and conceptual understanding (p. 121). The third strand, i.e. strategic competence, entails the "ability to formulate, represent, and solve mathematical problems" which is similar to what has been termed "problem solving" in mathematics education (p. 116). This strand includes the ability to first understand a problem situation and its key features, and then provide a numerical, symbolic, verbal or graphical representation. When students are faced with non-routine problems for which they do not immediately know a correct solution method, they are required to find a way to understand and solve the problem. Adaptive reasoning is the "capacity for logical thought, reflection, explanation, and justification" while the last strand, productive disposition, is the "habitual inclination to see mathematics as sensible, useful, and worthwhile, coupled with a belief in diligence and one's own efficacy" (p. 116).

2.4.2 Probabilistic Reasoning, Thinking and Understanding

Unlike mathematical understanding, probabilistic understanding has not been defined in the literature. Researchers defined terms such as *probabilistic reasoning* and *probabilistic thinking* which entail the idea of *understanding*. This section considers these definitions while the next section provides an overview of available theoretical frameworks developed as aids in describing students' probabilistic reasoning and thinking.

Jolliffe (2005), in agreement with Watson (2005), defined *probabilistic reasoning* as "understanding and being able to explain and justify probabilistic processes" (p. 326), while Garfield (2003) defined *reasoning about uncertainty* as

> Understanding and using ideas of randomness, chance, likelihood to make judgments about uncertain events; knowing that not all outcomes are equally likely; knowing how to determine the likelihood of different events using an appropriate method (such as a probability tree diagram or a simulation using coins or a computer program) (p. 25).

The second term that has received attention in the literature is *probabilistic thinking* which Jolliffe (2005), along with Langrall and Mooney (2005), defined as

> the way people reason with the ideas of probability and make sense of probabilistic information. … probabilistic thinking involves understanding how models are used to simulate random phenomena, how data are produced to estimate probabilities, and how symmetry and other properties of the situation enable the determination of probabilities. It also involves being able to understand and use context when solving a problem… (p. 326)

Jolliffe (2005) emphasizes that both probabilistic thinking and reasoning involve understanding. Notice that the definitions for probabilistic reasoning and thinking make frequent references to the term 'understanding'. However, the term has not been explicitly defined specific to the domain of probability.

2.5 Use of Understanding In This Study

Since the research questions posed in this study refer to students' *understanding of probability*, a definition of *understanding* as used here needs to be provided. Dewey (1933) indicated that "to understand is to grasp meaning" (p. 132) and this *meaning* may be acquired by seeing something "in its *relations* to other things" (p. 137). In a similar manner, Thompson and Saldanha (2003) defined understanding as "the thought which results from a person's interpreting signs, symbols, interchanges, or conversation - assigning meanings according to a web of connections the person builds over time" (p. 99). However, Thompson and Saldanha

point out that these connections are constructed through interactions that the person has with "with his or her own interpretations of settings and through interactions with other people as they attempt to do the same" (p. 99). Savery and Duffy (1996) pointed out that "understanding is in our interactions with the environment" and "cognitive conlict or puzzlement is the stimulus for learning" (p. 136).

The definition of understanding used in this study combines i) the definitions of understanding provided by Dewey (1933) and Thompson and Saldanha (2003) and ii) the definitions of probabilistic reasoning expressed by Garfield (2003) and Jolliffe (2005). Therefore, *students' understanding of probability* in this study means to grasp the meaning of probability concepts by viewing them in relation to other concepts and to be able to explain the processes used in the solution of probability problems or in the completion of probability activities as students interact with one another in small groups. Such an understanding includes, as Garfield stated, knowledge that not all outcomes have an equal probability of occurring and the ability to use an appropriate method to determine the probability of an event. In this study, the extended meaning of understanding given by Thompson and Saldanha (2003) of "assimilation to a scheme" is used; this meaning allows for correct as well as incorrect or inappropriate understandings (i.e. misconceptions) people may have. Researchers indicated that "even errors are often signs of intelligent, although partial understanding of basic concepts" (Resnik and Ford, 1981, p. 196). Therefore, in describing the understanding one may have we need to be "addressing two sides of the assimilation – what we see as the thing a person is attempting to understand and the scheme of operations that constitutes the person's actual understanding" (Thompson and Saldanha, 2003, p. 99; see also Liu & Thompson, 2007). As a

result, two things needed to be specified for each item on the pre-test and post-test used in this study with regards to probabilistic understanding:

iii) What is the goal of the item? That is, what probability content or concept is the item addressing and the student trying to understand? and

iv) What has the student actually understood or what misconception does he/she hold relative to the particular item?

In this dissertation *achievement* is measured quantitatively by considering students' responses to a probability pre-test and post-test. *Understanding* is measured in two ways: i) a distractor analysis of student responses to the multiple-choice items on the pre-test and post-test and ii) a qualitative analysis of audio-taped conversations as students worked on activities or problem sets in groups during class. A distractor analysis provides useful, but not sufficient, information on students' understanding since multiple-choice items do not provide access to a responder's reasoning and subsequently, to a responder's explanation as to why he/she chose a particular response. Without access to a responder's reasoning, an incorrect response to a multiple-choice item does not directly indicate the use of a particular misconception; it simply indicates the "application" of a given misconception (J. P. Smith, personal communication, December 14, 2011). Therefore, both a distractor analysis and a qualitative analysis of student conversations were carried out in the attempt to gain insight on students' understanding of probability in this study.

2.6 Available Models

2.6.1 Shaughnessy (1992): Stochastic Understanding

Shaughnessy (1992) modeled people's stochastic understanding developmentally; this understanding is characterized by the following four types:

1. *Non-statistical*. Indicators: responses based on beliefs, deterministic models, causality, or single outcome expectations; no attention to or awareness of chance or random events.
2. *Naïve-statistical*. Indicators: use of judgmental heuristics, such as representativeness, availability, anchoring, balancing; mostly experientially based and nonnormative responses; some understanding of chance and random events.
3. *Emergent-statistical*. Indicators: ability to apply normative models to simple problems; recognition that there is a difference between intuitive beliefs and a mathematized model, perhaps some training in probability and statistics, beginning to understand that there are multiple mathematical representations of chance, such as classical and frequentist.
4. *Pragmatic-statistical*. Indicators: an in-depth understanding of mathematical models of chance (i.e frequentist, classical, Bayesian); ability to compare and contrast various models of chance, ability to select and apply a normative model when confronted with choices under uncertainty; considerable training in stochastics; recognition of the limitations of and assumptions of various models (p. 485)

2.6.2 Jones, Thornton, Langrall & Tarr (1999): Probabilistic Reasoning

Jones, Langrall, Thornton and Mogill (1997) provided a framework for elementary students' probabilistic thinking across four constructs: *sample space*, *probability of an event*, *probability comparisons*, and *conditional probability*. Around the same time, Tarr and Jones (1997) described a framework for assessing middle school students' thinking about conditional probability and independence across four levels of thinking: *Level 1 Subjective*; *Level 2 Transitional*; *Level 3 Informal Quantitative*; and *Level 4 Numerical*. Subsequently, a framework was developed by Jones, Thornton, Langrall, and Tarr (1999) which included the constructs and levels of reasoning mentioned above, but in addition, included the construct of *experimental probability*. The framework was tested and validated through the middle grades. The framework may be used "as a filter for analyzing and classifying students' oral and written responses" (Jones et al., 1999, p. 153). Students' level of understanding of a construct is revealed through their ability to demonstrate certain behaviors when dealing with situations that involve conditions of uncertainty.

Table 2.1

Framework for Students' Probabilistic Reasoning (Jones et al., 1999, p. 15)

Construct	Level 1 Subjective	Level 2 Transitional	Level 3 Informal Quantitative	Level 4 Numerical
Sample Space	-lists an incomplete set of outcomes for a 1-stage experiment	-lists a complete set of outcomes for a 1-stage experiment and sometimes for a 2-stage experiment	-consistently lists the outcomes of a 2-stage experiment using a partially generative strategy	-adopts & applies a generative strategy that enables a complete listing of the outcomes for 2- and 3-stage cases
Experimental Probability of an Event	-regards data from random experiments as irrelevant and uses subjective judgments to determine the most or least likely event -indicates little or no awareness of any relationship between experimental and theoretical probabilities	-puts too much faith in small samples of experimental data when determining the most or least likely event; believes that any sample should be representative of the parent population -may revert to subjective judgments when experimental data conflict with preconceived notions	-begins to recognize that more extensive sampling is needed for determining the event that is most or-least likely -recognizes when a sample of trials produces an experimental probability that is markedly different from the theoretical probability	-collects appropriate data to determine a numerical value for the experimental probability -recognizes that the experimental probability determined from a large sample of trials approximates the theoretical probability -can identify situations in which the probability of an event can be determined only experimentally
Theoretical Probability of an Event	-predicts most/least likely event on the basis of subjective judgments -recognizes certain and impossible events	-predicts most/least likely event on the basis of quantitative judgments but may revert to subjective judgments	-predicts most/least likely events on the basis of quantitative judgments -uses numbers informally to compare probabilities	-predicts most/least likely events for 1-and simple 2-stage experiments -assigns numerical probability to an event (either a real probability or a form of odds)

Table 2.1 (cont'd)

Probability Comparisons	-uses subjective judgments to compare the probabilities of an event in two different sample spaces -cannot distinguish "fair" probability situations from "unfair" ones	-makes probability comparisons on the basis of quantitative judgments-not always correctly -begins to distinguish "fair" probability situations from "unfair" ones	-uses valid quantitative reasoning to explain comparisons and invents own way of expressing the probabilities -uses quantitative reasoning to distinguish "fair" and "unfair" probability situations	-assigns numerical probability and makes a valid comparison
Conditional Probability	-following one trial of a 1-stage experiment, does not always give a complete listing of possible outcomes for the second trial -uses subjective reasoning in interpreting with and without replacement situations	-recognizes that the probabilities of some events change in a without replacement situation; however, recognition is incomplete and is usually restricted to events that have previously occurred	-recognizes that the probability of all events changes in a without replacement situation -can quantify changing probabilities in a without replacement situation	-assigns numerical probabilities in with replacement and without replacement situations -uses numerical reasoning to compare the probability of events before and after each trial in with replacement and without replacement situations
Independence	-has a predisposition to consider that consecutive events are always related -has a pervasive belief that one can control the outcome of an experiment	-begins to recognize that consecutive events may be related or unrelated -uses the distribution of outcomes from previous trials to predict the next outcome (representativeness)	-can differentiate independent and dependent events in with and without replacement situations -may revert to strategies based on representativeness	-uses numerical probabilities to distinguish independent and dependent events

A very similar framework to the one developed by Jones, Thornton, Langrall, and Tarr (1999) was constructed by Polaki, Lefoka and Jones (2000). In this latter framework, the researchers included *probability of an event* as a single construct instead of separating *experimental probability of an event* and *theoretical probability of an event* like Jones et al. (1999) did.

2.7 Theoretical Framework For This Study

The aforementioned models provide a set of key constructs of probabilistic reasoning demonstrated in terms of students' behaviors or learning goals that can be used to make sense of students' understanding of probability. Note that Shaughnessy's (1992) model described people's understanding of stochastics in general in a developmental manner, whereas Jones et al. (1999) focused on students' reasoning along specific probability constructs.

The study presented here aimed to examine students' understanding of experimental and theoretical probability. Participants completed a probability pre-test and post-test along with a set of in-class group activities (treatment group) or problem sets (control group). The activities and problem sets focused on the following probability constructs: Law of Large Numbers, experimental probability, theoretical probability, conditional probability, independence, discrete probability distributions and the Binomial distribution. Since the data included both oral and written student responses, a framework was needed to qualitatively analyze these responses. Given that the instruments involved items on the constructs mentioned above, a framework was needed that focused on these constructs and examined students' level of understanding of them. Therefore, the framework developed by Jones et al. (1999) was used in this study. The way in which the framework was used to qualitatively analyze the data along with the results of the qualitative analysis are presented in Chapter 6.

2.8 Summary

During the last few decades, probability has been gaining importance as an area that students need to have experience with in order to be well-informed citizens. However, the literature suggests that "there is no simple story about how students reason about chance" (Konold et al., 1993, p. 413) making probability instruction a challenging task. Research has revealed that both children and adults frequently hold intuitions about probability that come at odds with theory. These problematic intuitions may explain why probability seems to be difficult to learn. "One root of the trouble with probability is lack of experience with the long-term regularity that the mathematics purports to describe" (Moore, 1997, p. 3). According to Konold et al. (1993) one of the major reasons that probability is difficult to teach is that students do not carry one but a variety of preconceptions regarding probability which they bring to the classroom. Moreover, "[P]robability does not consist of mere technical information" and a set of procedures but instead, requires a unique way of thinking that is different from other mathematical domains (Fischbein and Schnarch, 1997, p. 104).

Shaughnessy (1981) suggests that instruction uses activity-based learning and considers both an experimental and theoretical approach to the study of probability while the ASA advocates active learning in introductory statistics courses (Franklin and Garfield, 2006; Aliaga et al., 2005). Active learning may be fostered through the use of activities that involve group work, cooperative learning, and class discussions (GAISE, 2005; see Hall & Rowell, 2008). Furthermore, Steinbring (1991) supports that "empirical and theoretical probability should be developed concurrently in the classroom" (as cited in Mojica, 2006, p. 10).

This dissertation study considers the recommendations set forth by the ASA and statistics educators and examines the effects of two instructional methods on students' understanding of

experimental and theoretical probability. The literature review helped identify instructional characteristics advocated by researchers in mathematics and statistics education - such as group work, active learning through the use of experiments that generate real data, and cooperative learning – which were adopted in this study. Moreover, students' probabilistic difficulties, misconceptions and use of heuristics specified in the literature were considered in the development of the instruments used in this study as well as in the analysis of student responses on the instruments.

CHAPTER 3

CYPRUS EDUCATIONAL SYSTEM

As the current dissertation takes place in a tertiary education institution in Cyprus, this chapter aims to provide information on the educational system of the country including the intended mathematics curriculum. Consideration is given to the inclusion of probability both at the secondary and tertiary levels since the participants in this dissertation study were beginning college students in an introductory statistics course. Also, this chapter provides an overview of the results of students in Cyprus on international studies specific to the domain of probability and the consequences of these results on the Cyprus educational system. Moreover, a section is included which considers the current educational reform efforts taking place in Cyprus and their effect on the teaching of probability.

3.1 Control of the Educational System

Since 1960, the island of Cyprus has been an independent republic with a democratic governmental system in which executive power is exercised by the president and legislative authority is exercised by a house of representatives (Papanastasiou, 1997). The educational system is highly centralized and controlled by the Ministry of Education and Culture which was founded in 1965. Education in Cyprus is compulsory up to and including grade 9. Public as well as private elementary, secondary and tertiary education institutions operate on the island and are all liable to supervision by the Ministry of Education and Culture. Public education at all levels is free whereas students attending private education institutions pay tuition. Entrance into most private secondary schools is based on competitive exams in the Greek language and in mathematics.

Until 1980 the influence of the Greek educational system on the Cyprus educational system was "direct and unquestionable" with identical curriculum materials being used in elementary and secondary schools in both countries (Lordou-Kaspari, 2003, p. 271). In 1980 "a new system of specializations was introduced in the Cyprus Lyceum" (high school) (p. 271). This meant that new curriculum materials were needed for grades 10-12. With this in mind, the Ministry of Education and Culture established the Curriculum Development Service which in turn developed the new curricula and textbooks needed for high school. In this respect, since 1980, the Cyprus educational system gained some degree of autonomy "although the links between the Cyprus and the Greek education are still strong" (p. 271).

Currently, the Ministry of Education and Culture of Cyprus controls the curriculum, the textbooks and other resources needed to deliver public education. Specifically, the textbooks used in public elementary and lower secondary schools are either locally produced by the curriculum development unit or donated by the Greek government (Pashiardis, 2007). The curriculum in public schools is the same for all students up to the end of grade 9. Beyond grade 9 students may select some of the school subjects they take based on their interests.

In public educational institutions the teaching staff is appointed and promoted by the Education Service Committee which operates within the Ministry of Education and Culture. Private schools are operated and managed by individuals or bodies, yet they are supervised by the Ministry of Education and Culture. Inspectors from the Ministry of Education and Culture visit public as well as private educational institutions at all levels a few (unannounced) times during the school/academic year (Pashiardis, 2007). The role of these inspectors is to attend class sessions, observe whether students attend class and have acquired the course textbook, and assess teacher performance. At the tertiary level, colleges and universities determine course

requirements for students with consultation with the Ministry of Education and Culture. All private tertiary education institutions must register with the Ministry. If such an institution would like to have its programs accredited then, in addition to registering with the Ministry, the institution has to undergo an accreditation process.

Currently, higher education in Cyprus is provided by public and private tertiary education institutions as follows: i) public or private universities; ii) non-university public higher-level education institutions or iii) non-university private higher-level education institutions (colleges). There are currently three public national universities (the University of Cyprus, the Cyprus University of Technology and the Open University of Cyprus) and five private universities (European University of Cyprus, Frederick University, Neapolis University, University of Central Lancashire-Cyprus, and the University of Nicosia) in Cyprus. In addition, there are six public colleges: the Higher Technical Institute offering courses in electrical, mechanical and civil engineering; the Higher Hotel Institute; the Mediterranean Institute of Management; the Cyprus Forestry College; the School of Nursing; and the Cyprus Police Academy (Lordou-Kaspari, 2003). Moreover, there are 24 private colleges offering diplomas or four-year university-level degrees each specializing in certain areas of study including art, beauty therapy, accountancy, banking, business administration, music, secretarial studies, technology, and hotel and catering. It should also be noted that due to the geographic position of Cyprus, the island has become a center of international business and shipping. As a result, most local higher-education institutions offer courses in such areas as business administration, management, marketing, finance, and accounting, attracting both home and overseas students.

3.2 Language of Instruction in Schools in Cyprus

The constitution of Cyprus identifies Modern Greek and Turkish as the two official languages of the island. A Greek-Cypriot dialect exists which is unique to the island and it is used in everyday spoken conversations by the, approximately, 700,000 Greek-Cypriots on the island (Papapavlou, 2001). This dialect is closely related to Modern Greek but contains influences from various languages including Latin, English and Turkish.

The economy of Cyprus depends highly on tourism, especially during the summer months, and from the UK. Due to the geographic position of Cyprus at the crossroads of the Middle East, Africa and Europe, several offshore companies operate on the island and employ many Cypriots at their offices. Given these and the entry of Cyprus to the European Union in 2004, English has become a language that is used to a great extent on the island. Most job openings require that candidates have a good knowledge of the English language; that they are able to converse and write in English (Pavlou, 2000; as cited in Ministry of Education and Culture, Cyprus, 2004). English is widely used in all government departments, in courts, and in the banking sector (Papapavlou, 2001). Moreover, English frequently appears on shop signs and billboards whereas local newspapers and television shows make frequent use of English words. In schools, the teaching of English begins in the elementary grades with students taking English language classes for two periods per week in grades 4-6 (Ministry of Education and Culture, Cyprus, 2004).

In public secondary schools, instruction is carried out in Modern Greek but English is a second language for which students take compulsory classes in grades 7 through 10 (Ministry of Education and Culture, Cyprus, 2004). In contrast, in most private secondary schools, instruction is carried out in English and students take compulsory classes in the Modern Greek language. At

most tertiary education institutions the language of instruction is English (Ministry of Education and Culture, Cyprus, 2004) which was also the case at the research site for this study.

3.3 Intended Curriculum

The intended curriculum for all subjects taught in public elementary and secondary schools in Cyprus is determined by the Ministry of Education and Culture. In the case of private elementary and secondary education, the individual institutions determine the intended curriculum with approval from the Ministry of Education. Mathematics is compulsory for all students up to and including grade 12. This is the case in both public and private elementary and secondary education institutions.

Specific to public schools, all students take the same mathematics course in the lower secondary grades (grades 7-9/gymnasium). Upon entering upper secondary school (grades 10-12/lyceum) students must register for electives, in addition to the compulsory common core courses. Mathematics is offered as a common core course in grades 10-12. All students take common core mathematics in grade 10. However, in grades 11 and 12 they may fulfill their mathematics requirement by taking 'core mathematics' or by taking 'advanced mathematics'.

At the secondary level, the aim of the mathematics curriculum is to develop in students the ability to think logically, to analyze situations presented to them in the form of problems, to understand concepts and their properties, and to use the language of mathematics (Papanastasiou, 1997).

3.3.1 Probability in the Intended Secondary Curriculum

Probability first appeared in the Cyprus curriculum during the school year 1984-1985 in the algebra course for grade 9 (Lordou-Kaspari, 2003). Currently, probability is included in the secondary school curriculum as part of the core mathematics course and the 'advanced

mathematics' course in grade 12 (Ignatiou and Zotos, 2007a; 2007b). The textbook used in the core mathematics course (consisting of 198 pages) and the textbook used in the 'advanced mathematics' course (consisting of 255 pages) each include a chapter on probability. In both textbooks, the chapter on probability is rather short with the core mathematics textbook devoting only 16 pages (i.e. 8% of the textbook) to this domain and the 'advanced mathematics' textbook devoting 29 pages (i.e. 11% of the textbook) to it.

An examination of these two textbooks reveals that the intended curriculum on probability in public secondary schools in Cyprus includes the following: i) definition of randomness; ii) definition and identification of sample space and experimental outcomes; iii) definition of probability and of various events including their properties (i.e. probability is a number between 0 and 1; probability of a certain event is 1 and of an impossible event is 0; the sum of the probabilities of complementary events is 1; if events A and B are mutually exclusive then $P(A \cup B) = P(A) + P(B)$); iv) definition and computation of the probability of simple events; v) definition and computation of the probability of the union and intersection of events including pictorial representations using Venn diagrams; vi) computation of the probability of an event which involves the use of combinations and vii) tree diagrams and their use in solving probability problems. The content described above is included in both the core mathematics and the 'advanced mathematics' textbooks for grade 12 (Ignatiou & Zotos, 2007a, 2007b). In addition, the textbook for 'advanced mathematics' includes a section on conditional probability and a section on independence which are not part of the probability chapter in the grade 12 core mathematics textbook.

With regards to probability problems, the core mathematics textbook for grade 12 includes thirteen problems in the probability chapter whereas the 'advanced mathematics'

textbook includes a total of forty-eight problems in the corresponding chapter. In addition, each of these two textbooks contains five problems on probability in the review exercises found at the end of the book. Table 2.1 provides an overview of the content covered by these probability problems.

Table 3.1

Probability Content In Grade 12 Mathematics Textbooks Used In Public Schools in Cyprus

Topic	Number of Problems	
	Core Math Text	Advanced Math Text
Find the sample space	3	3
Compute the probability of an event using probability properties	3	11
Compute the probability of simple and joint events	6	13
Compute the probability of an event using combinations	6	7
Compute conditional probabilities	0	12
Indicate whether two events are independent; compute the probability of independent events	0	7
Total Number of Problems	18	53

3.3.2 Probability at the Tertiary Level

Probability content in tertiary-level courses in Cyprus is taught i) in a statistics course for majors in the natural science, majors in the applied sciences or business, majors in the social sciences, and education majors; ii) in a quantitative methods course for majors in business and majors in education; and iii) as a research tool in programs in mathematics, statistics, education and other disciplines (European University Cyprus, 2007; University of Cyprus, 2007; Frederick University, 2008; University of Nicosia, 2008; Neapolis University, 2011; Cyprus University of Technology, 2011; Open University of Cyprus, 2011; University of Central Lancashire-Cyprus 2012). Such courses are part of an undergraduate or graduate degree.

Note that among all tertiary education institutions in Cyprus, only the University of Cyprus (which was the first university established on the island and began operating in 1992 as a public university) and the newly established private University of Central Lancashire-Cyprus offer undergraduate and graduate programs in mathematics and statistics. In other tertiary education institutions in Cyprus, probability is only taught as part of service courses in mathematics and statistics (including methods courses) for programs not in the natural sciences.

3.4 Performance of Secondary School Students on Probability on TIMSS

Cyprus has been a member of the International Association for the Evaluation of Educational Achievement (IEA) since 1990 and has participated in the Third International Mathematics and Science Study in 1995 and in almost all activities of Trends in International Mathematics and Science Study (TIMSS) since then. Throughout the TIMSS studies, Cyprus has been performing poorly on mathematics assessments. When the results of the 1995 Third International Mathematics and Science Study were announced, the "exceedingly low achievement of Cypriot students" sent a shock wave through the Cypriot society and the Ministry of Education and Culture (Papanastasiou, 2002, p. 231). Furthermore, a comparison between the 1995 and 2007 TIMSS results as well as between the 1999 and 2007 TIMSS results indicated a decrease in the mathematics achievement of 8^{th} grade Cypriot participants (Mullis, Martin, & Foy, 2008).

One of the content domains that TIMSS 8^{th} grade participants are examined on is that of data and chance. In the 2007 TIMSS the average scale score for 8^{th} grade students in Cyprus on data and chance was 464 (standard error 1.6) which was significantly lower than the TIMSS scale average of 500 in this domain (Mullis, Martin, & Foy, 2008, p. 121). Cypriot 8^{th} grade girls

averaged 474 (standard error 2.4) on data and chance which was significantly higher than the average of 454 achieved by Cypriot 8th grade boys (standard error 2.5) in this content domain (p. 140). At the 8th grade in Cyprus, only 3% of class time is devoted to data and chance. Topics on this domain taught up to and including grade 8 are considered to be only for the more able students and as a result, only 3% of Cypriot 8th graders are taught such topics. That is, this 3% of Cypriot 8th graders receive the 3% of instructional time devoted to data and chance. This is not surprising since TIMSS 2007 and the intended curriculum discussion in section 3.3.1 revealed that these topics are intended for grade 12 (Mullis, Martin, & Foy, 2008; Ministry of Education and Culture & Pedagogical Institute, 2002).

3.5 Recent Developments in the Educational System of Cyprus

At the time of the announcement of the TIMSS 1995 results, the Ministry of Education and Culture in Cyprus attempted to find excuses for the low achievement of students.

> Certain circles in the Ministry of Education and Culture suggested that Cyprus ought to have withdrawn from the international test for mathematics and science in good time before it sank to the bottom, indicating the naïve attitude that a problem does not exist if you do not see it (O Filelefthеros, 1996; see Papanastasiou, 2002, p. 231)

For some though the results were not surprising since the educational system was in need of reform (Papanastasiou, 2002). In 1997 the International Institute for Education Planning conducted a study of the educational system of Cyprus (Vrasidas and McIsaac, 2001). The results of this study indicated that indeed the system was in need of reform in order for the quality of public secondary education to be improved.

In view of these results, it is encouraging that since 2005 the government of Cyprus initiated an Education Reform Program (The Ministry of Education and Culture, Republic of

Cyprus, 2008). A report by the Ministry of Education and Culture (2008) describes the plans and areas in need of reform which include revisions in the national curriculum and the establishment of a center for research. Among the key goals listed with regards to curriculum revisions, is to help students develop into active citizens,

> to enhance their critical thinking and research capabilities and to include a variety of teaching methodologies and introduce flexibility in the school program, so that the teacher may use the most appropriate approach for the particular class (The Ministry of Education and Culture, Republic of Cyprus, 2008, p. 36)

Relative to teaching approaches, among the problems of the Cypriot educational system listed in the UNESCO (United Nations Educational, Scientific, and Cultural Organization) Report on Education in Cyprus, which was released after the TIMSS 1995 results, was that teachers do not involve students in the learning process (Papastylianou, 1997; see also Papanastasiou, 2002). In addition, based on the results of the study carried out in 1997 by the International Institute for Education Planning the following should be integrated into the Cyprus educational system: mixed ability classroom teaching, cooperative learning, and the use of technology (Vrasidas and McIsaac, 2001).

In 2010, a detailed report comprising the reformed intended curriculum for each subject area taught in grades K-12 was published by the Ministry of Education and Culture in collaboration with the Pedagogical Institute and the Curriculum Development Service. These reports, including one for mathematics, specify the content, procedures, applications and experiences that students are expected to acquire while in school. Curriculum materials reflecting the recommendations included in these reports were published in 2010. A small-scale pilot of these materials was carried out during the school year 2010-2011 in grades K-9. Based on the

pilot results, the curriculum materials were to be revised by the end of the school year 2010-2011 and implemented in stages during the school year 2011-2012 in all grades.

The 475 page-long mathematics report of the intended curriculum is subdivided into five domains: Number, Measurement, Geometry, Algebra and Statistics-Probability. The report specifies that mathematics concepts should be taught in a way that advances students' interest and curiosity while simultaneously placing emphasis on problem solving. To this end, the intended curriculum includes activities which require students to explore and discuss mathematical ideas (Ministry of Education and Culture, Pedagogical Institute and Curriculum Development Service, 2010). The general goals of mathematics education include that students should: i) appreciate the value of mathematics and its use in all aspects of human activity; ii) develop the ability to solve problems in multiple ways; and iii) develop the knowledge and skills required in the workplace and for further studies in areas in which the use of mathematics is necessary (p. 5). Moreover, the mathematics intended curriculum aims to provide opportunities for students to: i) develop flexibility and creativity in the application of mathematics concepts in problem situations; ii) solve problems cooperatively, express their ideas and respond to the ideas of their classmates; iii) learn through the correct as well as incorrect responses given by them and their classmates; and iv) develop their verbal, written and presentation skills so as to be able to express and define mathematics concepts (p. 6). Relative to the domain of probability, Tables 3.2 - 3.4 provide the breakdown of skills that students in grades 10-12 are expected to acquire (Ministry of Education and Culture, Pedagogical Institute and Curriculum Development Service, 2010, p. 440-466) (trans.):

Table 3.2

Probability in the Reformed Intended Curriculum for Grade 10 in Cyprus

Grade 10 – Probability – Student Skills
1. Represent events using Venn diagrams (union, intersection, difference, complement) 2. State and apply the Kolmogorov axioms. Reason through them using suitable examples and Venn diagrams. Apply the consequences of the Kolmogorov axioms (($P(A'), P(A \cup B), P(A - B)$) in solving problems. 3. Study random experiments involving two or more steps and write down the sample space using tables and tree diagrams. 4. Understand and apply the Counting Principle in random experiments. 5. Convert tree diagrams into probability tree diagrams and compute the probability of compound and independent events. 6. Compute permutations and combinations and apply these in the computation of probabilities.

Table 3.3

Probability in the Reformed Intended Curriculum for Grade 11 in Cyprus

Grade 11 – Probability – Student Skills
1. Distinguish events as conditional and independent and compute their probabilities. 2. Understand and apply the rule of Total Probability. 3. Understand the meaning of random variable. 4. Find the probability density function and the distribution function of a discrete random variable. 5. Compute the mean and standard deviation of a discrete random variable. 6. Study the Binomial distribution.

Table 3.4

Probability in the Reformed Intended Curriculum for Grade 12 in Cyprus

Grade 12 – Probability – Student Skills
1. Understand the meaning of a probability density function and distribution function of a continuous random variable. 2. Compute probabilities under specific intervals; find the mean and standard deviation of a continuous random variable. 3. Know the properties of the Normal distribution. Solve problems by applying the table of the standard normal distribution and the Central Limit Theorem.

3.6 Research Site for this Study

In view of the various statements made by the Ministry of Education and Culture of Cyprus and the problem identified in the UNESCO report, as well as considering that "demands for dealing with data in an information age continue to grow" (Franklin & Garfield, 2006, p. 363), it is important to study the effects of different teaching approaches in statistics classrooms. A first step into such an endeavor – since there is a lack of research in this area specific to Cyprus, especially at the secondary and post-secondary levels - was carried out through this dissertation study.

The research site for this study was a specialized private business college operating since 1983 as a tertiary education institution in Cyprus with approval from the Cyprus Ministry of Education. Instruction at this college is carried out in English. Most students attending this college are native Cypriots who have graduated from a public high school and for whom English is a second language.

At this college, probability is taught i) in an introductory business statistics course which is compulsory for all freshmen and ii) in an Operations Management course required by junior Business Computing majors and which involves applications of mathematics, statistics and probability to the business world.

The study took place in the introductory statistics course at this college during spring 2010. Until that time, the course was being taught using a lecture format with complete reliance placed on the course textbook. As mentioned in Chapter 1, the MAA (1998) and research in the STEM fields promote the use of teaching techniques which involve less lecturing, increased group work and student interactions in undergraduate classrooms. Moreover, it has been recommended that statistics classrooms use a frequency-based approach to teaching probability

along with an active learning approach that includes the use of activities (Cobb, 2000; Watson, 2006). This study aimed to adopt these recommendations in the aforementioned introductory statistics course. The methods used in accomplishing this are described in detail in the Chapter 4: Methods.

CHAPTER 4

METHODS

4.1. Overview of the design

A mixed methods design was used to address the research questions for this study. The design included treatment and control groups, each comprised of students in sections of an introductory statistics class, and utilized a pre-test and post-test design. During the study students in the treatment group worked in small groups on four in-class activities about experimental and theoretical probability, and students in the control group worked in small groups on solutions to four sets of probability problems. In each group, the conversations of specific students were audio-recorded across the four occasions that they worked in groups. These data sources allowed for both quantitative and qualitative analyses of data.

4.2. Research Site

4.2.1 College

The college at which the study took place is situated in Larnaca, one of the main cities in Cyprus. It has been operating since 1983 as a tertiary education institution with approval from the Cyprus Ministry of Education. The college is a specialized business school at which students may pursue a four-year Bachelor's degree in one of the following areas: Business Administration, Business Computing, Accounting, or Banking. Alternatively, students may pursue a two-year diploma in business administration or computing and information systems. In addition, the college offers one-year certificates in law or business and information technology. The 4-year Bachelor's degree programs in Business Administration and Business Computing have already been approved by the Cyprus Ministry of Education and most students attending the college major in one of these two areas. During the academic year 2009-2010, all freshmen

pursued one of these two degrees. Approval for the rest of the programs offered at this college is currently being sought.

At this college, the mathematics and statistics requirements for students in the Business Administration program differ slightly from those for students majoring in Business Computing. Table 4.1 below presents listings of the requirements for the two programs in the order in which students take the courses.

Table 4.1

Mathematics Course Requirements For Students At Research Site

Business Administration	Business Computing
MAT 101 – Calculus	MAT 101 – Calculus
MAT 201 – Statistics I	MAT 201 – Statistics I
MAT 202 – Statistics II	MAT 210 – Discrete Mathematics
MAT 203 – Quantitative Methods	MAT 203 – Quantitative Methods
	MGT 312 – Operations Management

Most of the students attending this college are native Cypriots. A majority of students understands and speaks Modern Greek and most of them are using English as a second language. The official language of instruction for all programs is English.

The academic year officially begins in the first week of October and finishes at the end of May. During the entire month of September, freshmen take a foundations course in each of the following subjects: mathematics, English, and accounting. At the end of September these students take a placement exam in each of these subjects. The aim of these courses is to help incoming students attain the basic skills in these subjects that will be required during their academic studies. The material covered in the mathematics foundations course includes real numbers and their properties, rules of exponents, simplifying algebraic expressions, and solving linear and quadratic equations.

The college offers morning classes as well as evening classes. For example, students may take a 4-credit course, such as MAT 201 – Statistics I, by enrolling in a morning section that meets for four 50-minute periods a week between the hours of 8:15am-1:30pm or by enrolling in an evening section of the course which meets once a week from 5:50-8:20pm. Evening sections tend to be smaller than morning sections. Courses are carried out in classrooms that can hold up to 30 students. All classrooms in which the introductory statistics course is accommodated are equipped with an overhead projector as well as a LCD projector.

College students in Cyprus are required by the Ministry of Education to acquire and use a textbook for every course they take. At the college where the study took place, all course textbooks are written in English and are selected by the course instructor in consultation with previous instructors of the same course, the department head and the course coordinator.

4.2.2. Course

The study took place in an introductory statistics course taught by the researcher in spring 2010 at the aforementioned college. This course (MAT 201 – Statistics I) is offered in the spring semester of every academic year and it is compulsory for all freshmen attending the college. Students taking the course are split into two morning sections (Sections A and B) and one evening section (Section C). Since each morning section of the course meets for four 50-minute sessions each week whereas the evening section meets for one 150-minute session each week, students in the morning sections spend 50 minutes more per week in the course. These meeting times are kept constant throughout the spring semester which lasts for 15 weeks (13 weeks of teaching and 2 weeks of final exams). During the spring semester 2010, to control for the factor of contact time for this study, students in the evening section were asked to attend four extra

class sessions from 5:50 – 8:20 pm. An arrangement was made between the instructor of MAT 201 and the students in Section C based on the students' class and work schedules.

The introductory statistics course (MAT 201) covers the following topics:

i. Basic concepts of statistics (population, sample, parameter, statistic, categorical / numerical / discrete / continuous data);
ii. Tables and charts for presenting categorical or numerical data (bar charts, pie charts, histograms, contingency tables, and scatter plots);
iii. Measures of central tendency and variation;
iv. Covariance and correlation;
v. Basic probability concepts (sample space, simple and joint events, Venn diagrams, union and intersection of events, finding the probability of events);
vi. Conditional probability, independence, and Bayes' Theorem;
vii. Counting rules;
viii. Discrete probability distributions (expected value and variance of discrete random variables; binomial distribution);
ix. and the normal distribution.

The main course textbook for MAT 201 was a U.S. publication of *Business Statistics: A First Course* (Levine, Krehbiel & Berenson, 2010). During this study, problems from this textbook were selected by the researcher/instructor and were assigned as homework to the students in *all* sections of the course.

4.2.3 Timeline and Procedures

During the summer of 2009, the instruments that were to be used in the study were developed. For the purposes of this, the researcher carried out a thorough search in statistics

education and mathematics education journals, conference proceedings, statistics textbooks, statistics education and mathematics education textbooks, instruments used in international studies in mathematics education, and websites developed by mathematics and statistics educators. Through this search it became evident that finding items and especially appropriate activities specific to probability was a challenging task, since these were not commonly found and many were not suitable for the purposes of this study. Information provided by a committee member led to contacting the researchers involved in the ARTIST (Assessment Resource Tools for Improving Statistical Thinking) project who in turn gave access to the researcher of this study to the item bank as well as the instruments developed and tested by the ARTIST group. Once the instruments for this study were developed, they were reviewed by committee members and by a mathematics education graduate student at Michigan State University. Items were piloted in late January 2010, and revised shortly thereafter.

Instruction on probability formally began in week 7 (Monday, March 22nd, 2010) and continued until week 13 (Wednesday, May 21st, 2010) of the Spring 2010 semester. The pre-test and background questionnaire were administered during week 7. In the treatment group, the administration of the four activities on experimental and theoretical probability commenced during week 7 and ended in week 13. On the same weeks that students in the treatment group worked on the probability activities, students in the control group worked on sets of probability problems during class. The post-test was embedded in the course final examination which was administered at the end of week 15 and specifically on June 4th.

In order to motivate students to take the completion of the instruments seriously, completing each instrument counted towards the student's final course grade. Course policy at

the college where this study took place specifies that the final exam should count for 70% of the course grade. The other 30% is allotted, according to the instructor's judgment, to class participation, a midterm, and a formal course assignment common to all sections of the course. The formal course assignment for all sections of MAT 201 consisted of problems on statistics only, counted for 8% of the course grade, and was due at the beginning of week 7, the same week that students worked on Activity 1 and formal instruction on probability began. The probability items on the post-test (embedded in the course final exam) were used to determine individual mastery of the material covered on probability.

In all three sections of MAT 201, the course grade was split as follows:

Course Assignment on Statistics	8%
Class Participation	12%
Midterm exam	10%
Final Exam	70%

The percentage allotted to class participation (12%) was split as follows: In the case of the treatment group (Section A and Section C), students received 2% for each of the four activities on probability they participated in, for a total of 8%; in the case of the control group (Section B), students received 2% for each of the four occasions they worked in groups to solve probability problems from the textbook, for a total of 8%; students who completed the pre-test received 4% to count towards their course grade.

During summer 2010, the researcher coded student responses to the background questionnaire, the pre-test and the multiple-choice items on the post-test and entered these data in SPSS. Quantitative data analysis of these data was carried out during summer 2010. With regards to the free-response items on the post-test, rubrics were created during fall 2010. Student

responses to the free-response items were then coded and quantitatively analyzed using SPSS. Data from the audio recorded group conversations were transcribed in December, 2011 and January, 2012 and were then qualitatively analyzed.

4.2.4 Pilot

As previously mentioned, all instruments were piloted in January 2010. The pilot aimed to reveal any possible misinterpretations caused by the instructions to the probability problems on the instruments or the instructions to the probability activities and any obstacles faced by the students due to specific vocabulary used in these problems and activities. The pilot also helped in determining the length of time it would take to complete each activity, and how to handle technical aspects of recording conversations of groups of students.

Piloting began on January 18th, 2010, once the study received IRB approval, and lasted for two weeks of instructional time at the college. This time period marked the end of the fall semester at the college (February 8th was the commencement of the spring semester). Participants were 25 (out of 28) students taking MAT 202 Statistics II and four students taking MAT 210 Discrete Mathematics that semester (i.e. Fall 2009). These students had received instruction on the probability concepts covered on the instruments during the spring semester of 2009 in MAT 201 Statistics I. The researcher was not the instructor for MAT 202 but was the instructor of MAT 210 during the pilot process. The completion of the instruments on the pilot study did not count towards the course grade for any of the participants in the pilot. In order to motivate participants in the pilot to work on the instruments, the researcher treated participants from MAT 202 on two occasions to coffee and breakfast treats at the college cafeteria and gave the course handbook for MAT 210 to the four students taking that course as a gift.

Once the study received IRB approval, the researcher visited the sections of MAT 202 and MAT 210, talked to the students about the research study and asked for volunteers. Since it was the end of the semester, and the mathematics instructor of MAT 202 had completed instruction on the course material specified on the syllabus, he allowed the researcher to use the remaining of the class sessions of MAT 202 prior to the end of the fall semester for the purposes of the pilot. At the beginning of the first data collection meeting, the consent form (see Appendix A) was read to the students, the study was explained once more, and participants were asked to sign the consent form if they agreed to participate. The consent forms were placed in an envelope and returned by one of the students in each course to the course coordinator. The envelope was given to the researcher after semester course grades for MAT 210 and MAT 202 had been submitted. All four students in MAT 210 and 25 of the 28 students in MAT 202 had volunteered to participate.

The background questionnaire and pre-test were piloted during a 50-minute session. With regards to the probability activities, two or three groups of volunteers worked on each activity and the conversations of all groups were recorded. Audio recordings took place in regular classrooms where the participants and the researcher met. This aimed to create conditions similar to those that the researcher was to encounter in spring 2010 when recording group conversations in the regular classroom. Such recordings during the pilot study helped the researcher identify noise levels during group conversations and their effect on the quality of audio recording. This indicated to the researcher how groups should be positioned in the classroom during spring 2010 to ensure better quality recording.

In addition to piloting the instruments with the students in MAT 202 and MAT 210, expert reviewers were asked to provide feedback on the instruments during fall 2009. This included

expert reviewers who were members of the researchers' dissertation committee. Based on this feedback, in October 2009 the instruments underwent a first round of revisions with regards to content and format. Relative to the pre-test and post-test, experts' recommended revisions were sorted from worst item to best item. These recommendations were then used to modify or delete items. In addition, in December 2009, a mathematics graduate student at Michigan State University provided further feedback on the instruments. Upon the completion of the pilot at the end of January 2010, the instruments underwent further revisions with regards to content, format, and duration, based on the pilot data.

4.3. Participants

Prior to the start of the academic year 2009-2010, participants were split into three sections (Sections A, B, and C) by the course coordinator in consultation with the academic board at the college. During the academic year 2009-2010, the academic board decided to try a new approach to placing freshmen in groups, based on their English language proficiency. As a result, Section A (morning section) included the students of moderate/high English language proficiency whereas Section B (morning section) included the students of low English language proficiency. Based on the students' course selection, all four of the students who majored in Business Computing were placed in Section A, although two of these students were of low English language proficiency. Section C (evening section) included students of all levels of English language proficiency. Volunteers for this dissertation study came from all three sections of MAT 201 – Statistics I who had been split into three sections: Section A (morning) which consisted of 22 students, Section B (morning) which consisted of 19 students and Section C (evening) with 7 students. Out of these 48 students, 44 agreed to participate in the study by signing a consent form: 20 from Section A, 17 from Section B, and 7 from Section C.

Participants were aged 18-27 years old with a mean age of 19.5. Twenty of them were male (45.5%) and twenty-four were female (54.5%). Most of these students (42 out of 44) had attended a public secondary school in Cyprus whereas one of them had attended a private secondary school in Cyprus and one of them a secondary school abroad. Most of the participants (40 out of 44) were native Cypriots with two from Russia, one from Georgia and one from Moldova. All of them were able to communicate orally in Modern Greek, the native language in Cyprus, and 41 of them were using Modern Greek as their first language and English as their second language. Thirty six of the participants indicated that they had received instruction on statistics and probability in high school whereas 7 of them indicated that they had not (1 student did not respond). Forty three were freshmen whereas one of the participants was a sophomore. This student (sophomore) had taken MAT 201 for the first time in spring 2009, with the same instructor, but had not passed the course then. With regards to the various programs of studies, 32 of the 44 participants majored in Business Administration, 8 majored in Accounting, and 4 majored in Business Computing.

4.4. Instruction

The researcher was also the course instructor of MAT 201 in spring 2010. For the purposes of this study, during the spring semester of 2010, one of the morning sections (Section A) and the evening section (Section C) of MAT 201 – Statistics I were taught using an instructional method that combined lectures and small-group cooperative learning sessions with students performing experiments that generated real data and completing probability activities. These two sections of MAT 201 formed the treatment group. The second morning section (Section B) acted as the control group. In the control group, the researcher used an instructional method that

combined lectures with small-group cooperative learning sessions during which students worked in class on solving probability problems from the course textbook.

Upon enrollment in the course, students did not know that different instructional methods were to be used in the course sections. During the first session of MAT 201 in spring 2010, the researcher/instructor talked to the students in the three sections of the introductory statistics course about this dissertation study. The researcher indicated to the students that participation was voluntary and explained the benefits as well as any risks regarding their participation. Furthermore, the researcher pointed out to the students the purpose of the study, the necessary IRB procedures and the need of signing a consent form in case they agreed to participate (see Appendix A). According to the IRB, the researcher/instructor should not have access to the signed consent forms until final course grades had been submitted. As a result, after talking to the students about the study during the first class session of the course, the researcher/instructor left the room and the signed consent forms were placed in an envelope and returned to the course coordinator by one of the students in each section.

Meanwhile, the three sections of MAT 201 needed to be assigned as treatment or control. In order for this assignment to be determined, the two morning sections which were of an almost equal size (20 students in Section A and 17 students in Section B) were randomly assigned as treatment or control by tossing a coin to determine what Section A would act as (heads for treatment and tails for control). The evening section (Section C – 7 students) was assigned as a treatment group.

4.4.1. Treatment Group

During weeks 7 – 12 (March 22nd – May 14th) students in the treatment group worked in small groups during class to complete four activities on probability which covered the following

topics: i) Law of Large Numbers; ii) Conditional probability and independence; iii) Discrete probability distributions; and iv) the Binomial distribution. The activities aimed to bring about the bidirectional relationship between experimental and theoretical probability and to prepare students for the study of theoretical probability. Each activity required the students to perform a probability experiment, to collect and analyze data and reach some conclusions. Students in the treatment group received 2% for working on each of the four activities, for a total of 8% to count towards their course grade. All activities appear in Appendix C.

The aim of Activity 1 (McConnell et al., 1998; see Appendix C) was to provide students with a hands-on experience in viewing the connection between experimental and theoretical probability, leading to the Law of Large Numbers. Two dice of different colors were given to each group and students were asked to roll them 50 times. The group's frequency and relative frequency for each sum was recorded in a table. Students had already received instruction on frequency, relative frequency and frequency histograms, so they were familiar with the terms used in the activity. Once students carried out the experiment, they then examined the theoretical probability of each sum of two dice by considering a pictorial representation of the sample space that was provided in the activity. Next, the results of the 50 rolls of each group were recorded on the class board and each group used these results to calculate the relative frequency for the entire class. Students compared their group's experimental results with the theoretical results. Also, they compared the combined class results from all groups with the theoretical results, in order to see that as the number of simulations increases the more the relative frequency approaches the theoretical probability.

In Activity 2 (see Appendix C), students investigated the ideas of conditional probability and independence. This was done through an exploration of the sum of three dice. Students were

asked to roll three dice of different colors simultaneously 50 times and record their results along with the sum of the three dice in a table. Based on their group's results, they were then asked to compute the relative frequency of a simple event (i.e. one of the three dice comes up 3) and then the relative frequency of the same event this time based on a pre-set condition (i.e. the relative frequency of one of the dice comes up 3 given that the sum is 6). The students then explored the probabilities of these two events through a theoretical approach by responding to a set of questions which were ordered in such a way as to carefully guide them to the theoretical conditional probabilities and the idea of independence.

Activity 3 (Khazanov, 2008; see Appendix C) allowed students to investigate discrete probability distributions. The ideas presented here were adopted from a proposed activity provided by Khazanov (2008). This activity involved a game of chance that made use of dice and chips. Students had to place a set of 12 chips given to each of them along a number line, on which numbers 1 through 15 were written, so as to increase their chances of having them removed first – and thus win the game – based on the sum of two dice. A designated student rolled the two dice, the sum of the numbers showing on the dice was calculated, and any student in the group that had chips placed above the number represented by the sum removed one of those chips. The winner was the student who had all of his/her chips removed first. Students were asked to provide reasons for the way they each distributed the chips on the number line. They were also asked to indicate the mistake one could make to eliminate their chance of winning. In the last part of the activity, the students explored the discrete probability distribution of the sum of two dice theoretically. In this last part, a table was provided to students listing the 36 possible outcomes when two dice are rolled. Students were asked to use this table to compute the probability of each sum from 2 to 12 when a pair of dice is tossed. This procedure aimed to

help students understand why they may or may not have made poor choices when placing their chips on the number line, and to realize that if they placed more chips towards the center of the distribution they would have a higher chance of winning.

In the last activity (Shaughnessy et al., 2004), students explored the binomial distribution first through the use of a table of random numbers and then theoretically. The researcher used Excel to generate sets of random numbers from 1 to 10. A different set of such numbers was given to each group. Students were instructed on how to use their corresponding table of random values and were asked to use it to indicate data on 50 sets of three free-throw attempts for a basketball player with a 70% free-throw average. On the basis of these 50 sets of simulated data, students then computed the probability that the player will make at least two baskets in three free throws. Next, students compared their results with those of one other group in class. In the meantime, the instructor recorded the results of at least two baskets in 50 sets of three free-throw attempts of all groups in class. These combined class results were used by the students to compute the probability of making at least two baskets and compare their group's answer to that derived based on the class results.

During the four occasions of working on activities, students were placed in groups of 2 - 4 as determined by the instructor. An attempt was made to include in each group students of varying mathematical ability in order to create opportunities for scaffolding (Vygotsky, 1978; as cited in Giraud, 1997). Group members were assigned roles by the researcher/instructor so that all members would have an opportunity to participate. These roles were rotated on every occasion that students worked on an activity. The role of each student was indicated in written form on the top of the first page of each activity handed to the group. Three roles were used in the probability activities: performer of experiment (e.g. the person simulating rolls of a die as

required by the experiment), recorder of data and of group responses to activity questions, and checker of data recordings and group responses to activity questions. In the case where a group consisted of two students, one of the students was assigned a dual role: performer of experiment and checker. In the case where a group consisted of 4 students, two of the students were assigned the same role. At the beginning of each activity, the researcher briefly described the meaning of each role to the whole class. All students in each group were encouraged to help the members of their group and discuss their ideas on how to perform the experiment, how to record and analyze the data and discuss what they thought were possible conclusions they could draw from their data. The researcher/instructor circulated in the classroom as the students worked in their groups and tried to monitor and encourage students to participate. Students were asked to submit a copy of their group's responses to the activity to the instructor at the end of class. Feedback was provided to the students at the beginning of the next class session.

The class session following the completion of an activity began with providing feedback to the students on their group work, reviewing the conclusions that could be drawn from the activity, and looking at the bidirectional relationship between experimental and theoretical probability as a whole class. These were completed during the first 15-20 minutes of the 50-minute class session. The remaining class time was devoted to the study of probability terminology and theoretical probability on the probability topic of that week using a lecture format.

Following the lecture session on probability terminology and theory, the instructor solved a few probability problems from the course textbook on the board relative to the probability topic under study that week. The students were then assigned 2-3 problems from the course textbook

to work on individually as homework. These were corrected by the instructor and handed back to the students.

4.4.2. Control Group

During weeks 7-12 (March 22nd – May 14th) students in the control group worked in small groups on four sets of probability problems assigned from the course textbook. The problems covered the following topics: i) basic concepts of probability; ii) conditional probability and independence; iii) discrete probability distributions; and iv) the Binomial distribution (See Appendix C). Students in the control group received 2% for each set of problems they worked on, for a total of 8% to count towards their course grade.

Similar to the treatment group, student groups in the control group were set up by the instructor and consisted of 2-4 students. An attempt was made to include in each group students of varying mathematical ability in order to create opportunities for scaffolding (Vygotsky, 1978; as cited in Giraud, 1997). Group members were assigned roles by the instructor so that all students would have an opportunity to participate. These roles were rotated on every occasion that students worked in groups. A group member undertook one of the following three roles: recorded the group's solution to a problem, checked solutions, or asked questions to stimulate group conversation. The first two correspond to the roles of recorder and checker used in the treatment group. Instead of a student who would ask questions to stimulate group conversation, the role of performer of the experiment was used in the treatment group. On the occasions that students worked in groups, the researcher briefly described the meaning of each role to all students at the beginning of class. Students were encouraged to help their group members and discuss their ideas on how to solve a problem. Apart from the difference in one of the roles used, the set up and structure of groups was the same in the control and treatment groups.

The instructional sequence followed in the control group was different to that used in the treatment group. In the control group, the study of a probability topic began with looking at the connection between the probability topics previously learned and the probability topic to be studied that week. This was followed by a study of probability terminology and theoretical probability on the topic of that week using a lecture format which included looking at the bidirectional relationship between experimental and theoretical probability as a whole class using examples from the textbook. This was different than the instructional sequence followed in the treatment group in which students first worked in small groups on a probability activity prior to formal instruction on a topic.

Following the above instructional sequence, the instructor solved a few probability problems from the textbook on the particular topic studied. Then, students were placed in groups and were asked to solve problems from the textbook. The instructor circulated around the classroom during this time. Students submitted a copy of their group's solutions to the instructor at the end of class and feedback was provided to them during the next class session. Finally, students were assigned 2-3 problems from the textbook to work on individually as homework. These were corrected by the instructor and handed back to the students.

4.4.3. Commonalities between Treatment and Control Groups

In all course sections (control and treatment groups), students were provided with opportunities to work with their classmates during class. Since, "students do not always enact the roles as the teacher would like" (Anderson et al., 1997; Ross et al., 1996; as cited in Esmonde, 2009), the researcher/instructor circulated in the classroom as the students worked in their groups and tried to monitor and encourage students to participate. In this manner, the instructor was also able to respond to student questions as they worked in their groups. The students in the treatment

group were given opportunities to work cooperatively in small groups at the beginning of the study of a probability topic through working on a probability activity, whereas the students in the control group had such opportunities at the end of the study of a probability topic through working on probability problems from the course textbook. In all sections, part of class time was devoted to lecturing on probability terminology and theoretical probability as well as to whole-class discussion.

In the two treatment sections as well as in the control group, group structuring was determined by the instructor with students being placed in groups of 2-4 students. Member roles were rotated between each group session so that students would have the opportunity to carry out all roles by the end of the course. An attempt was made to include in each group students of varying mathematical ability in order to create opportunities for scaffolding (Vygotsky, 1978; as cited in Giraud, 1997). In placing students into groups, the students' high-school mathematics grade (upon entrance to the college), their pre-test score, and their English language skills were taken into consideration.

In order to be able to gather information regarding students' responses to the group activities or problem sets as well as their reasoning and understanding of the probability concepts, the conversations of students in both the treatment group and the control group were audio-recorded. Since the researcher was also the course instructor, it would have been difficult to be aware of the development of students' reasoning and understanding of probability concepts without a means of recording the students' conversation. Overall, 19 audio recordings (10 from the treatment group and 9 from the control group) were collected during the Spring 2010 semester as students worked either on an activity or problem set. Information relative to the audio-recordings is provided in Table 4.2.

Table 4.2

Number Of Audio Recordings Collected In Each Group

	Treatment Group # of Audio Tapes		Control Group # of Audio Tapes
Activity 1	2	Problem-set 1	2
Activity 2	2	Problem-set 2	2
Activity 3	3	Problem-set 3	3
Activity 4	3	Problem-set 4	2
Total	10		9

An audio recorder was provided to each group of students that was to have their conversations recorded. At the end of the day, the audio recordings were downloaded to the researcher's personal computer which could be accessed through a password known only to the researcher. Audio recordings were deleted from the recorders once they had been downloaded. The audio recordings were transcribed during December 2011 and January 2012.

Although the official language of instruction at the college is English, since many students have difficulties with the English language, instruction in all sections of MAT 201 in spring 2010 was carried out in both English and Modern Greek. Through conversations with colleagues and with the course coordinator at the college, it became known to the researcher/instructor that students taking MAT 201 each year tended to have similar difficulties with the English language. The instructor of MAT 201 in spring 2010 was a native Greek speaker who is fluent in spoken and written English. Statistics and probability terminology were taught to students in English but word problems were read to students in English and then translated into Modern Greek. Furthermore, during lecture, whatever explanations, examples, and procedures were provided in English (in written and verbal form) they were also given orally in Modern Greek. The aim of this was to speak in the language that the students were more

comfortable with in an attempt to help them better understand the statistics and probability concepts presented in class.

The final exam for MAT 201 took place on June 4th, 2010, it was graded by the researcher/instructor and final course grades were submitted a week later. The envelope with the signed consent forms was handed to the researcher during the second week of June 2010.

4.5. Instruments

Three instruments were developed for the study: a student background questionnaire, a probability pre-test, and a probability post-test. All three were used in each of the three course sections.

4.5.1. Student Background Questionnaire

The background questionnaire was administered during the first week of classes in all sections of MAT 201. It consisted of 11 items and was used to specify such information as the students' age, Modern Greek and English competency, class level, major field of study, and prior exposure to probability (see Appendix B). The students did not receive any points to count towards their course grade for completing the background questionnaire.

4.5.2. Probability Pre-test

The probability pre-test was administered during the first week of classes (February 9th – 10th) of the spring 2010 semester in all sections of MAT 201. It was used in order to identify whether the students in the control group and the students in the treatment group had equivalent initial probability knowledge. In turn, this helped in identifying the type of analysis that needed to be carried out in order to address the research questions.

The pre-test comprised of a set of multiple-choice items from the web ARTIST (Assessment Resource Tools for Improving Statistical Thinking) project and in particular from the Probability Scale created by the ARTIST investigators (delMas et al., 2006; Garfield et al., 2006), as well as multiple-choice items on probability from the TIMSS studies for grade 8 (TIMSS & PIRLS International Study Center, 1995, 2001, 2007, 2009). There exist 11 online ARTIST topic scales each consisting of 7-15 multiple-choice items. The 11 topic scales are: data collection, data representation, measures of center, measures of spread, normal distribution, probability, bivariate quantitative data, bivariate catergorical data, sampling distributions, confidence intervals, and significance tests. The ARTIST Probability Scale covers material that is identified in the intended probability curriculum in Cyprus. Therefore, it was expected that the students would be able to respond to these items. Moreover, since Cyprus participated in all TIMSS activities for grade 8 through the years, some of the TIMSS items on probability for grade 8 were included on the pre-test.

A total of 14 multiple-choice items made up the pre-test, eight from the web ARTIST project (Items 1-6, 8 and 9) and six from the TIMSS studies (Items 7 and 10-14) (See Appendix B). Table 4.3 provides a listing of the sources of the pre-test items along with information made available for some of the TIMSS items regarding the international percentage of students responding correctly to an item.

Some of these items underwent modifications based on feedback from experts and based on participant responses in the pilot study. This feedback was also used to help reduce the number of items on the initial version of the pre-test so that all items on the pre-test could then be embedded on the final exam in spring 2010 as part of the post-test. Students were given 4% towards their course grade for completing the pre-test.

Table 4.3

Pre-Test Items - Sources

Item(s)	Source
1-6, 8, 9	ARTIST Probability Scale
7	TIMSS 1995 (TIMSS & PIRLS International Study Center, 1995)
10	TIMSS 1995 (TIMSS & PIRLS International Study Center, 1995)
11	TIMSS 1999 (TIMSS & PIRLS International Study Center, 2001)
12	TIMSS 2007 (TIMSS & PIRLS, 2009)
13	TIMSS 2003 (TIMSS & PIRLS International Study Center, 2007)
14	TIMSS 2007 (TIMSS & PIRLS, 2009)

4.5.3. Probability Post-test

The post-test was embedded in the course final exam which was administered on June 4th, 2010 and was used to assess students' knowledge and understanding of probability after instruction. When constructing the probability post-test, the intended curriculum for MAT 201 – Statistics I was considered. All items on the pre-test were embedded on the post-test. In addition, the post-test included a multiple-choice item on discrete probability distributions and two free-response items: one on conditional probability and one on the Binomial distribution. The post-test is included in Appendix B.

Grading policy at the research site specified that the final exam for a course should count for 70% of the course grade and should have duration of 3-4 hours. The final exam for MAT 201 in spring 2010 had duration of 3.5 hours. The post-test was the part of the final exam that

comprised of 15 multiple-choice items and 2 free-response items which examined the topics under consideration in this study.

Several items, on the pre-test and on the post-test, required students to use theoretical probability. For example, in items 4 and 5 students were provided with the sample space of rolling two fair dice and were asked to determine equally likely and non-equally likely events. Item 17 required students to determine the probabilities of events following the binomial distribution and to compute the mean and variance of the distribution. Moreover, item 18 involved finding conditional probabilities.

Some items on both the pre-test and post-test, described experimental situations. For example, item 1 dealt with buying lottery tickets and recognizing equally likely, independent events. Item 3 involved flipping a fair coin and indicating the chance of getting a head on the next flip after getting five consecutive heads. In addition, item 6 described a situation in which the experimental approach would need to be used to estimate the probabilities of events. Moreover, items 8 and 9 described experimental situations involving two containers filled with various quantities of red and blue marbles, in which students needed to recognize equally likely and non-equally likely events. Item 12 involved a probability situation of selecting marbles from a bag without replacement and indicating what is the likely color of the next marble.

4.6 Data Analysis Plan

4.6.1 Quantitative Analysis

In order to be able to carry out quantitative analyses, student responses on the pre-test and post-test items were coded. The pre-test comprised of multiple-choice items only and the codes used in SPSS were: 0 for incorrect; 1 for correct; and 999 for missing data. The same codes were used for the multiple-choice items on the post-test. For each free-response item on the post-test a

scoring rubric was created (See Appendix D). The codes used in this case were: 0 for incorrect; 1 for partially correct; 2 for completely correct; and 999 for missing data. The researcher consulted with the Center for Statistics on the MSU campus and the members of the committee when deciding which method was best to use in quantitatively analyzing the data.

Relative to the research questions, first a comparison of *gain scores* within each group (control and treatment) from pre-test to post-test was performed in order to determine the effect of each instructional method on students' *achievement* in probability. That is, the pre-test scores of the treatment group were compared to the post-test scores of the same group in order to determine if the instructional treatment had an effect on students' achievement. Also, the control group's pre-test and post-test scores were compared to determine if the instructional method used in this group had an effect on students' achievement. Second, a comparison of *normalized gain scores* was performed in order to establish whether the instructional method used in the treatment group (Instructional Method B) had a better effect on students' achievement on probability than the instructional method used in the control group (Instructional Method A). Moreover, a comparison of post-test scores on the open-ended items and of post-test total scores was carried out.

In each of these comparisons, descriptive measures were first computed for student scores in the control and treatment groups and next, tests were performed to determine whether the data followed the normal distribution. Normality tests helped determine the type of test (parametric versus non-parametric) to be used when comparing student scores in the treatment and control groups. In the cases that data followed the normal distribution, Welch's test was used to compare student scores in the treatment and control groups; this test does not assume equal variances between two samples and so, it is a more general method to use than ANOVA. If data

did not follow the normal distribution, the Mann-Whitney non-parametric test was used to carry out such comparisons. "The Mann-Whitney test is used for testing differences between means when there are two conditions and different subjects have been used in each condition" (Field, 2000, p. 49). In the case of analyzing gain scores, if data did not follow the normal distribution, then the Wilcoxon Signed-Ranks non-parametric test was used. "The Wilcoxon test is used in situations in which there are two sets of scores to compare, but these scores come from the same subjects" (Field, 2000, p. 54). Furthermore, items in which students performed substantially worse on the post-test than on the pre-test were identified as this seemed to be an odd occurrence after formal instruction on probability. Data on these individual items are presented in Chapter 5; possible explanations for this decrease in performance are provided in Chapter 7.

In addition to an analysis of scores relative to achievement, a distractor analysis of multiple-choice items was carried out in order to determine the effects of the instructional treatment on students' *understanding* of probability. As a first step, the percent-correct responses on each multiple-choice item on the pre-test and post-test were computed in order to determine how difficult students found these items to be. The correct response to each multiple-choice item on the pre-test and post-test is provided in Table 4.4. The multiple-choice items that students found to be less difficult on the post-test compared to the pre-test, as well as those that they found to more difficult on the post-test compared to the pre-test were identified.

As a second step in the distractor analysis, the heuristic or misconception associated with each distractor on each multiple-choice item was identified by the researcher. This was based on information on heuristics and misconceptions that have been defined in the literature such as the equiprobability bias, positive and negative recency, outcome approach and representativeness heuristic. The list of item distractors and associated heuristics or misconceptions was then

examined by the researcher's dissertation committee members and a meeting was held to discuss any disagreements. Following this, a revised list of item distractors and associated heuristics or misconceptions was created. The percentage of students that selected each distractor on the pre-test as well as on the post-test was determined (see Appendix D). Then, item parts assessing the same heuristic or misconception were grouped together and the mean percentage of students who applied the particular heuristic or misconception on the pre-test and on the post-test in each group was computed. Moreover, the heuristics or misconceptions for which the mean percentage of students who applied them in each group increased or decreased were indicated.

Table 4.4

Content – Heuristic – Misconception Assessed By Each Multiple-Choice Item

Item	Correct Response Content Assessed by Item	Incorrect Response Heuristic/Misconception Assessed
1	c - Student recognizes equally likely independent events	a - Negative recency/Gambler's Fallacy b - Positive recency
2	c - Student expresses an understanding of the meaning of probability	a - Misconception that 'chance' means 'being lucky' b - Deterministic approach to probability
3	b - Student recognizes equally likely independent outcomes	a - Negative recency/Gambler's Fallacy c - Positive recency
4	d - Student recognizes equally likely events	a – Representativeness b - Representativeness c – Representativeness
5	b - Student correctly computes probability using combinatorial reasoning	a – Equiprobability bias c - Inability to use sample space provided to compute probability d - Inability to use combinatorial reasoning or sample space provided
6	b - Student expresses an understanding of the Law of Large Numbers	a - Equiprobability bias c - Equiprobability bias

Table 4.4 (cont'd)

7	d - Student is able to use the sample space to compute the probability of the union of events	a - Belief that the union (i.e. 'or') of events means considering only one of the events b – P(A U B) = [P(A) + P(B)]/2 c - Belief that the union (i.e. 'or') of events means considering only one of the events e - P(AUB) = P(A) + P(B)
8	b - Student understands that probability of a compound event is lower than or equal to the probability of its parts (i.e. the simple events that make it up)	a - Conjunction fallacy c - Equiprobability bias
9	c - Student is able to i) use a two-way frequency table to compute probabilities and ii) use the relative size of outcomes to compare probabilities	a - Reliance on absolute size instead of relative size when comparing probabilities b - Reliance on absolute size instead of relative size when comparing probabilities
10	c - Student is able to use a two-way frequency table to compute the probability of a compound event	a - When computing the probability of a compound event using a two-way table student divides by column total b - When computing the probability of a compound event using a two-way table student divides by row total d - Misconception that since *one* item is selected at random, the numerator must be 1 and probability = 1/number of favorable outcomes

Table 4.4 (cont'd)

11	c - Student is able to use the sample space to compute the probability of an event	a - Misconception that since *one* item is selected at random, the numerator must be 1 and probability = 1/number of favorable outcomes b - Misconception that since *one* item is selected at random the numerator is 1 and the denominator is the characteristic associated with the outcome i.e. P(multiple of 3) = 1/3 d - Misconception that probability = # of favorable outcomes/# of remaining outcomes (i.e. use odds)
12	a - Student reasons about a probabilistic situation that involves selections 'without-replacement'	b - Positive recency c - Equiprobability bias d - Ignores information about a 'without-replacement' situation. Outcome approach.
13	b - Student recognizes non-equally likely outcomes given in absolute/frequency terms	a – Student is off task c - Confuse 'least likely' with 'most likely' d – Student is off task
14	a - Student expresses an understanding of the concept of probability when information is given in absolute/frequency terms	b - Misconception that a larger number of outcomes in sample space means higher probability c - Equiprobability bias d - Outcome approach
15 (Post-Test only)	b - Student is able to find the expected value of a discrete probability distribution	a - Ignores probability associated with each outcome and thinks that expected value is median value c – Unweighted average of outcomes d – Expected value / number of categories

4.6.1.1 Initial Equivalence of Groups

In order to be able to investigate whether the two instructional methods under consideration had a significant effect on students' achievement and understanding of probability,

it was necessary to first identify whether the students in the three course sections had comparable initial probability knowledge through an analysis of pre-test scores. The pre-test comprised of 14 multiple-choice items (See Appendix B). In SPSS, a correct response to an item on the pre-test received a value of 1 and an incorrect response a value of 0. Missing data was coded as 999. Therefore, the minimum score on the pre-test was 0 and the maximum 14.

Descriptive measures of pre-test scores by course section are provided in Table 4.5. The morning treatment section had the highest mean and median whereas the evening treatment section had the lowest mean and median. Also, the control group had the smallest degree of variation from the mean whereas the evening treatment section had the highest variation from the mean as well as the highest standard error.

Table 4.5

Descriptive Measures Of Pre-Test Scores By Course Section

	N	Mean	Median	Standard Deviation	Standard Error of the Mean
Control (Section B / morning section)	17	8.7	9	1.9	0.5
Treatment 1 (Section A / morning section)	20	9.6	10	2.4	0.5
Treatment 2 (Section C / evening section)	7	7.3	8	3.4	1.3

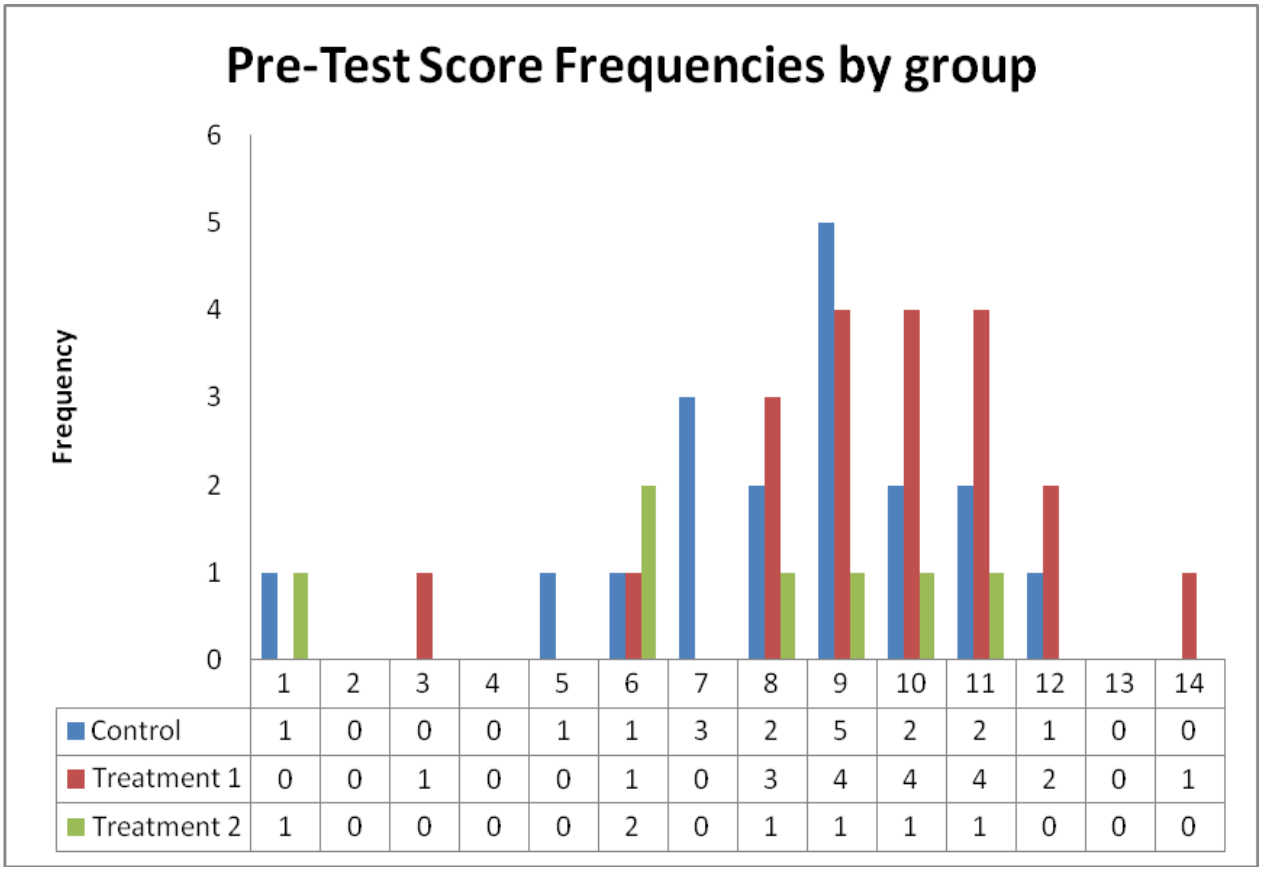
A further breakdown of the pre-test data by score gave rise to the following frequencies as presented in Figure 4.1. This breakdown indicates that the mode for the control group was 9, for the morning treatment section the modes were 9, 10 and 11 whereas for the evening treatment section it was 6. When data from the three course sections were combined, the mode was 9. Overall, 30 of the 44 participants received scores between 8 and 11 on the pre-test. Of these 30

participants, 11 were from the control group, 15 from the morning treatment section and 4 from the evening treatment section.

Figure 4.1

Pre-Test Score Frequencies By Course Section

For Interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Control	1	0	0	0	1	1	3	2	5	2	2	1	0	0
Treatment 1	0	0	1	0	0	1	0	3	4	4	4	2	0	1
Treatment 2	1	0	0	0	0	2	0	1	1	1	1	0	0	0

Following the computation of descriptive measures, an analysis was carried out to identify whether the pre-test results followed a normal distribution. When the data from all three sections was considered together, both the Kolmogorov-Smirnov and the Shapiro-Wilk tests for normality resulted in p-values lower than 0.05 (0.005 and 0.023 respectively for each test). This meant that the distribution of the sample was significantly different from a normal distribution.

Since the treatment occurred in two different sections (Sections A and C) which had considerably different class sizes (20 versus 7 students respectively) and occurred at different times during the day (morning versus evening), it was necessary to check for class effects prior to carrying out analyses to determine whether the three course sections were initially equivalent with respect to their knowledge of probability. So, an analysis was first carried out for the three sections separately using a non-parametric method due to the small sample sizes of 20, 17 and 7 participants in Sections A, B and C respectively, and due to the fact that the sample data did not follow a normal distribution. In particular, the Kruskal-Wallis test for three independent samples was carried out using SPSS with codes assigned as follows: 0 for participants in the control group i.e. Section B; 1 for participants in the first treatment section i.e. Section A (morning); and 2 for participants in the second treatment section i.e. Section C (evening). Following this, the Kruskal-Wallis test was carried out for k = 2 independent samples. In the case of two independent samples, two sections were compared at a time: control versus treatment1 (i.e. Section B versus Section A); control versus treatment2 (i.e. Section B versus Section C); and treatment1 versus treatment2 (i.e. Section A versus Section C). Table 4.6 shows the results of these tests.

Table 4.6

Results Of The Kruskal-Wallis Test For k-independent Samples

		Mean Rank			
	Sample Size	k = 3	k = 2 Control vs. Treatment1	k = 2 Control vs. Treatment2	k = 2 Treatment 1 vs. Treatment2
Control (Section B-morning)	17	20.4	16.1	13.3	
Treatment1 (Section A-morning)	20	26.5	21.5		15.5
Treatment2 (Section C-evening)	7	16.3		10.6	9.6
p-value		0.13	0.13	0.40	0.09

Since in all cases the p-value was higher than 0.05, there were no significant class effects. That is, Section B (control) did not differ significantly from either Section A (treatment) or Section C (treatment) with respect to pre-test scores. In addition, the two treatment sections (Section A and Section C) did not differ significantly from each other.

Based on the above results, the data from the two treatment sections were then grouped together to form one treatment group which was compared to the control group. The comparison of the treatment versus control pre-test scores was performed using the Kruskal-Wallis non-parametric test for two samples since, as previously mentioned, the data differed significantly from the normal distribution and sample sizes were small. Note that the Kruskal-Wallis test is used in the place of a one-way ANOVA when the data is not normally distributed (Montgomery, 1997). In particular, the test resulted in a non-significant p-value of 0.37. In conclusion, the participants in the control group had comparable initial probability knowledge to the students in the treatment group.

4.6.2 Qualitative Analysis

For the purposes of analyzing students' *understanding* of probability, six students were selected (three from the treatment group and three from the control group) and their conversations were fully transcribed: one student of high mathematical ability, one student of moderate mathematical ability and one student of low mathematical ability from each group. The selection criterion was students' mathematics grade in the last year of high school. This information was obtained from the background questionnaire (question 8; see Appendix B). In order to perform this selection, students in the treatment group and in the control group were placed into three categories: those who received a mathematics grade lower than or equal to 10; those who received a mathematics grade between 11 and 15; and those who received a mathematics grade between 16 and 20 in their last year of high school. Note that the maximum grade one could receive in a high school subject is 20 with 10 being the passing grade. Next, students in the treatment group and students in the control group were assigned a number and one student from each category was randomly selected using the random generator function in Excel. Data collected in the form of audio tapes were qualitatively analyzed.

The audio-taped student conversations were transcribed in December, 2011 and January, 2012 and then underwent a qualitative analysis during February, 2012. The researcher, along with the help of a hired transcriber, completed the transcriptions of the audio-taped data. The hired transcriber was a Cypriot college senior student majoring in elementary education who had completed two courses in mathematics education and a course on research methods, and who was fluent in both Modern Greek and English. Recall that overall 19 audio recordings were collected; 10 in the treatment group and 9 in the control group. The researcher transcribed the 9 audio recordings collected in the control group and the hired transcriber the 10 audio recordings

collected in the treatment group. Once transcriptions were completed, the researcher and the hired transcriber exchanged audio tapes and their corresponding transcripts in order to check each other's work.

As mentioned in Chapter 2 the framework developed by Jones, Thornton, Langrall and Tarr (1999) was used to analyze student conversations. This framework included six constructs (sample space, experimental probability of an event, theoretical probability of an event, probability comparisons, conditional probability, and independence) across four levels of reasoning (Level 1: Subjective; Level 2: Transitional; Level 3: Informal Quantitative; and Level 4: Numerical) as specified in Table 2.1 in Chapter 2.

Once transcriptions of audio tapes were completed, the first step in the qualitative analysis was to read the transcripts to become acquainted with the data. After the first reading of the transcripts, the researcher read the transcripts a second time and noted all of the vignettes which related to a particular construct under examination in this dissertation and which involved participation of one of the six students under study. During a third reading of the transcripts, the researcher mapped each of these vignettes to the levels of reasoning on the framework. Then, the researcher read the transcripts once more to confirm her impressions and to check whether there were any instances that she disagreed with her first attempt of mapping a vignette to a level of reasoning (intra-rater reliability). No such instances were noted.

At this stage, the (same) hired transcriber was called to help with the qualitative analysis. A part of one of the transcripts for Activity 1 was used as an example to demonstrate to the transcriber how to analyze the data. Then, clean hard copies of all transcripts were provided to him. Once the hard copies of the transcripts were returned to the researcher, they were checked for agreement on i) the number of vignettes related to the constructs under study and ii) the

levels of reasoning associated with each vignette. Then, inter-rater reliability was computed. The researcher and the second coder (transcriber) met once more to discuss and resolve any instances of disagreement in coding. The repeated readings of the transcripts by the researcher, the analysis carried out by a second coder (transcriber) and examination of students' written work provided triangulation in the analysis of the data.

In qualitatively analyzing the data, two types of codes were used. These codes are specified in Table 4.7 and are used throughout this chapter.

Table 4.7

Types of Codes Used In Qualitative Analysis

Type of Code	
Person Involved in Transcript Excerpt	Level of Reasoning
THA – Treatment/High Ability Student	L1 – Level 1: Subjective Reasoning
TMA – Treatment/Moderate Ability	L2 – Level 2: Transitional Reasoning
TLA – Treatment/Low Ability	L3 – Level 3: Informal Quantitative Reasoning
CHA – Control/High Ability	L4 – Level 4: Numerical Reasoning
CMA – Control/Moderate Ability	
CLA – Control/Low Ability	
OGM – Other Group Member	
IR – Instructor/Researcher	

A major part of the qualitative analysis was to decide the level of reasoning represented by excerpts of student conversations. Table 4.8 provides a demonstration of the analysis methods used in matching excerpts to levels of reasoning with regards to the construct of theoretical probability of an event. This table indicates the detailed description of each level of reasoning as specified in the Jones et al. (1999) framework, and presents examples of responses selected from students' oral justifications to questions in the activities (treatment group) or problem sets (control group) as representing each of the four levels of reasoning particular to this construct.

Table 4.8

Theoretical Probability: Levels of Reasoning and Examples of Representative Excerpts

Level 1 Subjective	Level 2 Transitional	Level 3 Informal Quantitative	Level 4 Numerical
-predicts most/least likely event on the basis of subjective judgments -recognizes certain and impossible events	-predicts most/least likely event on the basis of quantitative judgments but may revert to subjective judgments	-predicts most/least likely events on the basis of quantitative judgments -uses numbers informally to compare probabilities	-predicts most/least likely events for 1-and simple 2-stage experiments -assigns a numerical probability to an event (either a real probability or a form of odds)
Activity 3: Treatment OGM: Why did you place chips on the high values? TMA: OGM did. She will never win. …. TMA: What mistake could you make to eliminate your chances of winning? Put all your chips on 15. TMA: Mmmm, and 13, 14 as well.	Activity 3: Treatment TMA: 8. No chips. OGM: Oh, come on! TMA: 5. No chips. … TMA: Come on OGM, make it happen! … OGM: Who won the game? … Why do you think he won? TMA: Because he was smart! … TMA: And he didn't place chips on numbers over 12.	Activity 1: Treatment OGM: Based on the theoretical probabilities that you have computed in 17 above, which sums give the best chance of winning the game you have just played in your groups? THA: Number 7. Which is six times. And we should write the pairs that give the number 7. 1-6, 6-1, 5-2, 2-5 THA: 3-4, 4-3…	Activity 1: Treatment THA: To get three is 2 out of 36 so 1 out of 18? OGM: Yes. THA: To come up four it is 1-3, 2-2, 3-1? So we have 3, 3 out of 36, so 1 out of 12? OGM: Yes.

The two excerpts provided as examples of L1 and L2 reasoning involved the participation of the moderate ability student in the treatment group, and originated from Activity 3 on discrete

probability distributions. Note that initially the TMA student was able to recognize impossible events (i.e. getting a sum of 13, 14 or 15 when two dice are rolled) and so, the particular excerpt was labeled as representing L1 reasoning. At a later stage in the activity and once the game was over, students were asked to specify their group's winner and indicate why that person won. Prior to providing a valid reason as to why the particular group member won (i.e. he had not placed any chips on numbers over 12 whereas the rest of the group members did), the TMA student made a couple of statements that were subjective in nature. In particular, the TMA student i) seemed to believe that the result of rolling two dice depends on *who* rolls the dice ("Come on OGM, make it happen!); and ii) stated that the OGM won "because he was smart!". Since, the TMA student reverted to subjective reasoning, the second excerpt taken from Activity 3 was labeled as representing L2 reasoning.

The excerpts provided as examples of L3 and L4 reasoning originated from Activity 3 and Activity 1 respectively in which students dealt with the sum of two dice. In Activity 3 students played a game in which they placed chips on a number line that had the numbers 1 – 15 written on it. Students rolled two dice on which the numbers 1-6 were written, and computed the sum of the faces; on each roll, a student that had placed a chip over the number represented by the sum removed one of his/her chips. The winner was the student who had his/her chips removed first. In the excerpt provided in Table 4.8 as representing L3 reasoning, the THA student correctly specifies that the sum of 7 gave the best chance of winning (i.e. most likely event) since it had the largest number of outcomes associated with it. The THA student also correctly listed the outcomes associated with this sum. However, the THA student did not use numerical probabilities to make valid comparisons between the various sums and so, this student exhibited L3 reasoning.

In the passage provided under L4 Reasoning, the high ability student attempted to find the probability of i) getting a sum of three and ii) getting a sum of four when two dice are rolled. In each case the student correctly considered the outcomes that result in each of these sums and assigned a correct numerical probability to the event under consideration, therefore the excerpt was labeled as representing L4 Reasoning.

4.7 Summary

This dissertation made use of a mixed-methods design that utilized a probability pre-test and post-test along with audio-taped student conversations to examine students' achievement and understanding of probability. Table 4.9 presents a summary of the data sources associated with each measure considered in this study.

Table 4.9

Data Sources and Associated Measures

	Measure	
Data Source	Achievement	Understanding
Pre-test scores (Multiple-choice items, Open-ended items, Total Scores)	✓	
Post-test scores (Multiple-choice items, Open-ended items, Total Scores)	✓	
Post-test multiple-choice item distractor analysis		✓
Audio recordings		✓

The results of the quantitative analysis relative to students' achievement and understanding of probability are presented in Chapter 5 whereas the results of the qualitative analysis of audio recordings with respect to students' understanding of probability are provided in Chapter 6.

CHAPTER 5

STUDENTS' ACHIEVEMENT AND UNDERSTANDING OF PROBABILITY RESULTS OF QUANTITATIVE ANALYSIS

This chapter presents the quantitative results that reflect the effects of the instructional treatment on students' achievement and understanding of probability. The chapter includes i) a comparison of gain scores within each group (treatment and control) and of normalized gain scores between groups relative to the multiple-choice items that were common to the pre-test and post-test; ii) a comparison of post-test scores on the open-ended items and iii) a comparison of post-test total scores. Moreover, the effects of the instructional treatment on understanding are presented through a distractor analysis of multiple-choice items. For the purposes of analysis SPSS-PASW Statistics 18 was used.

5.1 Effects of Instructional Treatment on Achievement

Recall that the pre-test included 14 multiple-choice items that were common to the post-test. The pre-test and post-test were administered to three sections of an introductory statistics course. Two of these sections served as the treatment group and one as the control group. Participants in the control group had comparable initial probability knowledge to participants in the treatment group (see Chapter 4).

5.1.1 Comparison of (Normalized) Gain Scores

5.1.1.1 Comparison of Gain Scores Within Each Group

In the comparison of gain scores, only the 14 multiple-choice items that were common to the pre-test and post-test were considered. The post-test included an additional multiple-choice item on finding the mean of a discrete probability distribution and two open-ended items which

were not included in the analysis of gain scores. In SPSS, each correct response to a multiple-choice item received a value of 1 and each incorrect response a value of 0. Therefore, for the purposes of analysis of gain scores, the minimum score that could be achieved was 0 and the maximum 14.

Descriptive statistics for the treatment and control groups indicate that while the mean and median multiple-choice score for the treatment group increased from pre-test to post-test, the mean and median for the control group decreased.

Table 5.1

Changes in Descriptive Statistics from Pre-Test to Post-Test

	Control Group	Treatment Group
Pre-Test		
Mean	8.7	8.96
Standard Error	0.45	0.53
Median	9	9
Post-Test		
Mean	6.47	9.93
Standard Error	0.55	0.40
Median	6	10

Based on the results of the analysis, the median pre-test score on the multiple-choice items was 9 in both the control and treatment groups. However, the median post-test score on the multiple-choice items in the control group decreased by 3 score points whereas in the treatment group it increased by 1 score point in comparison to the pre-test. Moreover, whereas 25% of students in the control group scored lower than 7 on the pre-test multiple-choice items, 75% of students in this group scored lower than 7 on the corresponding post-test items. That is, 50% more students in the control group scored below 7 on the post-test multiple-choice items in comparison to the corresponding pre-test items. In the case of the treatment group, 50% of

students scored below 9 on the pre-test multiple-choice items whereas 25% scored below 9 on the corresponding post-test items. Considering the above information relative to post-test multiple-choice scores, 75% of students in the control group scored below 7 while 75% of students in the treatment group scored above 9. Furthermore, 25% of students in the treatment group scored above 12 on the post-test multiple-choice items whereas none of the students in the control group achieved such scores.

Since the data of pre-test multiple-choice scores did not follow a normal distribution (see Chapter 4), the Wilcoxon Signed-Ranks (non-parametric) test was used to examine the score gains of students on the multiple-choice items. "The Wilcoxon test is used in situations in which there are two sets of scores to compare, but these scores come from the same subjects" (Field, 2000, p. 54). In this case, the two sets of scores that were compared were the pre-test and post-test multiple-choice scores. The same students who took the pre-test also completed the post-test. The control group data included 12 negative ranks meaning that 12 out of the 17 students in this group received a lower score on the post-test multiple-choice items compared to the corresponding pre-test items. Only two students (i.e. 12%) in the control group performed better on the post-test multiple-choice items in comparison to the pre-test multiple-choice items. The data from the treatment group included only 4 negative ranks whereas 14 out of the 27 students (i.e. 52%) in this case received higher scores on the post-test multiple-choice items than on the pre-test multiple-choice items.

Table 5.2

Wilcoxon Signed-Ranks Tests for Multiple-Choice Gain Scores

	Control Group	Treatment Group
Negative Ranks	12	4
Positive Ranks	2	14
Ties	3	9
Total (i.e. N)	17	27
z-score	-2.52	-1.71
z-score basis	Positive ranks	Negative ranks
Significance	0.012	0.087

Relative to the z-scores generated by the analysis, these were significant only in the case of the control group. The negative z-score (-2.52) in this case was based on the positive ranks, meaning that student scores on the multiple-choice items moved in the opposite direction (i.e. decreased) from pre-test to post-test. Since this z-score was significant ($p = 0.012 < 0.05$), this means that the multiple-choice scores of students in the control group were significantly lower on the post-test compared to the pre-test. In the case of the data from the treatment group, the negative z-score (-1.71) was based on the negative ranks, meaning that student scores on the multiple-choice items moved in the same direction (i.e. increased) from pre-test to post-test. However, this z-score was not significant ($p = 0.087 > 0.05$), meaning that student scores on the multiple-choice items of students in the treatment group did not differ (increase) significantly from the pre-test to the post-test.

Overall, the results of the Wilcoxon Signed-Ranks tests indicated that i) students in the control group performed significantly lower on the post-test multiple-choice items and ii) students in the treatment group did not perform significantly different on the post-test multiple-choice items compared to the corresponding pre-test items.

In addition to the above results relative to score gains, figures were created using Excel that demonstrate the percent-correct responses on each multiple-choice item on both the pre-test and post-test (Figure 5.2 and Figure 5.3 for the control and treatment groups respectively). Furthermore, Excel was used to construct a figure which shows the differences in percent-correct responses between the pre-test and post-test for each multiple-choice item (Figure 5.4 on score gains).

As can be seen in Figure 5.1, in the case of students in the control group, there was an increase in the percent-correct responses on some items, a reduction in percent-correct responses on other items while the percentage of correct responses remained stable in the case of two items. In particular, in the control group gain scores were as follows:

Increase in percent-correct: Items 5, 7, and 10.

Percentage increase: 5.9%, 5.8% and 5.9% respectively.

Stable percent-correct: Items 2 and 9.

Decrease in percent-correct: Items 1, 3, 4, 6, 8, 11, 12, 13 and 14.

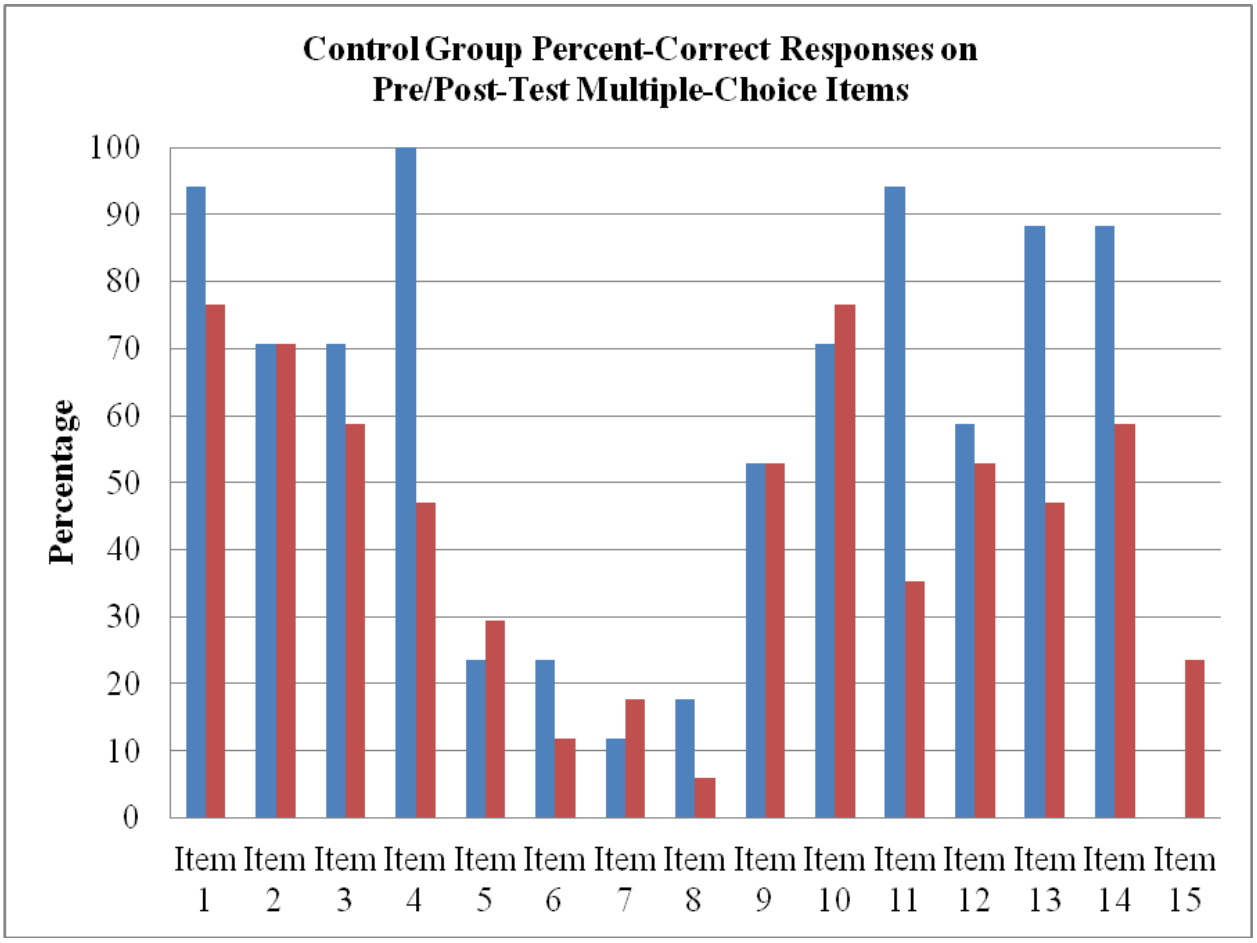
Percentage decrease: 17.6%, 5.9%, 52.9%, 11.7%, 11.7%, 58.8%, 5.9%, 41.1%, and 29.4% respectively.

These results indicate that for a majority of items, students in the control group had a lower performance on the post-test compared to their performance on the pre-test. Moreover, all of the percentage descreases are higher compared to the percentage increases. This is in agreement with the Wilcoxon test results which indicated that the post-test multiple-choice scores of students in the control group were significantly lower than their multiple-choice scores on the pre-test. In particular, the most substantial *descreases* in percent-correct responses were with regards to Items 4, 11 and 13. A discussion of the possible explanations for these

considerable percent-correct descreases is provided in Chapter 7. Note that in the case of Item 4, responses exhibited a ceiling effect since all students in the control group provided the correct answer to this item on the pre-test and so, it was not surprising that percent-correct responses decreased on the post-test.

Figure 5.1

Control Group % Correct Responses on Pre/Post-Test Multiple-Choice Items



As Figure 5.2 shows, students in the treatment group demonstrated an increase in percent-correct responses on some items, a reduction in percent-correct responses on other items while the percent-correct responses remained stable in the case of three items.

Increase in percent-correct: Items 1, 3, 5, 7, 10 and 11

Percentage increase: 7.4%, 25.9%, 22.3%, 18.6%, 48.2% and 3.7% respectively.

Stable percent-correct: Items 8, 12 and 14.

Decrease in percent-correct: Items 2, 4, 6, 9 and 13.

Percentage decrease: 11.1%, 7.4%, 3.7%, 3.7% and 3.7% respectively.

Therefore, students in the treatment group had a percent-correct increase on six multiple-choice items and a percent correct decrease on five multiple-choice items. Moreover, most of the percentage increases were higher than the percentage decreases. The most substantial *increases* in percent-correct responses were with regards to Items 3, 5, and 10. Note that, similar to the results in the control group, the percent-correct responses on Item 8 remaimed low on both the pre-test and post-test, indicating that this was an *impossible* item for students.

Figure 5.2

Treatment Group % Correct Responses Pre/Post-Test Multiple-Choice Items

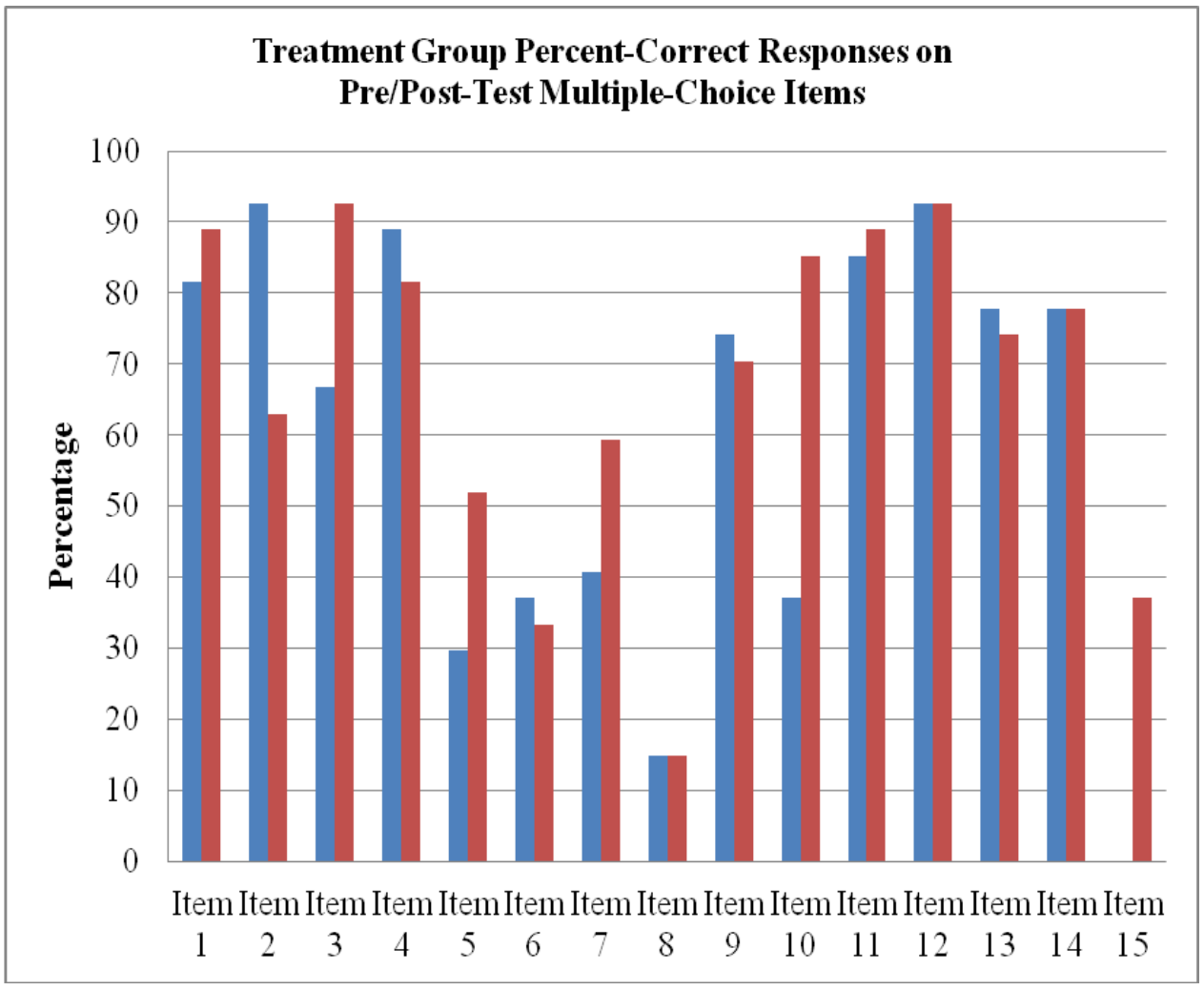


Table 5.3 provides an additional demonstration of what has been previously discussed on the basis of Figures 5.2 and 5.3. The results shown in Table 5.3 indicate that students in the control group had a positive gain on three multiple-choice items; stable percent-correct responses on two multiple-choice items and a negative gain on nine multiple-choice items. Specific to the treatment group, students had a positive gain on six multiple-choice items; stable percent-correct responses on three multiple-choice items and a negative gain on five multiple-choice items.

Table 5.3

Changes in Percent-Correct Responses on Multiple-Choice Items

Item	Control Group % Change	Treatment Group % Change
1	-17.6	7.4
2	0	-11.1
3	-5.9	25.9
4	-52.9	-7.4
5	5.9	12.3
6	-11.7	-3.7
7	5.8	18.6
8	-11.7	0
9	0	-3.7
10	5.9	48.2
11	-58.8	3.7
12	-5.9	0
13	-41.1	-3.7
14	-29.4	0

5.1.1.2 Comparison of Gain Scores Between Groups

In addition to the descriptive measures computed and the analysis of gain scores for each of the two groups, a final piece of analysis was carried out with respect to gain scores which directly addressed the research questions i.e. whether Instructional Method B (treatment) had a better effect on students' achievement on probability than Instructional Method A (control). In order to carry out this part of the analysis, *normalized* gain scores were used (Bao, 2006; Hakes, 1998). That is, instead of simply subtracting pretest scores from posttest scores to produce the raw gain, the raw gain was divided by the possible gain to get a ratio, g. This ratio was computed for each individual student in each of the two groups (control and treatment) using the following formula (Bao, 2006):

$$g = \frac{posttest \quad score - pretest \quad score}{\max imum \quad score - pretest \quad score}$$

Based on this formula, the value of the normalized gain would, at most, be equal to 1. A problem is caused when a student scores perfectly on the pre-test since in that case, the denominator in the formula becomes zero (undefined). In this study, only one student (in the treatment group) received a perfect score on the pre-test multiple-choice items so, a negligible amount of data was lost due to this effect. In particular, the data collected from this student was not used when carrying out tests to compare the normalized gain scores between the two groups.

Positive as well as negative normalized gains were possible. A student exhibited a negative normalized gain score when the score received on the multiple-choice items was lower on the post-test than on the pre-test. Particular to this study, 12 students in the control group and 4 students in the treatment group demonstrated negative normalized gain scores. The mean normalized gain in the control group was -0.64 (standard error 0.23) whereas in the case of the treatment group it was 0.066 (standard error 0.11).

Normality tests were then carried out to determine whether the data followed the normal distribution. Both the Kolmogorov-Smirnov ($p = 0.011$ and $p = 0.000$ for the control and treatment groups respectively) and the Shapiro-Wilk tests ($p = 0.032$ and $p = 0.000$ for the control and treatment groups respectively) resulted in p-values lower than 0.05, indicating that the data differed significantly from the normal distribution.

Subsequently, the Mann-Whitney non-parametric test was used to compare the normalized gain scores of the control and treatment groups. The test resulted in a p-value of 0.001 (< 0.05) indicating that the normalized gain scores of the treatment group were significantly different from the normalized gain scores of the control group. Note than in the Mann-Whitney test, scores are ranked from lowest to highest. The test generated a mean rank of 14.41 for the control group and a mean rank of 26.96 for the treatment group. This means that

the control group had a bigger number of lower normalized gain scores as compared to the treatment group. Likewise, the treatment group had a bigger number of higher normalized gain scores in comparison to the control group. In summary, Instructional Method B (treatment) had a significantly better effect on students' achievement on probability than Instructional Method A (control).

5.1.2 Comparison of Scores on Additional Post-Test Items

Apart from the 14 multiple-choice items involved in the analysis of gain scores, the post-test included an additional multiple-choice item on discrete probability distributions and two open-ended items: i) Problem 16 on simple and joint probabilities, conditional probability and independence and ii) Problem 17 on the binomial distribution (see Appendix B).

With regards to the multiple-choice item on discrete probability distributions (i.e. problem 15 on the post-test; see Appendix B), 23.5% of students in the control group and 48.1% of students in the treatment group responded correctly.

For the purposes of comparing the open-ended item scores of the control and treatment groups, scoring rubrics were created (see Appendix C) which allotted numerical values to student responses. Based on these rubrics, responses were coded by receiving a numerical value of 0 (incorrect), 1 (partially correct) or 2 (completely correct). This resulted in the first open-ended item (Problem 16 on the post-test) receiving 18 points and the second open-ended item (Problem 17 on the post-test) receiving 6 points. Therefore, the minimum score on the open-ended items that a student could receive was 0 and the maximum 24.

Descriptive statistics for the data generated by the two open-ended items are presented in Table 5.4. In particular, the mean for the treatment group was 17.48, the standard error of the mean was 0.77, and the standard deviation 3.98 (see Table 5.4). In comparison to the descriptive

statistics generated for the control group, the treatment group had a higher mean (12.24 for the control group), a lower standard error (1.34 for the control group) and a lower standard deviation (5.51 for the control group) (see Table 5.3). That is, the mean of the treatment group on the post-test open-ended items was 5.24 points higher than the mean of the control group. In addition, the scores of the control group on the open-ended items had a higher variation from the mean in comparison to the scores of the treatment group.

Table 5.4

Descriptive Statistics for the Open-Ended Item Scores

Statistic	Control Group	Treatment Group
N	17	27
Mean	12.24	17.48
Standard Error	1.34	0.77
Median	14	18
Standard Deviation	5.51	3.98

Based on the results, the control group had a median of 14, meaning that half of the students in the control group received a score lower than 14 (out of 24) on the post-test open-ended items. However, with the exception of four students in the treatment group, the remaining 23 students in this group scored higher than 14. That is, 85% of the students in the treatment group received a score higher than 14 on the post-test open-ended items. Moreover, whereas 75% of students in the control group scored lower than 16, only 25% of students in the treatment group scored lower than 16 on the post-test open-ended items.

Following the computation of descriptive statistics, tests were carried out to determine whether the data from the post-test open-ended items were normally distributed. Both the Kolmogorov-Smirnov and the Shapiro-Wilk tests resulted in non-significant p-values for the control group data (i.e. 0.11 and 0.3 respectively) and in significant p-values for the treatment

group data (i.e. 0.04 and 0.001 respectively). This meant that the distribution of the sample in the case of the control group was not significantly different from a normal distribution but in the case of the treatment group it was significantly different.

Subsequently, the Mann-Whitney non-parametric tests were used to compare the post-test results from the open-ended items in the control and treatment groups. This resulted in a p-value of 0.001 (< 0.05) indicating that the scores of the treatment group on the post-test open-ended items were significantly different from the scores of the control group. Note that in the Mann-Whitney test, scores are ranked from lowest to highest. The test generated a mean rank of 14.15 for the control group and a mean rank of 27.76 for the treatment group. This means that the control group had a bigger number of lower scores as compared to the treatment group. Likewise, the treatment group had a bigger number of higher scores in comparison to the control group.

In summary, students in the treatment group performed better on the additional post-test multiple-choice item on discrete probability distributions. Moreover, based on the results of the Mann-Whitney non-parametric test, the achievement of students in the treatment group on the post-test open-ended probability items was significantly higher than the achievement of students in the control group on the corresponding items. Therefore, Instructional Method B (using lectures and small-group cooperative learning sessions that involve the use of activities that generate real data) was successful in producing significantly higher achievement scores on the post-test open-ended items compared to Instructional Method A (using lectures and small-group cooperative learning sessions during which students solved probability problems).

5.1.3 Comparison of Post-Test Total Scores

As previously mentioned, a student could receive a maximum score of 15 on the multiple-choice items and a maximum score of 24 on the open-ended items on the post-test. Thus, overall, the maximum possible score on the post-test was 39.

Descriptive statistics for the data generated by the post-test items are presented in Table 5.5. Particular to the treatment group the mean was 27.89, the standard error of the mean was 1.01, and the standard deviation 5.26. In comparison to the descriptive statistics generated for the control group, the treatment group had a higher mean (18.94 for the control group), a lower standard error (1.61 for the control group) and a lower standard deviation (6.64 for the control group). That is, the mean of the treatment group was 8.95 score points higher than the mean of the control group. In addition, the post-test scores of the control group had a higher variation from the mean in comparison to the post-test scores of the treatment group.

Table 5.5

Descriptive Statistics for the Open-Ended Item Scores

Statistic	Control Group	Treatment Group
N	17	27
Mean	18.94	27.89
Standard Error	1.61	1.01
Median	20	29
Standard Deviation	6.64	5.26

Based on the results, the median for the control group was 20 whereas the median for the treatment group was 29. Moreover, in the control group 75% of the data was below 22 whereas in the treatment group 74% of the data was above 27 (7 students in the treatment group received scores 27 or lower).

Following the computation of descriptive statistics, tests were carried out to determine whether the post-test data were normally distributed. Both the Kolmogorov-Smirnov and Shapiro-Wilk tests resulted in p-values higher than 0.05 in the case of the control group (i.e. 0.2 and 0.9 respectively for each test) and in p-values lower than 0.05 in the case of the treatment group (i.e. 0.00 for both tests). Based on these results, the data for the control group did not differ significantly from the normal distribution whereas the data for the treatment group deviated significantly from the normal distribution.

Subsequently, non-parametric tests were used to compare the total post-test scores of the control and treatment groups. In particular, the Mann-Whitney test was used. This resulted in a p-value of 0.00 (< 0.05) indicating that the post-test scores of the treatment group were significantly different from the post-test scores of the control group. The test generated a mean rank of 12.53 for the control group and a mean rank of 28.78 for the treatment group. This means that the control group had a bigger number of lower scores compared to the treatment group. Likewise, the treatment group had a bigger number of higher scores compared to the control group.

In summary, the post-test achievement of students in the treatment group was significantly higher than the post-test achievement of students in the control group. Therefore, Instructional Method B was successful in producing significantly higher post-test scores compared to Instructional Method A.

5.2 Effects of Instructional Treatment on Understanding

As mentioned in Chapters 1 and 2, the meaning of understanding given by Thompson and Saldanha (2003) of "assimilation to a scheme" is used in this study. This allows for correct as well as incorrect or inappropriate understandings (i.e. misconceptions) people may have. In this

study, students' understanding of probability includes both appropriate as well as inappropriate understandings. Therefore, two things were specified for each multiple-choice item on the pre-test and post-test used in this study with regards to probabilistic understanding:

i) The goal of the item (correct choice). That is, the probability content or concept that the item addressed and the student tried to understand and
ii) The heuristic or misconception associated with each distractor (incorrect choice) and which the student might have applied on a particular item.

In this dissertation, students' understanding of probability is measured through:

i) a distractor (quantitative) analysis of student responses to the multiple-choice items on the pre-test and post-test (Chapter 5) and
ii) a qualitative analysis of audio-taped conversations as students worked in groups on activities (treatment) or problem sets (control) (see Chapter 6).

5.2.1 Multiple-Choice Items: Content Assessed and Difficulty Level

As a first step in the distractor analysis, the percent-correct responses for each multiple-choice item were computed to determine how difficult the 44 students who completed the pre-test and post-test found these items to be. If less than 30% of participants provided the correct response to an item then participants found the item to be difficult. If 30-80% of participants provided the correct response to an item then participants found the item to be of moderate difficulty, and if more than 80% of participants responded correctly then participants found the item to be easy. Table 5.6 provides the percent-correct responses for each multiple-choice item on the pre-test and post-test used in this study (see also Appendix E), along with the difficulty level exhibited by students.

As Table 5.6 shows, there was a change in the difficulty level exhibited by students in 11 of the 14 multiple-choice items that were common to the pre-test and post-test. In particular, student difficulty level changed from Moderate to Easy in the case of 2 items, whereas in the case of 6 items it changed from Easy to Moderate. In addition, student difficulty level changed from Difficult to Moderate in the case of 2 items whereas for 1 item it changed from Moderate to Difficult. Moreover, the difficulty level exhibited by students in Items 1, 8, and 9 remained stable while students found item 15, which was only included on the post-test, to be difficult.

Table 5.6

Percent-Correct Responses on Multiple-Choice Items (N = 44)

Item	Pre-Test % Correct	Student Difficulty Level	Post-Test % Correct	Student Difficulty Level
1	86.4	Easy	84.1	Easy
2	84.1	Easy	77.3	Moderate
3	68.2	Moderate	81.8	Easy
4	93.2	Easy	68.2	Moderate
5	27.3	Difficult	43.2	Moderate
6	31.8	Moderate	25	Difficult
7	29.5	Difficult	43.2	Moderate
8	15.9	Difficult	11.4	Difficult
9	65.9	Moderate	63.6	Moderate
10	50	Moderate	81.8	Easy
11	88.6	Easy	68.2	Moderate
12	79.5	Moderate/Easy	77.3	Moderate
13	81.8	Easy	63.6	Moderate
14	81.8	Easy	70.5	Moderate
15			38.6	Difficult

Table 5.7 indicates the content assessed by the four items (Items 3, 5, 7, and 10; see Appendix B) in which there was a positive shift in the difficulty level exhibited by students i.e. Moderate to Easy or Difficult to Moderate as well as the percentage change in correct responses in the control and treatment groups. Notice that in the case of the control group, changes in

percent correct responses on these four items varied from -5.9% to +5.9% whereas in the treatment group such changes varied from 18.6% to 48.2%. The highest overall increase in percent correct responses occurred in the treatment group with regards to Item 10 which involved using a two-way frequency table to compute the probability of a compound event. In the case of the control group the highest increase in percent correct responses was at 5.9% in Item 5 (computing probability using combinatorial reasoning) and in Item 10. A breakdown of percent-correct responses to these multiple-choice items on the pre-test and post-test for each course section is provided in Appendix E.

Table 5.7

Multiple-Choice Items with Positive Shift in Student Difficulty Level

Item	Content Assessed	Change in Difficulty Level	Change in Percent Correct
3	Recognize equally likely independent outcomes	From Moderate to Easy	Overall: + 13.6% Control: -5.9% Treatment:
5	Compute probability using combinatorial reasoning	From Difficult to Moderate	Overall: + 15.9% Control: +5.9% Treatment: +22.3%
7	Use sample space to compute probability of union of events	From Difficult to Moderate	Overall: + 13.7% Control: +5.8% Treatment: +18.6%
10	Use a two-way frequency table to compute probability of a compound event	From Moderate to Easy	Overall: + 31.8% Control: +5.9% Treatment: +48.2%

Table 5.8 specifies the content covered in the seven multiple-choice items (Items 2, 4, 6, 11, 12, 13, and 14) in which there was a negative shift in the difficulty level exhibited by students i.e. Easy to Moderate or Moderate to Difficult as well as the percentage change in correct responses in each group relative to these items. Notice that the highest overall decrease (-25%) in percent

correct responses occurred in Item 4 which involved recognizing equally likely events, while the second highest overall decrease (-20.4%) occurred in Item 11 in which students were asked to use the sample space to compute the probability of an event. These were the two items in which the control group's percent correct responses had the biggest decreases (-52.9% in Item 4 and -58.8% in Item 11). In the case of the treatment group the highest loss in percent correct responses was in Item 2 which related to the meaning of probability. A breakdown of percent-correct responses on every multiple-choice item on the pre-test and post-test in each course section is provided in Appendix E.

Table 5.8

Multiple-Choice Items with Negative Shift in Student Difficulty Level

Item	Content Assessed	Change in Difficulty Level	Change in Percent Correct
2	Understanding of the meaning of probability	Easy to Moderate	Overall: -6.8% Control: 0% Treatment: -11.1%
4	Recognition of equally likely events	Easy to Moderate	Overall: -25% Control: -52.9% Treatment: -7.4%
6	Understanding of the Law of Large Numbers	Moderate to Difficult	Overall: -6.8% Control: -11.7% Treatment: -3.7%
11	Use sample space to compute the probability of an event	Easy to Moderate	Overall: -20.4% Control: -58.8% Treatment: 3.7%
12	Reason about a 'without-replacement' probabilistic situation	Easy to Moderate	Overall: -2.2% Control: -5.9% Treatment: 0%
13	Recognize non-equally likely outcomes given in absolute/frequency form	Easy to Moderate	Overall: -18.2% Control: -41.1% Treatment: -3.7%
14	Understand concept of probability when information is given in absolute/frequency terms	Easy to Moderate	Overall: -11.3% Control: -29.4% Treatment: 0%

5.2.2 Multiple-Choice Items: Distractor Analysis

As previously mentioned, for each multiple-choice item on the pre-test and post-test, the heuristic or misconception associated with each distractor was specified (see Table 4.4 in Chapter 4). A table is included in Appendix E that indicates the percentage of students in each course section that selected each distractor on each multiple-choice item on the pre-test as well as on the post-test. For the purposes of distractor analysis, item parts assessing the same heuristic or misconception were grouped together. The percentages of students selecting the particular distractor on each of the grouped items were added up and the mean percentage of students applying the particular heuristic or misconception on the pre-test and post-test in each group was computed. Table 5.9 provides this information for heuristics or misconceptions that have been defined in the literature such as the equiprobability bias, positive and negative recency, outcome approach, representativeness, and use of absolute size when computing probabilities.

As the results in Table 5.9 show, in the case of the control group, the mean percentage of students who applied the aforementioned heuristics or misconceptions either increased or remained the same from the pre-test to the post-test. In particular, the mean percentage of students who applied the outcome approach remained stable at 5.9%; the same was true with regards to using the absolute size instead of relative size when computing probabilities (percentage remained stable at a high 20.6%). In addition, there was a slight increase in the mean percentage of students who applied the equiprobability bias which remained high on both the pre-test and the post-test (28.4% and 29.4% respectively). Moreover, there was a small increase in the mean percentage of students who applied the negative recency (by 2.9%) and the positive recency (by 3.9%). The biggest change was in the application of the representativeness heuristic (17.7% increase).

Unlike the results of the control group, the mean percentage of students in the treatment group who applied the heuristics or misconceptions listed in Table 5.9 mostly decreased. In particular, the mean percentage of students who used the equiprobability bias, negative recency, positive recency and outcome approach decreased by 3.4%, 9.25%, 4.9% and 1.85% respectively. Only in the case of the representativeness heuristic and use of absolute size instead of relative size when computing probabilities the mean percentage increased by 3.7% and 2% respectively.

Table 5.9

Mean Percentage Of Students Applying Particular Heuristics or Misconceptions

Heuristic or Misconception	Item Distractor	Control		Treatment	
		Pre-Test	Post-Test	Pre-Test	Post-Test
Equiprobability Bias	5a, 6a, 6c, 8c, 12c, 14c	28.4	29.4	20.4	17
Negative Recency	1a, 3a	11.8	14.7	11.1	1.85
Positive Recency	1b, 3c, 12b	5.9	9.8	8.6	3.7
Outcome Approach	12d, 14d	5.9	5.9	3.7	1.85
Representativeness	4a, 4b, 4c	0	17.7	2.5	6.2
Use absolute size instead of relative size when computing probabilities	9a, 9b	20.6	20.6	11.1	13.1

Next, Table 5.10 considers students' responses relative to other types of misconceptions represented by item distractors. With regards to the first misconception listed in Table 5.10, the mean percentage of students who believed that a larger number of outcomes in the sample space implies a higher probability of occurrence increased in the case of the control group (from 0% to 5.9%) and decreased in the case of two treatment group (by 3.7%). Second, the mean percentage of students who applied division by an incorrect total when computing probabilities using a two-

way table increased by 3% in the case of the control group and by 11.1% in the case of the treatment group. Regarding the misconception that probability = 1/number of favorable outcomes when one item is selected at random, the mean percent of students who applied this misconception remained stable at 5.6% in the case of the treatment group but increased by 23.6% in the case of the control group.

Notice that three of the misconceptions listed in Table 5.10 relate to the union of events. The mean percentage of students who applied the first of these misconceptions (i.e. the term 'or' means considering only one of the events) was higher in the case of the control group on both the pre-test and post-test while in both groups the percentage of students who applied this misconception slightly increased from pre-test to post-test. Similarly, the mean percentage of students who applied the misconception that $P(A \cup B) = \frac{P(A) + P(B)}{2}$ was higher and remained stable at 29.4% in the case of the control group on both the pre-test and post-test whereas it decreased by 7.4% in the case of the treatment group (from 14.8% to 7.4%). Unlike the first two misconceptions relating to the union of events, the mean percentage of students who applied the misconception $P(A \cup B) = P(A) + P(B)$ decreased in the case of the control group (by 5.9%) and increased in the case of the treatment group (by 3.7%).

Table 5.10

Mean Percentage of Students Who Applied Particular Misconceptions

Misconception	Item Distractor	Control		Treatment	
		Pre-Test	Post-Test	Pre-Test	Post-Test
Larger number of outcomes in sample space means higher probability	14b	0	5.9	3.7	0
Division by incorrect total when computing probability using a two-way table	10a 10b	5.9	8.9	0	11.1
Union of events (i.e. 'or') means considering only one of the events	7a 7c	11.8	14.7	3.7	5.6
$P(A \cup B) = \frac{P(A) + P(B)}{2}$	7b	29.4	29.4	14.8	7.4
$P(A \cup B) = P(A) + P(B)$	7e	23.5	17.6	18.5	22.2
Selecting *one* item at random means prob. = 1 / number of favorable outcomes	10d 11a	0	23.6	5.6	5.6

5.3 Summary

In summary, based on the analysis of gain scores, the achievement of students in the treatment group on the post-test multiple-choice probability items was significantly higher than the achievement of students in the control group on the corresponding items. Second, based on the results of the Mann-Whitney non-parametric test, the achievement of students in the treatment group on the post-test open-ended probability items was significantly higher than the achievement of students in the control group on the corresponding items. Moreover, the overall post-test achievement of students in the treatment group was significantly higher than the post-test achievement of students in the control group. Therefore, Instructional Method B (using lectures and small-group cooperative learning sessions that involve the use of activities that

generate real data) resulted in significantly better effects on students' achievement on probability than Instructional Method A (using lectures and small-group cooperative learning sessions during which students solved probability problems).

Specific to the multiple-choice items *gain scores*, the results of the Wilcoxon Signed-Ranks tests indicated that i) students in the control group performed significantly lower on the post-test multiple-choice items compared to the corresponding pre-test items; and ii) students in treatment group did not perform significantly different on the post-test multiple-choice items compared to the corresponding pre-test items. Further analysis indicated that the *normalized gain scores* of the treatment group were significantly higher than the normalized gain scores of the control group meaning that Instructional Method B had a significantly better effect on students' achievement in probability than Instructional Method A.

Results relative to the multiple-choice items indicated that for a majority of the multiple-choice items, students in the control group had a lower performance on the post-test compared to their performance on the pre-test. Moreover, all of the percentage descreases were higher compared to the percentage increases in performance. This was in agreement with the Wilcoxon test results which indicated that the post-test multiple-choice scores of students in the control group were significantly lower than their multiple-choice scores on the pre-test. In particular, the most substantial *descreases* in percent-correct responses were with regards to Items 4, 11 and 13. A discussion of the possible explanations for these considerable percent-correct descreases is provided in Chapter 7. In the case of the treatment group, students had a percent-correct increase on six multiple-choice items and a percent correct decrease on five multiple-choice items. Moreover, specific to students in the treatment group, most of the percentage increases were higher than the percentage decreases in performance.

In addition to the quantitative analysis carried out to measure students' achievement on probability, a distractor analysis was carried out to measure students' understanding of probability. Initially, student responses revealed the difficulty level of each multiple-choice item on the pre-test and post-test. Then, the mean percentage of students who applied heuristics and misconceptions defined in the literature was computed.

In the case of the control group: i) the mean percentage of students who applied the *outcome approach* remained stable at 5.9%; ii) the mean percentage of students who used the *absolute size* instead of relative size when computing probabilities remained stable at a high 20.6%; iii) there was a slight increase in the mean percentage of students who applied the *equiprobability bias* which remained high at 28.4% and 29.4% on the pre-test and post-test respectively; iv) there was a small increase in the mean percentage of students who applied the *negative recency* (by 2.9%) and the *positive recency* (by 3.9%); and v) the biggest change was in the application of the *representativeness* heuristic (17.7% increase from the pre-test to the post-test).

Unlike the results of the control group, in the case of the treatment group, the mean percentage of students who applied the heuristics or misconceptions listed in Table 5.9 mostly decreased. In particular, the mean percentage of students who used the equiprobability bias, positive recency and outcome approach decreased slightly (by 3.4%, 4.9% and 1.85% respectively). The most substantial decrease occurred relative to the use of the negative recency i.e. 9.25%. Only in the case of the representativeness heuristic and use of absolute size instead of relative size when computing probabilities the mean percentage increased by 3.7% and 2% respectively.

A further presentation of results relative to the measurement of students' *understanding* of probability is provided in Chapter 6 which presents the results of the qualitative analysis of audio-taped conversations as students worked in small groups while completing probability activities or sets of probability problems.

CHAPTER 6

STUDENTS' UNDERSTANDING OF PROBABILITY

RESULTS OF QUALITATIVE ANALYSIS

The purpose of this chapter is to discuss the effects of the instructional treatment on students' *understanding* of probability through a *qualitative* analysis of results. Recall that Chapter 5 provided a presentation of the effects of the instructional treatment on students' *achievement and understanding* of probability through a *quantitative* analysis that included a comparison of pre-test and post-test scores as well as a distractor analysis.

This chapter begins by providing information relative to the activities and problem sets used in the treatment and control groups respectively. Following this is a discussion of the results of the qualitative analysis specific to students' understanding of basic probability concepts such as sample space, experimental probability of an event and theoretical probability of an event. Next, a section on students' understanding of conditional probability and independence is included. Then, the chapter presents students' understanding of discrete probability distributions and last, their understanding of the binomial distribution.

For the purposes of *qualitative* analysis, transcripts of audio-taped student conversations were coded using the levels of reasoning described in the Jones et al. (1999) framework that was presented in Chapter 2. Six students were selected and monitored, three from the treatment group and three from the control group, and their conversations were fully transcribed: one student of high mathematical ability, one student of moderate mathematical ability and one student of low mathematical ability from each group. The selection was carried out based on students' mathematics grade in the last year of high school as described in Chapter 4.

Recall that two types of codes were used when analyzing the transcript data: codes depending on the person speaking in the transcript excerpt (e.g. THA – Treatment/High Ability Student; CMA – Control/Moderate Ability Student) and codes based on the level of reasoning as presented in the framework (e.g. L1 – Level 1: Subjective Reasoning).

6.1 Information on Activities and Problem Sets

Table 6.1 provides information regarding the four activities completed by students in the treatment group and the four problem sets completed by students in the control group. Detailed descriptions of the activities and problem sets can be found in Chapter 4: Methods whereas the activities and problem sets are provided in Appendix C.

Note that in the cases of Problem Set 2 and Problem Set 4, students in the control group only worked on one problem in groups. These problem sets originally consisted of two problems on which students were asked to work in small groups. However, they ended up completing only one problem in each of these two problem sets since i) they took 40 minutes or more to complete one problem alone; ii) all groups required complete translation into Greek of the context provided in the problems and iii) all groups seemed to find the problems quite challenging. Once a copy of students' group work was collected, the second problem that had originally been assigned as part of group work was solved on the board using a whole-class discussion approach.

Table 6.1

Information On Activities And Problem Sets

	Activity 1 (Treatment)	Problem Set 1 (Control)
Average Duration	54 minutes	44 minutes
Concepts Covered	Sample space; Experimental probability of simple and joint events; Theoretical probability of simple and joint events; Probability comparisons; Law of Large Numbers.	Sample space; Simple events, joint events, complementary events; Probability of intersection of events; Probability of union of events. Problems 4.2, 4.8 and 4.9 (See Appendix C)
	Activity 2 (Treatment)	Problem Set 2 (Control)
Average Duration	57 minutes	40 minutes
Concepts Covered	Experimental and theoretical probability of simple events; Sample space; Conditional probability; Independence.	Conditional probability; Independence. Problem 4.23 (See Appendix C)
	Activity 3 (Treatment)	Problem Set 3 (Control)
Average Duration	26 minutes	38 minutes
Concepts Covered	Impossible events; Sample space; Probability distribution of a discrete random variable.	Expected value of a discrete random variable; Sample space; Probability distribution of a discrete random variable. Problems 5.3a and 5.4 (See Appendix C)
	Activity 4 (Treatment)	Problem Set 4 (Control)
Average Duration	51 minutes	43 minutes
Concepts Covered	Sample space; Experimental and Theoretical probability of joint events; Law of Large Numbers; Binomial Distribution.	Expected value and standard deviation of a discrete random variable that follows the binomial distribution; Binomial distribution. Problem 5.13

6.2 Inter-Rater Reliability

As mentioned in Chapter 4, transcripts were coded independently by the researcher of this study and by a second rater. A hard copy of each transcript was provided to the researcher and to the second coder who i) read the transcripts once to become acquainted with the data; ii) read the transcripts a second time and noted all of the excerpts which related to a particular construct under examination and which involved participation of one of the six students under study and iii) mapped each of these excerpts to the levels of reasoning on the framework. Once

the coding of transcripts was completed, they were checked by the researcher for agreement on the number of excerpts related to the constructs under study and the levels of reasoning associated with each excerpt. The researcher and second coder met once more to discuss and resolve any instances of disagreement in coding. The percent-agreement for each concept is presented in Table 6.2.

Note that Table 6.2 does not include the concepts of discrete probability distribution and binomial distribution. This is due to the fact that the framework does not accommodate for these concepts. So, students' level of reasoning as exhibited in excerpts associated with discrete probability distributions (in Activity 3/Problem Set 3) and the binomial distribution (in Activity 4 /Problem Set 4) could not be *directly* coded using the framework. However, in Activity 3/Problem Set 3 and in Activity 4/Problem Set 4 students needed to make connections to basic probability concepts; reasoning on such concepts is discussed in the framework. Excerpts in which students made connections to basic probability concepts are accounted for in Table 6.2 when computing the inter-rater reliability. Any excerpts relative to the concepts of discrete probability distributions and the binomial distribution in Activity 3/Problem Set 3 and in Activity 4/Problem Set 4 respectively in which students did not make connections to basic probability concepts are discussed in sections 6.4 and 6.5 of this chapter.

Table 6.2

Percent-Agreement Between Coders of Transcript Excerpts

Concept	Number of Excerpts Associated with Concept	Percent-Agreement in Coding of Reasoning Level
Sample Space	14	13/14 = 93%
Theoretical Probability	54	44 / 54 = 81.5%
Experimental Probability	26	22 / 26 = 84.6%
Conditional Probability	11	9 / 11 = 81.8%
Independence	8	100%

6.3 Effects of Instructional Treatment on Understanding of Basic Probability Concepts

In order to examine the effects of the instructional treatment on students' understanding of basic probability concepts, the first four constructs included in the Jones et al. (1999) framework i.e. sample space, experimental probability of an event, theoretical probability of an event, and probability comparisons were considered when analyzing transcripts of student conversations. The audio-taped conversations of the students monitored in the treatment and control groups as they worked in small groups on *all* activities and problem sets were qualitatively analyzed with respect to basic probability concepts.

6.3.1 Sample Space

Overall, the four activities used in the treatment group provided more opportunities for students to express their reasoning regarding the concept of sample space than the problems in the course textbook that students in the control group worked on. Particularly, in problem sets 1, 2, and 4, transcripts of student conversations revealed no instances during which the three students in the control group reasoned about sample space.

As shown in Table 6.3 the high ability student in the treatment group (THA) exhibited Level 4: Numerical Reasoning with regards to sample space across all four activities. The high ability student in the control group (CHA) expressed reasoning relative to sample space in Problem Set 3 only in which case that reasoning was at L4 as well. That is, both students were able to apply a generative strategy that enabled a complete listing of outcomes. For example, when students were asked at the beginning of Activity 1 which sum they believed is most probable when two dice were rolled the THA student responded as follows:

16 THA: So, we should write the pairs first. 1-1…

17 THA and OGM together: 1-2, 1-3, 1-4, 1-5, 1-6

18 THA: 2-1, 2-2, 2-3, 2-4, 2-5, 2-6, 3-1, 3-2, 3-3, 3-4, 3-5, 3-6, 4-1, 4-2, 4-3, 4-4,

19 4-5, 4-6, 5-1, 5-2, 5-3, 5-4, 5-5, 6-1, 6-2, 6-3, 6-4, 6-5, 6-6. 12 is out of the

20 question, same with 2 because they only show up once.

In addition, in Activity 2, when students were asked to determine the number of outcomes when three dice are rolled the THA student reasoned in the following manner:

80 OGM: What does it mean by how many outcomes?

81 THA: It could be 1-1-1; 1-2-1; got it?

82 OGM: So, how many are there?

83 THA: I don't know.

84 OGM: Infinitely many?

85 THA: It can't be infinitely many because a die has numbers on it up to 6.

86 OGM: Wait, I have my class handout from last time when we were looking at two dice.

87 THA: If we have two dice then the number of outcomes is 36.

88 OGM: Yes, it's 36.

89 THA: So, here it should be 6 x 6 x 6; 216.

Moreover, the THA student was able to construct the sample space for three free-throw attempts of a basketball player in Activity 4 (See Appendix C).

With regards to Activity 3 and Problem Set 3 in which students dealt with discrete probability distributions, both the THA student and the CHA student were able to identify that i) there are 36 possible outcomes when two dice are rolled; ii) alternating faces count as different outcomes (e.g. 1-2 and 2-1 are two different outcomes); and iii) the possible sums when two dice are rolled are 2 to 12. In particular, the CHA student reasoned as follows:

76 IR: So, how many outcomes give a sum of 2?

77 CHA: 1

78 IR: To get a sum of 3, how many outcomes give us that?

79 CHA: Three

80 IR: Which three?

81 CHA: 1-2, 2-1 and 3-0; oops there is no 0. So, only two. To get a sum of four

82 we have 2-2, 1-3, 3-1….

…..

89 CHA: For a sum of 5

90 OGM: 4-1, 1-4

91 CHA: 2-3

92 CHA: and 3-2

93 CHA: So, four.

94 OGM: For sum of six we have 1-5, 5-1

95 CHA: Yes

96 OGM: 3-3

97 CHA: Yes

98 CHA: 2-4, 4-2

99 CHA: Five. Oh I see how it goes. The next one should have six outcomes.

100 OGM: Then, 7, 8, 9, 10, 11

101 CHA: Wait, are you sure? Maybe we should write down the results to see if

101-102 that is the case. You know what we should do? We should add them up

102 to see if they are 36 altogether.

103 OGM: 1 plus 2, three; plus 3, six; plus 4, ten; plus 5, 15; plus 6, 21; … 66

104 CHA: Then it's not like that. We are ok up to here. Lets see how many there

104-105 are for a sum of 7. We have 1-6, 6-1, 5-2…

As shown in Table 6.3, with regards to sample space, the moderate ability student in the treatment group (TMA) exhibited L4 reasoning in Activity 1, L2 reasoning in Activity 2, and L4 reasoning once again in Activity 4. Particular to Activity 1, the TMA student was able to provide a complete listing of the outcomes that generate each sum of two dice. In Activity 2, the TMA student reasoned as follows in response to a question regarding the number of outcomes when three dice are rolled:

106 IR: …When you have one die, how many possible outcomes are there?

107 TMA and OGM: 6

108 IR: When we have two dice?

109 TMA: 12

110 OGM: 36!

111 OGM: So, in this case it is 6 to the power of 3.

112 TMA: 36 times 6; 216.

The above excerpt provides evidence that the TMA student correctly indicated the number of possible outcomes when one die is rolled, but only through the aid of another group member could this student point out the number of possible outcomes when two or three dice are rolled. In Activity 4 the TMA student "adopts and applies a generative strategy that enables a complete listing of the outcomes" (Jones et al., 1999, p. 15) for three free-throw attempts of a basketball player thus exhibiting L4 reasoning with regards to sample space:

11 TMA: How many outcomes are there? Well, he can make it, he can make it,

11-12 not make it. He is making three shots ok?

…

16 OGM: So, how many combinations there are?

17 TMA: How many possible outcomes.

18 OGM: He can make the basket three times; he can make it twice and lose once.

19 TMA: Yes, that's right. To make it twice and lose once, that's given, it's BBN.

20 Lets start off by writing down, to make it three times.

21 OGM: It might be BBN ooops BNN.

22 TMA: BBB. BBN is given…

23 OGM: NNN

24 TMA: BNB…NBB…NNN

25 OGM: He could make it once and lose twice.

26 OGM: Or lose twice and make it once.

27 TMA: BNN…NNB

28 OGM: NBN

29 TMA: That's right. …

The CMA student expressed the same sequence of reasoning (L4 → L2 → L4) relative to sample space in Problem Set 3 only. The CMA student was initially able to adopt a generative strategy to list the possible outcomes that relate to a particular sum when two dice are rolled but later included outcomes that were not part of this sample space (i.e. 1-7 and 7-1 as outcomes relating to a sum of 8 and 2-7 and 7-2 as outcomes relating to a sum of 9 when two dice are rolled) thus exhibiting L2: Transitional Reasoning. However, the CMA student soon realized that pairs which include numbers higher than 6 cannot be included as outcomes since a regular die

does not have a face value higher than 6 and continued to provide a complete listing of outcomes associated with each sum. In this respect, the CMA student transitioned from L2 to L4 reasoning with regards to sample space. The following excerpt demonstrates the aforementioned levels of reasoning exhibited by the CMA student. Prior to this excerpt, students correctly indicated the list of outcomes that resulted in sums of 2 to 7 when two dice are rolled, and were then involved in the process of finding the number of outcomes that result in sums of 8 to 10.

128 OGM: It goes 1, 2, 3, 4, 5, 6 and then it will be 7, 8, 9, 10 … you'll see. Eight.

1-7

129 CMA: 7-1, 4-4, 5-3, 3-5, 6-2, 2-6

130 OGM: There are more

131 CMA: That's it! Seven

…

136 OGM: For nine … 1-8, 8-1

137 OGM: It will be eight of them, you'll see.

138 CMA: 5-4, 4-5, 2-7, 7-2. Wait.

139 OGM: Let's look at them again. 1-8, 8-1

140 CMA: Does a die have an 8?

141 OGM: Ooops I forgot.

142 CMA: We made a mistake before as well.

143 OGM: Yeah, you are right, we made a mistake.

144 CMA: Let's start from 1 each time. 1-7, no. 2-6, ok.

145 OGM: 6-2

146 CMA: 3-5, 5-3, 4-4. That's it. Five

…

149 OGM: Ok, for nine.

150 CMA: Nothing with 1.

151 OGM: 6-3, 3-6

152 CMA: 4-5, 5-4

153 OGM: Now it will go in reverse order.

154 OGM: Yes

155 CMA: Four

156 OGM: For ten. 5-5 only.

157 CMA: What? No, there are more.

158 OGM: 6-4, 4-6, 7-3, 3-7

159 CMA: There is no 7 on the die! 3 over 36

160 OGM: For eleven?

161 CMA: 5-6 and 6-5. 2 over 36. And the last one is 1 over 36.

As evident from this excerpt, for a while the CMA student, along with the other group members, incorrectly considered face values that do not exist on a regular six-sided die but soon realized their mistake. The CMA student then reverted to listing correct outcomes for sums of 8 to 10.

As shown in Table 6.3 the TLA student exhibited L4 reasoning in Activities 2 and 4 with regards to sample space whereas the CLA student expressed L2 reasoning in Problem Set 3 only. In Activity 2, the TLA student correctly indicated that there are 6 possible outcomes when one die is rolled, 36 possible outcomes when two dice are rolled and "6 to the power of 3" (line 111) possible outcomes when three dice are rolled. In Activity 4, the TLA student was able to use a

generative strategy to provide a complete listing of the eight possible outcomes for three free-throw attempts of a basketball player:

10 TLA: So, we need to find all the possible outcomes.

11 OGM: And this one counts as one of the outcomes right?

12 TLA: Yes.

13 OGM: So, BBN and then

14 TLA and OGM: BNB

15 OGM: NBB

16 TLA: Wait a second, you confused me. BBN, BNB, NBB … BNN

17 OGM: BBB

18 TLA: NNN

19 OGM: Is that it?

20 TLA: Is there anything else? NNB

21 TLA is repeating out loud the outcomes they already listed.

22 TLA: NBN. Did we include it? No.

As previously mentioned, the CLA student exhibited L2 reasoning in Problem Set 3 with regards to sample space. This student was able to provide a partial set of outcomes associated with each sum of two dice; the list was completed by the other group members.

Table 6.3 indicates the levels of reasoning and transitions between levels of reasoning that the six students demonstrated relative to the concept of sample space. Overall, the activities provided students in the treatment group with more direct as well as indirect opportunities to reason about sample space compared to the problem sets used in the control group. In particular, as indicated on Table 6.3, in problem sets 1, 2 and 4 students in the control group did not reason

at all relative to sample space. In addition, the reasoning exhibited by students in the treatment group relative to sample space was more stable (i.e. it involved fewer transitions between levels of reasoning) compared to reasoning exhibited by students in the control group. Moreover, students mostly exhibited a high level of reasoning regarding the concept of sample space.

Table 6.3

Levels of Reasoning Exhibited by Students Relative to Sample Space

Treatment Group					Control Group		
Activity	THA	TMA	TLA	Problem Set	CHA	CMA	CLA
1	L4	L4		1			
2	L4	L2	L4	2			
3	L4			3	L4	L4 ↓ L2 ↓ L4	L2
4	L4	L4	L4	4			

6.3.2 Theoretical Probability of an Event

The framework by Jones et al. (1999) considers *theoretical probability* and *probability comparisons* as two separate constructs. Since i) as stated in the framework, the construct of probability comparisons involves making quantitative judgments (which are part of the construct of theoretical probability) and ii) in many cases that students were asked to compute the theoretical probability of an event in an activity or problem they were also asked to compare this to the probability of another event, theoretical probability and probability comparisons were considered together for the purposes of qualitative analysis in this study. Table 6.4 indicates the levels of reasoning demonstrated by the six students relative to theoretical probability and probability comparisons.

In Activity 1, the TMA and THA students that were monitored in the treatment group transitioned from L2 reasoning to L4 reasoning while the TLA student remained stable at L2 reasoning with regards to theoretical probability. The first question posed in Activity 1 asked students to state which sum they believed was most probable when two dice are rolled. Notice that in Activity 1, the TLA student and the TMA student were in the same group.

2 TLA: 7

3 TMA and OGM: Number 7

4 TMA: And the reason is? It comes up through 6-1

5 TMA and TLA: 5-2 and 4-3

6 TMA: Number 7 … write it down dear.

The TLA and TMA students correctly responded that the sum of 7 is the most probable sum when two dice are rolled. However, they did not consider all possible outcomes that result in a sum of 7 (i.e. they did not consider alternating faces of dice). On the other hand, the THA student correctly listed all possible outcomes when two dice are rolled (including alternating faces) but incorrectly responded that the sum of 6 was the most probable sum when two dice are rolled "because it is halfway through 2 and 12" (line 30). The reasoning of all three students at the beginning of Activity 1 indicates that they made comparisons based on quantitative judgments. However, their reasoning was not entirely correct, thus exhibiting L2 reasoning. The THA student then continued in the attempt to convince the group members that the sum of 6 was the most probable sum by pointing to the five outcomes that result in a sum of 6 and indicating that there are only four outcomes that result in a sum of 5. Once again, the THA student exhibited L2 reasoning since this student did not consider the outcomes associated with each possible sum but only focused on two of the possible sums. At a later stage, students were asked to compute the

theoretical probability of each sum from 2 to 12 when two dice are rolled using a pictorial representation of the sample space that was provided to them in the activity (see Appendix C). Both the TMA student and the THA student correctly computed these probabilities by considering the outcomes associated with each sum and dividing by 36, thus exhibiting L4 reasoning with regards to theoretical probability (lines 226-247 and lines 272-303 in the respective transcripts). Moreover, the TMA student added up the probabilities of the possible sums to make sure that they resulted in 100% (lines 248-249).

Note that in Activity 1, the TMA student and the TLA student were in the same group. Although the TLA student participated actively in the group work by i) providing very brief responses, ii) converting fractions into percentages for the purposes of reporting probabilities and iii) volunteering to construct one of the histograms, the oral responses provided by this student were not elaborate enough to clearly indicate the level of reasoning carried with respect to theoretical probability. A similar situation arose in the control group. The CHA student and the CLA student were in the same group while working on Problem Set 1. Although the CLA student participated actively in the group work by i) providing very brief responses and ii) asking questions, the oral responses provided by this student were not elaborate enough to clearly identify the reasoning level carried with respect to theoretical probability.

In the case of Problem Set 1, the three students that were monitored in the control group initially exhibited L1 reasoning with respect to theoretical probability. The CLA student remained stable at this level of reasoning, whereas the reasoning patterns followed by the CHA and CMA students during group work were similar (CHA: L1 → L4 → L2; CMA: L1 → L4 → L2 → L4). The first two problems in Problem Set 1 required students to provide examples of simple events, joint events and complementary events. All three students could identify

complementary events with ease. However, they had difficulties providing examples of simple and joint events.

20 IR: You only need to provide one event but that event should be a simple one.

21 What does simple event mean?

22 CHA: That it has one characteristic.

23 IR: What is one characteristic that the ball could have?

24 CLA: It could be either red or green.

25 OGM: A simple event is the color.

26 CLA: One red or one green.

27 CHA: Event red or event green.

28 CLA: So, I will write Event A …

29 CHA: Not like that.

30 CLA: Red

31 CHA: or green. In part b it's what OGM said before.

32 OGM: What did I say?

33 CHA: That if it's not red then it will be green.

34 CLA: Here it says red so the answer is green ball.

35 CHA: Green balls … no ball. Only one will be chosen.

36 CLA: Oh come on! Details!

In their written work this group of students specified that "Event red or green balls" is a simple event. On the other hand, the CMA student, who worked in a different group, could easily express simple events verbally and in written form (i.e. "The ball is red"). When it came to providing an example of a joint event the CHA student and the CMA student correctly identified

such an event both verbally and in written form. The CLA student had difficulty identifying such an event but, with the help of the CHA student, understood how to use the contingency table provided in the problem to identify two characteristics relative to the problem context which when taken together may form a joint event (lines 88-128). The framework by Jones et al. (1999) indicates that if a student "recognizes certain and impossible events" then he/she is exhibiting L1 reasoning with respect to theoretical probability. This could be extended to include other types of events such as simple, joint, and complementary events. In that case, the three students in the control group exhibited L1 reasoning at the beginning of Problem Set 1.

In the last problem in Problem Set 1 (i.e. problem 4.9; see Table 6.4), students were asked to compute the probability of simple and joint events.

Table 6.4

Problem 4.9 (Levine, Krehbiel and Berenson, 2010)

	Income Level		
U.S. Tax Code	Less than $50,000	More than $50,000	Total
Fair	225	180	405
Unfair	280	320	600
Total	505	500	1,005

Referring to the contingency table, if a respondent is selected at random, what is the probability that he or she
a. thinks the tax code is unfair?
b. thinks the tax code is unfair and makes more than $50,000?
c. thinks the tax code is unfair or makes more than $50,000?
d. Explain the difference in the results in (b) and (c).

The three students implemented L4 reasoning when working on finding the probability of a simple event by assigning a correct numerical probability to it. When students were asked to compute the probability of a joint event initially none of them could correctly carry out this computation or identify the correct frequencies on the contingency table that needed to be used.

102 CMA: In part b we need the probability that ‘thinks the tax code is unfair and

103 makes more than 50000’. So, 320 over 500.

104 IR: Why over 500?

105 OGM: Over 1005?

106 IR: Whatever you think is correct.

107 OGM: 320 over 500.

This excerpt provides evidence that the CMA student used quantitative reasoning to compute the probability of an event but this was carried out in an incorrect manner, thus exhibiting L2 reasoning. At the end of the group work, the same student realized that this computation must have been incorrect. This realization came about when the probability computed in part b needed to be used in part c. At that point, the CMA student indicated that “In part b it should be 320 over 1005 as well” (line 155). By assigning a correct probability to an event, the CMA student moved to L4 reasoning with regards to theoretical probability. Similarly, the CHA student and the CLA student were confused about which number should be used in the denominator of the computation in part b.

196 CHA: 320

197 IR: Good.

198 CHA: Out of 600?

199 IR: Why out of 600?

200 CHA: Or over 1005? Or over 500? I am confused! Come on guys, help a bit!

201 CHA: 320 … 320 over 1005

202 CLA: Why 1005? Do we take the total?

203 CHA: That is the total but on the other hand it says ‘more than’ so we might need

204 to divide by 500.

205 OGM: It must be over 500.

206 CHA: CLA who is going to write this?

207 CLA: You can be the recorder.

208 CHA: So, I should be doing everything.

209 CLA: I don't mind writing the answers but you should tell me what to write.

The written response that these students ended up providing was $\frac{600}{1005} - \frac{280}{1005} = \frac{320}{1005}$ which was correct. In their verbal communications students exhibited L2 reasoning; they were using quantitative judgments to compute the probability of joint events; however, they were not doing so in a completely correct manner. The CHA's statement that the answer should be "320 … 320 over 1005" (line 201) along with the fact that the written response provided was correct and the CLA student stated to the CHA student that "you should tell me what to write" implies that the CHA student moved to L4 reasoning with regards to theoretical probability in Problem Set 1. The same though cannot be concluded about the CLA student who seemed to remain 'confused' and depended entirely on the CHA student in providing the group's written response. Note that at the beginning of the session the IR specified that the CLA student would be the group's recorder.

Another type of computation that students were required to carry out involved finding the probability of the union of events (part c of problem 4.9). In this case the CMA student again moved from L2 to L4 reasoning. The CHA student though exhibited L2 reasoning throughout the computations involved in part c of the problem and there were no instances of conversation involving the CLA student during that part of the group work. In particular, the CHA student stated that "[I]n part c it says 'or'. So, I will consider first only 'unfair' so 600 out of 1005. And

then I will consider ‘more than’ which is 500 out of 1005” (lines 160-161). This corresponds to L2 reasoning since the CHA student used quantitative judgments to compute the probability; however, this computation was not carried out correctly since the CHA student did not subtract the intersection of the two events. At the end of the group work another statement made by the CHA student was that “[S]ince the denominators are the same we don’t need to write them right? I mean the over 1005. On top it should be 320 here” (lines 219-220). A copy of the student’s written work revealed that their response to part c was 600 + 320 – 280 = 640. Since the computation for the probability of the union of events was incorrect the CHA student’s level of reasoning remained at L2.

On the other hand, the CMA student started off by exhibiting L2 reasoning when computing the probability of the union of events. Help received by the other group members along with an examination of class notes helped the CMA student move to L4 reasoning with regards to this type of computation.

112 CMA: Unfair … 600. More than 50000 … 500. So we should consider the 600 and the 500.

113 OGM: Aren’t we supposed to use 320?

…

121 CMA: Let’s take 320 … then 280.

…

124 CMA: So part c is 780 over 1005? …

126 OGM: Wait guys. Here it is. It’s 600 over 1005. In the class example that she

127 showed us before it was red and black. In our case it’s fair and unfair right?

128 CMA: No, it’s ‘less than’ and ‘more than’.

129 OGM: What are you talking about?

130 CMA: Whatever

131 OGM: Write 320 over 500; the other one is 320 over 600. And then subtract …

132 CMA: No way! It's not out of 600; it's out of 1005.

133 OGM: Then we subtract 320 over 1005.

134 OGM: We should write 320 over 1005; then 600 over 1005.

135 CMA: No, we only need the 500 and the 600. The 320 involves both characteristics.

136 OGM: Yes, so we subtract 320 over 1005.

137 CMA: What are you doing? Why are you using 280?

138 OGM: Because it's about 'unfair'. We need to take all of those who said 'unfair'.

139 CMA: Yes, 600.

140 OGM: Ok wait, wait … we should write 600 over 1005

141 CMA: Plus 500 over 1005

At the beginning of Activity 3 all three students monitored in the treatment group could easily identify impossible events (i.e. sum of two dice being 1, 13, 14, or 15) thus exhibiting L1 reasoning. When asked which sums give the best chance of winning the game that students were playing in Activity 3 (i.e. which sums are most likely to occur when rolling two dice) the THA student indicated that "[N]umber 7. Which is six times. And we should write the pairs that give the number 7: 1-6, 6-1, 5-2, 2-5 … 3-4, 4-3" (lines 281-284). This indicates that the THA student "uses valid quantitative reasoning to explain comparisons" but did not assign numerical values to probabilities, thus exhibiting L3 reasoning with regards to theoretical probability. The THA student then asked the IR "on the table are we supposed to write the pairs? The results?". At that

time the IR responded that students should write down the probability associated with each sum. The THA student then went ahead to divide the number of outcomes associated with each sum by 36 thus assigning a correct probability to each sum (L4 reasoning).

In Activity 3 the TMA student and the TLA student moved across three levels of reasoning (L1 → L2 → L4). As previously mentioned, both students were able to identify impossible events at the beginning of the activity thus exhibiting L1 reasoning with regards to theoretical probability. Later, when asked to compute the probability of each sum when two dice are rolled, the TMA student correctly started enumerating the outcomes associated with each sum but used an incorrect denominator when assigning a probability to each sum (L2 reasoning). That is, instead of dividing by 36, the TMA student divided by 11 (the number of possible sums when two dice are rolled). With the aid of another group member, the TMA student soon realized that the denominator should be 36 thus assigning the correct probability to each sum and indicated that a sum of 7 gave the best chance of winning the game (L4 reasoning). The TLA student seemed to need more help that the TMA student did from other group members in enlisting the outcomes associated with each sum (lines 309-360). At first, the TLA student listed the number of possible outcomes associated with each sum instead of assigning a probability to each sum (L2 reasoning) but with the aid of another group member realized that each of these numbers should have been divided by 36. At the end of the activity, the TLA student correctly indicated that the sum of the probabilities of dice sums should be 36/36 and that a sum of 7 gives the best chance of winning the game (L4 reasoning).

In Problem Set 3 the three students monitored in the control group worked on two problems on discrete probability distributions (5.3a and 5.4; See Appendix C). Problem 5.4 required students to construct the probability distribution associated with each of three different

scenarios of a game. Students had to use basic probability concepts such as sample space and theoretical probability of the sum of two dice. Transcripts revealed no instances in which the CLA student reasoned about theoretical probability. The CHA student and the CMA student both moved from L2 reasoning to L4 reasoning with regards to theoretical probability.

65 IR: In the first method the player wins if the sum is less than 7.

66 OGM: So, 2 up to 6. So, the probability has to do with getting one of the numbers

67 from 2 to 6. That is, five numbers, So, five out of

68 CHA: 12. Or is it out of 6 since there are six numbers on a die?

69 IR: How many dice are involved here?

70 CHA: Two. So, there are 36 probabilities. So, here we have 5 out of 36?

71 IR: We have two dice. What outcome will give us a sum of 2? What needs to

72 show up on the dice?

73 CHA: 1-1

74 IR: Anything else?

75 CHA: No.

76 IR: So, how many outcomes give a sum of 2?

77 CHA: One

78 IR: How many outcomes give us a sum of 3?

79 CHA: Three

80 IR: Which three?

81 CHA: 1-2, 2-1 and 3-0; ooops there is no 0! So, only two. To get a sum of four we

82 have 2-2, 1-3, 3-1. So, we need to find all the possible results associated with sums 2 to 6.

As the above excerpt indicates, the CHA student was initially confused; the student attempted to use quantitative reasoning to deal with the situation at hand but the probabilities this student provided were incorrect (L2 reasoning). Once the IR started asking questions, the CHA student started to provide correct responses and realized that in order to be able to solve the problem, the group needed to find the probability associated with each sum that was of interest in the particular situation (i.e. sums 2 to 6). The CHA student then went ahead to correctly list the outcomes associated with each of these sums and correctly divided the number of outcomes by 36 to arrive at the probability of each sum (L4 reasoning; lines 71-133).

In Problem Set 3 the CMA student exhibited L2 reasoning at first. When asked about the probability of getting a sum of 2 when two dice are rolled, he stated "2 over 36 … because there are two dice" (lines 51 and 54). The IR reminded the CMA student that what is of interest is the *sum* of the dice and the CMA student responded "Oh, the sum! Then 1 over 36 because we can only get a sum of 2 if we get 1-1 on the dice" (line 56). This statement indicates L4 reasoning since the student assigned a correct numerical probability to an event. When asked about the probability of getting a sum of 3, the CMA student once again moved from L2 (line 67) to L4 reasoning (lines 74, 76, and 78).

66 IR: Why 2 over 36?

67 CMA: Well, we have 1-2 and … no, 1 over 36.

68 IR: Why 1 over 36?

69 OGM: There is no other outcome.

70 IR: What if the dice are of different colors?

71 OGM: What does that have to do with the numbers on the dice?

72 IR: What outcome did you mention before that gives a sum of 3?

73 OGM: If the first die shows a 1 and the second die a 2.

74 CMA: What about the reverse?

75 IR: What do you mean?

76 CMA: So, 2-1 as well.

77 OGM: Oh, I see.

78 CMA: So, the probability for a sum of 3 is 2 over 36.

At first, the CMA student did not consider alternating faces thus providing an incorrect probability for the sum of 3 when two dice are rolled (L2 reasoning). Later, the CMA student went on to consider these and provide the correct probabilities associated with each sum (L4 reasoning).

As evident in Table 6.5, students in both the treatment and control groups mostly moved from a lower to a higher level reasoning with regards to theoretical probability. In some cases, students in the treatment group performed a higher number of transitions between levels of reasoning compared to students in the control group and in other cases they performed a smaller number of transitions.

Table 6.5

Levels of Reasoning Exhibited by Students Relative to Theoretical Probability

	Treatment				Control		
Activity	THA	TMA	TLA	Problem Set	CHA	CMA	CLA
1	L2 ↓ L4	L2 ↓ L4	L2	1	L1 ↓ L4 ↓ L2	L1 ↓ L4 ↓ L2 ↓ L4	L1
2	L2 ↓ L4	L4	L4 ↓ L2	2	L4		
3	L1 ↓ L3 ↓ L4	L1 ↓ L2 ↓ L4	L1 ↓ L2 ↓ L4	3	L2 ↓ L4	L2 ↓ L4 ↓ L2 ↓ L4	
4	L4	L2	L4				

6.3.3 Experimental Probability of an Event

The activities used in the treatment group provided direct opportunities for students to express their reasoning regarding experimental probability. However, the problems in the course textbook on which students in the control group worked in groups, did not directly lend themselves to discussions regarding experimental probability. Transcripts of student conversations revealed no instances during which students in the control group reasoned about experimental probability.

In the treatment group, the three students whose conversations were qualitatively analyzed exhibited various levels of reasoning relative to experimental probability. These reasoning levels varied within an activity as well as between activities. As evident from Table

6.6, the THA student and the TMA student followed a similar pattern of reasoning in Activity 1, moving from a lower level reasoning (L1 and L2 respectively) to L4 reasoning. Activity 1 began by asking students to indicate the sum they *believe* is most likely to occur when two dice are rolled. In response to this first question, the THA student reasoned as follows:

26 THA: I believe the answer is 6.

27 OGM: No.

28 THA: Why?

29 OGM: One, two, three … three. Four …

30 THA: Because it is also halfway through 2 and 12

31 OGM starts rolling the dice.

32 OGM: I want to see it in action.

33 THA: What you are doing is a different thing. You simply need to say which

33-34 one shows up more times.

35 OGM: Ok then let's put 6.

36 THA: All of the pairs have the same chances of coming up. It is not that the

36-37 more times you throw them …

This excerpt shows that the THA student "indicates little or no awareness of any relationship between experimental and theoretical probabilities" (Jones et al., 1999, p. 15), thus pointing towards L1 reasoning. Later on in Activity 1 students were asked to determine how close their group's experimental probability for each dice sum was to the theoretical probability and then, whether their group's experimental probabilities or the whole class experimental probabilities were closer to the theoretical probability of each dice sum. The THA student indicated that their

group's experimental probabilities were "relatively close" to the theoretical probabilities (lines 309-312) and later reasoned that:

418 THA: Columns 4 and 5 are almost the same. These are closer than these …

418-419 compared to ours these are closer.

420 OGM: Yes, yes

421 THA: So, here where it asks which set of relative frequencies those from your

421-422 group or those from the whole class are closer to the theoretical we will put

423 THA and OGM: the whole class results

424 OGM: are most close

425 THA: are closer to the theoretical probabilities

This excerpt indicates that the THA student "recognizes that the experimental probability determined from a large sample of trials approximates the theoretical probability" (Jones et al., 1999, p. 15) thus exhibiting L4 reasoning with respect to experimental probability. In the same activity, the TMA student indicated that 7 is the most likely sum when two dice are rolled as a response to the first activity question. When experimental data came into conflict with this preconceived notion, the TMA student believed that the roller was to blame and requested that they switch roles:

47 TMA: You and the number 7 are not in good terms?

48 OGM: Yes. 5 … 3

…

56 OGM: 5 … 9

57 TMA: Dude what are you doing?

…

60 TMA: Give them to me to roll them.

61 OGM: You are not a roller. When it says you should be a roller you can roll

62 OGM: 7

63 TMA: 7

64 OGM: See? And you were getting stressed!

65 TMA: Goodness! It takes you soooo long to roll a 7!

…

200 TMA: We said at the beginning that sum of 7 would be coming up more, didn't

200-201 we? Now, based on our results, what patterns do we see?

201 What predictions do we have now?

202 OGM: 6 and 9

203 TMA: Yeah, 6 was lucky

It is evident that the TMA student reverts to subjective reasoning (i.e. believing that generating a particular outcome depends on *who* rolls the dice) when experimental data conflicts with preconceived notions about probability, thus demonstrating L2 reasoning with respect to experimental probability. Later, when asked to compare their group's experimental results to the theoretical probabilities, the TMA student indicates that these are close (lines 256-274). A comparison of the group experimental probabilities to the theoretical probabilities and whole class experimental probabilities to the theoretical probabilities, provided the means for the TMA student to realize that the experimental probabilities obtained from the whole class data were closer to the theoretical probabilities than the group results alone (lines 321-348) thus exhibiting L4 reasoning.

Note that in Activity 1, the TMA student and the TLA student were in the same group. Although the TLA student participated actively in the group work by i) providing very brief responses, ii) converting fractions into percentages for the purposes of reporting probabilities and iii) volunteering to construct one of the histograms, the oral responses provided by this student were not elaborate enough to clearly indicate the level of reasoning carried with respect to experimental probability.

In Activity 2 the three students in the treatment group initially demonstrated L4 reasoning with regards to experimental probability however, at a later stage in the activity, the THA student and the TMA student exhibited L2 reasoning while the TLA student demonstrated L1 reasoning. At the beginning of Activity 2 the three students were able to assign a numerical value for the experimental probability of an event using the data collected, thus demonstrating L4 reasoning. At a later stage in the activity, students were asked to compute the theoretical probability that the sum is 6 when three dice are rolled. The THA student as well as the TMA student initially used the experimental data to determine this probability (lines 180-186 and lines 92-94 respectively) indicating a belief "that any sample should be representative of the parent population" (Jones et al., 1999, p. 15) which corresponds to L2 reasoning. Both students though later realized that they should have determined these probabilities theoretically and not using the experimental data. They then proceeded to compute this probability correctly.

The three students in the treatment group exhibited low-level reasoning (THA: L1; TMA: L2; and TLA: L1) with regards to experimental probability in Activity 3. Once the game winner was determined, students were asked why they thought the particular group member won. The THA student responded that "it was out of luck" (line 157) indicating "little or no awareness of any relationship between experimental and theoretical probabilities" (Jones et al., 1999, p. 15)

which corresponds to L1: Subjective Reasoning. In a similar manner, the TLA student addressed the winner as "you lucky thing!" (line 187) thus, also exhibiting L1 reasoning regarding experimental probability. On the other hand, the TLA student indicated that "we should have placed chips on 7. That's the one that comes up most often" (line 58). However, the TLA student made this statement after only nine rolls of the two dice, three of which resulted in a sum of 7. This indicated that the TLA student "puts too much faith in small samples of experimental data when determining the most or likely event; believes that any sample should be representative of the parent population" (Jones et al., 1999, p. 15).

In Activity 4 the THA student and the TMA student followed alternating levels of reasoning (THA: L4 → L1 → L4; TMA: L4 → L2 → L4) while the TLA student went from L2 to L4 reasoning with regards to experimental probability. The THA student was initially able "to determine a numerical value for the experimental probability" using the data (line 171), thus demonstrating L4 reasoning. Students were later asked to compare their results with those of another group and to indicate

209 THA: Why do you think the probabilities differ? Because they are

209 probabilities!

210 IR: Why did you have differences?

211 OGM: Because the numbers are random.

212 THA: So, let's write that the random numbers are different from the random

212-213 numbers of the other group. Also, write down that our result was 80%

213 whereas the other group got 76%.

The response of the THA student that probabilities differ "[B]ecause they are probabilities!" "indicates little or no awareness of any relationship between experimental and theoretical

probability" thus exhibiting L1 reasoning (Jones et al., 1999, p. 15). However, when another group member (OGM) related group differences to the concept of randomness, the THA student agreed and went on to provide an explanation based on this concept. At a later stage the THA student was able to determine the experimental probability based on the whole class results (lines 246-273) thus exhibiting L4 reasoning. Similar to the THA student, the TMA student was initially able to provide a numerical value for an experimental probability thus demonstrating L4 reasoning. When asked to determine an experimental probability based on the whole class results, the TMA student was once more able to provide a numerical value for it thus exhibiting L4 reasoning:

196 TMA: Based on the results from all groups, what is the probability that the
196-197 player makes at least two baskets in three free-throw attempts? So, we
197-198 should find the mean. 40 plus 38 plus 37 plus 44 plus 36 plus 38
198 divided by 6. Miss can you come over?

…

205 IR: The total number of times that this experiment was performed by each
205 group was what?
206 OGN: 50
207 IR: Where did you include the 50?
208 OGM: What do you mean?
209 TMA: Divide by 50.
210 OGM: And after that, all divided by 6.
211 TMA: Divide by 50.
212 IR: What should be divided by 50?

213 TMA: 40 over 50; 38 over 50. And then the result divided by 6.

The low ability student in the treatment group (TLA), first demonstrated L3 reasoning with respect to experimental probability. In particular, when asked to determine the probability of making at least two baskets using the experimental data, the TLA student responded "one, two, three, four, … 44. So, we will write P of at least two baskets is 44" (line 142); that is, the student provided a frequency value when asked to provide a probability. At a later stage, the TLA student was able to determine a numerical value for the experimental probability based on the whole class results thus exhibiting L4 reasoning with regards to experimental probability:

172 OGM: So, do we add everything up and divide by 50?

173 TLA: Don't we need to add everything up, divide by 6 and then by 50?

174 OGM: Ok. Would you like to write this or shall I write this?

175 IR: What is the total number of outcomes for the whole class?

176 TLA: 300. Oh, so do we take this total and divide by 300?

177 IR: Yes.

178 TLA: It comes out to 233.

179 OGM: And divide that by 300.

180 TLA: Miss, 0.77.

181 OGM: So, 77%.

Table 6.6 indicates the levels of reasoning demonstrated by the three students in the treatment group relative to the construct of experimental probability. Note that the problem sets did not provide students in the control group with any direct or indirect opportunities to reason about experimental probability and so, since there was no data on these students, Table 6.6 does not include a presentation of levels of reasoning exhibited by them.

As can be seen in Table 6.6, the three students monitored in the treatment group moved from a lower to a higher level reasoning in Activity 1 and Activity 4; moved from a higher to a lower level reasoning in Activity 2; and remained stable at a low level reasoning in Activity 3.

Table 6.6

Levels of Reasoning Exhibited by Students Relative to Experimental Probability

	Treatment		
Activity	THA	TMA	TLA
1	L1 ↓ L4	L2 ↓ L4	
2	L4 ↓ L2	L4 ↓ L2	L4
3	L1	L2	L1
4	L4 ↓ L1 ↓ L4	L4	L3 ↓ L4

6.3.4 Summary

An examination of Tables 6.3, 6.5 and 6.6 regarding students' reasoning levels relative to sample space, theoretical probability, and experimental probability respectively, revealed the following: i) The activities provided students with more direct as well as indirect opportunities to reason about sample space and experimental probability compared to the problem sets; ii) The problem sets did not provide any direct or indirect opportunities for students in the control group to reason about experimental probability and so, there was no data on these students relative to this construct; iii) Students' reasoning levels were the most stable with regards to sample space in comparison to theoretical or experimental probability; iv) Students exhibited higher levels of reasoning with regards to sample space (mostly L4) compared to theoretical or experimental

probability across all of the four activities and problem sets; v) With regards to experimental probability, the three students monitored in the treatment group moved from a lower to a higher level reasoning in Activities 1 and 4; moved from a higher to a lower level reasoning in Activity 2; and remained stable at a low level reasoning in Activity 3; and vi) With regards to theoretical probability students in both the treatment and control groups mostly moved from a lower to a higher level reasoning.

6.4 Effects of Instructional Treatment on Understanding of Conditional Probability and Independence

In order to examine the effects of the instructional treatment on students' understanding of conditional probability and independence, the audio-taped conversations of the three students selected in the treatment group and the three students selected in the control group as they worked in small groups on Activity 2 and Problem Set 2 (see Appendix C) respectively, were qualitatively analyzed. The six students exhibited the levels of reasoning shown in Tables 6.8 and 6.9 with regards to conditional probability and independence respectively. These two concepts are represented by two constructs on the Jones et al. (1999) framework i.e. conditional probability and independence.

6.4.1 Conditional Probability

Specific to conditional probability, both Activity 2 as well as Problem Set 2 provided students with direct opportunities to reason about conditional probability since both of these instructional materials focused on this concept. Activity 2 required students to roll a set of three dice of different colors 50 times, record the outcome and sum of the dice on each roll in a table, and compute various probabilities. In this activity, the THA and TMA students both demonstrated L4 reasoning with regards to conditional probability. An examination of the

transcripts revealed no conversational instances in which the TLA student expressed reasoning relative to conditional probability. Once students performed the experiment, they were asked to use their results to i) compute the relative frequency of getting a sum of 6 when three dice are rolled and ii) if the sum is 6, to determine the relative frequency that one of the dice results in 3.

31 THA: How many times we got a sum of 6?

…

35 THA: Two

36 THA and OGM: Twice out of 50.

37 THA: Based on your results, if the sum is 6 what is the relative frequency that one

38 of the dice results in 3?

…

40 THA: From these two times that we put as an answer here, did we have a three in any of them?

41 OGM: Let's check.

42 THA: Here. So, once.

A copy of the group's written work revealed that students assigned a correct numerical answer (i.e. ½ = 50%) for the question that required a conditional probability. So, the THA student exhibited L4 reasoning relative to conditional probability. The TMA student also demonstrated L4 reasoning in the same situation.

32 TMA: 7 times? Based on your results, what is the relative frequency to get a sum

33 of 6 when you roll three dice?

34 TMA and TLA: 7 over 50.

35 TMA: Yes. Based on your results, if the sum is 6, what is the relative frequency

36 that one of the dice results in 3?

37 TMA: So, if we got a sum of 6, how many of those times did we get a 3?

…

39 OGM: If we had a 3, how many of those times did we have a sum of 6?

40 TMA: If one of the dice was 3 … wait! If the sum was 6…

…

43 TMA: One…

44 OGM and TMA: Two, three

45 TMA: Four … 4 out of 7.

A copy of this group's written work indicated that they provided a correct value for the relative frequency required (i.e. 4/7 = 57%). So, the TMA student demonstrated L4 reasoning with regards to conditional probability.

Later on in Activity 2, students were asked "to find the theoretical probability that, when three dice are rolled, one of the dice results in 3 given that the sum is 6" (see Appendix C) by responding to a set of questions. These questions acted as a structured set of steps that aimed to help students arrive at the required theoretical conditional probability. Both the THA student and the TMA student correctly identified that there are 216 possible outcomes when three dice are rolled (line 89 and line 112 in the respective transcripts) and correctly indicated that the probability to get a sum of 6 when three dice are rolled is 10/216 by considering the outcomes that result in a sum of 6 (lines 96-112, 173 and lines 82-91, 151 in the respective transcripts). Next, both students correctly indicated that the probability of getting a sum of 6 (i.e. 10/216) should be used in the denominator of the required conditional probability and that the numerator (i.e. probability that one of the dice results in 3 and the sum is 6) should be 6/216.

188 THA: …The probability to get a 3 with a sum of 6 is 6 out of 216. The probability

189 that the sum is 6…

190 OGM: One sixth

…

204 OGM: But we said it is 10.

205 THA: 10 over 216. Hmmm, so, 6 over 216

206 OGM: 0,03

207 OGM: And the probability for the sum to be 6

208 THA: We found that … 0,05

Similarly the TMA student reasoned as follows:

146 TMA: So, write here, one of the dice is 3. And there, sum is 6. Over sum is 6.

147 OGM: One of the dice is 3 … 3 to the power of 3.

148 TLA: 3 to the power of 3 is 27.

149 TMA: The probability to get a sum of 6 is this over here.

150 OGM: 27 out of 216. Isn't it?

151 TMA: No, 10 over 216.

…

161 IR: How many outcomes resulted in a sum of 6 and contained a 3?

162 OGM: 6

163 OGM: Six out of?

164 OGM: 10. So, the numerator is 6 over 10.

…

171 TMA: Hey, it should be 6 over 216 here, not over 10.

Since both the THA and TMA students were able to assign these numerical probabilities the level of reasoning of reasoning exhibited by these students with regards to conditional probability was L4.

In Problem Set 2 the three students monitored in the control group worked on problem 4.23 from the course textbook which made use of the frequency contingency table provided in Table 6.7 and asked for the following conditional probabilities:

Table 6.7

Problem 4.23 (Levine, Krehbiel, and Berenson, 2010)

	Income Level		
U.S. Tax Code	Less than $50,000	More than $50,000	Total
Fair	225	180	405
Unfair	280	320	600
Total	505	500	1,005

a. Given that a respondent earns less than $50,000, what is the probability that he or she said that the tax code is fair?
b. Given that a respondent earns more than $50,000, what is the probability that he or she said that the tax code is fair?

All three students monitored in the control group ended up providing correct numerical probabilities for the two conditional probabilities required in problem 4.23. However, initially they all exhibited L2 reasoning since they all provided an incorrect numerical response to part a. In particular, the CHA and the CMA students ignored the conditioning event and stated that the answer to part a was 225/1005 (CHA: lines 7 and 44-46; CMA: line 10) while the CLA student claimed that the answer to part a was 225/405 (line 34). Following this initial statements though, all students went on to reason at L4 and ended up assigning correct numerical probabilities for both parts a and b of problem 4.23.

59 CHA: The question says to look at those who earn less than 50000 and then to

look for the tax code being fair.

OGM: What did we put after the vertical line?

CMA: Less than 50000.

CHA: So, being fair will be placed in the denominator.

CMA: Less than 50000 goes in the denominator and the other one on top. Both of these together go on the numerator.

…

CMA: Less than 50000 is 505 over 1005. This goes in the denominator.

CHA: Ok. And on top? 225 over 1005.

The CLA student reasoned as follows in the first part of problem 4.23.

CLA: So, 225 over 405

OGM: What did you do there?

CLA: I divided by the total. What we did with ‘give that’, shouldn’t it provide some other piece of information about something that happened first?

OGM: Wait.

CLA: Here there are those who believe it is fair and those who think it is unfair.

OGM: No, there are those who earn less than 50000 or more than 50000.

IR: How many people are there who earn less than 50000?

CLA: 225

OGM: 505

CLA: Oooh, 505

IR: Out of those, how many think the tax code is fair?

CLA: So, 225 over 505.

Lines 36-37 indicate that although the CLA student recognized that the probability of events change when we deal with conditional probability, this "recognition is incomplete and is … restricted to events that have previously occurred" (Jones et al., 1999, p. 15). In this manner, the CLA student demonstrated L2 reasoning with respect to conditional probability. However, through communication with other members of the group and the IR, the CLA student ended up assigning a correct numerical probability to part a of the problem, and later to part b as well.

Table 6.8 provides a summary of the reasoning levels exhibited by the six students monitored in the treatment and control groups relative to conditional probability.

Table 6.8

Levels of Reasoning Exhibited by Students Relative to Conditional Probability

	Treatment				Control		
Activity	THA	TMA	TLA	Problem Set	CHA	CMA	CLA
2	L4	L4		2	L2 ↓ L4	L2 ↓ L4	L2 ↓ L4

6.4.2 Independence

Activity 2 and Problem Set 2 provided students with direct opportunities to reason about independence. Activity 2 required students to determine whether getting a sum of 6 and one of the dice resulting in 3 when three dice are rolled were independent. Students were asked to indicate this twice during the activity; first following the performance of an experiment with three dice and later, after computing the theoretical probability of rolling a 3 on one of the three dice given that the sum was 6. Following the completion of the experiment, the THA student said that

57-58 THA: If one of them is 3 then the other two dice must be 1-2 or vice versa in

order to get a sum of 6.

59 IR: Ok. If the answers to 2a and 2b were the same what would that mean?

60 THA: That they don't affect each other. So, in this case, they affect each other.

This excerpt indicates that the THA student exhibited L3 reasoning with regards to the concept of independence since the student could "differentiate independent and dependent events" but at that time had not reached L4 reasoning which involved using numerical probabilities to distinguish such events. In contrast, the TMA student stated that since the answers to 2a (i.e. Based on your results, what is the relative frequency that you will get a sum of 6 when you roll three dice) and 2b (i.e. Based on your results, if the sum is 6, what is the relative frequency that one of the dice results in 3?) were not the same "they would not be affected. I mean they are not the same thing", implying that the events were independent (lines 48-49). The reasoning provided by the TMA student indicated that this student could not differentiate independent and dependent events and pointed towards L1 reasoning.

At the end of Activity 2, students were once more asked to determine whether getting a sum of 6 and one of the dice resulting in 3 were independent. The THA student was able to use numerical probabilities to show that the two events were dependent. Although one of the probabilities computed was incorrect (i.e. probability that one of the dice results in 3), the steps followed in identifying whether the events were independent as well as the remaining computations, and conclusion were correct. Therefore, the THA student ended up exhibiting L4 reasoning relative to the concept of independence. In contrast, the TMA student, along with the other members of his group, found this task to be too challenging and gave up without asking for help from the IR. So, the TMA student remained at L1 reasoning with respect to the concept of independence.

In problem 4.23 (Table 6.7) the three students monitored in the control group were asked to indicate whether income level was independent of attitude about whether the tax code was fair. The three students checked whether P(A|B) = P(A) in order to determine if the events were independent. All three students were able to use numerical probabilities correctly to conclude that the events were dependent thus exhibiting L4 reasoning with regards to the concept of independence.

Table 6.9 provides a summary of the reasoning levels exhibited by the six students monitored in the two groups with regards to the concept of independence.

Table 6.9

Levels of Reasoning Exhibited by Students Relative to Independence

	Treatment				Control		
Activity	THA	TMA	TLA	Problem Set	CHA	CMA	CLA
2	L3 ↓ L4	L1		2	L4	L4	L4

6.4.3 Summary

Both Activity 2 and Problem Set 2 provided students with direct opportunities to reason about conditional probability and independence. Moreover, both of these instructional materials afforded opportunities in which students needed to draw on previous knowledge of basic probability concepts (sample space, theoretical probability of simple and joint events, experimental probability) in order to compute conditional probabilities and decide whether events were independent.

Overall, i) students' reasoning with regards to conditional probability and independence was stable; ii) students exhibited high levels of reasoning with regards to both concepts; iii) in

the case of conditional probability the TLA and TMA students demonstrated L4 reasoning whereas in the case of independence, all three students monitored in the control group exhibited L4 reasoning; and iv) with regards to conditional probability all three students monitored in the control group moved from L2 to L4 reasoning.

6.5 Effects of Instructional Treatment on Understanding of Discrete Probability Distributions

In order to examine the effects of the instructional treatment on students' understanding of discrete probability distributions, the audio-taped conversations of the six students selected in the treatment and control groups as they worked in small groups, on Activity 3 (see Appendix C) and Problem Set 3 (see Appendix C) respectively, were examined. Note that the framework by Jones et al. (1999) did not specifically include discrete probability distributions as a construct. However, in both of these sets of instructional materials students needed to make connections to basic probability concepts; reasoning on such concepts was accounted for in the framework.

Activity 3 involved playing a game in which students rolled a pair of dice and each time removed a chip they might have placed over the number corresponding to the sum of the two dice. The winner was the person who had all of his/her chips removed first. Therefore, when distributing their chips over the number line on which the numbers 1 through 15 were written and with the aim of winning in mind, it would have been ideal if students made connections to the theoretical probability associated with each sum when two dice are rolled i.e. the probability distribution of rolling two dice.

In Activity 3, the term 'discrete probability distribution' was only mentioned in the title and in the last part of the activity. Throughout the rest of the activity students dealt with other basic probability concepts when responding to questions posed and at times, they dealt only

indirectly with discrete probability distributions. For example, once students placed the 12 chips given to each of them on the number line, they were asked to indicate how many chips they had placed above each number and provide an explanation for the way they had placed their chips. This aimed to provide students with an opportunity to indirectly deal with discrete probability distributions. Table 6.10 indicates how students distributed their chips on the number line along with the explanation provided by each as to the way they distributed their chips. This information was available on a copy of students' written work. Notice that, in their written explanations, only the TMA student made connections to the theoretical probability associated with each sum when deciding how to place the dice on the number line. The TMA student predicted the most likely events using informal quantitative judgments (i.e. did not assign numerical probabilities to each sum) however, he reverted to subjective judgments once the group started rolling the dice by expressing a belief that the outcome depends on *who* rolls the dice. Thus, he demonstrated L2 reasoning with regards to theoretical probability. The TLA student also used subjective reasoning (L2) with regards to theoretical probability when placing chips on the number line (i.e. the student placed chips on even numbers because "I like even numbers").

Table 6.10

Activity 3: Distribution of Chips By Students and Explanation

Number on the Number Line (Dice Sum)	THA	TMA	TLA
1			
2		1	3
3			
4		1	1
5	2		
6	3	2	1
7		2	
8	3	2	4
9	1	2	
10	1	1	1
11	1		
12	1	1	2
13			
14			
15			
Explanation	"I placed my chips randomly."	"More chips towards the center (number 6-9). Numbers have higher probabilities."	"I placed my chips on even numbers because I like even numbers."

At the end of Activity 3 students in the treatment group dealt directly with the probability distribution for the sum of two dice. Students were provided with a table on which the first column was a list of the possible sums of rolling two dice and were required to fill in the second column by providing the numerical probability corresponding to each sum. Next, they were asked to sum up the probabilities with the aim of helping them realize that the probabilities should sum up to 1. All three students monitored in the treatment group were able to compute the theoretical probability associated with each sum by considering the number of outcomes that give rise to each sum and dividing this number by 36. Moreover, all groups indicated that the sum of the probabilities should be 36/36 = 1. In the case of the THA and TLA students, initially

the probabilities provided by their corresponding groups did not sum up to 1. These students received this as an indication that they must have made a mistake on the table and went back to check and correct the probabilities provided for the sums of two dice. In this manner, all three students exhibited L4 reasoning with regards to theoretical probability (and indirectly, with regards to discrete probability distributions).

With regards to Problem Set 3, all three students monitored in the control group were able to compute the expected value of a discrete random variable with ease when a table listing the possible outcomes and associated probabilities was provided to them (problem 5.3; see Appendix C). However, all three students found problem 5.4 quite challenging. In this problem, students were required to construct the probability distribution for a discrete random variable themselves. The problem involved a game scenario on which three types of bets could be placed. All three students realized that they should create a table on which one column should list the outcomes associated with the bet being placed and a second column should indicate the probability associated with each outcome. The CHA student said that

38 CHA: I think we are supposed to create a table here with two columns and carry
39 out the multiplications like we did in the previous problem.

while the CMA student indicated that

34 CMA: Ok so we have three methods of playing the game. In the first one we win
35 if 2 through 6 come up. We need to find the probability associated with
36 each game method. We need to have a table like this one.

and the CLA student realized that

28 CLA: We will have to create the table ourselves here.
…

33 CLA: Well, there should be ‘under’ and ‘over’ on the table. The player wins or
34 loses. There is a bet. So, we should have ‘under 7’ so let’s write the values
below 7 and ‘over 7’.

Following these statements, both the CHA and CMA students led their groups in constructing the probability distribution for the sum of two dice by listing the outcomes associated with each sum and dividing the number of outcomes by 36. In this manner they demonstrated L4 reasoning with respect to theoretical probability. Moreover, they labeled outcomes as ‘win’ or ‘lose’ and correctly computed the probability of winning or losing under each of the three game scenarios. On the other hand, although the CLA student realized that a table should be constructed which should accommodate for the various outcomes of interest, the student then stated that “I don’t understand what the problem is asking for” (line 63). The other members of the group stated that they should list the outcomes that give rise to each sum; the CLA student participated in this process however, it was evident that the student called out these outcomes without understanding why this process had to be carried out (line 89: “I don’t understand this at all”). The group ended up providing only the frequencies associated with each sum and did not go on to convert these into probabilities.

6.6 Effects of Instructional Treatment on Understanding of Binomial Distribution

In order to examine the effects of the instructional treatment on students’ understanding of the binomial distribution, the audio-taped conversations of the six students monitored in the treatment and control groups as they worked in small groups, on Activity 4 (see Appendix C) and Problem Set 4 (see Appendix C) respectively, were examined. Note that the framework by Jones et al. (1999) did not specifically include binomial distribution as a construct. However, in

both of these instructional materials students needed to make connections to basic probability concepts; reasoning on such concepts is accounted for in the framework.

In Activity 4 students were asked to compute the probability of making at least two baskets in three free-throw attempts, in the case of a basketball player who has a constant probability of making the basket P(B) of 70% and a constant probability of not making the basket, P(N), of 30%. Students were asked to compute this probability in three different ways: i) Construct the sample space for three free-throw attempts, compute the probability of each outcome in the sample space, and then use these individual probabilities and the addition law of probability to compute the probability of making at least two baskets; ii) use a table of random numbers to indicate data on 50 sets of three free-throw attempts and use these data to compute the experimental probability of making at least two baskets; and iii) use the formula for the binomial distribution to compute the probability of making at least two baskets.

With regards to method i) described above, the three students monitored in the treatment group were able to correctly construct the sample space for three free-throw attempts thus exhibiting L4 reasoning with regards to sample space. Next, the THA and the TLA students demonstrated L4 reasoning with regards to theoretical probability (and subsequently with regards to the binomial distribution) since they were able to assign correct numerical probabilities to each outcome in the sample space and then, correctly computed the probability of making at least two baskets. The TMA student demonstrated L2 reasoning with respect to theoretical probability since the student used quantitative judgments to compute the probability of each outcome but did so incorrectly:

56 TMA: Do we do 70 plus 70 plus 30 divided by 3? Times 100?

Relative to method ii) the THA and the TMA students were able to use the table of random numbers to calculate the experimental probability of making at least two baskets thus exhibiting L4 reasoning with regards to experimental probability (and subsequently with regards to the binomial distribution). The TLA student was also able to use the table of random numbers but instead of assigning a numerical probability the student provided a frequency value for the experimental probability of making at least two baskets; thus, the TLA student demonstrated L3 reasoning with respect to experimental probability (and subsequently with respect to the binomial distribution). In the case of method iii) in which students were asked to use the general formula for computing a probability from the binomial distribution, a copy of each group's written work indicated that students were able to correctly compute the required probabilities from the binomial distribution (Question 6, Activity 4; See Appendix C). However, transcripts of student conversations revealed that all three students monitored in the treatment group had difficulties applying the general formula in the context of the activity.

In Problem Set 4 the three students monitored in the control group dealt directly with the binomial distribution by working on problem 5.13 in the course textbook. All three students were able to define n (the number of observations in the sample space), X (the number of events of interest), and p (the probability of an event of interest) as well as to compute the mean and standard deviation of the binomial distribution with ease. Moreover, all three students were able to apply the general formula for computing the probability from the binomial distribution and assigned correct numerical probabilities as responses to parts b and c of problem 5.13.

6.7 How Results Address The Research Questions

Recall that the research questions posed in this study aimed to examine the effects of two instructional methods on college students' achievement and understanding of probability.

Instructional Method A was used in the control group and involved the use of lectures and small-group cooperative learning sessions during which students solved probability problems. Instructional Method B was used in the treatment group and involved lectures and small-group cooperative learning sessions during which students used activities involving probability experiments that generate real data to make connections between experimental and theoretical probability.

The tables constructed in this chapter demonstrate the levels of reasoning that the six students monitored in the control and treatment groups exhibited relative to the concepts of sample space (Table 6.3), theoretical probability (Table 6.5), experimental probability (Table 6.6), conditional probability (Table 6.8) and independence (Table 6.9). The aim of constructing these tables was to be able to compare the levels of reasoning exhibited by students in the control and treatment groups in the attempt to identify which of the two instructional methods had a better effect on students' understanding of probability.

As evident from Tables 6.3 and 6.6, the activities used in the treatment group provided students with more direct as well as indirect opportunities to reason about sample space and experimental probability compared to the problem sets used in the control group. In particular, the problem sets did not provide any direct or indirect opportunities for students in the control group to reason about experimental probability. Given these, it is not surprising that, as specified in Chapter 5, students in the treatment group performed better on the multiple-choice items relating to the Law of Large Numbers (Item 6 on the pre-test and post-test) and the concept of sample space (Items 7 and 11 on the pre-test and post-test) than students in the control group. Specific to Item 6 the percent-correct responses decreased by 11.7% in the control group and by 3.7% in the treatment group from pre-test to post-test. On Item 7 the percent-correct responses

increased by 5.8% in the case of the control group and by 18.6% in the treatment group from pre-test to post-test. Last, on item 11, the percent-correct responses decreased by a substantial 58.8% in the case of the control group and increased by 3.7% in the treatment group. Considering these, Instructional Method B (treatment) produced better results than Instructional Method A relative to experimental probability and the concept of sample space.

An examination of Tables 6.5, 6.8 and 6.9 which demonstrate the levels of reasoning exhibited by students relative to theoretical probability, conditional probability and independence respectively, indicate that students in both the control and treatment groups were provided with opportunities to reason about these concepts. As evident from Table 6.5, i) students in both the control and treatment groups mostly transitioned from a lower to a higher level reasoning with regards to theoretical probability; ii) in some cases students in the control or treatment group exhibited a stable level of reasoning; and iii) students in the control group performed 0-3 transitions between levels of reasoning while students in the treatment group performed 0-2 transitions between levels of reasoning relative to theoretical probability. The results presented in Tables 6.8 and 6.9 reveal that students in the control and treatment groups either remained at L4 level of reasoning or performed one transition from a lower to a higher level reasoning regarding conditional probability and independence. Based on these results, it is difficult to identify whether one group performed better than the other relative to these concepts and so, which instructional method had a better effect on students' understanding of theoretical probability, conditional probability and independence.

6.8 Summary

This chapter presents the results of the qualitative analysis of the data collected through audio-taped student conversations of the six students monitored in the treatment and control

groups. Students' reasoning relative to basic probability concepts (sample space, experimental probability of an event and theoretical probability of an event), as well as relative to conditional probability and independence was evaluated using the Jones et al. (1999) framework in a direct manner since the framework accounted for these concepts. However, the framework does not accommodate for the concepts of discrete probability distributions and binomial distribution which were also examined in this study. So, students' level of reasoning exhibited in excerpts associated with discrete probability distributions (in Activity 3/Problem Set 3) and the binomial distribution (in Activity 4 /Problem Set 4) could not be *directly* coded using the framework. In Activity 3/Problem Set 3 and in Activity 4/Problem Set 4 students needed to make connections to basic probability concepts; reasoning on such concepts is discussed in the framework. Excerpts in which students made connections to basic probability concepts were analyzed using the Jones et al. (1999) framework. Any excerpts relative to the concepts of discrete probability distributions and the binomial distribution in Activity 3/Problem Set 3 and in Activity 4/Problem Set 4 respectively, in which students did not make connections to basic probability concepts were discussed in separate sections (sections 6.4 and 6.5 of this chapter).

With regards to students' reasoning levels relative to basic probability concepts, these were the most stable (i.e. involved fewer transitions between levels of reasoning) in the case of the concept of sample space compared to theoretical or experimental probability. Students exhibited higher levels of reasoning with regards to sample space (mostly L4) compared to theoretical or experimental probability across all of the four activities and problem sets. With regards to experimental probability, the three students monitored in the treatment group moved from a lower to a higher level reasoning in Activities 1 and 4; moved from a higher to a lower level reasoning in Activity 2; and remained stable at a low level reasoning in Activity 3. With regards

to theoretical probability students in both the treatment and control groups mostly moved from a lower to a higher level reasoning. In some cases, students in the treatment group performed a higher number of transitions between levels of reasoning relative to theoretical probability compared to students in the control group and in other cases they performed a smaller number of transitions.

Relative to conditional probability and independence, students' reasoning was stable and students exhibited high levels of reasoning with regards to both concepts. In the case of conditional probability the TLA and TMA students demonstrated L4 reasoning whereas in the case of independence, all three students monitored in the control group exhibited L4 reasoning. With regards to conditional probability all three students monitored in the control group moved from L2 to L4 reasoning.

In the case of discrete probability distributions, the three students monitored in the treatment group were able to make connections to theoretical probability when computing the probability distribution of a discrete random variable. These students were able to compute the theoretical probability associated with the each sum when rolling two dice. Overall, the three students in the treatment group exhibited L4 reasoning with regards to theoretical probability (and indirectly, with regards to discrete probability distributions). In the control group, all three students were able to compute the expected value of a discrete random variable with ease when a table listing the possible outcomes and associated probabilities was provided to them. However, when asked to construct the probability distribution for a random variable themselves, all three students in the control group found such a task to be challenging.

The last concept examined in this study was that of binomial distribution. The three students in the treatment group were able to construct the sample space for a random variable

that follows the binomial distribution thus exhibiting L4 reasoning with regards to sample space. The THA and the TLA students demonstrated L4 reasoning with regards to theoretical probability (and subsequently with regards to the binomial distribution) since they were able to assign correct numerical probabilities to each outcome in the sample space and then, correctly computed the probability of an event. The TMA student demonstrated L2 reasoning with respect to theoretical probability since the student used quantitative judgments to compute the probability of each outcome but did so incorrectly. In addition, the THA and the TMA students were able to use a table of random numbers to calculate the experimental probability of an event thus exhibiting L4 reasoning with regards to experimental probability (and subsequently with regards to the binomial distribution). The TLA student was also able to use the table of random numbers but instead of assigning a numerical probability the student provided a frequency value for the experimental probability of the event; thus, the TLA student demonstrated L3 reasoning with respect to experimental probability (and subsequently with respect to the binomial distribution). When it came to using the general formula for computing a probability from the binomial distribution, a copy of each group's written work indicated that students were able to correctly compute the required probabilities from the binomial distribution. However, transcripts of student conversations revealed that all three students monitored in the treatment group had difficulties applying the general formula in the given context. The three students monitored in the control group were able to define n (the number of observations in the sample space), X (the number of events of interest), and p (the probability of an event of interest) as well as to compute the mean and standard deviation of the binomial distribution with ease. Moreover, all three students were able to apply the general formula for computing the probability from the binomial distribution and assigned correct numerical probabilities as responses to textbook problems.

Chapter 7 provides a discussion of the quantitative and qualitative results presented in Chapters 5 and 6 respectively, including possible explanations regarding students' performance. In addition, Chapter 7 identifies the conclusions drawn from this study along with implications for further research.

CHAPTER 7

SUMMARY, DISCUSSION AND RECOMMENDATIONS

This study was designed to examine the role of tasks on college students' achievement and understanding of probability. This chapter is organized into five sections: i) summary of the study; ii) summary of the findings; iii) discussion of findings; iv) strengths and limitations; and v) implications and recommendations for further research.

7.1 Summary of the Study

Since the late 1950s there has been a strong call for an increase in the inclusion of probability in mathematics curricula in the USA as well as in Europe (Exarchakos, 1988; Lordou-Kaspari, 2003; Jones, 1970; MAA, 1998; NCEE, 1983; NCSM, 1977; NCTM, 2000). Over the past couple of decades, there have been various reform initiatives concerning the content and means of instruction in mathematics classrooms at the college level with recommendations set forth that lectures be replaced by more active learning methods in which group work is used (MAA, 1998). Specific to the area of statistics and probability, many researchers have recommended that there be a change in the way statistics courses are taught (Chance, 1997; Garfield, 1994; Shaughnessy, 1981; see also Keeler & Steinhorst, 2001). "One area that has received less focus in this literature is the teaching of probability" (Keeler & Steinhorst, 2001, Retrieved April 24, 2009 from http://www.amstat.org/publications/jse/v9n3/keeler.html). Much has been written about people's misconceptions and use of heuristics regarding judgment under uncertainty but there has been a lack of research on solutions to this phenomenon (Keeler & Steinhorst, 2001).

7.1.1 Purpose

With the above issues under consideration, this study examined the role of particular tasks on college students' achievement and understanding of probability. The research questions addressed in this study involved the use of two instructional methods:

Instructional Method A: Using lectures and small-group cooperative learning sessions during which students solve probability problems and

Instructional Method B: Using lectures and small-group cooperative learning sessions during which students use activities involving probability experiments that generate real data to make connections between experimental and theoretical probability.

Given Instructional Method A and Instructional Method B, the research questions addressed were:

3) What are the effects of using each of these instructional methods on college students' achievement on probability and on their understanding of experimental and theoretical probability?
4) Does Instructional Method B have a better effect on college students' achievement on probability and on their understanding of experimental and theoretical probability than Instructional Method A?

7.1.2 Methods

A mixed methods design was used to address the research questions for this study. The design included treatment and control groups, each comprised of students in three sections of an introductory statistics course taught by the researcher in spring 2010 at a college in Cyprus. Formal instruction on probability occurred during the second half of the semester and lasted for seven weeks. A pre-test comprising of 14 multiple-choice items and a post-test comprising of 15

multiple-choice items and 2 open-ended items were utilized. During the study, students in the treatment group worked in small groups on four in-class activities about experimental and theoretical probability, and students in the control group worked in small groups on solutions to four sets of probability problems. Three students in the treatment group and three students in the control group were monitored by having their conversations audio-recorded as they worked in groups.

Participants included 44 students across the three sections of the introductory statistics course. Students were not randomly assigned to these sections but instead were placed in them by the course coordinator in consultation with the academic board at the college based on students' English language proficiency. Two of the sections were combined to form the treatment group which was taught using Instructional Method B; one section included students of moderate/high English language proficiency while the other included students of all levels of English language proficiency. The third group acted as the control group and was taught using Instructional Method A; students in this group were of low English language proficiency. Due to this type of placement it was challenging to disentangle the extent to which treatment effects reflected in the results of the study were due to English language proficiency or the treatment itself.

In order to investigate whether the two instructional methods under consideration had a significant effect on students' achievement and understanding of probability, first an analysis was carried out to identify whether the students in the three course sections had comparable initial probability knowledge. The results of the analysis indicated that the students in the control group had comparable initial probability knowledge to the students in the treatment group.

Quantitative as well as qualitative analysis were carried out to address the research questions. First, a comparison of *gain scores* within each group from pre-test to post-test was performed in order to determine the effects of each instructional method on students' *achievement* in probability. That is, the pre-test scores of the treatment group were compared to the post-test scores of the same group in order to determine if the instructional treatment had an effect on students' achievement. Also, the control group's pre-test and post-test scores were compared to determine if the instructional method used in this group had an effect on students' achievement. Second, a comparison of *normalized gain scores* was performed in order to establish whether the instructional method used in the treatment group (Instructional Method B) had a better effect on students' achievement in probability than the instructional method used in the control group (Instructional Method A). Moreover, a comparison of post-test scores on the open-ended items and a comparison of post-test total scores between the two groups were carried out.

In addition to an analysis of scores relative to *achievement*, analyses were performed to determine the effects of the instructional treatment on students' *understanding* of probability. These analyses included i) a distractor analysis of multiple-choice items and ii) qualitative analysis of transcripts of the conversations of the six students monitored in the control and treatment groups. Recall that three students were selected from the control group and three from the treatment group: one student of high mathematical ability, one student of moderate mathematical ability and one student of low mathematical ability from each group. The selection criterion was students' mathematics grade in the last year of high school. The framework developed by Jones, Thornton, Langrall and Tarr (1999) was used to qualitatively analyze the transcripts. This framework included six constructs (sample space, experimental probability of an event, theoretical probability of an event, probability comparisons, conditional probability,

and independence) across four levels of reasoning (Level 1: Subjective; Level 2: Transitional; Level 3: Informal Quantitative; and Level 4: Numerical). Transcripts and their coding were completed by the researcher and by a hired transcriber/second coder.

7.2 Summary of the Findings

7.2.1 Effects of Instructional Treatment on Students' Achievement in Probability

Based on the results of the Wilcoxon Signed-Ranks test which was carried out to analyze gain scores, the multiple-choice scores of students in the control group were significantly lower on the post-test compared to the pre-test. The results indicated that for a majority of these items, students in the control group had a lower performance on the post-test compared to the pre-test. Moreover, all of the percentage descreases were higher compared to the percentage increases. In particular, the most substantial *descreases* were with regards to Items 4, 11 and 13 in which percent-correct responses decreased by 52.9%, 58.8% and 41.1% respectively .

In the case of the treatment group, student scores on the multiple-choice items did not differ (increase) significantly from the pre-test to the post-test. Students in the treatment group had a percent-correct increase on six multiple-choice items and a percent correct decrease on five multiple-choice items. Moreover, most of the percentage increases were higher than the percentage decreases. The most substantial *increases* were with regards to Items 3 and 10 in which percent-correct responses increased by 25.9% and 48.2% respectively.

In addition to the analysis of raw gain scores for each of the two groups, a final piece of analysis was carried out which involved *normalized* gain scores (Bao, 2006; Hakes, 1998) and which directly addressed the research questions i.e. whether Instructional Method B (treatment) had a better effect on students' achievement on probability than Instructional Method A (control). Positive as well as negative normalized gains existed in both groups. In particular, 12

students in the control group and 4 students in the treatment group exhibited negative normalized gain scores. The Mann-Whitney test resulted in a p-value of 0.001 (< 0.05) indicating that the normalized gain scores of the treatment group were significantly different from the normalized gain scores of the control group. The test generated a mean rank of 14.41 for the control group and a mean rank of 26.96 for the treatment group. This means that the control group had a bigger number of lower normalized gain scores compared to the treatment group. Likewise, the treatment group had a bigger number of higher normalized gain scores in comparison to the control group. In summary, Instructional Method B had a significantly better effect on students' achievement on probability than Instructional Method A.

Apart from the 14 multiple-choice items involved in the analysis of gain scores, the post-test included an additional multiple-choice item on discrete probability distributions and two open-ended items: i) one on simple and joint probabilities, conditional probability and independence and ii) one on the binomial distribution. With regards to this additional multiple-choice item, 23.5% of students in the control group and 48.1% of students in the treatment group responded correctly. For the purposes of comparing the open-ended item scores of the control and treatment groups, scoring rubrics were created which allotted numerical values to student responses.

The Mann-Whitney test was used to compare the results from the open-ended items in the control and treatment groups. The test resulted in a p-value of 0.001 indicating that the scores of the treatment group on the post-test open-ended items were significantly different from the scores of the control group. The test generated a mean rank of 14.15 for the control group and a mean rank of 27.76 for the treatment group. This means that the control group had a bigger number of lower scores compared to the treatment group. Likewise, the treatment group had a bigger number of higher scores in comparison to the control group. Therefore, the achievement

of students in the treatment group on the post-test open-ended probability items was significantly higher than the achievement of students in the control group on the corresponding items. Therefore, Instructional Method B was successful in producing significantly higher achievement scores on the post-test open-ended items compared to Instructional Method A.

Non-parametric tests were also used to compare the total post-test scores of the control and treatment groups. In particular, the Mann-Whitney test resulted in a p-value of 0.00 (< 0.05) indicating that the post-test scores of the treatment group were significantly different from the post-test scores of the control group. The test generated a mean rank of 12.53 for the control group and a mean rank of 28.78 for the treatment group. This means that the treatment group had a bigger number of higher scores compared to the control group. In summary, the post-test achievement of students in the treatment group was significantly higher than the post-test achievement of students in the control group. Therefore, Instructional Method B was successful in producing significantly higher post-test scores compared to Instructional Method A.

7.2.2 Effects of Instructional Treatment on Students' Understanding of Probability

In this dissertation, students' understanding of probability was measured through:

i) a distractor (quantitative) analysis of student responses to the multiple-choice items on the pre-test and post-test (Chapter 5) and

ii) a qualitative analysis of audio-taped conversations as students worked in groups on activities (treatment group) or problem sets (control group) (Chapter 6).

For the purposes of distractor analysis, item parts assessing the same heuristic or misconception were grouped together. The percentages of students selecting the particular distractor on each of the grouped items were added up and the mean percentage of students

applying the particular heuristic or misconception on the pre-test and post-test in each group was computed.

In the case of the control group, the mean percentage of students who applied the outcome approach remained stable at 5.9%; the same was true with regards to using the absolute size instead of relative size when computing probabilities (percentage remained stable at a high 20.6%). In addition, there was a slight increase in the mean percentage of students who applied the equiprobability bias which remained high on both the pre-test and the post-test (28.4% and 29.4% respectively). Moreover, there was a small increase in the mean percentage of students who applied the negative recency (by 2.9%) and the positive recency (by 3.9%). The biggest change in the case of the control group was in the application of the representativeness heuristic (17.7% increase).

Unlike the results of the control group, the mean percentage of students in the treatment group who applied the aforementioned heuristics or misconceptions mostly decreased. In particular, the mean percentage of students who used the equiprobability bias, negative recency, positive recency and outcome approach decreased by 3.4%, 9.25%, 4.9% and 1.85% respectively. Only in the case of the representativeness heuristic and use of absolute size instead of relative size when computing probabilities the mean percentage increased by 3.7% and 2% respectively.

Considering students' responses relative to other types of misconceptions represented by item distractors, the mean percentage of students who believed that a larger number of outcomes in the sample space implies a higher probability of occurrence increased in the case of the control group (from 0% to 5.9%) and decreased in the case of the treatment group (by 3.7%). Second, the mean percentage of students who applied division by an incorrect total when computing

probabilities using a two-way table increased by 3% in the case of the control group and by 11.1% in the case of the treatment group. Regarding the misconception that probability = 1/number of favorable outcomes when one item is selected at random, the mean percent of students who applied this misconception remained stable at 5.6% in the case of the treatment group but increased by 23.6% in the case of the control group.

Some misconceptions related to the union of events. The mean percentage of students who believed that the term 'or' means considering only one of the events was higher in the case of the control group on both the pre-test and post-test while in both groups the percentage of students who applied this misconception slightly increased from pre-test to post-test. Similarly, the mean percentage of students who applied the misconception that $P(A \cup B) = \frac{P(A) + P(B)}{2}$ was higher and remained stable at 29.4% in the case of the control group on both the pre-test and post-test whereas it decreased by 7.4% in the case of the treatment group (from 14.8% to 7.4%). Unlike the first two misconceptions relating to the union of events, the mean percentage of students who applied the misconception $P(A \cup B) = P(A) + P(B)$ decreased in the case of the control group (by 5.9%) and increased in the case of the treatment group (by 3.7%).

For the purposes of qualitatively analyzing audio-taped data, transcripts of student conversations of the six students monitored in the control and treatment groups as they worked in groups were coded independently by the researcher and by a second rater. Once the coding of transcripts was completed, they were checked by the researcher for agreement on the number of excerpts relating to the constructs under study and the associated levels of reasoning. The researcher and second coder met once more to discuss and resolve any instances of disagreement in coding.

Recall that the framework used in this study does not include the concepts of discrete probability distribution and binomial distribution. So, students' level of reasoning as exhibited in excerpts associated with discrete probability distributions (in Activity 3/Problem Set 3) and the binomial distribution (in Activity 4 /Problem Set 4) could not be *directly* coded using the framework. However, in Activity 3/Problem Set 3 and in Activity 4/Problem Set 4 students needed to make connections to basic probability concepts; reasoning on such concepts is discussed in the framework. Excerpts in which students made connections to basic probability concepts were accounted for in the analysis of data relative to students' understanding of basic probability concepts. Any excerpts relative to the concepts of discrete probability distributions and the binomial distribution in Activity 3/Problem Set 3 and in Activity 4/Problem Set 4 respectively in which students did not make connections to basic probability concepts were briefly discussed in separate sections in chapter 6.

An examination of the results of the qualitative analysis regarding students' reasoning levels relative to sample space, theoretical probability, and experimental probability respectively, revealed the following: i) The activities provided students with more direct as well as indirect opportunities to reason about sample space and experimental probability compared to the problem sets; and ii) The problem sets did not provide any direct or indirect opportunities for students in the control group to reason about experimental probability and so, there was no data on these students relative to this construct. Given these, it is not surprising that, as specified in Chapter 5, students in the treatment group performed better on the multiple-choice items relating to the Law of Large Numbers (Item 6 on the pre-test and post-test) and the concept of sample space (Items 7 and 11 on the pre-test and post-test) than students in the control group. Specific to Item 6 the percent-correct responses decreased by 11.7% in the control group and by 3.7% in the

treatment group from pre-test to post-test. On Item 7 the percent-correct responses increased by 5.8% in the case of the control group and by 18.6% in the treatment group from pre-test to post-test. Last, on item 11, the percent-correct responses decreased by a substantial 58.8% in the case of the control group and increased by 3.7% in the treatment group. Considering these, Instructional Method B (treatment) produced better results than Instructional Method A (control) relative to experimental probability and the concept of sample space.

In addition, the results of the qualitative analysis regarding students' reasoning levels relative to sample space, theoretical probability, and experimental probability respectively, revealed that: i) Students' reasoning levels were the most stable with regards to sample space in comparison to theoretical or experimental probability; ii) Students exhibited higher levels of reasoning with regards to sample space (mostly L4) compared to theoretical or experimental probability across the four activities and problem sets; and iii) with regards to experimental probability, the three students monitored in the treatment group moved from a lower to a higher level reasoning in Activities 1 and 4; moved from a higher to a lower level reasoning in Activity 2; and remained stable at a low level reasoning in Activity 3.

With regards to theoretical probability, students in both groups mostly transitioned from a lower to a higher level reasoning whereas in some cases students in the control or treatment group exhibited a stable level of reasoning. In addition, students in the control group performed 0-3 transitions between levels of reasoning while students in the treatment group performed 0-2 such transitions. Based on these results, it is difficult to identify which instructional method had a better effect on students' understanding of theoretical probability.

An examination of the results of the qualitative analysis relative to the concepts of conditional probability and independence revealed that both Activity 2 and Problem Set 2

provided students with direct opportunities to reason about conditional probability and independence. Moreover, both of these instructional materials afforded opportunities in which students needed to draw on previous knowledge of basic probability concepts (sample space, theoretical probability of simple and joint events, experimental probability) in order to compute conditional probabilities and decide whether events were independent. Overall, i) students' reasoning with regards to conditional probability and independence was stable; ii) students exhibited high levels of reasoning with regards to both concepts; iii) in the case of conditional probability the low-ability and moderate-ability students in the treatment group demonstrated L4 reasoning whereas in the case of independence, all three students monitored in the control group exhibited L4 reasoning; and iv) with regards to conditional probability all three students monitored in the control group moved from L2 to L4 reasoning. Based on these results, it is difficult to identify which instructional method had a better effect on students' understanding of conditional probability and independence.

In the case of discrete probability distributions, the three students monitored in the treatment group were able to make connections to theoretical probability when computing the probability distribution of a discrete random variable. Overall, the three students in the treatment group exhibited L4 reasoning with regards to theoretical probability (and indirectly, with regards to discrete probability distributions). In the control group, all three students were able to correctly compute the expected value of a discrete random variable when a table listing the possible outcomes and associated probabilities was provided. However, when asked to construct the probability distribution themselves, all three students in the control group found such a task to be challenging.

The last concept examined in this study was that of binomial distribution. The three students monitored in the treatment group were able to construct the sample space for a random variable that follows the binomial distribution thus exhibiting L4 reasoning with regards to sample space. The high-ability and low-ability students in the treatment group demonstrated L4 reasoning with regards to theoretical probability (and subsequently with regards to the binomial distribution) since they were able to assign correct numerical probabilities to each outcome in the sample space and then, correctly computed the probability of an event. The moderate-ability student in the treatment group demonstrated L2 reasoning with respect to theoretical probability since the student used quantitative judgments to compute the probability of each outcome but did so incorrectly. In addition, the high-ability and moderate-ability students in the treatment group were able to use a table of random numbers to calculate the experimental probability of an event thus exhibiting L4 reasoning with regards to experimental probability (and subsequently with regards to the binomial distribution). The treatment low-ability student was also able to use the table of random numbers but instead of assigning a numerical probability the student provided a frequency value for the experimental probability of the event thus demonstrating L3 reasoning with respect to experimental probability (and subsequently with respect to the binomial distribution). When it came to using the general formula for computing a probability from the binomial distribution, a copy of each group's written work indicated that students were able to correctly compute the required probabilities from the binomial distribution. However, transcripts of student conversations revealed that all three students monitored in the treatment group had difficulties applying the general formula in the given context. The three students monitored in the control group were able to define n (the number of observations in the sample space), X (the number of events of interest), and p (the probability of an event of interest) as well as to compute

the mean and standard deviation of the binomial distribution. Moreover, all three students monitored in the control group were able to apply the general formula for computing the probability from the binomial distribution and assigned correct numerical probabilities as responses to textbook problems.

7.3 Discussion of Findings

Based on the findings relative to students' *achievement* in probability, scores on the multiple-choice items of students in the treatment group did not increase significantly from the pre-test to the post-test. Surprisingly, the corresponding scores of students in the control group decreased significantly from the pre-test to the post-test. Recall that the two groups were initially equivalent with respect to pre-test multiple-choice item scores. Consideration of these results leads one to wonder: Why were the post-test multiple-choice item scores of the control group significantly lower compared to the pre-test multiple-choice item scores?

The results of the quantitative analysis provided in Chapter 5, revealed that in both the control and treatment groups, student responses exhibited a ceiling effect on some items i.e. more than 80% of the students in each group responded correctly to these items. In the treatment group, student responses exhibited a ceiling effect on items 1, 2, 4, 11 and 12 on both the pre-test and post-test, indicating that students in this group found the particular items to be easy. In the case of the control group, a ceiling effect was observed on pre-test items 1, 4, 11, 13 and 14. Interestingly, when it came to the post-test, there was a dramatic decrease in percent-correct responses on items 4, 11 and 13 by 52.9%, 58.8% and 41.1% respectively. Overall, in the control group, percent-correct responses decreased in nine out of the fourteen multiple-choice items that were common to the pre-test and post-test. It should also be noted, that unlike the aforementioned items, item 8 proved to be a very difficult item for students in both the control

and treatment groups since less than 30% of students in either group responded correctly to it on the pre-test and post-test.

The information presented in the previous paragraph points towards the need for a closer examination of the substantial decreases in percent-correct responses on items 4, 11 and 13 from the pre-test to the post-test in the case of the control group. Considering item 4, in which students were provided with a pictorial representation of all possible outcomes when two fair six-sided dice are rolled and were asked to respond to the following,

> You are about to roll 2 fair six-sided dice, hoping to get a double. (A double = both dice show the same value on top). Which double will occur the <u>least</u> often?
> a) 6 and 6
> b) 1 and 1
> c) 1 and 1, and, 6 and 6 are both least likely to occur.
> d) All doubles are equally likely.

41.2% of students in the control group selected option c) as the correct answer on the post-test whereas *all* students had correctly selected option d) on the pre-test. A possible explanation for this phenomenon might lie in the instructional materials used in the two groups in this study. The activities used in the treatment group involved experiments in which students rolled two or three fair dice and computed probabilities of various events. However, the problem sets provided in the course textbook on which students in the control group worked in groups during class did not provide any hands-on opportunities for students to work with dice and so, these students were not provided with as many opportunities to overcome their misconceptions relating to dice outcomes. Students' selection of option c) is supported by results of other studies reported in the literature (Green, 1983; Konold et al., 1993) in which participants tended to prefer middle numbers on a die and avoid the numbers 1 and 6, thinking that the middle numbers have a higher chance of occurring.

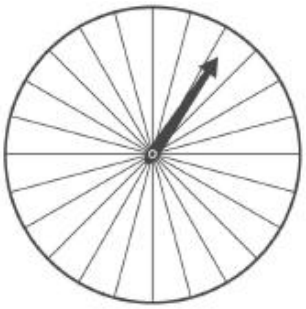
The next item under consideration is item 11 in which students were asked to respond to the following:

> The eleven chips shown below are placed in a bag and mixed. Chelsea draws one chip from the bag without looking.
>
> Figure 7.1
>
> *Pre/Post-Test Item 11 Coin Figure*
>
> 
>
>
> What is the probability that Chelsea draws a chip with a number that is a multiple of three?
>
> a) $\frac{1}{11}$
>
> b) $\frac{1}{3}$
>
> c) $\frac{4}{11}$
>
> d) $\frac{4}{7}$

On the pres-test, 94.1% of students in the control group had correctly selected option c). Surprisingly, on the post-test, 47.1% of these students selected option a) (whereas *none* of them had made this selection on the pre-test). A possible explanation may be that problems in the course textbook tend to reinforce the idea of equally likely outcomes and so, as reported in the literature, students believed that in order for an experiment to be "fair", all outcomes needed to be equally likely (Jacobs, 1999; Shaughnessy, 2003). Another explanation might involve the issue of language. When responding to this item on the pre-test during regular class time, many participants (especially in the control group) were not aware of the meaning of 'multiple of' and asked for a translation of the term in Greek at which time I (instructor/researcher) wrote the

translation on the classboard. Recall that the post-test was embedded in the course final exam which did not take place during regular class time. Students had the right to ask a supervisor during the final exam for a translation of a few words in Greek. It is possible that many students in the control group could not remember the meaning of 'multiple of', did not ask an invigilator for its meaning, and instead focused on the idea of drawing *one* out of the eleven chips in the bag.

The third item in which percent-correct responses exhibited a substantial decrease from pre-test to post-test was item 13 which is presented below:

> The figure below shows a spinner with 24 sectors. When someone spins the arrow, it is equally likely to stop on any sector.
>
> Figure 7.2
>
> *Pre/Post-Test Item 13 - Spinner*
>
> 
>
> 3 of the sectors are blue, 1 is purple, 12 are orange, and 8 are red.
> If a person spins the arrow, on which color sector is the spinner LEAST likely to stop?
> a) Blue
> b) Purple
> c) Orange
> d) Red

Relative to this item, 35.3% of students in the control group selected option c) on the post-test. It seems that students in the control group confused the meaning of 'least' with the meaning of 'most' likely when responding to this item, pointing towards students' difficulties with the English language.

Recall that participants in the introductory statistics course in which this study took place were split into three sections by the course coordinator in consultation with the academic board at the college, based on students' English language proficiency. Two of the sections served as the treatment group and one section as the control group. One of the treatment sections included students of moderate/high English language proficiency whereas the other included students of all levels of English language proficiency. The control section included students of low English language proficiency. Course instructors were not aware of this type of placement until the commencement of the academic year. Moreover, due to this placement, it is challenging to identify the extent to which the results are attributed to difficulties with the English language or the instructional method. As noted in items 11 and 13, possible explanations for substantial the decreases in percent-correct responses in the case of the control group may be (at least partially) explained by difficulties with the English language. Researchers suggest that linguistic difficulties associated with the terminology used for probability are prevalent among students (Green 1982a, Konold, 1988; see Ulep, 1990). Adding to this, students' English language difficulties, creates further obstacles to students' understanding and achievement in probability.

Further research needs to be carried out in order to examine the factors that might have led to these changes in multiple-choice item scores. The fact that students completed the pre-test during a regular classroom session whereas the post-test was administered under exam conditions (it was embedded in the course final exam) in combination with the low English proficiency of students in the control group, might have led to the significant decrease in post-test multiple-choice item scores in the case of the control group. Exam conditions may add to students' stress level and in combination with English language difficulties may impede their comprehension of probability items stated in English. The moderate/high English language proficiency of students

in one of the treatment sections may have helped students better demonstrate their actual understanding of probability concepts since such level of proficiency might have helped students better comprehend the context of each multiple-choice item. In this manner, students having less difficulties with the English language might make the results of the two treatment sections more directly attributable to the instructional treatment used.

Similarly, in the case of the open-ended items included on the post-test, the scores of the treatment group were significantly higher than the scores of the control group. As in the case of the multiple-choice items, here as well, the instructional method used and the language issue might have both played a role in the results exhibited by participants. The activities used in the treatment group included both a hands-on experiment and a theoretical component so students were given the opportunity to study a concept in both manners and make connections between the experimental and theoretical approach to probability. Such an approach has been shown to aid students in overcoming probability misconceptions (Shaughnessy, 1977, 1981) and has been supported by researchers (Steinbring, 1984; 1991; see Jones and Thornton, 2005). The problem sets that students in the control group worked on did not include an experimental component; in these problems students needed to apply knowledge they had gained during lectures on a probability concept to merely solve the problem at hand. However, according to Shaughnessy (1981), "an initial formalistic approach to probability is unlikely to help students overcome misconceptions" (p. 95). The fact that the course textbook did not provide students in the control group with direct opportunities to work with experimental probability indicated a preference for a classical approach to probability. This comes in contrast to current calls to teach probability in conjunction with data collection and analysis (Shaughnessy, 2003). Moreover, the low English language proficiency of students in the control group might have led to difficulties in

comprehending the problem context and in turn might have caused the significantly lower open-ended item results in comparison to the results of the treatment group on the corresponding items.

Results of the distractor analysis relative to students' understanding of probability are in line with research on students' probabilistic misconceptions. As evident in the literature, probabilistic misconceptions are resistant to change (Konold, 1995) and in general, "there is no simple story about how students reason about chance" (Konold et al., 1993, p. 413). Specific to the distractor analysis carried out in this study, the results indicated that the percentage of students applying some of the probabilistic heuristics, increased in some cases whereas it decreased in others. This was true in both the treatment and control groups. However, with regards to the 12 heuristics presented in Tables 5.9 and 5.10 in Chapter 5, the percentage of students in the control group who applied these increased in the case of 8 (67%) of them, remained stable in the case of 3 (25%) of them and decreased in the case of 1 (8%) of them. On the other hand, the percentage of students in the treatment group who applied these heuristics, increased in the case of 5 (42%) of them, remained stable in the case of 1 (8%) of them, and decreased in the case of 6 (50%) of them. These results are in line with past research that indicated that the use of activity-based instruction may help students with respect to probabilistic misconceptions (Shaughnessy, 1977, 1981).

7.4 Strengths and Limitations

7.4.1 Strengths

Although the sample of students who participated in this study cannot be regarded as representative of statistics students in general, the reasoning and understanding abilities of these students and the role that different types of tasks have on their learning and understanding of

probability may aid educators gain knowledge about students who study statistics in a college-level service course.

The fact that the class sections (treatment and control) were taught at the same setting, by the same instructor, using the same textbook and language of instruction, may be considered strengths. Controlling for these factors allowed the results of the study to be more clearly attributed to the different tasks used in the study, thus more directly addressing the research questions.

For the purposes of data collection, audio recorders were used to record verbatim the responses students gave as they worked in groups on probability problems or activities which involved open-ended probes or questions. Using audio recordings as a research tool allows the researcher to obtain more data and to have easy reach to students' verbal interactions and responses thus aiding in addressing the research questions. Given that the researcher was also the instructor it would have been impossible to keep track of students' group discussions and reasoning as they worked in class without a means of data collection such as audio recording.

Related to the above issue, the fact that the instruments included both multiple-choice and open-ended questions can be considered an advantage of the study. Each of these two types of questions has advantages and disadvantages. Although multiple-choice questions do not allow respondents to express their reasoning freely, open-ended questions do so. Moreover, it is more challenging to analyze open-ended questions however, multiple-choice items are easier and quicker to quantify. Therefore, using a combination of multiple-choice and open-ended items on the instruments adds to the strength of the study.

7.4.2 Limitations

A limitation of the study is that the interpretation of the data might involve a degree of bias since the instructor was also the researcher (Zimmerman, 2002). In order to control for such bias a second coder was hired; this coder helped code participant responses to questions on the activities or in the problem sets.

Second, "[A] criticism of qualitative research methods is that it is very difficult to make and justify generalizations that apply to other settings." (Jaworski, 1998, p. 119). This study was performed in a single college at a particular town in Cyprus. More diverse student populations may respond differently than the participants in this study. Therefore, the results may only be generalized to students with similar demographics.

Third, the design of the study was complicated by the fact that students were placed in the course sections by the course administrator and academic board based on their English language proficiency. This placement caused obstacles in the interpretation of results since it was challenging to determine the extent to which the study's results were attributable to the method of instruction or to language issues.

Last, the search for a framework to be used in analyzing students' understanding of probability proved to be a very challenging task. Any available frameworks I came across to related to grades K-12 (Jones et al., 1997, 1999; Jones and Thornton, 2005). In this study, the framework used to analyze students' understanding of probability was that by Jones, Thornton, Langrall and Tarr (1999) which included six constructs (sample space, experimental probability of an event, theoretical probability of an event, probability comparisons, conditional probability, and independence) across four levels of reasoning (Level 1: Subjective; Level 2: Transitional; Level 3: Informal Quantitative; and Level 4: Numerical). The framework is limited in that i) it

did not directly relate to *college* students' understanding of probability; and ii) it did not accommodate for all probability constructs under examination i.e. discrete probability distributions and the binomial distribution. Moreover, the type of classification used in the framework assumes that i) "an individual's thinking about a specific probabilistic construct will be consistent across different contexts related to that construct" and ii) "that an individual's thinking will be consistent across the constructs themselves" (Rubel, 2007). However, as evident from the results, students' reasoning regarding the probability constructs under examination did not model such consistencies. This lack of consistency in reasoning is evident in the tables presented in Chapter 6 which reveal that students transitioned between levels of reasoning during the activities or problem sets. Similar to this study, Jones et al. (1999) also came across "these types of instability within their data, especially following instruction, and speculated that students' probabilistic thinking might be tightly woven within the features of the context itself" (Rubel, 2007). Although the framework by Jones et al. (1999) has its limitations, it was selected because it distinctly addressed each of the constructs of experimental and theoretical probability in a cognitive manner (thus aiding in addressing the issue of *understanding*) and these two constructs were vital elements of the study.

7.5 Implications and Recommendations for Further Research

The findings of this study may help mathematics instructors and curriculum developers by providing valuable information on the effect of particular tasks on students' achievement and understanding of probability. While the sample used in this study may not be viewed as representative of statistics students in general, the perceptions of these participants may help educators understand tertiary level students who study probability as part of a service course rather than by choice.

As the Curriculum Principle (NCTM, 2000) states, curriculum materials should emphasize the connections between mathematical ideas. An examination of the course textbook used in the course in which this study took place, indicated that it did not afford students with opportunities to study experimental probability, particularly through modeling. However, researchers (e.g. Steinbring, 1984; 1991; see Jones and Thornton, 2005; Shaughnessy, 2003) recommend that instruction in probability should help students view the connections between theoretical and experimental probability. Therefore, curricular materials should involve the modeling of real-world situations through simulations. Moreover, during the process of selecting the instructional materials to be used in this study, I realized that there is a lack of available activities that could be used during instruction on probability in introductory statistics courses at the tertiary level. Given this, curriculum developers should ensure that such activities are developed.

In Cyprus, research on students' achievement and understanding of probability has not been emphasized, especially at the secondary and post-secondary levels. Furthermore, the role of such research on the teaching and learning of probability, and in curriculum development has not been widely applied. Current practices are under review as reform initiatives are being implemented in mathematics classrooms at all levels (see Chapter 2). Reports on the reformed mathematics curriculum place increased emphasis on new topics that have rarely been taught at the school level in Cyprus, including probability, and in the way these topics are taught i.e. using increased student involvement in the learning process, cooperative learning and technology (Papastylianou, 1997; see also Papanastasiou, 2002; Vrasidas and McIsaac, 2001). In making such changes to the curriculum, it is important to be aware of available research. The findings of this study suggest that curriculum developers and teachers promote the use of activity-based

instruction that helps increase students' achievement on probability. Teachers and curriculum developers should make use of real data and activities as means in encouraging students to express their reasoning and modify their beliefs with regards to probability concepts (Garfield & Ahlgren, 1988; Shaughnessy, 1992).

The results of this study also inform *my* future probability instruction. If another opportunity arose to teach a college-level course in probability in which I would have total control of student placement in the various course sections (unlike the situation that arose in this study), then, based on these results I would i) assign students to course sections randomly and not based on their English language proficiency; and ii) use the instructional method implemented in the treatment group in this study since it had a positive effect on students' achievement in probability.

The results of this study give rise to recommendations for further research. First, let us consider research that could be carried out as a follow-up to this study. The findings relative to students' *achievement* indicate that the instructional treatment of using activities that generate real data while having students work in small groups during probability instruction, significantly increased students' achievement in probability. Further research needs to be carried out with the aim of revealing the factors that help explain these significant differences in achievement between the treatment and control groups in this study. Factors such as gender, the role of language, students' confidence in mathematics, and students' feelings regarding the use of group work should be examined. The limitations identified relative to the framework used in this dissertation give rise to the following needs: i) the development and validation of a framework that can be used to assess *college* level students' understanding of probability (including the construction of a clear definition of the term *probabilistic understanding*); ii) the development

and validation of a framework that can be used to assess students' understanding of discrete probability distributions and the binomial distribution. In addition, findings from a larger group of students at various tertiary education settings could produce trends that were not evident in this study. In addition, testing a larger sample of students could provide more insight into the prevalence of the resulting trends.

Particular to the language factor and given that instruction at the research site was carried out in English whereas the students' first language was Greek, further research is needed into the relationship between classroom discourse relative to probability carried out in Greek and that in English, and whether students' difficulties arise from these relationships. Relative to this issue, the effectiveness of allowing students to discuss probability in their home language during instruction, needs to be investigated as well as the role played by language in the assessment of achievement in probability; in particular, the effects of tests written in the students' first language on achievement in probability should be investigated.

A question that remains is the role that additional reflection might play in students' understanding of probability from the activities. In this study students were not directly assigned any out of class work related to the activities nor were they directly assessed on exams on the content presented in the activities using questions that resembled the activities. So, one may assume that students spent a minimal amount of time outside of class reflecting on the activities. The question remains as to whether additional reflection might help students develop and retain an understanding of the probability concepts presented in the activities.

APPENDICES

APPENDIX A: CONSENT FORM AND PERMISSION FOR USE OF REPRINTS

MICHIGAN STATE
U N I V E R S I T Y

1/15/10

Title of Research Study: College Students' Achievement and Understanding of Experimental and Theoretical Probability: The Role of Tasks

Dear Student:

I am contacting you to ask whether you would be willing to participate in a research project relating to mathematics education. In particular, the project attempts to examine the role of different tasks on college students' achievement and understanding of probability.

If you agree to participate, then I would kindly ask that you fill out a background questionnaire, a pre-test and post-test of probability problems. You may also be asked to work on a set of classroom group activities or problem sets. As you will be working on the group activities your group conversations will be audio-taped and these tapes will be transcribed by me after the final course grades have been completed and submitted.

Division of Science and Mathematics Education
Michigan State University
North Kedzie Hall
East Lansing, MI
48824-1034

Such a study holds benefits as well as some risks for participants. Educational research has revealed that active learning and communication in mathematics classrooms can strengthen students' reasoning and clarify their understanding of mathematics concepts. Moreover, research has shown that students engage more with the material and more of them are able to complete the introductory statistics course when cooperative learning is used during instruction. Such instructional methods will be part of this study and so, you may benefit through your participation. However, there is also some small risk in that you might feel embarrassed if you do not know the answer to a problem or do not know how to proceed with some step of a probability experiment which you will be performing as part of this study. Also, thinking out loud and/or being in a cooperative learning environment in the classroom might be an unfamiliar process to you and you might feel somewhat uncomfortable at first. Always feel free to express any difficulties you encounter as your work on the group activities and to seek help from your classmates or your instructor.

The results of this study might be published in scientific research journals and professional publications for teachers. Only I will use the tapes or transcripts for analysis and that will be done after final grades have been completed and submitted. Access to the tapes and transcripts will be protected according to university regulations and up to the maximum extent allowable by law. No data that is collected will identify the student from which it was collected, other than to members of this research team. Results of the study will be made available to you if you request them.

Participation in this study is voluntary and you may withdraw from participation at any time. Your consent to participate or your decision to withdraw will not be made known to me until after the final course grades have been completed and submitted.

The data materials (background questionnaire, pre-test and post-test) will be collected by the course coordinator. I will not examine these materials until after the final course grades have been completed and submitted. If you decide not to participate or if you withdraw from the study at any point, the data materials you complete as well as anything you say that might have been recorded on tape will not be used as data for this study.

If you are willing to participate, please sign this form and return it to the course coordinator.

Student's Printed Name _____

Student's Signature _____

Date_____

If you have any questions about this project, please do not hesitate to contact the researcher, Irini Papaieronymou, by phone (99784507), e-mail (papaiero@msu.edu) or regular mail: Corner of Faneromenis Ave. and Kalvou Str., P.O.Box 40763, 6307 Larnaca, Cyprus. If you have any questions or concerns about your rights as a participant in this study, you may contact Michigan State University's University Committee on Research Involving Human Subjects (UCRIHS) at 202 Olds Hall, Michigan State University, East Lansing, MI 48824-1046, Phone: (517) 355-2180, E-Mail:irb@msu.edu.

Sincerely,
Irini Papaieronymou
PhD Candidate in Mathematics Education, Michigan State University,USA
Mathematics Instructor, PA College, Larnaca, Cyprus

AUDIOTAPE CONSENT STATEMENT

Cooperative Learning in a Probability Project

One of the goals of this project is to help teachers and researchers better understand students' probabilistic thinking. To do this, we must give examples of students' work. One way to give examples is to provide transcripts of what students say as they are solving problems (with students not being identified, of course). If you would be willing for your audio-taped work to be shared with teachers and researchers, please give your consent below.

I agree that portions of audiotapes of my conversations with members of my group as we work on probability problems or experiments during class time may be used by Irini Papaieronymou in presentations and classes for teachers or researchers. I understand that I have the right to listen to the audiotapes before they are used. I have decided that I:

____want to listen to the audiotapes

____do not want to listen to the audiotapes

Sign now if you <u>do not</u> want to listen to the audiotapes. If you want to listen to the tapes, you will be asked to sign after listening to them.

_____
Student's Printed Name

_____ _____
Student's Signature Date

PERMISSION FOR USE OF REPRINTS (ARTIST PROJECT)

Dear Irini:

Thank you for registering for the Web ARTIST Assessment Builder. We hope you have had a chance to review and download some of our items.

We offer 11 online multiple choice tests on individual topics, and the CAOS (Comprehensive Assessment of Outcomes in a first Statistics Course) test. We encourage you to try one or more of our online topic tests which have been found to be especially helpful for reviewing material. We also encourage you to use the CAOS as a pretest and posttest to measure gains in student learning. All test results can be quickly and easily downloaded once students have completed taking each test. For more information and to sign up to administer a test, please go to: https://app.gen.umn.edu/artist/tests/index.html

We invite you to send us copies of any good assessment items or materials that you are willing to contribute to our website by sending them in an e-mail to Bob delMas (delma001@umn.edu).

We have a Question and Answer page where we post questions we receive that we think will be of interest to users, and we share our responses to these questions. So feel free to send questions.

Thanks in advance for your feedback, and let us know if you have any problems or questions.

Best regards,

Joan Garfield, Bob delMas, Beth Chance, and Ann Ooms

The ARTIST Team

Irini:

Now that you are registered to access the online tests, you can have a PDF file of the CAOS test sent to you. You can make copies of this to use in class. We ask that you collect the tests from the students and do not let students keep copies of the CAOS test. There is no need to send us an Excel file of the students' responses.

You can learn more about retrieving a PDF file of the CAOS test at:

https://app.gen.umn.edu/artist/tests/index.html

Read the instructions for Step 2.

Best regards,

Bob delMas

****

Robert C. delMas, Ph.D.

Associate Professor

Quantitative Methods in Education

Department of Educational Psychology

University of Minnesota

Phone: (612) 625-2076

Fax: (612) 624-8241

Irini:

Yes, you can definitely use the ARTIST Probability Scale for your dissertation research. For your Reference section, you can list the project as follows.

Garfield, J., delMas, R., Chance, B. (2002). Assessment Resources Tools for Improving Statistical Thinking (ARTIST). A National Science Foundation funded project (CCLI-ASA-0206571). https://app.gen.umn.edu/artist/

And you can cite it as Garfield, delMas and Chance (2002) in the body of your dissertation.

Bob

APPENDIX B: INSTRUMENTS

B.1 BACKGROUND QUESTIONNAIRE

Student Background Questionnaire

A. Background Information:

1. First Name _____ Last Name _____

2. Age in years _____

3. Town/Country of birth _____

4. What is your first language? (Check one)

   ___ Greek

   ___ English

   ___ Other (please specify) _____

5. Please indicate your level of English proficiency (Circle one in each column):

Table B.1

English language proficiency level

Language	Oral	Written	Reading
English	Poor Good Excellent	Poor Good Excellent	Poor Good Excellent

B. Educational Background:

6. Which of the following best describes the high school you graduated from?

_____ Public high school in Cyprus (continue to question 7)

_____ Private high school in Cyprus (continue to question 8)

_____ I did not graduate from a high school in Cyprus. I graduated from a high school in (specify country) _____

7. If you attended a public high school in Cyprus did you take *advanced* mathematics?

___ Yes

___ No

8. What was your grade in mathematics in your last year of high school? _____

9. Did you ever study statistics and probability in high school?

____ Yes

____ No

10. What year are you at P.A. College? (Check one)

___Freshman (1st year student)

___Sophomore (2nd year student)

___Junior (3rd year student)

___Senior (4th year student)

11. What is your major at P.A. College? (Check one)

___ Accounting (4 years, Bachelor of Arts)

___ Business Administration (4/5 years, Bachelor of Arts)

___ Business Computing (4/5 years, Bachelor of Science)

___ Banking (4 years, Bachelor of Arts)

___ Business Administration (2 years, Diploma)

___ Computing and Information Systems (2 years, Diploma)

___ Business and Information Technology (1 year, Certificate)

___ Law (1 year, Certificate)

B.2 PRE-TEST

Probability Pre-Conceptions

Please work individually on the following problems. In each problem, circle the answer that seems most reasonable to you. Your answers (whether correct or incorrect) will help your instructor to determine what you already know about probability, prior to instruction in this course on probability.

1. George and Mike each bought one ticket for a lottery each week for the past 100 weeks. George has not won a single prize yet. Mike won 20 euros last week, for the first time. Who is more likely to win a prize this coming week if they each buy only one ticket?
   a) George
   b) Mike
   c) They have an equal chance of winning

2. Suppose you read on the back of a lottery ticket that the chances of winning a prize are 1 out of 10. Select the best interpretation.
   a) You will win at least once out of the next 10 times you buy a ticket.
   b) You will win exactly once out of the next 10 times you buy a ticket.
   c) You might win once out of the next 10 times but this will not happen for sure.

3. Melita is flipping a fair coin with her eyes shut. Nora records the outcome and places the coin again in Melita's hand. Heads has just come up 5 times in a row! The chance of getting heads on the next flip is

   a) less than the chance of getting tails since we are expected to get tails.

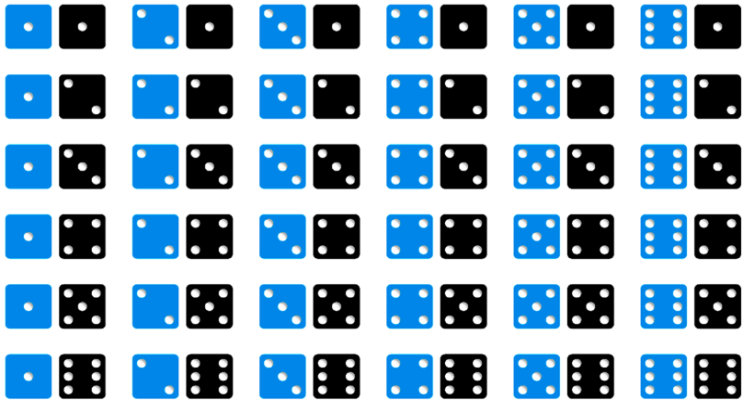
   b) equal to the chance of getting tails since the flips are independent and the coin is fair.

   c) greater than the chance of getting tails since heads seem to be coming up.

For problems 4 and 5 consider the following diagram which shows the 36 possible results when two six-sided dice are rolled.

Figure B.1

Pre-Test Items 4 and 5 – Sample Space for Rolling Two Fair Dice



4. You are about to roll 2 fair six-sided dice, hoping to get a double. (A double = both dice show the same value on top). Which double will occur the <u>least</u> often?

   a) 6 and 6

   b) 1 and 1

   c) 1 and 1, and, 6 and 6 are both least likely to occur.

   d) All doubles are equally likely.

5. Maria rolls two six-sided dice at the same time. Each side of each die is uniquely labeled with a number from 1 to 6. The following are two of the possible results that could occur when these two dice are rolled:

   Result 1: a 5 and a 6 are obtained in any order.

   Result 2: a 5 is obtained on each die.

   Which of the following statements is correct?

   a) The probability of obtaining each of these results is equal.

   b) There is a higher probability of obtaining Result 1 (a 5 and a 6 in any order).

   c) There is a higher probability of obtaining result 2 (a 5 on each die).

   d) It is impossible to give an answer.

6. A game company created a little plastic dog that can be tossed in the air. It can land either with all four feet on the ground, lying on its back, lying on its right side, or lying on its left side. However, the company does not know the probability of each of these outcomes. They want to estimate the probabilities. Which of the following methods is most appropriate?

   a) Since there are four possible outcomes, assign a probability of ¼ to each outcome.

   b) Toss the plastic dog many times and see what percent of the time each outcome occurs.

   c) Simulate the data using a model that has four equally likely outcomes.

7. A set of 24 cards is numbered with the positive integers from 1 to 24. The cards are shuffled and only one card is selected at random. What is the probability that the number on the card can be divided by 4 or 6?

a) $\frac{1}{6}$

b) $\frac{5}{24}$

c) $\frac{1}{4}$

d) $\frac{1}{3}$

e) $\frac{5}{12}$

8. Two containers, labeled A and B, are filled with red and blue marbles according to the quantities listed in the table below. Each container is shaken several times and marbles are then selected from the containers at random, without looking.

Table B.2

Pre-Test Item 8 – Marbles in Container

Container	A	B
Red	80	40
Blue	20	60

Which of the following outcomes has the <u>smallest</u> probability?

a) Obtaining a blue marble from container A.

b) Obtaining a blue marble from container A and a blue marble from B.

c) All of the above are equally likely.

9. Two containers, labeled A and B, are filled with red and blue marbles according to the quantities listed in the table below. Each container is shaken vigorously. After choosing one of the containers, you will reach in and, without looking, draw out a marble. If the marble is blue, you win 50 euros.

Table B.3

Pre-Test Item 9 – Marbles in Container

Container	A	B
Red	6	60
Blue	4	40

Which container gives you the best chance of drawing a blue marble?

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

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

c) Equal chances from each container

10. One thousand people selected at random were questioned about smoking and drinking. The results of this survey are summarized in the table below.

Table B.4

Pre-Test Item 10 Cross Tabulation

	Smokers	Non-smokers
Drinkers	320	530
Non-drinkers	20	130

What is the probability that a randomly selected respondent drinks and smokes?

a) $\frac{320}{340}$

b) $\frac{320}{850}$

c) $\frac{320}{1000}$

d) $\frac{1}{320}$

11. The eleven chips shown below are placed in a bag and mixed. Chelsea draws one chip from the bag without looking.

Figure B.2

Pre-Test Item 11 Coin Figure

What is the probability that Chelsea draws a chip with a number that is a multiple of three?

a) $\frac{1}{11}$

b) $\frac{1}{3}$

c) $\frac{4}{11}$

d) $\frac{4}{7}$

12. Sophie has a bag in which there are 16 marbles: 8 are red and 8 are black marbles. She shakes the bag and without looking she draws 2 marbles from it and does not put them back. Both marbles that she has drawn are black. She then draws a third marble out of the bag.

    What can you say about the likely color of this third marble?

    a) It is more likely to be red than black.

    b) It is more likely to be black than red.

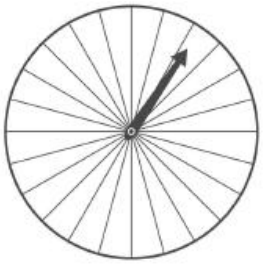
    c) It is equally likely to be red or black.

    d) You cannot tell if red or black is more likely.

13. The figure below shows a spinner with 24 sectors. When someone spins the arrow, it is equally likely to stop on any sector.

    Figure B.3

    Pre-Test Item 13 -Spinner



    3 of the sectors are blue, 1 is purple, 12 are orange, and 8 are red.

    If a person spins the arrow, on which color sector is the spinner LEAST likely to stop?

    a) Blue

    b) Purple

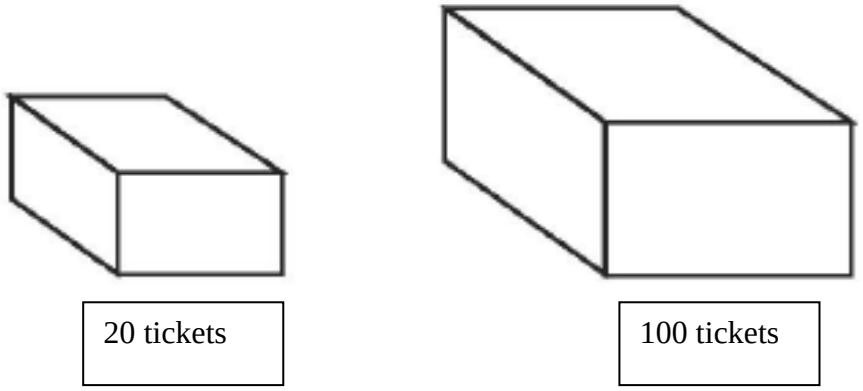
    c) Orange

    d) Red

14. The smaller box contains 20 tickets numbered from 1 to 20. The larger box contains 100 tickets numbered from 1 to 100.

Figure B. 4

Pre-Test Item 14 – Ticket Boxes



Without looking at them you can pick a ticket from either box. Which box would give you the greater chance of picking out a ticket with the number 17 on it?

a) The box with 20 tickets.

b) The box with 100 tickets.

c) Both boxes would give the same chance.

d) It is impossible to tell.

Sources:

Garfield, J., delMas, R., Chance, B. (2002). *Assessment Resources Tools for Improving Statistical Thinking (ARTIST)*. A National Science Foundation funded project (CCLI-ASA-0206571).

Mullis, I.V.S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L., and Smith, T. A. (1998). *Mathematics and Science Achievement in the Final Year of Secondary School: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.

International Association for the Evaluation of Educational Achievement. *TIMSS 1999: IEA's repeat of the Third International Mathematics and Science Study at the Eighth Grade: TIMSS mathematics items*. Chestnut Hill, MA: Boston College.

International Association for the Evaluation of Educational Achievement (2009). *TIMSS 2007 user guide for the international database*. TIMSS and PIRLS International Study Center, Lynch School of Education, Boston College.

International Association for the Evaluation of Educational Achievement (2007). *TIMSS 2003: Mathematics items released set: Eighth grade*. TIMSS and PIRLS International Study Center, Lynch School of Education, Boston College.

B.3 POST-TEST

1. George and Mike each bought one ticket for a lottery each week for the past 100 weeks. George has not won a single prize yet. Mike won 20 euros last week, for the first time. Who is more likely to win a prize this coming week if they each buy only one ticket?

   a) George

   b) Mike

   c) They have an equal chance of winning

2. Suppose you read on the back of a lottery ticket that the chances of winning a prize are 1 out of 10. Select the best interpretation.

   a) You will win at least once out of the next 10 times you buy a ticket.

   b) You will win exactly once out of the next 10 times you buy a ticket.

   c) You might win once out of the next 10 times but this will not happen for sure.

3. Melita is flipping a fair coin with her eyes shut. Nora records the outcome and places the coin again in Melita's hand. Heads has just come up 5 times in a row! The chance of getting heads on the next flip is

   a) less than the chance of getting tails since we are expected to get tails.

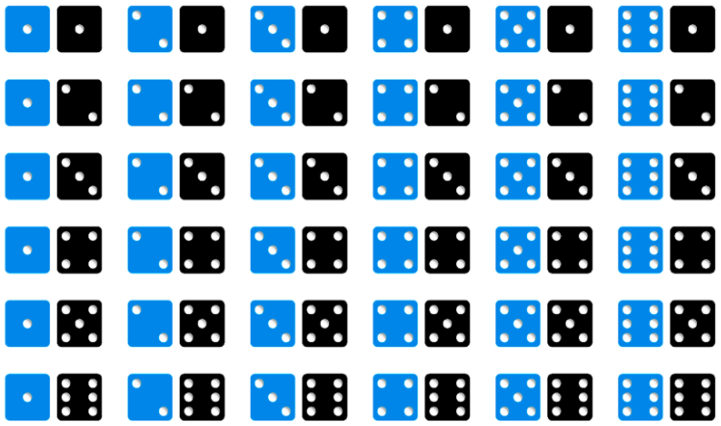
   b) equal to the chance of getting tails since the flips are independent and the coin is fair.

   c) greater than the chance of getting tails since heads seem to be coming up.

For problems 4 and 5 consider the following diagram which shows the 36 possible results when two six-sided dice are rolled.

Figure B. 5

Post-Test Items 4 and 5 – Sample Space for Rolling Two Fair Dice



4. You are about to roll 2 fair six-sided dice, hoping to get a double. (A double = both dice show the same value on top). Which double will occur the <u>least</u> often?

   a) 6 and 6

   b) 1 and 1

   c) 1 and 1, and, 6 and 6 are both least likely to occur.

   d) All doubles are equally likely.

5. Maria rolls two six-sided dice at the same time. Each side of each die is uniquely labeled with a number from 1 to 6. The following are two of the possible results that could occur when these two dice are rolled:

   Result 1: a 5 and a 6 are obtained in any order.

   Result 2: a 5 is obtained on each die.

   Which of the following statements is correct?

   a) The probability of obtaining each of these results is equal.

   b) There is a higher probability of obtaining Result 1 (a 5 and a 6 in any order).

   c) There is a higher probability of obtaining result 2 (a 5 on each die).

   d) It is impossible to give an answer.

6. A game company created a little plastic dog that can be tossed in the air. It can land either with all four feet on the ground, lying on its back, lying on its right side, or lying on its left side. However, the company does not know the probability of each of these outcomes. They want to estimate the probabilities. Which of the following methods is most appropriate?

   a) Since there are four possible outcomes, assign a probability of ¼ to each outcome.

   b) Toss the plastic dog many times and see what percent of the time each outcome occurs.

   c) Simulate the data using a model that has four equally likely outcomes.

7. A set of 24 cards is numbered with the positive integers from 1 to 24. The cards are shuffled and only one card is selected at random.

   What is the probability that the number on the card can be divided by 4 or 6?

   a) $\frac{1}{6}$

   b) $\frac{5}{24}$

   c) $\frac{1}{4}$

   d) $\frac{1}{3}$

   e) $\frac{5}{12}$

8. Two containers, labeled A and B, are filled with red and blue marbles according to the quantities listed in the table below. Each container is shaken several times and marbles are then selected from the containers at random, without looking.

Table B.5

Post-Test Item 8 – Marbles in Container

Container	A	B
Red	80	40
Blue	20	60

Which of the following outcomes has the <u>smallest</u> probability?

a) Obtaining a blue marble from container A.

b) Obtaining a blue marble from container A and a blue marble from B.

c) All of the above are equally likely.

9. Two containers, labeled A and B, are filled with red and blue marbles according to the quantities listed in the table below. Each container is shaken vigorously. After choosing one of the containers, you will reach in and, without looking, draw out a marble. If the marble is blue, you win 50 euros.

Table B.6

Post-Test Item 9 – Marbles in Container

Container	A	B
Red	6	60
Blue	4	40

Which container gives you the best chance of drawing a blue marble?

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

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

c) Equal chances from each container

10. One thousand people selected at random were questioned about smoking and drinking. The results of this survey are summarized in the table below.

Table B.7

Post-Test Item 10 Cross Tabulation

	Smokers	Non-smokers
Drinkers	320	530
Non-drinkers	20	130

a) $\frac{320}{340}$

b) $\frac{320}{850}$

c) $\frac{320}{1000}$

d) $\frac{1}{320}$

11. The eleven chips shown below are placed in a bag and mixed. Chelsea draws one chip from the bag without looking.

Figure B.6

Post-Test Item 11 Coin Figure

What is the probability that Chelsea draws a chip with a number that is a multiple of three?

a) $\frac{1}{11}$

b) $\frac{1}{3}$

c) $\frac{4}{11}$

d) $\frac{4}{7}$

12. Sophie has a bag in which there are 16 marbles: 8 are red and 8 are black marbles. She shakes the bag and without looking she draws 2 marbles from it and does not put them back. Both marbles that she has drawn are black. She then draws a third marble out of the bag.

What can you say about the likely color of this third marble?

a) It is more likely to be red than black.

b) It is more likely to be black than red.

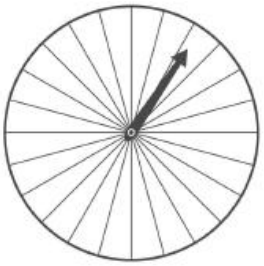
c) It is equally likely to be red or black.

d) You cannot tell if red or black is more likely.

13. The figure below shows a spinner with 24 sectors. When someone spins the arrow, it is equally likely to stop on any sector.

Figure B.7

Post-Test Item 13 - Spinner



3 of the sectors are blue, 1 is purple, 12 are orange, and 8 are red.

If a person spins the arrow, on which color sector is the spinner LEAST likely to stop?

a) Blue

b) Purple

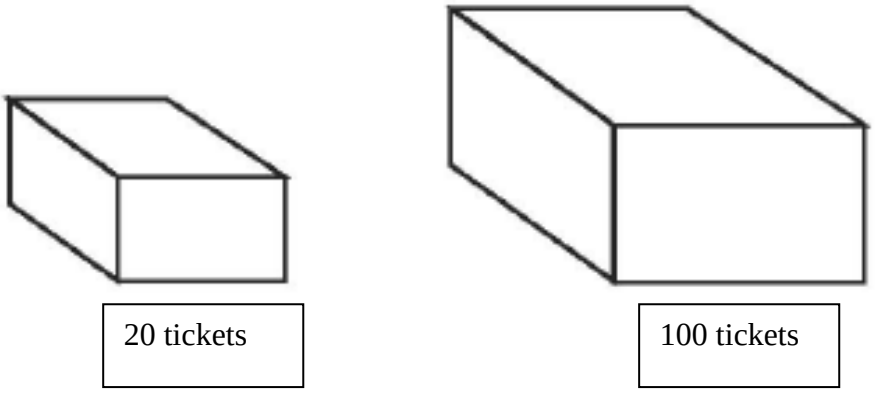
c) Orange

d) Red

14. The smaller box contains 20 tickets numbered from 1 to 20. The larger box contains 100 tickets numbered from 1 to 100.

Figure B. 8

Post-Test Item 14 – Ticket Boxes



Without looking at them you can pick a ticket from either box. Which box would give you the greater chance of picking out a ticket with the number 17 on it?

a) The box with 20 tickets.

b) The box with 100 tickets.

c) Both boxes would give the same chance.

d) It is impossible to tell.

15. A restaurant manager is considering a new location for her restaurant. She anticipates that the annual cash flow for the new location is:

Annual Cash Flow:	30,000	50,000	60,000	90,000	100,000
Probability:	0.05	0.15	0.30	0.40	???

The expected cash flow for the new location is:

a) 60,000

b) 73,000

c) 66,000

d) 14,600

16. An increasing number of employees are exploring the Internet for savings in business travel. A recent article reported on the results of a survey of 502 corporate travel managers. Suppose that a contingency table of whether employees research airline ticket prices and buy airline tickets on the internet revealed the following results:

Table B.8

Post-Test Item 16 Cross Tabulation

	Buy airline tickets on the internet		
Research airline ticket prices on the internet	Yes	No	Total
Yes	138	164	302
No	52	146	198
Total	190	310	500

a) Give an example of a simple event.

b) Give an example of a joint event.

c) Determine the probability that a manager selected at random:

   i. Buys airline tickets on the internet.

   ii. Researches airline ticket prices on the internet.

   iii. Researches airline ticket prices on the internet and buys airline tickets on the internet.

   iv. Researches airline ticket prices on the internet or buys airline tickets on the internet.

d) Calculate the probability that a randomly selected manager researches airline ticket prices on the internet given that he/she buys airline tickets on the internet.

e) Calculate the probability that a randomly selected manager buys airline tickets on the internet given that he/she researches airline ticket prices on the internet.

f) Determine whether researching airline ticket prices on the internet and buying airline tickets on the internet are independent.

17. The Cyprus Postal Service claims that it delivers packages on time 85% of the time. Consider a sample of 20 packages that need to be delivered by the Cyprus Postal Service.

a) Compute the probability that all of these 20 packages are delivered on time.

b) Compute the probability that at least 18 of these packages are delivered on time.

c) Compute the mean and standard deviation of this distribution.

APPENDIX C

Instructional Materials

C.1 ACTIVITY 1 (Treatment Group)

Activity Objectives:

i. Students make and justify predictions based on experimental and theoretical probabilities.

ii. Students conduct an experiment to determine experimental probabilities (relative frequencies).

iii. Students determine theoretical probabilities.

iv. Students compare experimental and theoretical probabilities.

v. Group experimental results are combined to illustrate the Law of Large Numbers.

Source:

McConnell, J. W., Brown, S., Usiskin, Z., Senk, S. L., Widerski, T., Anderson, S., Eddins, S., Feldman, C. H., Flanders, J., Hackworth, M., Hirschhorn, D., Polonsky, L., Sachs, L., and Woodward, E. (1998*). The University of Chicago School Mathematics Project: Algebra: Integrated Mathematics* (p. 368-369). Glenview, IL: Scott, Foresman and Company.

Activity 1: Relative Frequencies, Theoretical Probabilities & the Law of Large Numbers

Exploring the Sum of Two Dice

Instructions:

i. Please complete the following activity in groups as assigned by your instructor. Try to get everyone in your group involved as you work on the activity.

ii. Please look at the attached note regarding the role that each group member is assigned for the purposes of this activity. As you work on the activity, remember to carry out your role.

iii. Materials needed: Two six-sided dice and graph paper.

iv. If you have any questions regarding any part of the activity, you may ask any member of your group or your instructor for help.

v. Please keep the recorder ON at all times; speak clearly and close to the recorder.

vi. Once you finish, make sure you hand in your responses (including this handout and the graph paper) and the dice to your instructor. You only need to hand in one copy of your responses for the entire group.

Names of Group Members (First and Last Name):

_____

_____

_____

Read the instructions carefully as you work on the activity.

1. When two dice are rolled and the numbers on top are added, the sum can be any whole number from 2 to 12.
    i. Which sum do you believe is most likely to occur (i.e. would come up more often if you roll two dice)?

    ii. Why did you pick this number?

2. Your group's roller should roll the two dice you were given 50 times (Don't worry; it sounds a lot but it really isn't!!). The recorder should record the result in the column labeled "Frequency" in the table below. The checker should check that the result recorded is correct. At this time, only fill in Column 2 of the table below.

Table C.1

Activity 1 – Recording Results of Sum of Two Dice

Column 1 Sum	**Column 2 Frequency**	**Column 3 My Group's Relative Frequency**		**Column 4 Probability (Fraction, decimal or percentage)**	**Column 5 Class Relative Frequency**
		Fraction	**Decimal**		
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					

3. Construct a frequency histogram using your group's results from the 50 dice rolls as you indicated them in the table above. Use the graph paper provided or the space below.

4. After you have rolled the dice 50 times, calculate your group's relative frequency (experimental probability) of getting each sum. Write each relative frequency as both a fraction and a decimal in Column 3 in the table on the previous page.

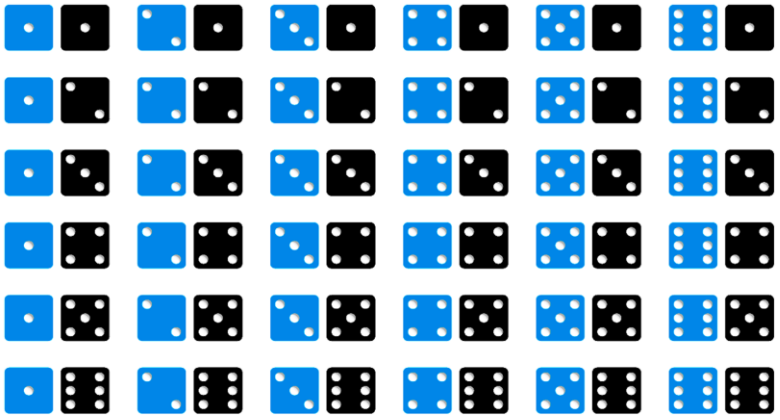
5. Once you are done with 4 above, have one of your group members go to the board and record your group's relative frequencies in the column corresponding to your group's number.

6. Based on your table on page 2, how, if at all, would you revise your prediction in part 1? What patterns do you notice?

7. The diagram below shows the 36 possible outcomes when two dice are rolled.

Figure C.1

Activity 1 – Sample Space for Rolling Two Fair Dice



Notice that there is only 1 way to get a sum of 2. We say that the *theoretical probability* of getting a sum of 2 is 1/36 ≈ 0,028. In contrast, there are 3 ways of getting a sum of 10 (6 and 4, 5 and 5, 4 and 6), so the probability of getting a sum of 10 is 3/36 ≈ 0,083.

i. Calculate the probability of getting each of the other sums from 2 to 12, and record these numbers (as a fraction, decimal or percentage) in Column 4 labeled "Probability" in part 2.

ii. How close are your group's relative frequencies (experimental probabilities) to the theoretical probabilities you just computed? i.e. how close are your answers in Column 3 to your answers in Column 4 on the table in part 2?

8. i. Combine your results with the results of the other groups in your class which are written on the board. Calculate the relative frequency (experimental probability) for each sum from 2 to 12 for the entire class. Fill in Column 5 labeled "Class relative frequency" on the table in part 2.

ii. Which set of relative frequencies – those from your small group, or those from the whole class – are closer to the theoretical probabilities? i.e. Which are closer to the results in Column 4: the results in Column 3 or the results in Column 5?

iii. Use the combined results of all the groups in the class to create another frequency histogram. You may use the graph paper provided or the space below.

iv. Compare the shape of this histogram to the histogram you constructed in #3 in part 2.

Is one graph more symmetrical than the other?

If yes, which one is more symmetrical: the histogram for your group's data or the histogram for the whole class data?

C.2 ACTIVITY 2 (Treatment Group)

Activity Objectives:

i. Conduct an experiment to determine experimental probabilities (relative frequencies).

ii. Construct frequency histograms.

iii. Students compare experimental and theoretical conditional probabilities.

iv. Group experimental results are combined to illustrate the Law of Large Numbers.

v. Explore a situation that relates to conditional probability and independence of events.

vi. Determine sample spaces.

vii. Find the probability of simple events and impossible events.

viii. Apply the multiplication principle for independent events to a sample space.

Source: Irini Papaieronymou

Activity 2 – Conditional Probability and Independence

Exploring the Sum of Three Dice

Instructions:

i. Please complete the following activity in groups as assigned by your instructor.

ii. Materials needed: Three six-sided dice of different colors.

iii. If you have any questions regarding any part of the activity, you may ask any member of your group or your instructor for help.

iv. Respond to all the questions asked on this handout in the best way you can and use this handout to fill in your responses.

v. Please keep the recorder ON at all times; speak clearly and close to the recorder.

Names of Group Members and Role:

(roller – rolls the dice) _____

(recorder – writes down the answers) _____

(checker – checks that answers are recorded correctly) _____

Today you will explore the sum of three dice by rolling a set of three fair six-sided dice. You will first perform an experiment and then will use your results to answer questions regarding *conditional probability and independence*. We will study conditional probability and independence in more detail next time in class.

1. Roller: Roll the three dice simultaneously (at the same time) 50 times.

   Recorder: Record the result showing on each die and the sum of the three dice on each roll in the table below.

   Checker: Check that the answers recorded are correct.

   Table C.2

   Activity 2 – Results of Rolling Three Fair Dice

Roll	**Result**			**Sum of the three dice**
	Red Die	**Green Die**	**Purple Die**	
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				

Table C.2 (cont'd)

31				
32				
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				
48				
49				
50				

2. Go back to your results of the set of 50 three-dice rolls in the table on the previous page.

   a. Based on your results, what is the relative frequency that one of the dice comes up 3?

   b. Based on your results, if the sum is 6, what is the relative frequency that one of the dice is 3?

   c. Based on your answers in 2a and 2b, would you say that the relative frequency of one of the dice coming up 3 is affected by knowledge that the sum is 6?
   (Hint: Is your answer in 2a equal to your answer in 2b?)

Note: The relative frequency you have computed in part 2b is based on the condition that one of the dice resulted in 3. Relative frequencies that are based on certain pre-set conditions can also be determined theoretically using *conditional probability*.

Let us begin to determine the relative frequency computed in part 2 ***theoretically***.

3.a. How many outcomes are possible when three dice are rolled?

b. What are the different ways in which three dice can give you a sum of 6?

c.What is the probability that you will get a sum of 6 when three dice are rolled?

4. Theoretically, the ***conditional probability*** **of an event A given an event B** is as follows:

$$\text{P(event A given event B)} = \frac{\text{P(A and B)}}{\text{P(B)}}$$

$$\text{or symbolically, } P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad \text{(Formula 2)}$$

a. Use Formula 2 above to find the theoretical probability that, when three fair dice are rolled, one of the dice comes up 3 given that the sum is 6.

Event A =

Event B =

P(A | B) =

b. How close is your answer in 4a to your answer in 2b?

In addition to computing conditional probabilities theoretically, we can also use probability theory to determine whether the probability of one event affects the probability of another event.

If the probability of one event *affects* the probability of another event, then the two events are said to be *dependent*.

If the probability of one event *does not affect* the probability of another event, then the two events are said to be *independent*.

Theoretically, statistical independence of two events can be computed as follows:

Let A and B be two events. Then, A and B are independent if and only if $P(A|B) = P(A)$

(Formula 3)

With regards to this activity:

Let Event A = one of the dice comes up 3 and Event B = get a sum of 6

5.a.Use Formula 3 on the previous page to determine whether Events A and B are independent. Show your work.

b. Does your conclusion in 5a agree with your answer to 2c?

C.3 ACTIVITY 3 (Treatment Group)

Activity Objectives:

i. Students make and justify predictions based on experimental and theoretical probabilities.

ii. Help students overcome probabilistic misconceptions: equiprobability and outcome orientation.

iii. Students study the probability distribution of a random variable.

iv. Students study sample spaces.

Source: Khazanov, L. (2008). Addressing students' misconceptions about probability during the first years of college. *Mathematics and Computer Education*, 42(3), 180-192.

The following is the original activity as described by Khazanov:

1. Instructor explains the rules of the game. Numbers 1 through 15 are written on a number line. Participants have 36 (or 48) chips to place above these numbers. The instructor or a designated student repeatedly rolls two dice. The sum of the numbers showing is calculated. If a student has a chip above the number on the line that matches the sum of the numbers on the dice, then she removes any chips. The winner is the person or the group that has all their chips removed ahead of any other group or person.
2. Once the game is over, engage students in a discussion. Ask them to provide a rationale for their distributions. Then ask which distribution won, and why. Some students could have chosen a uniform distribution from 2 to 12 (equiprobability bias), others might have placed all of their chips above 7 (outcome orientation misconception). It is important for these students to understand why they made poor choices. However, it is equally important to emphasize that some of the students who made poor choices stood a chance winning. They

were not destined to lose, but had a smaller chance of winning than those who placed more chips towards the center of the distribution. Students' understanding may be reinforced by constructing a 6*6 table listing 36 equally likely outcomes when two dice are tossed. Ask your students to compute the sum for each entry and observe that while the possible values for the sums are 2 to 12, the number of times these numbers occur is different.

3. A good additional question: what mistake could one make to totally eliminate one's chance of winning (make the probability of winning equal to 0)? The mistake, of course, is placing a chip over 1 or any number greater than 12.

Activity 3: Discrete Probability Distributions

A Game of Chance: Where Should The Chips Be Placed?

Instructions:

i. Please complete the following activity in groups as assigned by your instructor.

ii. Materials needed: Two dice of different colors, 36 chips of three different colors, and a poster on which a number line is drawn.

iii. Remember to keep the recorder on during the entire activity and to talk close to it.

iv. If you have any questions regarding any part of the activity, you may ask any member of your group or your instructor for help.

v. Please respond to all the questions asked on this handout in the best way you can and use this handout to fill in your responses.

Names of Group Members:

(roller: rolls the dice) _____

(recorder: writes down the question) _____

(checker: checks that the answers are correct) _____

Today you will play a game that relates to the sum of two dice. In this game, you will each pick a color of chips (12 chips for each person), place your chips over any numbers on the number line provided on the poster, throw a pair of dice, and after each roll, remove any chips you have over the number that corresponds to the sum of the two dice. The group member who has his/her chips removed first wins the game.

1. In your groups, pick one of the available colors of chips and gather all 12 chips of that color so that you can use them in this activity. Each group member should have 12 chips of the same color.
2. On the separate poster paper given to you by your instructor, a number line is provided with the numbers 1 through 15 written on it.
3. Place your 12 chips above any of the numbers on the number line. Distribute all of your chips in any way you want on the number line. You may place all of them over one number or you may spread them out over various numbers.
4. Before you start playing the game, fill in the table in part 5 and respond to question 6.

5. In the following table, indicate how many chips each member of your group has placed above each of the numbers on the number line.

Table C.3

Activity 3 – Placement of Chips on Number Line

Number on the Number Line	**Student Name** _____	**Student Name** _____	**Student Name** _____
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			

6. Provide an explanation for the way that each group member placed their chips on the number line.

Student Name and Explanation:

Student Name and Explanation:

Student Name and Explanation:

*Before you begin to play the game, read question 7 carefully.

7. Roller: Roll the two dice repeatedly at the same time and call out their sum so that the recorder can record this down on the table provided below.

   Recorder: In the table below, indicate the sum of the numbers on the dice on each roll.

   Checker: Check that the results recorded are correct.

   As the two dice are rolled, if you have any chips above the number on the line that matches the sum of the numbers on the two dice just rolled, then remove <u>one</u> of these chips. Continue rolling the dice, each time removing chips as indicated above and recording the sum in the table below. You may add more rows to the table below if necessary. Stop filling in the table below once a group member has all of his/her chips removed that is, once you have a winner!

Table C.4

Activity 3 – Recording Result of Rolling Two Fair Dice

Roll	Sum
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	

8. How many times did you roll the dice before you had a winner in your group? That is, on which roll did someone in your group have all of his/her chips removed? _____

9. Who won the game in your group? _____

10. Why do you think this group member won the game?

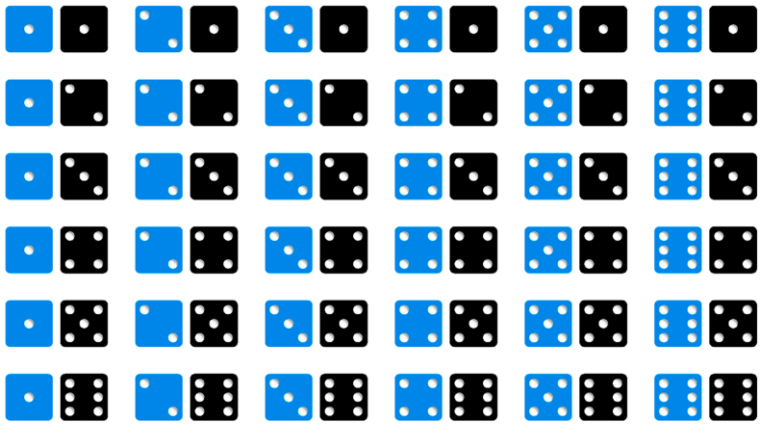
11. Were the other two group members destined to lose? That is, did they have NO chance of winning? Explain.

12. What mistake could you make to totally eliminate your chance of winning? That is, what mistake could you do to make your probability of winning equal to 0?

For the following questions on this activity, use the diagram provided below which shows the possible outcomes when two dice are rolled.

Figure C.2

Activity 3 – Sample Space for Rolling Two Fair Dice



13. How many possible outcomes are there when rolling two fair dice? _____

14. Which are the possible sums when two fair dice are rolled?

15. How is the number line connected to the sums you listed in question 14) above?

16. Given the list of possible sums that you have identified in 14) above, which numbers on the number line would eliminate one's chance of winning? i.e. which numbers would make the chance of winning 0?

17. In the table below, the left column indicates the possible sums when rolling two fair dice. Fill in the right column in the table, to indicate the theoretical probability for each of these sums. The two columns of this table taken together make up the *discrete probability distribution* for the sum of two fair dice. We will study discrete probability distributions next time in class (chapter 5)

    Table C.5

    Activity 3 – Probability Distribution for Sum of Two Dice

Sum of the numbers on two fair dice	Probability
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	

18. What is the sum of the values of the probability column in the table above?

19. Based on the theoretical probabilities that you have computed in 17) above, which sums give the best chance of winning the game you have just played in your groups?

C.4 ACTIVITY 4 (Treatment Group)

Activity Objectives:

i. Explore a binomial distribution problem.

ii. pproach a binomial probability problem through an investigation of relative frequencies and through an analysis of outcomes.

iii. Compare the experimental to the theoretical approach of solving a binomial distribution problem.

Source: Shaughnessy, M. J., Barrett, G., Billstein, R., Kranendonk, H. A., and Peck, R. (2004). What's the probability of a hit? In Lott, J. W., and House, P. A. (Eds.), *Navigating through probability in grades 9-12*, p. 40-42 and p. 94-97. Reston, VA: The National Council of Teachers of Mathematics, Inc.

Modifications done to original activity:

i) Context was changed. The original activity concerned a baseball player. Since baseball is not popular in Cyprus, the context was changed to be about a football player (football is the most popular type of sports game in Cyprus). So, the activity is about football goals not baseball hits, as was the case in the original.

ii) In the original activity, the baseball player had a batting average of 0.4. In the modified activity, the football player has a scoring average of 1/6. This was changed so that in later parts of the activity, the students can use a die to generate data.

iii) In the original activity, the students were asked to perform two sets of simulations: a set of 20 simulations and then a set of another 50 simulations. Due to time constraints, I believe that performing both of these sets of simulations would make the activity last longer than two 50

minute class periods but the goal is for all activities used to take up to two periods. So, in the modified version of the activity, the students are asked to do a set of 50 simulations only in their groups. After that, they will use the results of all groups which the instructor will record on the board instead of performing more simulations in their groups.

iv) Part 5)b) was added.

Activity4 – Binomial Distribution

What's the Probability of Making a Basket?

Instructions:

i. Please complete the following activity in groups as assigned by your instructor.

ii. Materials needed: a table of random numbers.

iii. Remember to keep the recorder on at all times (do NOT pause or stop it during the activity). Speak clearly and close to the recorder.

iv. If you have any questions regarding any part of the activity, you may ask any member of your group or your instructor for help.

v. Please respond to all the questions asked on this handout in the best way you can and use this handout to fill in your responses.

Names of Group Members:

(recorder: writes down the answers) _____

(asks questions) _____

(checker) _____

Suppose that a basketball player has a free-throw average of 70%. Thus, any time he attempts a free throw during a basketball game, this player has a constant probability of making the basket, or P(B), of 70% and a constant probability of not making the basket, or P(N), of 30%.

1)a) Let B = event that the player makes the basket

Let N = event that the player does not make the basket

Using the letters B and N construct the sample space (the set of all possible outcomes/ combinations) for three free-throw attempts.

b) Which outcome(s) in part 1a include at least 2 baskets?

2)a) Compute the probabilities (as a percentage) of each of the individual outcomes in your sample space in part 1a. Remember: the player's probability of making the basket is P(B) = 70% and his probability of not making the basket is P(N) = 30%.

2b) Use the probabilities in 2a above to calculate the theoretical probability that the basketball player would make <u>at least 2 baskets</u> in three free-throw attempts.

3) a) Use the table of random numbers given to you by your instructor to indicate data on 50 sets of three free-throw attempts for the player with a 70% free-throw average. Fill in the table below.

Table C.6

Activity 4 – Results of Free-Throw Basketball Attempts

Outcome (three free-throw attempt)	Number of baskets made

Table C.6 (cont'd)

3)b) On the basis of your 50 sets of simulated data for three free-throw attempts, what is the probability that the player will make <u>at least 2 baskets</u> in three free-throw attempts?

c) How does your result from 3)b) compare with your calculation in question 2b?

4) Turn to one of the other groups of students close to your group.

Name the students in that group:

_____

a) Compare the results of your simulations of 50 sets of three free-throw attempts with those of the other group. How close is your probability for <u>at least 2 baskets</u> in three free-throw attempts to the probability of the other group?

b) If the probabilities differ why do you think this is the case?

5) On the board, your instructor has recorded the results of at least two baskets in 50 sets of three free-throw attempts of all groups. Your group's results have also been recorded. Combine all class results to answer the following:

   a) Based on the results from all groups, what is the probability that the player makes at least 2 baskets in three free-throw attempts?

   b) How does your calculation in 5a) compare with your calculation in question 2b?

   c) How does your answer to 5a) compare with your group's simulated results of 50 sets of three free-throw attempts?

The probabilities that you have found in this activity can also be derived using probability theory. The ideas behind this activity are related to what is called a *binomial probability distribution*. The binomial probability distribution involves the following formula:

$$P(X \geq x) = \binom{n}{x} \cdot p^{x} \cdot (1-p)^{n-x}$$

where n is the number of observations, p is the probability of success (i.e. making the basket) and x is the number of baskets made.

So, in this activity, theoretically, the probability of the player making zero baskets in three free-throw attempts is:

$$P(X = 0) = \binom{3}{0} \cdot \left(\frac{4}{6}\right)^{0} \cdot \left(\frac{2}{6}\right)^{3} = 1 \cdot 1 \cdot \left(\frac{2}{6}\right)^{3} = \frac{8}{216} = \frac{1}{27}$$

6) Using the formula for the binomial distribution above:

a) Compute the probability that the player makes 2 baskets in three free-throw attempts.

b) Compute the probability that the player makes 3 baskets in three free-throw attempts.

c) Compute the probability that the player makes at least 2 baskets in three free-throw attempts.

C. 5 IN-CLASS PROBLEMS ASSIGNED TO CONTROL GROUP / GROUPWORK

Problems From the Course Textbook:

Levine, D. M., Krehbiel, T. C., and Berenson, M. L. (2010). *Business statistics: A first course (5th ed.)*. Upper Saddle River, NJ: Pearson Education, Inc.

p. 157-158

4.2 An urn contains 10 red balls and 8 green balls. One ball is to be selected from the urn.

a. Give an example of a simple event.

b. What is the complement of a red ball?

4.8 According to an Ipsos poll, the perception of unfairness in the U.S. tax code is spread fairly evenly across income groups, age groups, and education levels. In an April 2006 survey of 1,005 adults, Ipsos reported that almost 60% of all people said the code is unfair, whereas slightly more than 60% of those making more than $50,000 viewed the code as unfair ("People Cry Unfairness", *The Cincinnati Enquirer*, April 16, 2006, p. A8). Suppose that the following contingency table represents the specific breakdown of responses:

Table C.7

Problem 4.8 Cross Tabulation

Income Level			
U.S. Tax Code	Less than $50,000	More than $50,000	Total
Fair	225	180	405
Unfair	280	320	600
Total	505	500	1,005

a) Give an example of a simple event.

b) Give an example of a joint event.

c) What is the complement of "tax code is fair"?

d) Why is "tax code is fair *and* makes less than $50,000" a joint event?

4.9 Referring to the contingency table is Problem 4.8, if a respondent is selected at random, what is the probability that he or she

a. thinks the tax code is unfair?

b. thinks the tax code is unfair and makes more than $50,000?

c. thinks the tax code is unfair or makes more than $50,000?

d. Explain the difference in the results in (b) and (c).

p. 165

4.23 According to an Ipsos poll, the perception of unfairness in the U.S. tax code is spread fairly evenly across income groups, age groups, and education levels. In an April 2006 survey of 1,005 adults, Ipsos reported that almost 60% of all people said the code is unfair, whereas slightly more than 60% of those making more than $50,000 viewed the code as unfair ("People Cry Unfairness", *The Cincinnati Enquirer*, April 16, 2006, p. A8). Suppose that the following contingency table represents the specific breakdown of responses:

Table C.8

Problem 4.23 Cross Tabulation

Income Level			
U.S. Tax Code	Less than $50,000	More than $50,000	Total
Fair	225	180	405
Unfair	280	320	600
Total	505	500	1,005

a. Given that a respondent earns less than $50,000, what is the probability that he or she said that the tax code is fair?

b. Given that a respondent earns more than $50,000, what is the probability that he or she said that the tax code is fair?

c. Is income level independent of attitude about whether the tax code is fair? Explain.

p. 183

5.3 The manager of a large computer network has developed the following probability distribution of the number of interruptions per day:

Table C.9

Problem 5.3 Probability Distribution

Interruptions (X)	P(X)
0	0.31
1	0.34
2	0.18
3	0.10
4	0.04
5	0.02
6	0.01

Compute the expected number of interruptions per day.

5.4 In the carnival game Under-or-Over-Seven, a pair of fair dice is rolled once, and the resulting sum determines whether the player wins or loses his or her bet. For example, the player can bet $1 that the sum will be under 7 – that is, 2, 3, 4, 5, or 6. For this bet, the player wins $1 if the result is under 7 and loses $1 if the outcome equals or is greater than 7. Similarly, the player can bet $1 that the sum will be over 7 – that is, 8, 9, 10, 11 or 12. Here, the player wins $1 if the result is over 7 but loses $1 if the result is 7 or under. A third method of play is to bet $1 on the outcome 7. For this bet, the player wins $4 if the result of the roll is 7 and loses $1 otherwise.

a. Construct the probability distribution representing the different outcomes that are possible for a $1 bet on under 7.
b. Construct the probability distribution representing the different outcomes that are possible for a $1 bet on over 7.

c. Construct the probability distribution representing the different outcomes that are possible for a $1 bet on 7.

d. Show that the expected long-run profit (or loss) to the player is the same, no matter which method of play is used.

p. 190

5.13 When a customer places an order with Rudy's On-Line Office Supplies, a computerized accounting information system (AIS) automatically checks to see if the customer has exceeded his or her credit limit. Past records indicate that the probability of customers exceeding their credit limit is 0.05. Suppose that, on a given day, 20 customers place orders. Assume that the number of customers that the AIS detects as having exceeded their credit limit is distributed as a binomial random variable.

a. What are the mean and standard deviation of the number of customers exceeding their credit limits?

b. What is the probability that 0 customers will exceed their limits?

c. What is the probability that 1 customer will exceed his or her limit?

d. What is the probability that 2 or more customers will exceed their limits?

APPENDIX D

SCORING RUBRICS FOR POST-TEST OPEN-ENDED ITEMS

Table D.1

Scoring Rubric for Post-Test Item 16a

Code/Score	Response
1	**Correct Response:**
	Student specifies one of the following: i)Employee buys airline tickets on the internet ii)Employee does not buy airline tickets on the internet iii)Employee researchers airline ticket prices on the internet iv) Employee does not research airline ticket prices on the internet
0	**Incorrect Response:**
	Student provides a response different from any of the four correct statements provided above.
999	**Non-response:**
	Blank or incomplete sentence provided.

Table D.2

Scoring Rubric for Post-Test Item 16b

Code/Score	Response
1	**Correct Response:**
	Student specifies one of the following: i)Employee researches airline ticket prices on the internet and buys airline tickets on the internet. ii) Employee researches airline ticket prices on the internet and does not buy airline tickets on the internet. iii) Employee does not research airline ticket prices on the internet and buys airline tickets on the internet. Iv)Employee does not research airline ticket prices on the internet and does not buy airline tickets on the internet.
0	**Incorrect Response:**
	Student provides a response different from any of the four correct statements provided above.
999	**Non-response:**
	Blank or incomplete sentence provided.

Table D.3

Scoring Rubric for Post-Test Items 16c) i), 16c) ii) and 16c) iii)

Code/Score	Response
2	**Correct Response:**
	For part i) 190/500 or 19/50 or 0.38 or 38% For part ii) 302/500 or 0.604 or 60.4% For part iii) 138/500 or 0.276 or 27.6%
1	**Partially Correct Response:**
	Either the numerator or the denominator in the fraction provided is correct (but not both).
0	**Incorrect Response:**
	Neither the numerator nor the denominator in the fraction provided is correct or an incorrect decimal or percent is provided.
999	**Non-response:**
	Blank

Table D.4

Scoring Rubric for Post-Test Item 16c) iv)

Code/Score	Response
2	**Correct Response:**
	Student correctly uses the union rule for finding the probability of events and arrives at the correct answer: 302/500 + 190/500 – 138/500 = 352/500 or 0.704 or 70.4%
1	**Partially Correct Response:**
	Student either i)states the union rule correctly but does not carry out the computations or ii) states the union rule correctly but some computations are incorrect or iii)specifies the correct computation to be carried out i.e. 302/500 + 190/500 – 138/500 but does not provide a final answer or iv)provides a correct final answer but does not show any computations or v)provides incorrect intermediate computations
0	**Incorrect Response:**
	The answer provided is completely incorrect.
999	**Non-response:**
	Blank

Table D.5

Scoring Rubric for Post-Test Items 16d and 16e

Code/Score	Response
2	**Correct Response:**
	Student correctly identifies a conditional probability: For part d: 138/190 For part e: 138/302 The student may have used the rule for conditional probability correctly or might have correctly used the necessary cells from the table.
1	**Partially Correct Response:**
	Student either i)indicates that a conditional probability should be used and/or states the rule for conditional probability correctly but does not carry out the computations or ii)specifies the correct computation to be carried out i.e. (138/500) / (190/500) but does not provide a final answer or iii)provides a correct final answer but does not show any computations or provides incorrect intermediate computations or iv)provides the reverse fraction i.e. 190/138 (part d) or 302/138 (part e) or v) provides a fraction in which either the numerator or the denominator is correct (but not both)
0	**Incorrect Response:**
	The answer provided is completely incorrect.
999	**Non-response:**
	Blank

Table D.6

Scoring Rubric for Post-test Item 16f

Code/Score	Response
2	**Correct Response:**
	Student correctly i)indicates that the two events are not independent: and ii) shows this through computations that make use of conditional probability
1	**Partially Correct Response:**
	Student i)states that the two events are not independent but does not show any computations or provides incorrect computations or does not provide any reasoning for this conclusion or ii) student carries out correct computations that relate to the independence of events but does not indicate whether the events are independent or not or iii) student carries out correct computations that relate to the independence of events but incorrectly indicates that the two events are independent
0	**Incorrect Response:**
	Student indicates that the two events are independent and/or provides incorrect computations/reasoning when arriving at this conclusion.
999	**Non-response:**
	Blank

Table D.7

Scoring Rubric for Post-Test Item 17a

Code/Score	Response
2	**Correct Response:**
	$0.85^{20} = 0.039 = 3.9\%$
1	**Partially Correct Response:**
	Either the base or the exponent is incorrect Or Student provides correct computations but incorrect final answer
0	**Incorrect Response:**
	Incorrect computations and incorrect final answer.
999	**Non-response:**
	Blank

Table D.8

Scoring Rubric for Post-Test Item 17b

Code/Score	Response
2	**Correct Response:**
	i)Student recognizes that the binomial distribution should be used and that $P(X \geq 18)$ should be computed. and ii)Student carries out the computations correctly using the formula for finding the probability of a binomial distribution and iii)Student arrives at the correct final answer of 40.6%
1	**Partially Correct Response:**
	i)Student recognizes that the binomial distribution should be used and that $P(X \geq 18)$ should be computed but does not carry out any computations or the computations are incorrect or ii)Student carries out the computations correctly using the formula for finding the probability of a binomial distribution but does not arrive at a final answer or iii)Student only provides the final answer of 40.6% or iv)Student computes $P(X = 18)$ or $P(X \leq 18)$
0	**Incorrect Response:**
	Incorrect computations and incorrect final answer.
999	**Non-response**
	Blank

Table D.9

Scoring Rubric for Post-Test Item 17c

Code/Score	Response
2	**Correct Response:**
	Student correctly computes i) The mean i.e. 20 x 0.85 = 17 and ii) The variance i.e. 20 x 0.85 x 0.15 = 2.55
1	**Partially Correct Response:**
	i)Student computes either the mean or the variance correctly but not both. or ii)Student shows the computations for the mean and the variance but does not provide a final answer for either or iii)Student provides a correct final answer for both the mean and the variance but does not show any computations.
0	**Incorrect Response:**
	Incorrect computations and incorrect final answer.
999	**Non-response**
	Blank

Table D.10

Scoring Rubric for Post-Test Item 18a

Code/Score	Response
1	**Correct Response:**
	Student provides one of the following simple events: i) Textbook published is a success ii) Textbook published is break-even iii) Textbook published is a loser iv) Textbook received favorable reviews
0	**Incorrect Response:**
	Incorrect simple event provided.
999	**Non-response:**
	Blank

Table D.11

Scoring Rubric for Post-Test Items 18b, 18c, and 18d

Code/Score	Response
2	**Correct Response:**
	Student recognizes that conditional probability should be used and provides the correct answer: For part b: 80% For part c: 60% For part d: 20%
1	**Partially Correct Response:**
	Student recognizes that conditional probability should be used but does not provide a correct numerical final answer.
0	**Incorrect Response:**
	Student does not recognize that conditional probability should be used and does not provide a correct answer.
999	**Non-response:**
	Blank

Table D.12

Scoring Rubric for Post-Test Item 18e

Code/Score	Response
2	**Correct Response:**
	Student recognizes that Bayes' Theorem should be used and correctly applies the theorem to arrive at the correct final answer: P(Success|Favorable Reviews) = (0.8x0.3)/[(0.8x0.3)+(0.6x0.5)+(0.2x0.2)] = 0.24 / 0.58 = 0.414
1	**Partially Correct Response:**
	Student recognizes that Bayes' Theorem should be used and attempts to use it but not all computations are correct Or Student recognizes that Bayes' Theorem should be used, correctly uses it but arrives at an incorrect final answer.
0	**Incorrect Response:**
	Student recognizes that Bayes' Theorem should be used but computations and final answer are incorrect.
999	**Non-response:**
	Blank

APPENDIX E: STATISTICAL RESULTS

Table E.1

Percent-Correct Responses on Pre-Test and Post-Test Multiple-Choice Items

	Percent-Correct Responses							
	Control (Group B; morning) N = 17		Treatment1 (Group A; morning) N= 20		Treatment2 (Group C; evening) N = 7		Treatment1 & Treatment2 N = 27	
Item	PreTest	PostTest	PreTest	PostTest	PreTest	PostTest	PreTest	PostTest
1	94.1	76.5	80	85	85.7	100	81.5	88.9
2	70.6	70.6	95	85	85.7	71.4	92.6	81.5
3	70.6	64.7	75	90	42.9	100	66.7	92.6
4	100	47.1	90	80	85.7	85.7	88.9	81.5
5	23.5	29.4	30	55	28.6	42.9	29.6	51.9
6	23.5	11.8	45	30	14.3	42.9	37	33.3
7	11.8	17.6	50	55	14.3	71.4	40.7	59.3
8	17.6	5.9	15	15	14.3	14.3	14.8	14.8
9	52.9	52.9	85	70	42.9	71.4	74.1	70.4
10	70.6	76.5	40	95	28.6	57.1	37	85.2
11	94.1	35.3	95	90	57.1	85.7	85.2	88.9
12	58.8	52.9	95	95	85.7	85.7	92.6	92.6
13	88.2	47.1	80	75	71.4	71.4	77.8	74.1
14	88.2	58.8	80	75	71.4	85.7	77.8	77.8
15		23.5		40		71.4		48.1

Table E.2

Results of Distractor Analysis of Multiple-Choice Items (Correct Response in bold)

Item	Distractor Or Correct Choice	Control (Group B; morning) N = 17		Treatment1 & Treatment2 N = 27	
		Pre-Test (%)	Post-Test (%)	Pre-Test (%)	Post-Test (%)
1	A	5.9	17.6	3.7	3.7
	B	0	0	14.8	7.4
	C	**94.1**	**76.5**	**81.5**	**88.9**
2	A	23.5	11.8	7.4	3.7
	B	5.9	11.8	0	7.4
	C	**70.6**	**70.6**	**92.6**	**81.5**
3	A	17.6	11.8	18.5	0
	B	**70.6**	**64.7**	**66.7**	**92.6**
	C	11.8	17.6	7.4	3.7
4	A	0	11.8	0	3.7
	B	0	0	0	0
	C	0	41.2	7.4	14.8
	D	**100**	**47.1**	**88.9**	**81.5**
5	A	52.9	29.4	44.4	29.6
	B	**23.5**	**29.4**	**29.6**	**51.9**
	C	0	23.5	11.1	3.7
	D	23.5	17.6	7.4	14.8
6	A	58.8	35.3	44.4	51.9
	B	**23.5**	**11.8**	**37**	**33.3**
	C	17.6	47.1	14.8	11.1
7	A	11.8	29.4	7.4	11.1
	B	29.4	29.4	14.8	7.4
	C	11.8	0	0	0
	D	**11.8**	**17.6**	**40.7**	**59.3**
	E	23.5	17.6	18.5	22.2
8	A	64.7	82.4	70.4	81.5
	B	**17.6**	**5.9**	**14.8**	**14.8**
	C	11.8	5.9	11.1	3.7
9	A	23.5	23.5	11.1	11.1
	B	17.6	17.6	11.1	14.8
	C	**52.9**	**52.9**	**74.1**	**70.4**
10	A	5.9	11.8	0	7.4
	B	5.9	5.9	0	3.7
	C	**70.6**	**76.5**	**37**	**85.2**
	D	0	0	11.1	3.7

Table E.2 (cont'd)

11	A	0	47.1	0	7.4
	B	5.9	11.8	11.1	0
	C	**94.1**	**35.3**	**85.2**	**88.9**
	D	0	5.9	0	3.7
12	**A**	**58.8**	**52.9**	**92.6**	**92.6**
	B	5.9	11.8	3.7	0
	C	29.4	23.5	0	3.7
	D	5.9	11.8	0	3.7
13	A	5.9	0	3.7	0
	B	**88.2**	**47.1**	**77.8**	**74.1**
	C	5.9	35.3	14.8	25.9
	D	0	11.8	0	0
14	**A**	**88.2**	**58.8**	**77.8**	**77.8**
	B	0	5.9	3.7	0
	C	0	35.3	7.4	22.2
	D	5.9	0	7.4	0
15	A		29.4		29.6
	B		**23.5**		**48.1**
	C		11.8		18.5
	D		35.3		3.7

BIBLIOGRAPHY

BIBLIOGRAPHY

Aczel, A. D. and Sounderpandian, J. (2002). *Complete business statistics 5th ed*. New York, NY: McGraw-Hill Companies, Inc.

Albert, J. H. (2003). College students' conceptions of probability. *The American Statistician*, 57(1), 37-45.

Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, B., Utts, J., Velleman, R., and Witmer, J. (2005). *Guidelines for assessment and instruction in statistics education (GAISE) college report*. Retrieved April 27, 2008, from http://www.amstat.org/education/gaise/

Amit, M. and Jan, I. (2006). Autodidactic learning of probability concepts through games. In Novotná, J., Moraová, H., Krátká, M. and Stehlíková, N. (Eds.). *Proceedings of the 30th Conference of the International Group for the Psychology of Mathematics Education*, 2, 49-56. Prague: PME.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89, 369-406.

Antoine, W. (2000). *An exploration of the misconceptions and incorrect strategies liberal arts students use when studying probability*. Doctoral dissertation, Columbia University.

Aquilonius, B. C. (2005). *How do college students reason about hypothesis testing in introductory statistics courses?* Doctoral dissertation, University of California, Santa Barbara, CA.

Aspinwall, L. and Tarr, J. E. (2001). Middle school students' understanding of the role sample size plays in experimental probability. *Journal of Mathematical Behavior*, 20, 229-245.

Baldwin, R. G. (2009). The climate for undergraduate teaching and learning in STEM fields. *New Directions for Teaching and Learning*, 117, 9-17.

Bamberger, M. E. (2003). *Methods college students use to solve probability problems and the factors that support or impede their success*. Doctoral dissertation, Oregon State University.

Bao, L. (2006). Theoretical comparisons of average normalized gain calculations. *American Journal of Physics*, 74(10), 917-922.

Batanero, C., Henry, M., and Parzysz, B. (2005). The nature of chance and probability. In Jones, G. A. (Ed.), *Exploring Probability in School: Challenges for Teaching and Learning* (pp. 15-37). New York, NY: Springer Science and Business Media, Inc.

Batanero, C. and Sanchez, E. (2005). What is the nature of high school students' conceptions and misconceptions about probability? In Jones, G. A. (Ed.), *Exploring Probability in School: Challenges for Teaching and Learning* (pp. 241-266). New York, NY: Springer Science and Business Media, Inc.

Batanero, C., Serrano, L., & Garfied, J. B. (1996). Heuristics and biases in secondary school students' reasoning about probability. In L. Puig & A. Gutierrez (Eds.), *Proceedings of the 20th conference of the International Group for the Psychology of Mathematics Education*, 2, 51-58. Valencia, Spain: University of Valencia.

Batt, J. (2004). *Stollen innocence. A mother's fight for justice: The authorized story of Sally Clark*. Ebury Press.

Benko, P. (2006). *Study of the development of students' ideas in probability*. Doctoral dissertation, Rutgers, The State University of New Jersey.

Burdzy, K. (2009). *The search for certainty: On the clash of science and philosophy of probability*. Singapore: World Scientific Publishing Co. Pte. Ltd.

Carter, T. A. (2005). *Knowledge and understanding of probability and statistics topics by preservice PK-8 teachers*. Doctoral dissertation, Texas A & M University.

Chance, B. L. and Garfield, J. B. (2002). New approaches to gathering data on student learning for research in statistics education. *Statistics Education Research Journal*, 1, 38-44.

Cobb, G. (2000). Teaching statistics: More data, less lecturing. In T. L. Moore (Ed.), Teaching statistics: Resources for undergraduate instructors. *MAA Notes*, 52, 3-5. Washington, DC: Mathematical Association of America Inc.

Cohen, D. (Ed.) (1995). *Crossroads in mathematics: Standards for introductory college mathematics before calculus*. Memphis, TN: American Mathematical Association of Two-Year Colleges.

Cole, M., John-Steiner, V., Scribner, S. and Souberman, E. (Eds.) (1978). *L.S. Vygotsky; Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Cooper, J., and Mueck, R. (1990). Student involvement in learning: Cooperative learning and college instruction. *Journal of Excellence in College Teaching*, 1(1), 68-76.

Cyprus University of Technology, 2011. Retrieved October 20, 2011, from http://www.cut.ac.cy/

Dansereau, D. F. (1988). Cooperative learning strategies. In C. E. Weinstein, E. T. Goetz, & P. A. Alexander, (Eds.), *Learning and study strategies: Issues in assessment, instruction, and evaluation* (pp. 103-120). Orlando, FL: Academic Press.

Dees, R. L. (1991). The role of cooperative learning in increasing problem-solving ability in a college remedial course. *Journal for Research in Mathematics Education*, 22(5), 409-421.

delMas, R. C., Ooms, A., Garfield, J. B., & Chance, B. (2006). Assessing students' statistical reasoning. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved July 2, 2009, from http://www.stat.auckland.ac.nz/~iase/publications/17/6D3_DELM.pdf

Derry, S. J., Levin, J. R., Osana, H. P., Jones, M. S., and Peterson, M. (2000). Fostering students' statistical and scientific thinking: Lessons learned from an innovative college course. *American Educational Research Journal*, 37(3), 747-773.

Dewey, J. (1933). *How we think*. Lexington, MA: D. C. Heath.

Dewey, J. (1969). *Experience and education (11th printing)*. Toronto, Ontario: Collier-Macmillan Canada Ltd.

Diaz, C. and de la Fuente, I. (2007). Assessing students' difficulties with conditional probability and Bayesian reasoning. *International Electronic Journal of Mathematics Education*, 2(3), 128-148.

Effandi, Z. and Zanaton, I. (2007). Promoting cooperative learning in science and mathematics education: A Malaysian perspective. *Eurasia Journal of Mathematics, Science and Technology Education*, 3(1), 35-39.

Esmonde, I. (2009). Ideas and identities: Supporting equity in cooperative mathematics learning. *Review of Educational Research*, 79(2), 1008-1043.

European University Cyprus. (2007). *Programs of study 2007-2008*. Nicosia, Cyprus: Author.

Evans, J. and Pollard, P. (1985). Intuitive statistical inferences about normally distributed data. *Acta Psychologica*, 60, 57-71.

Even, R. & Tirosh (2008). Teacher knowledge and understanding of students' mathematical learning and thinking. In L. D. English (Ed.), *Handbook of International Research in Mathematics Education* (2nd Ed.) (pp. 202-222). New York, NY: Routledge.

Exarchakos, T. G. (1988). *Διδακτική των μαθηματικών* [Didactics of mathematics]. Athens, Greece: Ellinika Grammata.

Falk, R. (1986). Misconceptions of statistical significance. *Journal of Structural Learning*, 9, 83-96.

Falk, R. (1988). Conditional probabilities: Insights and difficulties. In R. Davidson & J. Swift (Eds.), *The Proceedings of the Second International Conference on Teaching Statistics*. Victoria, B.C.: University of Victoria.

Field, A. (2000). *Discovering statistics using SPSS for windows*. London, UK: SAGE Publications Ltd.

Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Dordrecht, Holland: D. Reidel Publishing Company.

Fischbein, E., and Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? *Educational Studies in Mathematics*, 15, 1-24.

Fischbein, E., Nello, M. S., & Marino, M. S. (1991). Factors affecting probabilistic judgments in children in adolescence. *Educational Studies in Mathematics*, 22, 523-549.

Fischbein, E. and Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education,* 28(1), 96-105.

Forman, E. A. (1996). Learning mathematics as participation in classroom practice: Implications of sociocultural theory for educational reform. In Steffe, L. P., Nesher, P., Cobb, P., Goldin, G. A., and Greer, B. (Eds.), *Theories of Mathematical Learning*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Franklin, C. A. and Garfield, J. B. (2006). The GAISE project: Developing statistics education guidelines for grades pre-k-12 and college courses. In Gail F. Burrill and Portia C. Elliott (Eds.), *Thinking and Reasoning with Data and Chance: Sixty-eighth Yearbook* (pp. 345-375). Reston, VA: The National Council of Teachers of Mathematics, Inc.

Frederick University. (2008). *Prospectus 2008-2009*. Nicosia, Cyprus: International and Public Relations Service, Frederick University.

Gal, I. (2004). Statistical literacy: Meanings, components, responsibilities. In D. Ben-Zvi and J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 47-78). Netherlands: Kluwer Academic Publishers.

Garfield, J. (1993). Teaching statistics using small-group cooperative learning. *Journal of Statistics Education*, 1(1).

Garfield, J. (1995). How students learn statistics. *International Statistical Review*, 63(1), 25-34.

Garfield, J. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, 2(1), 22-38.

Garfield, J. & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal of Research in Mathematics Education*, 19(1), 44-63.

Garfield, J., del Mas, R., and Chance, B. (2006). *Assessment resource tools for improving statistical thinking – ARTIST online tests*. Retrieved July 18, 2009 from https://app.gen.umn.edu/artist/tests/index.html

Gigerenzer, G. (1991b). How to make cognitive illusions disappear: Beyond "heuristics and biases." In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 2, pp. 83-115). London: Wiley.

Giraud, G. (1997). Cooperative learning and statistics instruction. *Journal of Statistics Education*, 5(3).

Green, D. R. (1979). The chance and probability concepts project. *Teaching Statistics*, 1(3), 66-71.

Green, D. R. (1982a). *Probability concepts in 11-16 year old pupils* (2nd ed.). Centre for Advancement of Mathematical Education in Technology, University of Technology, Loughborough.

Green, D. R. (1982b). A survey of probability concepts in 3000 pupils aged 11-16 years. In D. R. Grey, P. Holmes, V. Barnett, & G. M. Constable (Eds.), *Proceedings of the First International Conference on Teaching Statistics* (pp. 766-783). Sheffield, England: Teaching Statistics Trust.

Green, D. R. (1983). Shaking a six. *Mathematics in School*, 12, 5, 29-32.

Gunawardena, K. L. D. (1998). Introductory statistics: A cooperative learning approach. In *Proceedings of the International Conference on Teaching Statistics (ICOTS) 5*, Singapore. Retrieved June 14, 2011 from http://www.stat.auckland.ac.nz/~iase/publications/2/Topic2i.pdf

Hakes, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64-74.

Hall, M. R. and Rowell, G. H. (2008). Introductory statistics education and the national science foundation. *Journal of Statistics Education*, 16(2).

Hawkes, N. (2006). Scientists find the key to cot deaths. *The Times*. Retrieved November 1, 2006, from http://www.timesonline.co.uk

Hiebert, J., and Carpenter, T. P. (1992). Learning and teaching with understanding. In Grouws, D. A. (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65-97). New York, USA: Macmillan.

Hiebert, J., and Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In Hiebert, J. (Ed.), *Conceptual and procedural knowledge: The case of mathematics*, (pp. 1-27). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hiebert, J., Stigler, J., Jacobs, J., Givvin, K., Garnier, H., Smith, M., et al. (2005). Mathematics teaching in the United States today (and tomorrow): Results from the TIMSS 1999 video study. *Educational Evaluation and Policy Analysis*, 27, 111-132.

Hirsch, L. S. and O'Donnell, A. M. (2001). Representativeness in statistical reasoning: identifying and assessing misconceptions. *Journal of Statistics Education*, 9(2).

Ignatiou, K. and Zotos, E. (2007a). *Common core mathematics for grade 12 of the unified lyceum*. Nicosia, Cyprus: Ministry of Education and Culture, Pedagogical Institute, Program Development Services.

Ignatiou, K. and Zotos, E. (2007b). *Mathematics of choice for grade 12 of the unified lyceum*. Nicosia, Cyprus: Ministry of Education and Culture, Pedagogical Institute, Program Development Services.

Jacobs, V. R. (1999). How do students think about statistical sampling before instruction? *Mathematics Teaching in the Middle School*, 5, 240-246.

Jaworski, B. (1998). The centrality of the researcher: Rigor in a constructivist inquiry into mathematics teaching. In Teppo, A. R. (Ed.), Qualitative Research Methods in Mathematics Education. *Journal for Research in Mathematics Education Monographs*, 9, 112-127.

Johnson, D., and Johnson, R. (n.d.). *Cooperative learning*. Retrieved June 5, 2009 from http://www.co-operation.org/pages/cl.html

Johnson, D. W., and Johnson, R. T. (2004). *Assessing students in groups: Promoting group responsibility and individual accountability*. Thousand Oaks, CA: Corwin.

Johnson, R. T., and Johnson, D. W. (1985). Student-student interaction: Ignored but powerful. *Journal of Teacher Education*, 34(36), 22-26.

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., and Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin*, 89, 47-62.

Jolliffe, F. (2005). Assessing probabilistic thinking and reasoning. In Jones, G. A. (Ed.), *Exploring Probability in School: Challenges for Teaching and Learning* (pp. 325-344). New York, NY: Springer Science and Business Media, Inc.

Jones, D. L. (2004). *Probability in middle grades mathematics textbooks: An examination of historical trends, 1957-2004*. Unpublished doctoral dissertation, University of Missouri, Columbia.

Jones, P. S. (1970). *A history of mathematics education in the United States and Canada*. Washington, DC: National Council of Teachers of Mathematics.

Jones, C. J. and Harris, P. L. (1982). Insight into the law of large numbers: A comparison of Piagetian and judgment theory. *Quarterly Journal of Experimental Psychology*, 34A, 479-488.

Jones, G. A., Langrall, C. W., and Mooney, E. S. (2007). Research in probability: Responding to classroom realities. In Lester, F. K. Jr. (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning: A Project of the National Council of Teachers of Mathematics.* Charlotte, NC: Information Age Publishing.

Jones, G. A., Langrall, C. W., Thornton, C. A., and Mogill, A. T. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics*, 32, 101-125.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1999). Students' probabilistic thinking in instruction. *Journal for Research in Mathematics Education*, 30, 487-519.

Jones, G. A., and Thornton, C. (2005). An overview of research into the teaching and learning of probability. In Jones, G. A. (Ed.), *Exploring Probability in School: Challenges for Teaching and Learning* (pp. 65-94). New York, NY: Springer Science and Business Media, Inc.

Jones, G. A., Thornton, C. A., Langrall, C. W. and Tarr, J. E. (1999). Understanding students' probabilistic reasoning. In Stiff, L. V. and Curcio, F. R. (Eds.), *Developing Mathematical Reasoning in Grades K-12: 1999 Yearbook*. Reston, VA: The National Council of Teachers of Mathematics, Inc.

Kahneman, D., Slovic, P., and Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.

Kahneman, D., and Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430-454.

Kahveci, M. and Imamoglu, Y. (2007). Interactive learning in mathematics education: Review of recent literature. *Journal of Computers in Mathematics and Science Teaching*, 26(2), 137-153.

Kaplan, J. J. (2006). *Factors in statistics learning: Developing a dispositional attribution model to describe differences in the development of statistical proficiency*. Doctoral dissertation, The University of Texas at Austin.

Kaplan, M. and Kaplan, E. (2006). *Chances are … Adventures in probability*. NY: Penguin Group, Inc.

Keeler, C. M. and Steinhorst, R. K. (1995). Using small groups to promote active learning in the introductory statistics course: A report from the field. *Journal of Statistics Education*, 3(2).

Keeler, C. and Steinhorst, K. (2001). A new approach to learning probability in the first statistics course. *Journal of Statistics Education*, 9(3). Retrieved April 24, 2009 from http://www.amstat.org/publications/jse/v9n3/keeler.html

Khazanov, L. (2005). *An investigation of approaches and strategies for resolving students' misconceptions about probability in introductory college statistics*. Doctoral dissertation, Columbia University.

Khazanov, L. (2008). Addressing students' misconceptions about probability during the first years of college. *Mathematics and Computer Education*, 42(3), 180-192.

Knypstra, S. (2009). Teaching statistics in an activity encouraging format. *Journal of Statistics Education*, 17(2).

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, 6, 59-98.

Konold, C. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education*, 24(5), 392-414.

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education*, 3(1).

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., and Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in mathematics Education*, 24(5), 392-414.

Krause, E. C. (2001). *The effect of instruction in representation theory on student understanding of binomial distribution*. Doctoral dissertation, University at Albany, State University of New York.

Kvatinsky, T., and Even R. (2002). Framework for teacher knowledge and understanding about probability. In Phillips, B. (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics (ICOTS 6)*.

Langrall, C. W. and Mooney, E. S. (2005). Characteristics of elementary school students' probabilistic thinking. In Jones, G. A. (Ed.), *Exploring Probability in School: Challenges for Teaching and Learning* (pp. 95-120). New York, NY: Springer Science and Business Media, Inc.

Laplace, P. S. (1995). *Theorie analytique des probabilities* [Analytical theory of probabilities]. Paris: Jacques Gabay. (Original work published 1814).

Lecoutre, M. P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics*, 23, 557-568.

Leikin, R. and Zaslavsky, O. (1997). Facilitating student interactions in mathematics in a cooperative learning setting. *Journal for Research in Mathematics Education*, 28(3), 331-354.

Levine, D. M., Krehbiel, T. C., and Berenson, M. L. (2010). *Business statistics: A first course (5th ed.)*. Upper Saddle River, NJ: Pearson Education, Inc.

Li, J. (2000). *Chinese students' understanding of probability*. Doctoral dissertation, Nanyang Technological University.

Liu, Y. and Thompson, P. (2007). Teachers' understanding of probability. *Cognition and Instruction*, 25(2), 113-160.

Lordou-Kaspari, D. (2003). *Curriculum development in mathematics in Cyprus secondary schools (1960-1995)*. Nicosia, Cyprus: Livadiotis Ltd.

Mackisack, M. (1994). What is the use of experiments conducted by statistics students? *Journal of Statistics Education*, 2(1).

Manage, A. B. W. and Scariano, S. M. (2010). A classroom note on: Student misconceptions regarding probabilistic independence vs. mutual exclusivity. *Mathematics and Computer Education*, 44(1), 14-21.

Mathematical Association of America. (1998). *Quantitative reasoning for college graduates: A complement to the standards*. Retrieved May 10, 2009 from http://www.maa.org/past/ql/ql_toc.html

McConnell, J. W., Brown, S., Usiskin, Z., Senk, S. L., Widerski, T., Anderson, S., Eddins, S., Feldman, C. H., Flanders, J., Hackworth, M., Hirschhorn, D., Polonsky, L., Sachs, L., and Woodward, E. (1998*). The University of Chicago School Mathematics Project: Algebra: Integrated Mathematics*. Glenview, IL: Scott, Foresman and Company.

Ministry of Education and Culture, Cyprus. (2004). *Country report: Cyprus*. Retrieved April 26, 2009 from http://www.coe.int/T/DG4/Linguistic/Source/Country_Report_ Cyprus_ EN.pdf

Ministry of Education and Culture & Pedagogical Institute. (2002). *Post-secondary education in Cyprus*. Nicosia, Cyprus: Program Development Service.

Ministry of Education and Culture, Pedagogical Institute & Curriculum Development Service (2010). *Program of Studies: Mathematics*. Retrieved March 12, 2011 from http://www.moec.gov.cy/analytika_programmata/

Mises, R. von (1952). *Probability, statistics, and truth* (J. Neyman, O. Scholl, & E. Rabinovitch, Trans.). London: William Hodge and company. (Original work published 1928).

Mojica, G. F. (2006). *Middle school teachers' conceptions of empirical and theoretical probability: A study of pedagogy*. Dissertation, North Carolina State University, NC.

Montgomery, D. (1997). *Design and analysis of experiments* (4th ed.). New York, NY: John Wiley and Sons, Inc.

Moore, D. S. (1997). Probability and statistics in the core curriculum. In J. Dossey (Ed.), *Confronting the Core Curriculum* (pp. 93-98). Mathematical Association of America.

Mullis, I. V. S., Martin, M. O., and Foy, P. (2008). *International mathematics report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Retrieved October 25, 2006, from http://www.ed.gov/pubs/Nat AtRisk/recomm.html

National Council of Supervisors of Mathematics. (1977). Position paper on basic skills. *Arithmetic Teacher*, 25(1), 19-22.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics*. Reston, VA: Author.

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

National Research Council. (2003). *Evaluating and improving undergraduate teaching in science, technology, engineering and mathematics*. Washington, DC: National Academies Press.

National Research Council. (2001). *Adding + it up*. Kilpatrick, J., Swafford, J., and Findell, B. (Eds.). Washington, D.C.: National Academy Press.

Neapolis University. (2011). *Neapolis University Pafos*. Retrieved January 12, 2012 http://www.nup.ac.cy/Portals/0/Publications/NUP_Brochure_EN_lowRes.pdf

Nickerson, R. S. (2004). *Cognition and chance: The psychology of probabilistic reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Open University of Cyprus. (2011). Retrieved October 15, 2011 from http://www.ouc.ac.cy/web/guest/home

Papanastasiou, C. (1997). Cyprus. In D. Robitaille (Ed.), *National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS* (p. 91-97). Canada: Pacific Educational Press.

Papanastasiou, C. (2002). *TIMSS study in Cyprus: Patterns of achievements in mathematics and science. Studies in Educational Evaluation*, 28(2002), 223-233.

Papapavlou, A. N. (2001). Linguistic imperialism? The status of English in Cyprus. *Language Problems & Language Planning*, 25(2), 167-176.

Pashiardis, P. (2007). Cyprus. In Horner, W., Dobert, H., Kopp, B. von, & Mitter, W. (Eds.), *The Education Systems of Europe* (pp. 202-222). Downloaded September 12, 2009 from http://www.springerlink.com/content/hq08um126n636423/

Penas, L. M. (1987). Probability and statistics in midwest high schools. In *American Statistical Association 1987 Proceedings of the Section on Statistical Education*, (pp. 122). Alexandria, VA: American Statistical Association.

Pfaff, T. J. and Weinberg, A. (2009). Do hands-on activities increase student understanding? A case study. *Journal of Statistics Education*, 17(3).

Piaget, J., and Inhelder, B. (1975). *The origin of the idea of chance in children* (L. Leake Jr., P. Burrell, & H. D. Fischbein, Trans.). New York: Norton (Original work published 1951).

Polaki, M. V., Lefoka, P. J., and Jones, G. A. (2000). Developing a cognitive framework for describing and predicting Basotho students' probabilistic thinking. *Boleswa Educational Research Journal*, 17, 1-21.

Pollatsek, A., Well, A. D., Konold, C., & Hardiman, P. (1987). Understanding conditional probabilities. *Organizational Behavior and Human Decision Processes*, 40, 255-269.

Potthast, M. J. (1999). Outcomes of using small-group cooperative learning experiences in introductory statistics courses. *College Student Journal*, 33(1).

Pratt, D. (2000). Making sense of the total of two dice. *Journal for Research in Mathematics Education*, 31(5), 602-625.

Resnick, L. B., & Ford, W. W. (1981). *The psychology of mathematics for instruction*. Hillsdale, NJ: Erlbaum

Rubel, L. H. (2006). Students' probabilistic thinking revealed: The case of coin tosses. In G. F. Burrill & P. C. Elliott (Eds.), *Thinking and Reasoning with Data and Chance* (pp. 49-59). Reston, VA: National Council of Teachers of Mathematics.

Rubel, L. H. (2007). Middle school and high school students' reasoning on coin tasks. *Journal for Research in Mathematics Education*, 38(5), 531-556

Saenz, C. (1998). Teaching probability for conceptual change. *Educational Studies in Mathematics*, 35, 233-254.

Savery, J. R. and Duffy, T. M. (1996). Problem based learning: An instructional model and its constructivist framework. In B. G. Wilson (Ed.), *Constructivist Learning Environments: Case Studies in Instructional Design*. Englewood Cliffs, NJ: Educational Technology Publications.

Scheaffer, R. L., Watkins, A. E., & Landwehr, J. M. (1998). What every high-school graduate should know about statistics. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K-12*. Mahwah, NJ: Lawrence Erlbaum Associates.

Sedlmeier, P. and Gigerenzer, G. (1997). Intuitions about sample size: The empirical law of large numbers. *Journal of Behavioral Decision Making*, 10(1), 33-51.

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*, 8(3), 295-316.

Shaughnessy, M. (1981). Misconceptions of probability: From systematic errors to systematic experiments and decisions. In A. P. Shulte and J. R. Smart (Eds.), *Teaching statistics and probability* (pp. 90-99). Reston, VA: National Council of Teachers of Mathematics.

Shaughnessy, M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465-494). Reston, VA: National Council of Teachers of Mathematics.

Shaughnessy, M. (2003). Research on students' understanding of probability. In J. Kilpatrick, W. G. Martin, D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 216-226). Reston, VA: National Council of Teachers of Mathematics.

Shaughnessy, M. J., Barrett, G., Billstein, R., Kranendonk, H. A., and Peck, R. (2004). What's the probability of a hit? In Lott, J. W. and House, P. A. (Eds.), *Navigating through probability in grades 9-12*, p. 40-42 and p. 94-97. Reston, VA: The National Council of Teachers of Mathematics, Inc.

Shay, K. B. (2008). *Tracing middle school students' understanding of probability: A longitudinal study*. Doctoral dissertation, Rutgers, The State University of New Jersey.

Silver, E. (1986). Using conceptual and procedural knowledge: A focus on relationship. In Hiebert, J. (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 181-198). Hillsdale, NJ: Lawrence Erlbaum Associates.

Silver, E. A., Mesa, V. M., Morris, K. A., Star, J. R., and Benken, B. M. (2009). Teaching mathematics for understanding: An analysis of lessons submitted by teachers seeking NBPTS certification. *American Educational Research Journal*, 46(2), 501-531.

Skemp, R. (1978a). *The psychology of learning mathematics*. Hillsdale, NJ: Lawrence, Erlbaum.

Skemp, R. R. (1978b). Relational understanding and instrumental understanding. *Arithmetic Teacher*, 26(3), 9-15.

Slavin, R. E. (1995). *Cooperative learning: Theory, research, and practice* (2nd ed.). Boston: Allyn & Bacon.

Smith, K. A., Douglas, T. C. and Cox, M. F. (2009). Supportive teaching and learning strategies in STEM education. *New Directions for Teaching and Learning*, 117, 19-32.

Springer, L., Stanne, M. E., and Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, 69(1), 21-51.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Stanovich, K. E., and West, R. F. (2003). Evolutionary versus instrumental goals: How evolutionary psychology misconceives human rationality. In D. E. Over (Ed.), *Evolution and the psychology of thinking: The debate*. Retrieved May 10, 2009 from http://web.mac.com/kstanovich/iWeb/Site/Research%20on%20Reasoning.html

Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education*, 36(5), 404-411.

Tarr, J. E., and Jones, G. A. (1997). A framework for assessing middle school students' thinking in conditional probability and independence. *Mathematics Education Research Journal*, 9, 39-59.

The Ministry of Education and Culture, Republic of Cyprus. (2008). *Inclusion in the Cyprus educational system at the beginning of the twenty first century: An overview - national report of Cyprus*. Nicosia, Cyprus: Author.

Thompson, P. W., and Saldanha, L. A. (2003). Fractions and multiplicative reasoning. In J. Kilpatrick, G. Martin, & D. Schifter (Eds.), *A research companion to the principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

TIMSS International Study Center. (2000). *Mathematics teacher background with mathematics achievement: Main survey, Third International Mathematics and Science Study Repeat*. Retrieved July 4, 2009, from http://timss.bc.edu/timss1999i/questionnaires.html

TIMSS and PIRLS International Study Center. (1995). *TIMSS: IEA's third international mathematics and science study: Released item set for the final year of secondary school: Mathematics and science literacy, advanced mathematics, and physics*. Retrieved July 4, 2009, from http://timss.bc.edu/timss1995i/Items.html

TIMSS and PIRLS International Study Center. (2001). *TIMSS 1999 mathematics items: Released set for eighth grade*. Retrieved July 4, 2009 from http://timss.bc.edu/timss 1999i/study.html

TIMSS and PIRLS International Study Center (2007). *TIMSS 2003 mathematics items: Released set eighth grade*. Retrieved July 6, 2009 from http://timss.bc.edu/timss2003i/ released.html

TIMSS and PIRLS International Study Center (2009). *TIMSS 2007 user guide for the international database*. Retrieved July 6, 2009 from http://timss.bc.edu/TIMSS2007/ items.html

Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2nd ed.). Chicago: University of Chicago Press.

Travers, K. J., & Westbury, I. (1989). *The IEA study of mathematics I: Analysis of mathematics curricula*. New York: Pergamon Press.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.

Ulep, S. A. (1990). *Strategies preservice secondary mathematics teachers use in solving problems involving uncertainty*. Dissertation Abstracts International, (UMI No. 9116999).

University of Central Lancashire-Cyprus. (2012). *UCLAN Cyprus: The first British university in Cyprus*. Retrieved May 1, 2012 from www.uclancyprus.ac.cy

University of Cyprus. (2007). *Guide to undergraduate studies 2007-2008*. Retrieved May 18, 2009 from http://www.ucy.ac.cy/data/puof/undgradgr08.pdf

University of Nicosia. (2008). *Profile 2008-2009*. Nicosia, Cyprus: Author.

Utts, J. (2003). What educated citizens should know about statistics and probability. *The American Statistician*, 57(2), 74-79.

Vrasidas, C. and McIsaac, M. (2001). Integrating technology in teaching and teacher education: Implications for policy and curriculum reform. *Educational Media International*, 38(2), 127-132.

Watson, J. (2005). The probabilistic reasoning of middle school students. In Jones, G. A. (Ed.), *Exploring Probability in School: Challenges for Teaching and Learning* (pp. 145-170). New York, NY: Springer Science and Business Media, Inc.

Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. NJ: Lawrence Erlbaum Associates Inc.

Watson, J. M., & Moritz, J. B. (2002). School students' reasoning about conjunction and conditional events. *International Journal of Mathematics Education in Science and Technology*, 33(1), 59-84.

Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education*, 22(5), 366-389.

Webb, N. M., Tropper, J. D., and Fall, R. (1995). Constructive activity and learning in collaborative small groups. *Journal of Educational Psychology*, 87(3), 406-423.

Woolfson, M. M. (2008). *Everyday probability and statistics: Health, elections, gambling and war*. London, UK: Imperial College Press.

Zieffler, A., Garfield, J., Alt, S., Dupuis, D., Holleque, K., and Chang, B. (2008). What does research suggest about the teaching and learning of introductory statistics at the college level? A review of the literature. *Journal of Statistics Education*, 16(2).

Zimmermann, G. (2002). *Students' reasoning about probability simulations during instruction*. Dissertation, Illinois State University.