STUDENTS' REASONING ABOUT PROBABILITY SIMULATIONS DURING INSTRUCTION

Gwendolyn Zimmermann

212 Pages August 2002

This study examined the role instruction played in changing students' individual and collective reasoning and beliefs about probability simulation and what impact technology had in these changes.

APPROVED:

Date Graham A. Jones, Co-Chair

Date Cynthia Langrall, Co-Chair

Date Kenneth Berk

Date Edward S. Mooney

STUDENTS' REASONING ABOUT PROBABILITY

SIMULATIONS DURING INSTRUCTION

Gwendolyn Zimmermann

212 Pages August 2002

This study examined the role instruction played in changing students' individual and collective reasoning and beliefs about probability simulation and what impact technology had in these changes.

Twenty-three students from a high school Advanced Placement Statistics class participated in a 12-day whole-class teaching experiment. Students were administered a pre-, post-, and retention assessment to provide quantitative data to examine changes in student reasoning. Audio and videotapes of instructional sessions and interviews of 4 target students provided qualitative data related to their reasoning and beliefs.

Quantitative analysis revealed that the post- and retention assessment scores were significantly higher than the preassessment, but there was not a significant difference between the post- and retention assessment scores. Qualitative analysis discerned that following instruction, the frequency of valid responses increased in 5 of the 6 simulation components. More specifically, students made significant progress in their ability to use simulated outcomes to determine the probability of an event and to recognize the effect of repeated trials on the empirical probability. Students' use of the graphing calculator was

found to have a considerable impact on students' reasoning about probability simulation. Specifically, the syntax needed to operate the calculator focused students on the components of simulation, and the calculator provided a transparent medium for handling more complex problems that involved dependent events.

This study identified a number of helpful and problematic beliefs. Some of the helpful beliefs included the belief that assumptions are part of simulation and as the number of trials increased, the empirical probability approached the theoretical probability. Problematic beliefs were related to misconceptions, such as representativeness.

As students reasoned collectively, a number of sociomathematical norms emerged about how students justified valid simulation components. From these norms emerged one or more classroom mathematical practices, many of which became taken as shared by the students.

APPROVED:

Date	Graham A. Jones, Co-Chair
Date	Cynthia Langrall, Co-Chair
Date	Kenneth Berk
Date	Edward S. Mooney

STUDENTS' REASONING ABOUT PROBABILITY SIMULATIONS DURING INSTRUCTION

GWENDOLYN ZIMMERMANN

A Dissertation Submitted in Partial
Fullfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

Department of Mathematics

ILLINOIS STATE UNIVERSITY

2002

DISSERTATION APPROVED:

Date	Graham A. Jones, Co-Chair
Date	Cynthia Langrall, Co-Chair
Date	Kenneth Berk
Date	Edward S. Mooney

ACKNOWLEDGEMENTS

And they said it could not be done. With the support and assistance of many people, this dissertation study was completed within one year while teaching full-time. I would like to acknowledge the many people who made this feat possible.

I would like to express my sincere gratitude to the members of my dissertation committee for their vast commitment of time. In particular, I will forever be indebted to Dr. Graham A. Jones for his willingness to take on one last doctoral student. His insights were invaluable, his patience never ending, and his disposition always pleasant. He has been instrumental in helping me to become a researcher and a more reflective classroom teacher. I am grateful to Dr. Cynthia W. Langrall for providing another perspective that greatly enhanced this project. Thanks to Dr. Kenneth Berk for his help with the quantitative analysis of this study and for his comments that helped to provide clarity to my writing. I would like to thank Dr. Edward S. Mooney for his enthusiasm for this project and his willingness to make the drive to the Chicago-area for a few days to fill the role as the “other” researcher. My gratitude goes to Dr. Carol Thornton, who has advised, encouraged, and supported me throughout the entire doctoral program.

The timeline of this project would not have been possible without all the support and assistance from the mathematics teachers at WEGO. I am particularly indebted to Patty (Elvis) Arnold and Kevin Gimre. The humor and support they provided guided me through the times when I wondered if it was possible. Thanks to Vera Sorensen, my witness, who underestimates her role in this study. This study certainly would not have been possible without the students who participated in this study, and I would like to take this opportunity to extend my appreciation to them for their cooperation and willingness to be a part of my research.

I wish to thank John Carter. It was at his suggestion I enroll in the doctoral program at Illinois State University. His belief in me as a teacher and leader and his endless encouragement have been a source of strength and inspiration. I also want to thank the “Chicago Cohort.” Their comradery provided support and motivation, and our many discussions about mathematics education have contributed greatly to my knowledge and beliefs as an educator. I especially wish to thank Antoinette Seidelmann and Susan Brown. The combination of their humor, easy-going nature, and insightfulness into mathematics education made them the perfect college roommates.

I am most thankful to my two best friends who believe I can do anything (well, almost). In particular, I would like to express my sincerest appreciation to my brother, Greg Galeles, who has always been there, day or night, with his love and support; and to my loving husband, Jim Zimmermann, who married me in spite of the fact he would see little of me over the next three years. He always willingly took care of all the little details that make up life so I could concentrate on this program. I will be forever grateful to him for his never-ending love and support. He is truly the wind beneath my wings.

G. Z.

CONTENTS

TABLES

FIGURES

CHAPTER I

THE PROBLEM AND ITS BACKGROUND

As omnipresent as statistics is in everyday life, it is disturbing that much of the general public does not rely on statistical thinking. The public seems to rely heavily on anecdotal evidence, and when statistics are presented they are frequently misrepresented or misinterpreted (Kolata, 1997). Hence it is not surprising that statistical education has received enhanced emphasis in recent mathematical reforms (National Council of Teachers of Mathematics [NCTM], 2000). These reforms recognize that students need statistical skills that enable them to make informed decisions. Another sign of intensified attention to statistical reasoning has been the introduction of Advanced Placement Statistics in 1997 (College Board, 2000). The increased emphasis on statistical reasoning by such national organizations as NCTM and the College Board has not gone unheeded by teachers or students. In the first three years of implementing APS across U.S. high schools, students choosing to take the content specific test in statistics have increased from about 7600 students in 1997 to over 41,000 in the year 2001 (Straf, 2002), an increase of over 200 percent.

Probability and probability simulation are inherent parts of the content of a high school statistics course. Both the NCTM's <u>Principles and Standards of School Mathematics</u> (2000) and the College Board's APS curriculum (2000) recommend the inclusion of probability simulations as a tool to look at long run behavior patterns. Not only do simulations provide students the opportunity to make sense of probability concepts, they are also a natural vehicle by which to model contextual problems. Furthermore, a review of the literature reveals an increased interest in both curriculum and research in mathematical modeling (Biehler, 1991; Borovcnik & Peard, 1996; Doerr, 1998; Doerr & Tripp, 1999; Lesh, Amit, & Schorr, 1997; Wares, 2001). Modeling is relevant to this study because probability simulations enable students to develop the ability to formulate and analyze mathematical models. In fact, an understanding of probability and probability simulations will not only aid students in APS, it will help them to become more mathematically and statistically informed citizens (College Board, 2000).

As the call for statistically sophisticated students increases and the number of students enrolled in APS continues to grow, more research is needed on how students reason about probability simulations. Moreover, for teachers of statistics there is a need to know more about how instruction affects students' learning of probability simulations. This study was designed to address both of these needs.

Statement of the Problem

Having identified an increased emphasis on statistical education in general and probability simulation in particular, it is appropriate to examine what research tells us about high school students' reasoning in, and beliefs about, probability simulation. Although a substantial number of studies have investigated the probabilistic reasoning of students in the elementary and middle grades (e.g., Falk, 1983; Fischbein & Gazit, 1984; Jones, Langrall, Thornton, & Mogill, 1997; Jones, Thornton, Langrall, & Tarr, 1999; Piaget & Inhelder, 1975), fewer studies have examined the probabilistic reasoning of high school students (Batanero & Serrano, 1999; Fischbein, Nello, & Marino, 1991; Fischbein

& Schnarch, 1997; Green, 1983; Shaughnessy, 1977). Moreover, there appears to be limited research that has examined high school students' reasoning in relation to probability simulations (Benson & Jones, 1999; Zimmermann & Jones, 2002) and students' reasoning related to modeling (Biehler, 1991; Borovcnik & Peard, 1996; Doerr, 1998; Doerr & Tripp, 1999; Lesh et al., 1997). Finally, no studies have been located that have analyzed or evaluated instructional programs involving modeling and simulation. Given this gap in the research literature and the importance of cognitive research in informing instruction (Fennema, Franke, Carpenter, & Carey, 1993), there is a need for a study that examines the effect of instruction on students' individual and collective probabilistic reasoning and beliefs as they relate to simulation. This study addressed this void in the research literature by generating theoretical and practical knowledge to inform instruction on probability simulation.

Research Questions

The purpose of this study was to develop an understanding of how instruction affects high school students' reasoning and beliefs related to probability simulation. Further, this research sought to supplement what is known about students' individual reasoning and beliefs concerning probability simulation (Benson & Jones, 1999; Zimmermann & Jones, 2002) with what could be learned about the collective knowledge of students as they worked together in a classroom environment. In particular, this study addressed the following research questions:

1. How does high school students' individual and collective reasoning about probability simulation change during a whole-class teaching experiment? What role does technology play in this change?

2. How do high school students' beliefs about probability simulation change during a whole-class teaching experiment?
3. What kind of sociomathematical norms and classroom mathematical practices evolve during a whole-class teaching experiment that focuses on probability simulation?

Definition of Terms

Yates, Moore, and McCabe (1999) defined a probability simulation as an experimental process that models the probability elements of a real world context. For example, the distribution of pizza phone orders for a restaurant is about 60 percent for pizzas with meat and 40 percent for pizzas that are without meat. A probability simulation could be used to determine the probability that the next two phone orders in the restaurant are for pizzas with meat. In elaborating this procedure, Yates et al. identified a number of steps in the simulation process that are summarized in Figure 1.

Simulation Process

1. State the problem and list any assumptions.
2. Assign random digits to model problem outcomes.
3. Define a trial.
4. Repeat trial many times.
5. Determine empirical probability.

Figure 1. Simulation Process (adapted from Yates et al., 1999).

First, the problem should be stated and any assumptions noted. Second, random digits (or the outcomes of a probability generator) should be assigned to represent or model the outcomes of the problem context. Third, a trial should be defined to meet the conditions of the problem. Fourth, this trial should be repeated many times. Finally, the simulation data that has been collected should be used to determine the required empirical probability. In essence, the process outlined by Yates et al. provides a mathematical norm or framework for assessing the reasoning of high school students when they undertake a simulation.

In this study, theoretical probability is defined as the probability computed by dividing the number of outcomes in an event by the number of outcomes in the sample space. This is based upon the assumptions that each possible sample outcome is equally likely. Empirical (also known as experimental) probability is the "proportion of times the outcome would occur in a very long series of repetitions" (Yates et al., 1999, p. 314).

The term two-dimensional refers to probability activities or simulations that involve the performing of two random experiments or the performing of one random experiment twice. Zimmermann and Jones (2002) used the following definition of two-dimensional:

> In a two-dimensional probability activity or simulation one outcome is obtained from each random experiment; hence we are dealing with pairs of outcomes, in this case ordered pairs. Although less complex, the two-dimensional problems in this study essentially belong to a class of problems that mathematicians refer to as the joint probability density function of two random variables (Hogg & Tanis, 1997). By way of contrast a one-dimensional probability situation would involve just one random experiment and a sample space that consisted of single outcomes. English (1993) also uses the term two-dimensional when referring to the combinatorial problems that exhibit similar structure to the problems in this study.

Theoretical Framework

This research used a whole-class teaching experiment to study high school students' reasoning and beliefs as they experienced an instructional unit on probability simulation as part of the APS curriculum. Grounded in the work of Cobb (1999), a whole-class teaching experiment consists of a sequence of instructional lessons that are based on student knowledge and understandings as they emerge during the course of instruction (English, Jones, Lesh, Bussi, & Tirosh, 2002). The purpose of a whole-class teaching experiment is to provide a means by which to analyze and record student understanding as it develops.

As illustrated in Figure 2, a whole-class teaching experiment is comprised of two phases, the instructional development phase and the classroom-based analysis. During the process of a teaching experiment these phases cycle back on each other. Cobb (1999) terms the entire cyclic process the Developmental Research Cycle. The instructional development phase concerns the design of hypothetical learning trajectories that are informed by instructional theory and content-specific research. A hypothetical learning trajectory is made up of three components: learning goals for students, planned learning or instructional activities, and a conjectured learning process in which the teacher anticipates how students' learning and understanding might evolve. In some sense, the hypothetical learning trajectory is a set of lesson plans for the instructional program yet has special characteristics that differentiate it from a traditional lesson plan. During the second phase, classroom-based analysis, instruction is reformed and refined based on classroom events and activities. It is also during this phase that individual and collective student reasoning is traced. The classroom-based analysis utilizes Cobb's (2000) emergent perspective that views both an individual's cognitive characteristics and the social dynamics of the classroom as equally important in contributing to mathematical development. In other words, not only is it necessary to examine individual cognitive and psychological aspects of learning, it is also crucial to examine the established classroom norms and practices that are essential to understanding students' development. Therefore, this study examined both individual and collective cognitions of students as they reasoned through contextual problems involving probability simulation.

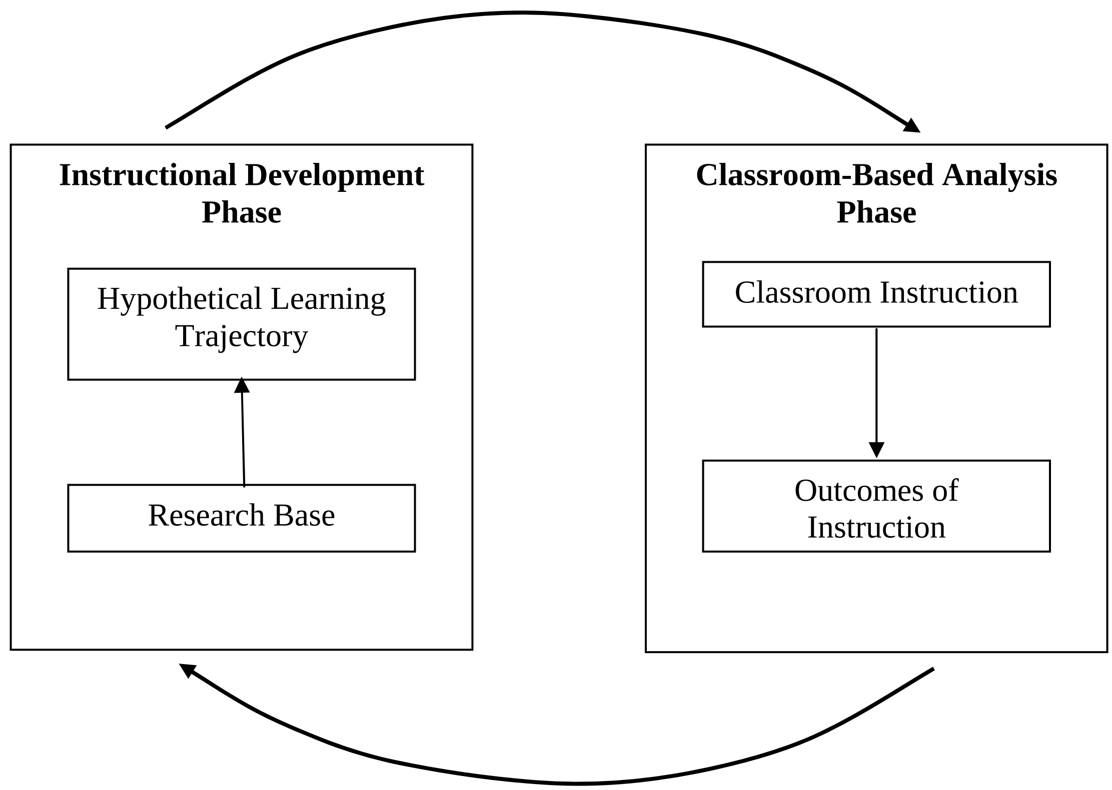

Figure 2. Developmental Research Cycle (adapted from Cobb, 1999).

Instructional Development Phase

The theoretical framework of this whole-class teaching experiment began with the instructional development phase (Figure 3). In this phase, instructional theory and content-specific research inform the development of the hypothetical learning trajectory and in turn design of the instructional sessions. Following is a more detailed description of each component of the instructional development phase.

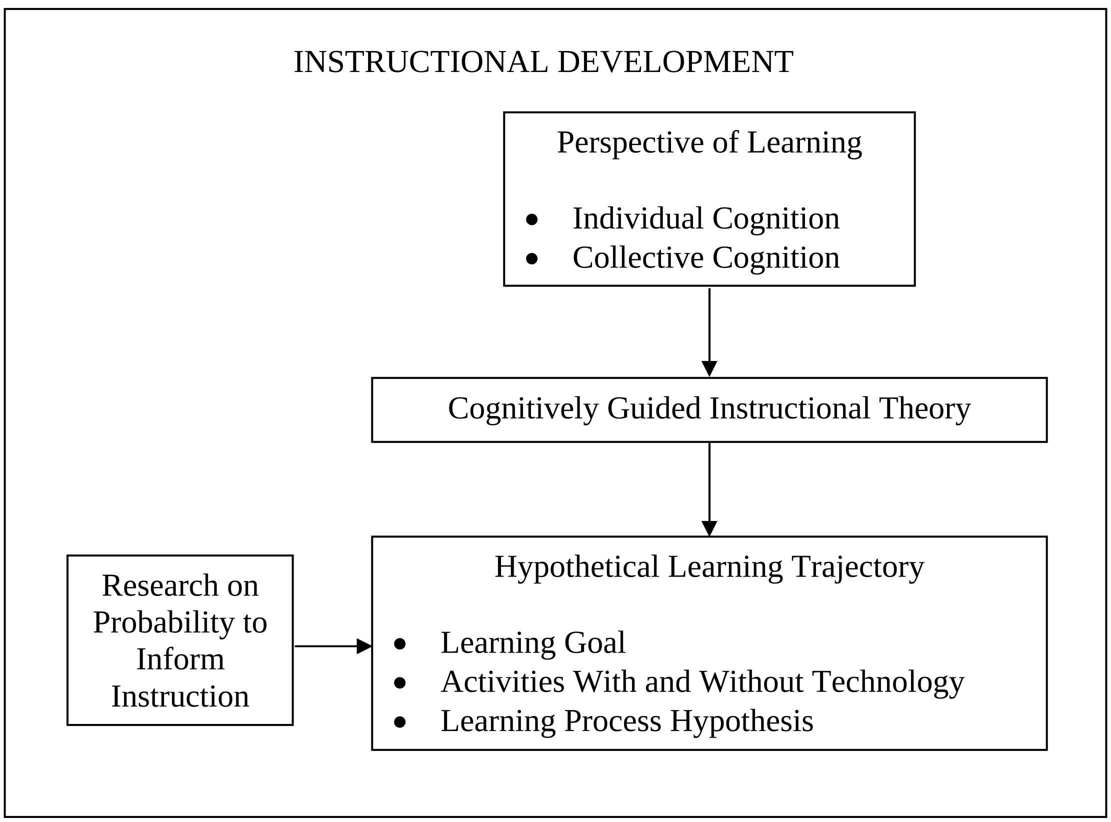

Figure 3. Instructional Development Phase.

<u>Perspective of Learning</u>

This component of the instructional development phase assumes that both individual and collective cognition is critical in guiding the development of instruction. In Cobb's (2000) view, the individual student interacts within the social structures of the class. During the process of instruction, the individual student develops her own reasoning and beliefs, but the social perspective of the classroom, mainly social norms, sociomathematical norms, and mathematical practices, influence this development. Cobb's perspective of learning can be viewed through the lens of symbolic interactionism (Blumer, 1969). According to the basic caveats of symbolic interactionism, people interact with each other and engage in various activities, and these interactions are part of a reciprocating process. In the classroom, students and teacher will interact, as will students with other students. Cobb holds that learning does not occur in isolation from the social perspective of the classroom environment and, therefore, both the individual and social cognitions must be examined to develop an understanding of how learning occurs.

As with all teaching experiments, the students' ongoing cognitions during the Developmental Research Cycle informed the instructional development phase. However, what distinguished the developmental research cycle of this whole-class teaching experiment from much of the earlier research (Steffe & Thompson, 2000; Tzur, 1999) using teaching experiments was the teacher-researcher's personal insight. Since the teacher-researcher was the daily classroom teacher of the students, she had first-hand knowledge of the classroom sociomathematical norms as well as the individual and collective cognitions of the class.

Cognitively Guided Instructional Theory

In accord with Cobb's (2000) recommendation that a whole-class teaching experiment requires an instructional theory, this whole-class teaching experiment was built on the theory of cognitively guided instruction (Figure 3). Cognitively guided instructional theory is based on the assumption that research-based knowledge of students' thinking can inform instruction (Fennema et al., 1993). Fennema et al. maintain that a teacher with knowledge of research on student mathematical thinking can use this information combined with individual student problem-solving strategies to select problems that will help develop student reasoning. Teachers who use cognitively guided instructional theory monitor student thinking so as to make sense of student learning in the context of their own knowledge. For the whole-class teaching experiment in this study, research related to probability, simulation and modeling, as well as technology was used to build the research knowledge base that the teacher-researcher used to inform instruction.

Hypothetical Learning Trajectory

Although the whole-class teaching experiment was founded on cognitively guided instructional theory, it was driven by a continuous revision of the hypothetical learning trajectory (Figure 3). As described earlier, a hypothetical learning trajectory (Simon, 1995) consists of a learning goal, a plan of instructional lessons and activities, and a hypothesis of the learning process for the class. The overall learning goal for this whole-class teaching experiment was the development of an understanding of probability simulations including the ability to use simulations to determine empirical probabilities. The plans of instructional lessons and activities were designed to provide problems that could be simulated and at the same time create a learning environment where students built mathematical ideas through interaction and reconceptualization of their prior knowledge (Yackel , Cobb, & Wood, 1990). Graphing calculator technology was utilized for most of the problems and activities, while manipulatives (e.g., candy, random number table, coins, etc.) were used for others. The hypothesized learning process, as formulated by the teacher-researcher, determined the exact nature of the activities and problems as well as any modifications to the learning process or activities. During the whole-class teaching experiment the hypothetical learning trajectory was continuously revised and reformulated as determined by the other phase of the Research Development Cycle, the classroom-based analysis. The hypothetical learning trajectory was informed by both the research literature as well as the classroom-based analysis. Before discussing the second phase of the whole-class teaching experiment, a review of the key content-specific research relevant to this teaching experiment is presented. Further content specific research will be discussed in Chapter 2.

Research on Probability to Inform Instructional Development

The instructional development phase was informed by research (Figure 3) associated with the specific mathematical content of the whole-class teaching experiment. In this case, a review of current literature included students' probability reasoning and beliefs, simulations and modeling, and technology. An analysis of teaching experiments involving probability and statistics has also been included in order to illustrate the usefulness of such a methodology in addition to providing helpful insights into the procedures that were used during the teaching experiments.

Students' Probability Reasoning

There is substantial literature on students' reasoning and beliefs about theoretical probability. In an overall sense, research indicates that children and adults have difficulty reasoning probabilistically (Fischbein et al., 1991; Garfield & Ahlgren, 1988; and Hawkins & Kapadia, 1984), and the factors affecting their probabilistic reasoning are varied. Some of these factors include students' tendency to rely on memorized procedures and lack of ability to reason proportionally (Garfield & Ahlgren, 1988; Green, 1983). Similarly, the concept of randomness seems to elude many students (Batanero & Serrano, 1999), and in spite of instructional intervention, some students are unable to construct valid arguments about concepts of randomness related to the regularity of patterns, frequencies of possible results, and the existence or absence of runs.

Students' Beliefs About Probability

The literature revealed that beliefs play a major role in students' reasoning about probability (Fischbein & Gazit, 1984; Fischbein et al., 1991; Fischbein & Schnarch, 1997; Garfield & Ahlgren, 1988; Piaget & Inhelder, 1975; Shaughnessy, 1992). That is, students' conceptions of probability often involve the use of judgment heuristics, such as representativeness, availability, and outcome approach, that are grounded in beliefs (Konold, 1991a; Konold, Pollatsek, Well, Lohmeier, & Lipson, 1993; Tversky & Kahneman, 1973, 1974). These heuristics are described in more detail in Chapter 2. According to Shaughnessy (1992), students' beliefs in such judgment heuristics indicate that they may not have developed a cognitive framework in which to understand the mathematical relevance of probability experiments. If this were the case, it would certainly have implications for students' understanding of probability simulations.

Much of the research points to the complex nature of students' beliefs, and the research seems to imply that neither instruction nor maturation necessarily help eliminate student misconceptions (Fischbein & Schnarch, 1987; delMas & Bart, 1989). In fact, delMas and Bart (1989) concluded that instruction could actually reinforce some misconceptions. Fischbein and Gazit (1984) hypothesized that instruction could be used to alter and develop new intuitive attitudes through participatory activities involving probability concepts. They concluded that student understanding did appear to increase with age, however, many of the concepts proved to be too difficult for the fifth-grade students in their study. It was also found that instruction had the "indirect" effect of overcoming some of the students' intuitive misconceptions.

In summary, many factors affect how students reason probabilistically. Research indicates that students with a weak understanding of proportions are unlikely to reason appropriately. Further, numerous studies reveal that individual beliefs have a substantial impact on how students reason probabilistically. And, although instruction appears to have limited effect on students' reasoning and beliefs, maturation seems to improve some concepts but has limited influence on the concept of randomness.

Simulations and Modeling

Given the importance of the role of simulation in the study of statistics, it is surprising that so little research has focused on this topic. Biehler (1991) recognized that computer generated simulations could serve as both a problem-solving tool and a process to explore the concept of empirical probability. Borovcnik and Peard (1996) questioned the use of simulations that employed computers prior to using actual manipulatives. Their concern was related to how students might perceive the "pseudo-randomness" of computers. They suggested that teachers commence instruction by using manipulatives and then compare simulation with manipulatives to simulations using technology.

There are more recent studies on students' reasoning with respect to probability modeling (Benson, 2000; Benson & Jones, 1999), a key aspect of simulation. Benson and Jones found that only 3 of the 7 students in their study (ranging from Grade 2 through college) were able to construct a valid model for a contextual problem involving two-dimensional probability. The researchers concluded that students' ability to construct two-dimensional models seemed to be closely related to their knowledge of two-dimensional theoretical probability. While the Benson and Jones study provided a glimpse of students' reasoning in relation to probability simulation, it should be noted that they did not look at the complete simulation process but only the first step; that is modeling the contextual situation.

Zimmermann and Jones (2002) focused on two-dimensional simulation problems like the following: given the probabilities that a radio station plays three kinds of music, what is the probability that hip-hop music is playing when the radio is turned on at two different times. In their study involving 9 high school students, Zimmermann and Jones found that the concept of two-dimensional simulations was difficult for most students. Students' inability to understand the concept of dimensionality also prevented most of them from recognizing that the trial in another simulation was one-dimensional when it should have been two-dimensional. However, most students were able to construct a valid probability generator, that is, a random device that produced the specified probability distribution in the given problem. Zimmermann and Jones also determined that students bring with them beliefs about simulations that can be both helpful and problematic to instruction.

Students' ability to construct models of problems is likely to factor into their ability to construct appropriate simulations. Despite this, only recently have research studies examined students' thinking in relation to mathematical modeling. According to Lesh et al. (1997), model-eliciting activities require that students mathematically interpret situations and problems. In the process of constructing models, students develop a multitude of unanticipated and effective mathematical models to solve problem (Lesh et al., 1997; Lesh & Clarke, 2000; Lesh & Lehrer, 2000). Construction of models tends to begin from an intuitive level and graduates to a more formal type reasoning that allows students to produce a more generalizable model.

Technology Use in Probability and Statistics

NCTM advocates the appropriate use of technology in the mathematics classroom (2000). Furthermore, the College Board expects that students will have regular access to both computers and calculators with statistical computation and graphics capabilities (Watkins, Roberts, Olsen, & Scheaffer, 1997). Therefore, a review of research related to simulation is germane to this study.

The College Board (Watkins et al., 1997) recognizes the role of technology in today's mathematics curriculum. Some of the advantages of technology include the ability to "store large data sets for interactive data analysis, provide variety, speed, and visualization of simulations, and thereby offer better understanding of 'in the long run' of sampling distributions" (Watkins et al., 1997, p. 11). It is precisely these advantages that make technology a useful cognitive tool in helping students to construct models in order to simulate random experiments (Biehler, 1991; Konold, 1991b). The ability of technology to store and quickly process large amounts of data coupled with the built-in capability of graphically representing data supports the use of technology in a teaching experiment that focuses on probability simulation.

The use of technology in the statistics classroom provides the teacher with additional resources to help students develop statistical reasoning in a more conceptual manner. Technology by itself will not give students immediate access to the solution of a problem. Instead, the power of technology, such as its speed, efficiency, and ability to do tedious manipulations, is what allows students the freedom to concentrate on the components of a simulation. It is the reasoning students do about the structure of a simulation that enables them to make generalizations and abstractions about simulations and probability.

Teaching Experiments Involving Statistical and Probabilistic Reasoning

Teaching experiments have become an important investigative tool into students' individual and collective reasoning. In particular, teaching experiments have been used to illustrate how students reason statistically and probabilistically in an instructional environment. Wares (2001) used a teaching experiment to examine how middle school students construct mathematical models that involved data. Because of the structure of the teaching experiment, Wares was able to design and modify activities to help students extend their reasoning to construct mathematical models. The teaching experiment also provided the opportunity for teacher-researcher questioning of student reasoning in order to plan the hypothetical learning trajectory. Classroom mathematical practices identified during the course of the teaching experiment included students' use of proportional reasoning to defend their mathematical model, their use of graphs for comparison, and the comparison of mathematical quantities using the concept of rate. Two sociomathematical norms were reported: one was that students should be able to defend and justify valid mathematical models, and the other related to the idea of a valid comparison of mathematical quantities.

Polaki (2000) used two versions of a teaching experiment to compare elementary students' growth in probabilistic reasoning. He found that both versions of the teaching experiment had a profound impact on students' reasoning. Further, mathematical practices and norms were generated during the study. The mathematical practices that emerged from the teaching experiment include the use of invented formal mathematical language, a systematic method for generating two-dimensional outcomes, and the use of the composition of the sample space to reason about probability problems. These mathematical practices were considered instrumental in building students' conceptual understanding of probability.

A review of Wares and Polaki's teaching experiments was beneficial because the studies provided a look at the procedures and processes that were utilized in order to examine individual and collective reasoning. In particular, both of these studies provided a picture of how sociomathematical norms and mathematical processes emerged during the course of a teaching experiment. In the light of this background research on the usefulness of teaching experiments, I will now examine the theoretical aspects of the next phase of a teaching experiment.

Classroom-Based Analysis Phase

The second phase of a whole-class teaching experiment is the classroom-based analysis (Figure 2). Whereas the instructional development phase is guided by discipline-specific instructional theory and the focus is on the broader picture of the entire instructional unit, the classroom-based analysis is guided by what is happening in the classroom on a daily basis (Cobb, 2000). It is during this phase that instruction takes place, and the outcomes of the teaching experiment are studied and used to inform the instructional development phase. More specifically, it is this phase of the whole-class teaching experiment that provides the data for constructing and modifying the hypothetical learning trajectory as well as data that is used to analyze individual and collective reasoning and beliefs. Finally, during the classroom-based analysis phase the sociomathematical norms and mathematical practices evolve and are identified by the researcher.

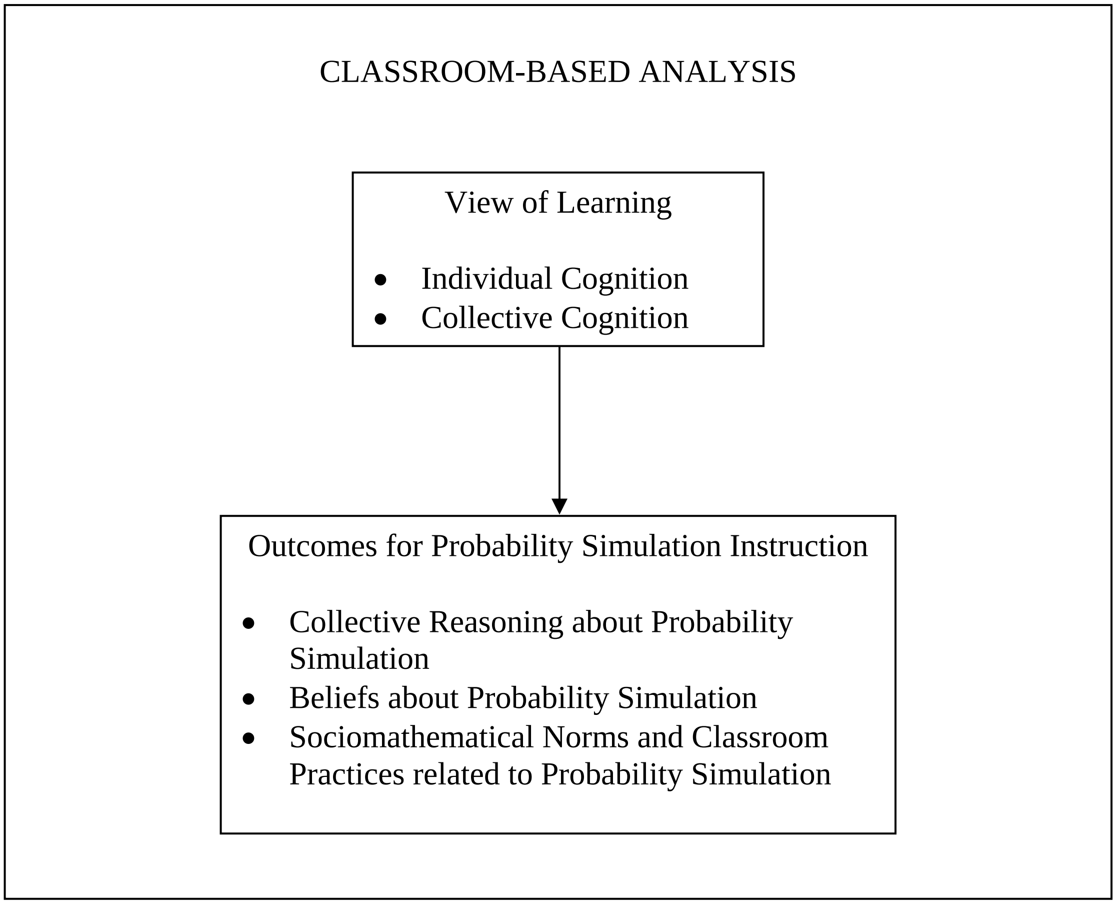

Figure 4. Classroom-Based Analysis Phase.

View of Learning

At the heart of the classroom-based analysis is an examination of individual and collective cognitions (Figure 4), and in order to study these views of learning Cobb's (2000) interpretive framework was used. The interpretive framework Cobb puts forth recognizes both the social and psychological perspectives. The social perspective consists of classroom social norms, sociomathematical norms, and classroom mathematical norms. Classroom social norms refer to the participation structure within the classroom and include the discourse related to student explanations and questioning of alternative perspectives. Examples of social norms may include the expectations of explaining and justifying solutions, making sense of other student explanations, and questioning alternative solutions that may not make sense to other students. Classroom social norms are not specific to mathematics; they are applicable across various disciplines. Sociomathematical norms, on the other hand, are specific to mathematics and refer to the validity of a mathematical solution or answer to a problem. An established sociomathematical norm may relate to what constitutes an efficient or insightful mathematical solution. Classroom mathematical practices consist of "taken-as-shared" ways of reasoning, discussing, and representing mathematical ideas. For instance, in the course of solving the problem 26 + 13, a student may write 20 + 10 = 30 and then 6 + 3 = 9, and finally 30 + 9 = 39. Initially, the student would be expected to justify her solution. However, as classroom discussion progresses this strategy may become "taken-as-shared" without the need for explanation. As such, it becomes a mathematical practice.

The psychological perspective refers to individual student conceptions and beliefs about the mathematical and social activity that occurs in the classroom. The interaction between the social perspective and the psychological perspective represents what Cobb (1999) calls the "emergent viewpoint". According to this perspective, students do not construct knowledge in isolation of the social processes of the classroom. Rather, individual conceptions and beliefs interact in a reflexive manner with social processes. According to the emergent viewpoint, both are considered to be equally important in the construction of knowledge.

Outcomes for Probability Simulation Instruction

The expected outcomes of the whole-class teaching experiment (Figure 3) were guided by the research questions. This study was designed to illustrate how both individual and collective reasoning about probability simulation develops. Moreover, as students reasoned through contextual problems the change in their beliefs related to probability simulations was examined. Finally, the whole-class teaching experiment revealed sociomathematical norms and classroom practices that evolved during instruction.

Significance of the Study

The findings of this study assist in bridging the gap between research and teaching practice and at the same time help to fill a void in the research related to probability simulation. Through the use of a whole-class teaching experiment on probability simulation, this research contributed to the knowledge base on how students develop reasoning and beliefs during instruction, both individually and collectively. Although whole-class teaching experiments have been used to examine other aspects of students' probabilistic and statistical thinking (Benson, 2000; Polaki, 2000; Wares, 2001), there is no evidence of any previous study that has focused on probability simulation. By using a whole-class teaching experiment, this study not only traced and documented how individual students develop conceptions and beliefs about probability simulation, it also examined how the social influences of the class impact collective learning.

Additionally, this study adds to the literature by providing more detail about students' probabilistic thinking at the high school level. There has been limited research that has examined high school students' probabilistic reasoning (Batanero & Serrano, 1999; Fischbein et al., 1991; Fischbein & Schnarch, 1997; Green, 1983) and even less research has focused on probability simulation (Zimmermann & Jones, 2002). This study adds to what little is known about high school students' probabilistic reasoning and at the same time adds to the literature on how students reason about probability simulations.

Finally, this study contributes to research by providing insight into the impact of instruction and technology on students' learning of probability simulations. To study the mathematical concept of probability simulation, it is helpful to conduct many trials and record the results. Today's technology provides the means to accomplish both these tasks, and thus can be used by students as a cognitive tool (Biehler, 1991; Konold, 1991b). In spite of the cited advantages of using technology to study probability simulation, few resources are available for teachers that have examined the role technology plays in the teaching and learning of probability simulation. This research study provides teachers with data about the effect of both instruction and the use of technology while teaching probability simulation, and at the same time provides insights into how the social structure of the classroom affects student learning of probability simulation.

CHAPTER II

REVIEW OF RELATED LITERATURE

The purpose of this chapter is to provide a review of literature that is relevant to this study. The first section provides an overview of research as it relates to students' conceptions and beliefs of probability, simulations, and technology. Research in this area spans all age levels; however, because this study examined the thinking and beliefs of high school students, the review will focus primarily on high school and college-aged subjects. Because of the integration of technology in today's mathematics classroom, the role and use of technology in probability simulation can be found in this section as well. The second and third sections relate the theoretical framework of this study. More specifically, the whole-class teaching experiment is examined as a research methodology in the second section. Thus, a description of a whole-class teaching experiment is provided, including details of the related components. The third section provides an overview of the theoretical framework used specifically for this study. The fifth and final section is a summary of Chapter 2.

Research on Students' Conceptions of Probability and Simulations

The following is a review of research literature that summarizes what is known about probabilistic thinking and simulation. It will focus primarily on high school and college level students with some reference to middle school students. In general, research indicates that regardless of age, probability is difficult to learn and understand. The reviews also reveal that in spite of the conceptual connections between simulation and probability, little research was found investigating how students reason about simulation.

Students' Probability Reasoning

In the past 20 years, many researchers have investigated how students reason when given a probabilistic situation or problem. Many of the researchers have concluded that both children and adults exhibit difficulty reasoning probabilistically (Fischbein et al., 1991; Garfield & Ahlgren, 1988; Hawkins & Kapadia, 1984), and that the factors that affect probabilistic reasoning are varied. In a paper that focused on difficulties both college and secondary students had in understanding probability, Garfield and Ahlgren (1988) reported that college students tend to rely on memorized procedural skills rather than on making sense of the problems. At the secondary level, it was students' lack of reasoning of both proportions and rational numbers that contributed to difficulties in probability reasoning. Part of the struggle students had with probability was attributed to the conflict that they may have encountered when asked to make sense of problems that run counter to their own personal experiences. In other words, Garfield and Ahlren's findings suggest that if students are unable to translate the problem, they will likely have problems applying probabilistic concepts.

In his survey of 3,000 students, Green's (1983) findings were consistent with Garfield and Ahlgren in that he reported that an understanding of ratios and proportional reasoning were key factors influencing students' ability to understand probability concepts. At the same time, Green concluded that students had problems related to linguistics. In particular, students had difficulty in distinguishing between events being "certain" versus "probable" as well as events of "low probability" versus those that are "impossible." What Green determined as a linguistic hurdle seems to support Garfield and Ahlgren's (1988) conclusion that students have trouble translating problems. However, what appears to be an issue related to vocabulary may actually have more to do with heuristics students use to reason probabilistically. More about heuristics will be discussed in the next section. Finally, Green determined that student maturation and intellectual ability had a positive effect on probabilistic reasoning, with the exception of randomness where no improvement related to age was evident. Green's study is noteworthy because of the vast number of students he was able to survey; however, the data lacks any insight into how students reason nor does he verify that correct answers in fact imply appropriate reasoning.

In a study of randomness, Batanero and Serrano (1999) examined high school students' concept of randomness before and after instruction. Some students were able to construct valid arguments about randomness using concepts related to the regularity of patterns, frequencies of possible results, and the existence or absence of runs. However, other students used these same concepts inappropriately when attempting to determine randomness. For example, when examining the outcomes of tossing a coin, some students argued that the outcome was not random because it contained too many heads or that it contained a long run of heads. Some of these students also tended to rely heavily on unpredictability by focusing on such features as luck rather than using quantitative reasoning. This pattern was more prevalent in older students. Additionally, students did not seem to have an intuitive understanding of runs and clusters. Depending on the length of the sequence of outcomes, runs of no more than three or four were expected and longer runs were considered a counterexample of randomness. As noted by Batanero and

Serrano, students' unwillingness to accept long runs or clusters may be related to the difficulty they have in understanding the concept of independence. Moreover, the inappropriate use of these ideas by students proved resistant to instruction. Confirming Green's findings, Batanero and Serrano determined that age was likely to have little influence on students' concepts of randomness. In other words, with respect to ability to consider randomness, the researchers found no significant difference between 14-year old students who had no probability instruction and 17-year old students who had received formal probability instruction.

In summary, many students find probabilistic reasoning challenging for a variety of reasons. Research findings indicated that these reasons were related to maturation, proportional reasoning, concepts of randomness, as well as difficulty in making sense of the problems. These studies helped to provide content-specific background for this study that involved instruction in probability simulation.

Students' Beliefs About Probability

A review of the probability literature on students' beliefs revealed that researchers used varying terms when referring to the dispositions subjects brought to studies on probability. The most common terms used were intuition and beliefs. In order to cast as wide a net as possible in relation to students' beliefs, Schoenfeld's (1985) characterization of beliefs was used. He views beliefs as a lens through which an individual sees and approaches mathematics. When analyzed according to Schoenfeld's characterization, the literature reveals that beliefs play a major role in students' reasoning about probability (Fischbein & Gazit, 1984; Fischbein et al., 1991; Fischbein & Schnarch, 1997; Garfield & Ahlgren, 1988; Piaget & Inhelder, 1975; Shaughnessy, 1992). That is, students' conceptions of probability often involve the use of intuitions, misconceptions, and judgmental biases that are grounded in beliefs (Konold, 1991a; Konold et al, 1993; Tversky & Kahneman, 1973, 1974). Judgmental heuristics provide a framework for interpreting students' probability reasoning, and as researchers have continued to investigate this cognitive area, it has become evident that a dualism exists between misconceptions and heuristics. Below is a description of common heuristics along with what is currently known about how people reason in a probabilistic situation using each cognitive process.

People who apply the representativeness heuristic are making judgments according to how well an outcome represents some characteristic of a parent population (Kahneman & Tversky, 1972). For example, given a fair coin, a person using representativeness would choose an outcome of HHTHTT to be more likely to occur than an outcome of HHHHTT since the first example illustrates a more even distribution of heads and tails. This response is compelling for students using representativeness even though the two outcomes both have probabilities of 1/64. Representativeness is also used to explain the negative recency effect ("gambler's fallacy") and the positive recency effect (Cohen as cited in Shaughnessy, 1992). Negative recency occurs when a person predicts the next playing card to be red because the last five cards have been black. Using the same example, a person who predicts the next card to be black would have reasoned using a positive recency effect. It appears that in a binomial experiment, subjects seem to expect the outcomes to represent a 50-50 distribution. Until more information is available about the distribution of outcomes, subjects seem to adopt a positive or negative recency heuristic. Even when limited information is available about the distribution of a sample, subjects will frequently rely too heavily on this data to make predictions about the entire population. Representativeness is also related to sample size in another way, such as when the size of the sample is not considered in situations of chance (Shaughnessy, 1992). Consider a population of balls that are 50% black and 50% white. Some people believe that the chance of getting at least 7 black balls in 10 draws is the same chance as getting at least 70 blacks in 100 draws (Schrage, 1983). They seem unaware that more extreme events are likely to occur in smaller sample sizes than in larger ones.

The availability heuristic is the determination of the likelihood of an event based on an easily recalled event or process. For example, if asked if it is more likely to die in a plane crash or a car crash, some people may argue that one is more likely to die in a plane crash because of the media coverage given to plane crashes versus car wrecks. In another study (Kahneman & Tversky as cited in Shaughnessy, 1977), subjects were asked if the letter R was more likely to appear in the first position of a word or in the third position. Even though R is more likely to appear in the third position, most subjects responded that it was more likely to appear in the first position. It appears that subjects were able to recall words beginning with R more easily than words with R as the third letter. Availability is a heuristic that is often used when a person is unable to use a normative method (Shaughnessy, 1992), and therefore, personal biases or easily recalled examples are relied upon. To illustrate, when a person is asked if it is possible to make more committees of 2 from a group of 10 or more committees of 8 from a group of 10, most people will answer that more groups of 2 are possible (Shaughnessy, 1977). It appears easier for people to form groups of 2 than groups of 8. According to Shaughnessy, “Availability is a heuristic which affects or reflects our own perception of relative frequency” (p. 472).

The final heuristic is that of the conjunction fallacy. Many people, both statistically naïve and trained, will claim that some types of conjunctive events are more likely to happen than the parent event. For example, Tversky and Kahneman (1973) reported that college subjects considered the event that a person was 55 and had a heart attack to be more likely than the event that a person had a heart attack, regardless of age. One possible explanation for this phenomenon may be that the person was using a representativeness or availability heuristic (Tversky and Kahneman, 1973). In other words, people naturally connect the attributes of age with heart attacks and, therefore, consider the conjunctive event more likely. Shaughnessy (1992) suggested that subjects were in fact confusing conjunction and conditional probabilities. That is, in interpreting the problem, subjects were considering the event that a person has a heart attack given the person is 55. Tversky and Kahneman discount this notion since a more explicit version of the conditional problem yielded an even higher percentage of subjects demonstrating reasoning based on the conjunction fallacy. Shaughnessy retorts that there is not enough evidence to support Tversky and Kahneman’s conclusion.

People who reason using the outcome approach (Konold et al., 1993) hold the belief that they are being asked to predict the outcome of a single event even when they are being asked to find the probability based on the distribution of occurrences. Konold et al. (1993) conducted a study to explore how participants reason in this non-normative way. Many participants correctly identified four sequences of coin tosses to be equally likely when asked which was most likely to occur. However, when asked which of the four sequences was least likely, only 38% of the respondents answered that all four sequences were equally unlikely. When subjects were probed about their reasoning in both situations, Konold et al. characterized their thinking as indicative of the outcome approach. The researchers reported that when being asked about the most likely outcome, some subjects responded as if they were asked to predict what would happen rather than determining the probability of an outcome. According to Konold et al., participants used an outcome approach for the first problem and the representativeness heuristic for the second problem. The source of inconsistencies seemed to stem from the use of multiple frameworks. Depending on certain maxims, a person may reason within a normative framework, an informal judgment framework relying on heuristics, or an outcome approach. This contradicts Kahneman and Tversky's (1972) findings that subjects used a representativeness heuristic.

The above heuristics provide valuable tools that help understand and explain common misconceptions or intuitions students have in developing and applying probabilistic reasoning. They can also prove to be beneficial when developing a framework to assess student understanding of probability. As an example of how heuristics have been used to study probabilistic reasoning, Fischbein and Schnarch (1997) discovered that contrary to their research hypothesis, intuitions of students ranging from 10 years old through college age did not necessarily stabilize with instruction. The effect of the negative recency decreased with age, and yet the conjunction fallacy remained strong through grade 9 and then decreased in older students. By way of contrast, the misconception manifested in the availability heuristic proved to gain strength over time. Fischbein and Schnarch revealed that a mixture of problem content, cognitive conceptions, and intuition influenced the complexity of the students' reasoning. It should be noted that the representativeness and availability heuristics could also prove to be useful tools in probabilistic reasoning. Shaughnessy (1992) suggests that availability can prove to be a useful way to organize information used in making decisions. Further, Borovcnik (as cited in Shaughnessy, 1992) recognizes that the concept of representativeness is fundamental to statistics in that statisticians look to make generalizations about a population from a random sample. Shaughnessy recommends that is our obligation to help our students discriminate between situations when heuristics could be helpful versus situations when they could be detrimental to our reasoning.

Like Fischbein and Schnarch (1987), other researchers found that the multifaceted nature of students' beliefs seems to complicate the effect instruction has on students' reasoning and beliefs. In a study of college students enrolled in an introductory probability course, delMas and Bart (1989) attempted to use instruction to change student misconceptions. The researchers' findings indicated that students possess schemas that they bring into class with them, and these schemas may or may not be accurate. These schemas act as a filter when processing information. If the information fits the schema then it may work to reinforce the process model. Otherwise, if the information is counter to existing schemas, the student may very well ignore it.

Some of the research that examined students' beliefs about probability focused on the effect of instruction. Because instruction was germane to this study, following is a review of some of those studies. In an earlier study, Fischbein and Gazit (1984) explored the effects of instruction on the probability conceptions of students in Grades 5, 6, and 7. It was the hypothesis of these researchers that instruction could be used to alter and develop new intuitive attitudes through participatory activities involving probability concepts. Student understanding did appear to increase with age, however, many of the concepts proved to be too difficult for the fifth-grade students. It was also found that instruction where students were involved in practical activities had the effect of improving some of the students' intuitive misconceptions. However, no assessment was conducted prior to instruction and results were based on comparison of the control and experimental groups.

Using a cognitive framework they developed, Tarr and Jones (1997) designed for middle school students an instructional probability program with particular focus on conditional probability and independence. Students made significant gains on preassessment scores indicating an improvement in probability reasoning. In particular, growth in conditional probability and independence was found. The researchers argue that the topics of conditional probability and independence are accessible to middle school students and as such should be part of the curriculum. What is surprising is that it was reported that many students who were assessed at the highest thinking levels in relation to probabilistic thinking used informal methods to correctly determine the probability of events in non-replacement situations.

Fast (1999) suggested that the use of analogies to teach probability could influence conceptual change. In his study, students completed version A of a probability-testing instrument and then immediately proceeded to version B. Version B was designed using analogies meant to elicit correct responses along with appropriate justifications. For example, in version A, students were asked which birth order, if either, was more likely to occur, that of BGGBG or BBBBB. In version B, students were provided the contextual situation of a 4-digit lottery and asked to compare the chances that a ticket with number 2222 would win with the chances that a ticket with number 2332 would win. In the interview stage, students who incorrectly answered a problem on version A but correctly answered the analogous situation in version B were provided the opportunity to correct their work in version A. Fast concluded that the analogy probe worked to build conceptual knowledge, and students were often able to realize and correct their original mistakes. However, most students did so only with probing questions and guidance by the researcher. The results tend to underscore the importance of the role of the teacher in helping students reflect on common misconceptions. Fast also reports that misconceptions demonstrated by the students were in fact true misconceptions and not errors in readability or understanding the problem as other researchers have implied (Fischbein et al., 1991; Green, 1983; Munisamy & Doraisamy, 1998). Certainly the idea of using analogies to aid students in correcting misconceptions seems to have merit, but with the caveat that further research seems desirable in order to examine the retention level that results from using this method.

In summary, many factors affect how students reason probabilistically. Research indicates that students with a weak understanding of proportions are unlikely to reason appropriately. Numerous studies also reveal that individual beliefs and judgment heuristics have a substantial impact on how students reason probabilistically. Finally, although instruction appears to have limited affect on students’ reasoning and beliefs, maturation seems to improve some concepts but has limited influence on the concept of randomness.

Simulations and Modeling

Simulation and mathematical modeling are inherently integrated. Simulation can be viewed as a specific type of mathematical model. Just as a mathematical model requires an understanding of the underlying mathematical principles and how these principles fit together to help explain some phenomenon, probability simulation also requires a similar understanding. As described in Chapter 1, a probability simulation process is composed of 5 steps that not only require an understanding of probability and randomness but also involve knowledge of how these steps are linked. Thus, in order to construct a probability simulation, a student must be able to construct the simulation model for the given problem. Because students must be able to build an appropriate mathematical model, their ability or inability to do so may help to explain how they reason about probability simulation. Therefore, the research literature on both simulation and modeling are relevant to this study.

Biehler (1991) recognized that computer generated simulations could serve as both problem-solving tools as well as the means to explore the concept of empirical probability. In a more general sense, Biehler (1993) viewed statistical software as an integral part of statistical practice. Statistical software as part of the curriculum can be a double-edged sword. On one hand, these technological tools may actually hamper the statistical knowledge a student learns and is able to apply. “Every statistical software tool introduces its own partly idiosyncratic language, concepts and objects which have to be learned before the tools can be reasonably used” (p. 79). In learning to use a software tool, students may get so caught up in using the software that they do not or cannot make sense of the underlying statistical concepts. On the other hand, Biehler views statistical software tools as a way to provide students with the ability to extend their statistical knowledge. Statistical software provides students with opportunities to extend and explore because of inherent features of the software, such as graphical displays and iterative and interactive ways of analyzing data. Students do not have to be constrained by one solution process and thus can build a conceptual understanding of statistics. More specifically, statistical software can serve as a useful tool for studying probability simulations. Statistical tools have the ability to generate random numbers and thus lend themselves to the design and implementation of random experiments (Biehler, 1993). To understand the "lawlike behavior underlying chance variation" (p. 96) in probability simulations, students must be able to analyze long run behaviors of random experiments.

Borovcnik and Peard (1996) admitted that simulation could be used to find a solution to a problem. However, they questioned the value of simulation in helping to explain why simulation produces an acceptable answer. Furthermore, Borovcnik and Peard raised concerns about the use of computer-generated simulations compared with simulations that use actual manipulatives. Their apprehension was related to the manner in which students might perceive the "pseudo-randomness" of computers. To circumvent any possible misconceptions on the part of the students, Borovcnik and Peard suggested that initially teachers should use manipulatives to present the concept of simulation and then compare simulated results with manipulatives to simulated results using technology. Thus, the manipulatives help the students to build cognitive models that, in turn, assist students in building more abstract representations. The use of manipulatives could help students make connections between the randomness inherent in such devices as dice, cards, colored balls, and the more abstract understanding of simulation and randomness built into technology, for example the graphing calculator. Manipulatives may also be necessary to help students make sense of a probability simulation for a contextual problem. First they must think about the components of the problem. Specifically, they need to choose an appropriate probability generator, and then visualize the sample space of the contextual problem in order to determine what a valid trial looks like. In this process, students are going beyond applying mathematical skills to solve the problem. Rather, they are beginning to abstract the processes required to reach a solution.

In a study in which 13 and 17 year old students used a simulation program to study conditional probability, Wilder (1994) reported the complexity of using a computer model. At the same time, Wilder concluded that the students' ability to make sense of the problems was the impetus to their understanding rather than the simulation itself. Early in the experiment, the students needed to model the simulation with actual cards before simulating the problem on the computer. After completing various simulations using both actual manipulatives and a computer-generated simulation, some students seemed to reach a point where they were so convinced of the accuracy of their computer model of the problem that they felt an actual simulation with manipulatives was unnecessary. In essence, they had built a more abstract understanding of the simulation problem. As a result, Wilder suggested that the process of modeling might prove beneficial to student understanding of probability. Moreover, he felt that the steps students needed to think through in order to actually collect data to determine the probability of a problem would help them make generalizations about probability concepts. Wilder's findings support Borovcnik and Peard's (1996) recommendations that simulations could be useful in helping students develop more abstract models and that manipulatives should be used prior to computer simulations.

The use of probability simulations was reported to increase the level of understanding of experimental probability in a group of sixth-grade students. In particular, Aspinwall and Tarr (2001) examined the effect an instructional program had on students' understanding of sample size in probabilistic situations. Researchers found that after an instructional program involving probability simulation using manipulatives, most students developed a higher level of conceptual understanding of the law of large numbers. More specifically, students were able to recognize the relationship between the number of trials and the probability of an event. However, Aspinwall and Tarr also reported that the results of individual simulations served to reinforce the misconceptions of some students. The researchers concluded that the use of instruction and probability simulation could be used to challenge students' thinking of experimental probability in order to develop conceptual understanding of the role of sample size.

In a research study investigating students' reasoning with respect to probability modeling (Benson, 2000; Benson & Jones, 1999), the researchers found that 6 of the 7 students interviewed (ranging from Grade 2 through college) were able to use an appropriate probability generator in a simulation activity. In fact, most of the students used a 1-1 correspondence to explain their choice of probability generators. That is, students were able to match the corresponding outcomes and their probabilities on the probability generator with those in the contextual problem. However, only 3 of the 7 students in their study were able to construct a valid model for a contextual problem involving two-dimensional probability. Two of the students who could construct a valid model were college-age and the third was in grade 12. Benson and Jones concluded that students' ability to construct two-dimensional models seemed to be closely related to their knowledge of two-dimensional theoretical probability. Note that while modeling the contextual problem is an important process in simulation, it is only the first step. In other words, students must be able to model the problem before they are able to proceed through the simulation process.

Supporting the findings of Benson and Jones (1999), Zimmermann and Jones (2002) found that, in their study on probability simulation involving 9 high school students, the concept of dimensionality again proved difficult for most students. For example, when asked to design a simulation to determine the probability of hearing hip-hip songs both times a radio was turned on, most students were unable to recognize that a trial consisted of an ordered pair (a two-dimensional trial). Specifically, the trial in this problem comprised the type of song heard at each of two times the radio was turned on. Students' inability to understand the concept of dimensionality also prevented most of them from recognizing that the suggested trial in another simulation was one-dimensional when it should have been two-dimensional. However, like Benson and Jones (1999), Zimmermann and Jones found that most students were able to construct a valid probability generator, that is, a random device that produced the specified probability distribution in the given problem.

Although little research was available that examined students' reasoning about probability simulation, even less was found that addressed beliefs students held related to simulation. Zimmermann and Jones (2002) determined that students bring with them beliefs about simulations that can be both helpful and problematic to instruction. One such helpful belief held by more than half the students in this study, was the belief that the problem contained unspecified assumptions. To some degree, all students held the belief that the probability generator should correspond to the probabilities stated in the problem context. More specifically, two-thirds of the students either explicitly or tacitly believed the stated probabilities in the problem should be matched to the probability generator. Six of the 9 students exhibited the helpful belief that the empirical probability will approach the theoretical probability as the number of trials increase. With respect to problematic beliefs, 5 of the 9 students, albeit with varying degrees, expressed the belief that a simulation could not or should not be used to model a real world problem. The researchers also found evidence of representativeness (Kahneman & Tversky, 1972) by students as they reasoned through the Pizza problem. Four of the students believed the outcomes of the simulation should always approximate the given probabilities; that is, that the outcomes of the sample should reflect the given probabilities of the population. Finally, 2 students held the problematic belief that the simulation was flawed unless the target outcome in the problem appeared in the first trial. In essence, they seemed to ignore the possibility that the other outcomes in the sample space might occur. Both the helpful and problematic beliefs found by Zimmermann and Jones proved to be beneficial in providing a framework for assessing student beliefs throughout a whole-class teaching experiment.

Students' ability to construct models of problems is likely to factor into their ability to construct appropriate simulations. Yet only recently have research studies examined students' thinking in relation to mathematical modeling. According to Lesh et al. (1997), model-eliciting activities require that students mathematically interpret situations and problems. In the process of constructing models, students develop a multitude of unanticipated and effective mathematical models to solve problem (Lesh et al., 1997; Lesh & Clarke, 2000; Lesh & Lehrer, 2000). Biehler (1994) argues that in random experiments, such as probability simulations, what is the focus of interest is the probability generator. Through the process of constructing or modeling the probability generator, information is gained about the distribution of the population. Construction of models tends to begin at an intuitive level and graduates to more sophisticated reasoning which in turn enables the students to produce a more generalizable model.

Technology

Probability simulation can help students to connect empirical and theoretical probabilities providing students have an opportunity to examine the long run behavior of a simulation (Watkins et al., 1997). However, time constraints in generating sufficient trials during a classroom session have generally limited students' ability to actually examine long runs. Now with the advent of technology, teachers possess the means to overcome this hurdle. Additionally, the use of technology is supported and encouraged by mathematics education groups. Both NCTM (2000) and the College Board (Watkins et al., 1997) have published expectations that students should have regular access to both computers and calculators with statistical computation and graphics capabilities.

The College Board (Watkins et al., 1997) recognizes the role of technology both in today's mathematics curriculum as well as outside academia. As a result, teachers of statistics can choose to focus on either or both of the following uses of technology: using it as a tool or using it conceptually. On the one hand, they can choose to instruct students on the statistical computing tools used by the professional community. On the other hand, teachers can utilize technology to facilitate conceptual understanding of statistical ideas. This research study will concentrate on the use of technology to help students develop a conceptual framework to reason statistically.

Not withstanding the caveats of Borovcnik and Peard (1996), technology offers advantages (Watkins et al., 1997) over the use of manipulatives. Some of these advantages include the ability to store large data sets that can then be manipulated and explored in order to help build a conceptual understanding. Furthermore, technology provides the advantage of speed coupled with the ability to visualize simulations. All of these advantages taken together help students to better understand the long run behavior of sampling distributions (Watkins et al., 1997), and at the same time provide the means by which to represent data graphically. Biehler (1991) and Konold (1991b) both support the use of technology in this way because they believe it enables students to construct models to simulate random experiments and, thus, develop an understanding of simulation. Technology provides students with a manner in which to conduct simulations, especially when the use of physical manipulatives can be limiting. Some of these limitations include the time and perseverance required for a large number of trials, the methods used by students to conduct these simulations, and the true “fairness” of manipulatives (Jiang & Potter, 1994).

The use of technology in the classroom goes beyond that of manipulating and analyzing data. Statistical software can act as a pedagogical tool that helps students to “bridge mathematics and ‘real’ life by opening access to modeling of concrete situations and real data” (Balacheff & Kaput, 1996, p. 478). Ben-Zvi (2000) suggests that such tools support learning and teaching by providing students with opportunities to actively construct knowledge, reflect on “observed phenomena,” and develop their own metacognitive processes. Because of the technology that is currently available, teachers have the opportunity to de-emphasize tedious computations in favor of greater “emphasis on statistical reasoning and the ability to interpret, evaluate, and flexibly apply statistical ideas” (Ben-Zvi, 2000, p. 130).

Dörfler (1993) views technology as a means for human cognitive activity. That is, technology can become a tool for the cognitive development of students’ reasoning and understanding of mathematical concepts. For example, student use of technology lends itself to what Ruthven (1996) terms “proximation strategies.” By proximation strategies, Ruthven means that when a student is unable to directly solve a problem, she may explore different strategies in search of a solution. Technology provides the student with more options and strategies for solving problems that can lead to mathematical understanding.

The above research is a clear indication of support for using technology in mathematics, and in particular, statistics. The use of technology in statistics can become a resource that helps students develop statistical reasoning beyond that of rote application of formulas and procedures. Technology by itself will not give students immediate access to the solution of a problem. Instead, the power of technology, such as its speed, efficiency, and ability to do tedious manipulations, is what allows students the freedom to concentrate on the conceptual components of a simulation. It is the reasoning students do about the structure of a simulation that enables them to make generalizations and abstractions about simulations and probability.

Teaching Experiments

The research methodology used for this study was a whole-class teaching experiment. A detailed description of the theoretical components of the teaching experiment follows, as well as how the teaching experiment provides the theoretical framework for this study in particular.

Theory of Teaching Experiments

As described in Chapter 1, a whole-class teaching experiment is a conceptual tool for recording and analyzing how student understanding develops for a specific content domain (English et al., 2002), such as probability simulation. Through the use of whole-class teaching experiments, researchers are able to capture a detailed picture of how student understanding develops through the interaction of teacher and students. In a whole-class teaching experiment, the researcher develops instructional lessons and activities including conjectures of how students may react to these lessons. By analyzing student thinking, the researcher then makes refinements and changes to the conjectured learning process in order to promote student understanding of the targeted content domain. It is through the analysis of student thinking that a working theory of the development of student thinking emerges.

Whole-Class teaching experiments go beyond providing a tool to describe student thinking as it develops. According to Cobb (2000) teaching experiments act as a mechanism for capturing how group interaction affects the mathematical development of students. This interaction can take place at the level of teacher and student or at the level of student and student. The examination of the social interaction that takes place within the classroom as revealed by the teaching experiment (Cobb, 2000) can provide valuable information for classroom teachers.

The Developmental Research Cycle (Figure 2) is the two-phased process that Cobb (2000) uses to direct the whole-class teaching experiment. The Developmental Research Cycle consists of the instructional development and classroom-based analysis phases. The instructional development phase comprises the sequence of instructional lessons and activities that are based on an instructional theory, and the classroom-based analysis phase is where student thinking and reactions are analyzed using an interpretative framework. Following is a detailed description of both phases of the Development Research Cycle.

Instructional Development

A whole-class teaching experiment begins with the instructional development phase (Figure 3). The initial step is the development of a hypothetical learning trajectory (Simon, 1995) that consists of three components: (a) the development of learning goals, (b) a plan of instructional lessons and activities, and (c) a conjecture or hypothesis of the learning process. According to Cobb (2000), the initial hypotheses are based on a domain-specific theory. This domain-specific theory is the instructional theory that the researcher uses to help with the planning and implementation of the whole-class teaching experiment.

The domain-specific theory that Cobb (2000) relied upon and is reflected in his emergent perspective was that of the Realistic Mathematics Education (RME) theory. As described in Chapter 1, Cobb’s emergent perspective perceives both the social structure of the classroom and individual cognitions as equally critical in students’ mathematical development. RME relies upon the principles that students learn mathematics in a constructivist manner and development of mathematical understanding is a result of a student's ability to reflect on mathematical problems in a realistic or meaningful way. RME theory assumes teacher awareness of student mathematical knowledge and ways of thinking from which a sequence of learning activities are developed. Cognitively guided instructional theory was the domain-specific theory utilized in this study, and is described in more detail toward the end of this section.

Classroom-Based Analysis

In this phase of the Developmental Researcher Cycle (Figure 2), Cobb (2000) requires an interpretive framework that is essential for analyzing and implementing a teaching experiment. The interpretive framework that defines the constructs Cobb uses in his emergent perspective for examining both student reasoning and interactions within the classroom fall within two perspectives; the social perspective and the psychological perspective. The three components of the social perspective include: (a) classroom social norms, (b) socio-mathematical norms, and (c) classroom mathematical practices.

The classroom social norms define a structure that is used to define and analyze the classroom interactions (Cobb, 2000). As described in Chapter 1, these interactions can be between students, whether in pairs, groups, or during whole-class discussions. These interactions also include communications between students and teacher. In a research study involving first-grade mathematics students, Cobb described the participation structure that developed when the teacher indicated her desire for students to share their thinking. The unfamiliar nature of this expectation caused a renegotiation of classroom social norms between the students and teacher. Thus, both students and teacher created a participation structure on which classroom discussions would evolve. The classroom social norms that evolved included what was expected in justifying a solution as well as how to disagree with another student's solution in an acceptable manner.

Whereas classroom social norms are more general in nature and applicable to any classroom environment, <u>socio-mathematical norms</u> are specific to students' mathematical activities and provide the means to examine the participation structure as related to mathematical discourse (Cobb, 2000). Examples of socio-mathematical norms include what students negotiate to be a mathematical solution, an "elegant" or sophisticated mathematical solution, and an alternative mathematical solution. Cobb explains that when a teacher asks for a different solution, students in turn must determine or renegotiate what constitutes a different solution. When the class accepts the renegotiated meaning of a different solution, this taken-as-shared meaning has become a socio-mathematical norm. Cobb views both the classroom social norms and the socio-mathematical norms as being critical to a teaching experiment. Together these norms inform the level to which students have achieved intellectual autonomy during classroom instruction.

The third component of the social perspective is the <u>classroom mathematical practices</u>. Classroom mathematical practices relate to solution strategies invented by individual students that then become taken-as-shared by the rest of the class. Classroom mathematical practices are ways of reasoning upon which students agree. For instance, consider the case of a student doing whole number addition. A student may choose to solve the problem 26 + 18 by rewriting the problem as 20 + 10 and 6 + 8 and then add the results. The classroom social norms and socio-mathematical norms require that a student provide a reasonable explanation of her solution. At some point, this strategy would become taken-as-shared by the class, and thus becomes a classroom mathematical practice because explanation is no longer needed.

Cobb's (2000) psychological perspective recognizes the contributions of the individual student to her own learning. These individual contributions include beliefs about the nature of a mathematical activity, the role of the individual and others in the activity, the individual's beliefs and values about mathematics, and mathematical conceptions. According to Cobb's (2000) emergent perspective, it is the role of the researcher to analyze how the social and psychological perspectives interconnect. In other words, it is the researcher's responsibility to examine how the social structure of the classroom and the individual students interact in order to develop a working theory of how students' reasoning of a mathematical concept evolves.

Cognitively Guided Instructional Theory

As stated earlier, the initial conjectures that are part of the instructional development phase of the developmental research cycle are guided by an instructional theory (Cobb, 2000). The instructional theory used for this study was informed by cognitively guided instructional theory (CGI) (Carpenter & Fennema, 1988). CGI is a theory that integrates what is known about students' cognitive processes with pedagogy. In other words, a basic tenet of CGI is that a teacher makes instructional decisions based on what is known about the student's thinking. In their study, Carpenter, Fennema, Peterson, Chiang and Loef (1989) implemented CGI by first informing teachers of current research on student reasoning of whole number arithmetic. Teachers were then supported to use this information and their own assessment of student knowledge and ability to inform their instruction. Carpenter et al. found that teachers who put into practice a CGI strategy were more likely to have focused on developing student knowledge through a problem-solving approach than were the control group. Additionally, the CGI teachers tended to promote the student use of multiple problem-solving strategies. CGI emphasizes the process of student solutions rather than the solutions themselves in order to provide the teacher with an opportunity to assess student thinking and at the same time encourage a variety of problem-solving strategies so students may connect to their own knowledge and thus build on it.

The intent of this section was to provide an overview of the general theory that underlies the teaching experiment. At the same time, it provided an explanation of the crucial components of a teaching experiment. The next section provides a review of some research utilizing a whole-class teaching experiment methodology that is similar to this study.

Research on Whole-Class Teaching Experiments

Teaching experiments have become an important investigative tool into students' reasoning. In particular, teaching experiments have been used to illustrate how students reason statistically and probabilistically in an instructional environment. Wares (2001) used a teaching experiment to examine how middle school students construct statistical mathematical models. Because of the structure of the teaching experiment, Wares was able to design and modify activities to help students extend their reasoning to construct mathematical models. It also provided the opportunity for teacher-researcher questioning of student reasoning in order to plan the hypothetical learning trajectory. Two sociomathematical norms appeared. One was that students should be able to defend and justify valid mathematical models. The other related to the idea of a valid comparison of mathematical quantities. Classroom mathematical practices identified during the course of the teaching experiment included students' use of proportional reasoning to defend their mathematical model, their use of graphs for comparison, and the comparison of mathematical quantities using the concept of rate.

Polaki (2000) used two versions of a teaching experiment to compare elementary students' growth in probabilistic reasoning. He found that both versions of the teaching experiment had a profound impact on students' reasoning. Further, the study generated mathematical practices. The mathematical practices that emerged from the teaching experiment included use of invented formal mathematical language, a systematic method for generating two-dimensional outcomes, and using the sample space to reason about probability problems. These mathematical practices were considered instrumental in building students' conceptual understanding of probability.

A review of Wares and Polaki's teaching experiments was beneficial because the studies provided a look at the procedures and processes that were utilized in order to examine the individual and collective reasoning that developed. In particular, these studies provided a picture of how sociomathematical norms and mathematical practices developed during the course of a teaching experiment. The next section focuses on the theoretical framework that drove this study. Similar to Wares' and Polaki's teaching experiments, this framework is an adaptation of the approach used by Cobb (2000).

Theoretical Framework

This study followed the Development Research Cycle (Figure 2) as adapted from Cobb (1999). This cyclic process consisted of the instructional development phase and the classroom-based analysis phase. The instructional development phase was driven by the hypothetical learning trajectory that was informed by an instructional theory. For this whole-class teaching experiment, the theory used was cognitively guided instructional theory (Carpenter & Fennema, 1988), which, as described previously, relies on research-based knowledge to inform instruction. Research on simulation (Zimmermann & Jones, 2002; Benson, 2000; Benson & Jones, 1999) and the use of technology for simulation purposes (Biehler, 1991; Konold, 1991b; Watkins et al., 1997) guided the instructional theory component, including the design and selection of tasks as well as the interpretation of student reasoning and beliefs. The instructional development phase was also predicated on the crucial role that both individual and collective cognition played in informing the hypothetical learning trajectories.

The intent of this study was to document changes in students' reasoning and beliefs throughout the course of a teaching experiment. Moreover, an integral part of examining student cognitions is the social structure of the classroom environment (Cobb, 2000). Cobb's interpretive framework was used to analyze the interactions within the classroom as they evolved throughout the whole class teaching experiment. Thus, emphasis was placed on documenting and analyzing the emergence and evolution of classroom social norms, socio-mathematical norms, and classroom mathematical practices. Changes in students' reasoning and beliefs about probability simulation were observed in order to encapsulate the psychological aspects of students' individual thinking.

Summary

A review of the literature reveals that probability is a difficult concept for all students. Further, studies suggest that a mixture of problem content, cognitions, and beliefs contribute to how students' reason probabilistically. Although many studies have examined student beliefs and reasoning about probability, little research has documented how students reason about probability simulation. The research that is available indicates that the two-dimensional nature of probability problems (Zimmermann & Jones, 2002; Benson, 2000; Benson & Jones, 1999) is a critical factor in students' abilities to reason about probability simulation. Furthermore, researchers seem to believe that technology can play a crucial role in helping students to reason about probability simulation. This study sought to address the void in the literature by examining the reasoning and beliefs of high school students about probability simulation, as well as the impact technology had on students' conceptual development of simulation. Additionally, this study examined the emergence and evolution of societal factors in the classroom that contributed to this learning environment.

CHAPTER III

THE METHODOLOGY

This study utilized a whole class teaching experiment methodology (Cobb, 2000) to examine the emergence and evolution of students' reasoning and beliefs about probability simulation. The teaching experiment methodology also provided the framework by which to examine and document the social environment of the classroom as it contributed to student cognitions. In particular this chapter describes the processes and procedures used to respond to the research questions on students' individual and collective cognitions about probability simulation. It includes specifics about the student sample, the instructional program, procedures of the whole-class teaching experiment, the data collection, and the methods of analysis.

Participants

The participants for this study were 23 students enrolled in an Advanced Placement Statistics (APS) class in a Midwest high school. The prerequisite to enroll in the class was a grade of "C" or better in an Advanced Algebra course. Thus, the level of mathematical ability of the students varied from average to high-level ability. Three of the students were at the junior-level and were taking APS concurrently with an honors-level precalculus course. The remaining 20 students were senior-level students having 3 years of mathematics prior to their senior year. Six of the 20 seniors were concurrently enrolled in an Advanced Placement Calculus course and 1 senior was enrolled in honors precalculus. The content of a student's previous course determined the level of probability instruction he or she had prior to this class. Thus, the extent of probability instruction ranged from simple probability problems involving tree diagrams and basic counting principles to introductory instruction involving permutations and combinations to a unit of probability involving conditional probabilities.

Four of the students were purposefully sampled (Miles & Huberman, 1994) in order to create a case-study analysis. The rationale of the case-study analysis was to provide a more in-depth view of student reasoning and beliefs as they evolved during the whole-class teaching experiment. Based on the results of a preassessment (discussed in more detail later), 1 student was selected from each of the upper and lower quartiles and 2 students were selected from the middle quartiles in order to provide a range of student abilities. Students were also chosen based on their ability to communicate their reasoning and beliefs in an effective manner.

As the teacher-researcher, I was also a participant in this whole-class teaching experiment. As the name implies, I was both the daily classroom teacher of the APS students and the researcher conducting the whole-class teaching experiment. I am a full-time high school mathematics teacher with 7 years experience in the classroom.

Finally, both a witness and an additional researcher were participants in this research study. The witness was present during the entirety of the whole-class teaching experiment. She is a former high school mathematics teacher with over 30 years experience, and her role was to act as another pair of eyes in the classroom. The witness recorded what transpired during the instructional lessons, as well as collaborated with me to develop and revise the hypothetical learning trajectories for the subsequent instructional sessions. She also coordinated the videotaping of the teaching experiment.

More specifically, the witness focused her attentions on the 4-student case study group and attempted to identify the emergence and evolution of classroom sociomathematical norms and mathematical practices. An additional researcher was brought in to conduct the student case-study interviews in order to protect against any bias students may have exhibited had their teacher or the witness interviewed them.

Procedures

A number of key components accompanied the teaching experiment: a) assessments and interviews, b) classroom setting, and c) follow-up activities.

Written pre-, post-, and retention assessments were given to all students. The case-study students participated in interviews following the pre- and post-assessments. Further, the interview sessions of the 4 case-study students were each audio and videotaped, and the additional researcher recorded field notes as well as collected student artifacts that were created during the interviewing process.

For the whole-class teaching experiment students were assigned to groups of 3 to 4 students based on student responses to the preassessment and teacher knowledge of the students. Each group was made as heterogeneous as possible according to their assessed knowledge of probability simulation. The whole-class teaching experiment consisted of twelve 55-minute period instructional days. Two video cameras and six audio recorders captured each instructional episode.

Debriefing/planning sessions were conducted after each instructional session where the teacher-researcher recorded impressions, recollections, and any other thoughts relevant to the whole-class teaching experiment. The purpose of these debriefing/planning sessions with the witness was to develop and refine the hypothetical learning trajectories. Finally, student artifacts, such as selected homework assignments, were also collected.

Instrumentation and Interviews

A combination of assessment instruments and interviews were used to capture and document changes in students' reasoning and beliefs about probability simulation.

Instrumentation

All students participating in the study were administered pre-, post-, and retention assessments (see Appendix A) during the course of the study. The preassessment was administered approximately 2 weeks prior to the commencement of the whole-class teaching experiment. This timeframe enabled the researcher to analyze the preassessments and select the 4 case-study students. At the conclusion of the 12-day whole-class teaching experiment, students were given a post-assessment, and approximately 4 weeks after the post-assessment, a retention assessment was administered. The paper and pencil assessments were an adaptation of an assessment protocol used in a previous study by the teacher-researcher (Zimmermann & Jones, 2002). The three assessments contained parallel, researcher-constructed items designed to assess students' abilities to assess and construct valid probability simulations. The intent of the pre-assessment was to provide baseline data to compare with later assessments, and it also provided the means by which the 4 case-study students were selected.

The focus of the assessment instruments was on two-dimensional contextual probability problems involving simulation. The contexts of the problems were chosen to be accessible to teenage students. All three of the assessment instruments were mathematically isomorphic; hence, in describing the assessments I will use the preassessment instrument as the exemplar. The preassessment consisted of two tasks. In the first task, the Pizza Problem, a fictional student, Stan, designed a simulation to determine the probability that the next two phone orders were for pizza with meat. Students were asked to determine the validity of Stan's simulation design. In the second part of the Pizza Problem, students were provided with Stan's simulation results and were asked to determine the probability that the next two phone orders were for pizza with meat. The intent of the first part was to determine the extent to which students were able to assess the validity of a simulation design. The purpose of the second part of the Pizza Problem was twofold: a) to determine if students were able to recognize the inappropriateness of the outcomes as they were recorded, and b) to ascertain if the students could determine whether or not the recorded results allowed them to calculate the required empirical probability. It should be noted that the outcomes were recorded as single outcomes and not as pairs as was required to solve the two-dimensional pizza problem. This continuous string provided students with the opportunity to recognize the inappropriate manner of recording the outcomes yet allowed them to determine the empirical probability of the problem.

In the second task, the Radio Problem, students were given a situation involving a school radio station with a specified airtime format for three different types of music: hip-hop, alternative, and country. Students were asked to design a simulation to determine the probability that hip-hop was played at two specified times when the radio was turned on. The next part of the Radio Problem asked students to make up data that they thought their simulation would produce and then to use this data to determine the probability that hip-hop was playing both times the radio was turned on. The goal of this second task was to determine if students could construct a simulation that incorporated a valid probability generator, an appropriately defined trial, and a properly determined empirical probability. Finally, students were asked about the effect of increasing the number of trials on the empirical probability for the Radio Problem. Students were encouraged to provide as much detail as possible and thoroughly explain their responses.

Interviews

Both before the teaching experiment began and within a week after the post-assessment was given, the additional researcher interviewed the 4 case-study students. Specifically, the students were questioned about their responses to each of the pre- and post assessments. This purpose of the interviews was to develop a more in depth picture of their reasoning and beliefs related to the assessment items.

The clinical interviews (Romberg, 1992) were conducted in a small, quiet room within the school, with the student situated across a table from the researcher. Each interview lasted approximately 25 to 45 minutes depending on the student's responses. In addition to paper and pencil, students were provided with probability generators: colored chips, dice, spinners, a random number table, and graphing calculator. During the interview, each student was provided with a copy of his or her assessment responses. Building from student responses, the researcher asked questions aimed at probing student thinking and beliefs related specifically to the assessment instrument and probability simulation in general. Each interview was audio and videotaped.

Developmental Research Cycle

Following the pre-assessment, the teacher-researcher conducted a 12-session whole-class teaching experiment. The Instructional Development was one phase of the

Developmental Research Cycle, and, as indicated in Figure 2, the research base on probability and simulations developed in Chapter 2 together with the Classroom-based Analysis helped drive the development and refinement of the hypothetical learning trajectories.

Instructional Development Phase

Prior to each of the 12 sessions, the teacher-researcher and witness collaborated to develop a hypothetical learning trajectory for the subsequent instructional session. The hypothetical learning trajectory consisted of learning goals, instructional activities, and conjectured learning processes. The teacher-researcher and witness started with the learning goals and then developed an instructional activity that would be used to introduce and develop the learning goals. Finally the teacher-researcher and witness used research-based knowledge of probability and simulation along with the classroom-based analysis to construct hypotheses about the learning processes. In other words, the teacher-researcher and witness would write down anticipated responses and actions students might exhibit as they worked through the instructional activity.

To better illustrate the development of the hypothetical learning trajectory, Session 1 of the whole-class teaching experiment will be used as an exemplar. Prior to instruction, the teacher-researcher and witness met after school to discuss the first session. To begin with, learning goals were determined. For session 1 (see Appendix B), the learning goals focused on the component parts of a simulation, including the ability to construct a valid probability generator, define a trial, use the probability generator to simulate the problem and analyze the results. Next the teacher-researcher and witness chose an activity titled “Counting Successes,” (Scheaffer, Gnanadesikan, Watkins &

Witmer, 1996). This activity was chosen because it met the learning goals as previously determined and the only prerequisite knowledge was a basic understanding of proportions. The activity, “Counting Successes,” was also chosen as it involved a simpler one-dimensional contextual situation rather than a more difficult two-dimensional problem. In the activity (see Appendix C1), students were given an article in which a scheme was suggested for eliminating the need for change when making purchases. The activity required the students to simulate the scheme given in the article. Students were familiar with using a random-number table, thus, it was utilized for this activity. Students worked through the activity in groups, and when groups had finished the problem, the teacher-researcher facilitated a whole-class discussion posing questions that encouraged students to share their strategies and reflect on the reasoning of other students.

As a part of the hypothetical learning trajectory, the teacher-researcher and witness hypothesized about the learning processes that students would experience and conjectured about anticipated problems students may encounter. For this instructional activity, it was conjectured that students would not have any difficulty constructing a valid probability generator using the random-number table. However, it was expected that some students would solve the problem using theoretical probability or some would question the randomness of their results if the outcomes of the simulation did not seem to match the stated probabilities of the problem.

The instructional development phase for subsequent sessions followed a similar pattern for developing the hypothetical learning trajectory. Table 1 shows a summary of the activities used by the teacher-researcher as well as the major concepts intended by the activity. Appendix B contains the detailed hypothetical learning trajectories for each of the 12 sessions.

Table 1

Outline of Instructional Sessions for Whole-Class Teaching Experiment

Session	Instructional Activity	Concepts
1	Counting Successes	1D, Design Simulation
2	True/False History Test	MD, Design Simulation
3	Free Throw Shooter and Blood Bank	MD, Design Simulation
4	Randomly-Generated Outcomes	1D, Randomness, Technology
5	Designing Simulations on the TI-83	MD, CS, Technology
6	Tree Diagrams	MD, Theoretical Probability
7	Venn Diagrams – Day 1	Theoretical Probability
8	Venn Diagrams – Day 2	Conditional Probability
9	A “False Positive” Aids Test	Theoretical vs. Empirical
10	What is Random Behavior?	Representativeness
11	What’s the Chance?	Independence of Events
12	Are these simulation designs valid?	Validity of Simulation Design

Note. 1D: One-Dimensional Sample Space
MD: Multiple-Dimensional Sample Space

For this research study, the major learning goal for students was to develop a conceptual understanding of probability simulation that would enable them to recognize and construct valid probability simulations. The planned learning or instructional activities targeted two-dimensional probability simulations similar to the ones in the assessment. However, the activities began with one-dimensional situations in order to build student understanding more effectively. Figure 5 shows some of the tasks used during instruction. For example, WCTE1 (whole-class teaching experiment – Session 1) refers to an activity involving a one-dimensional outcome. Sessions 4, 10, and 11 also required students to consider one-dimensional outcomes, whereas sessions 2, 3, and 5 involve two-dimensional outcomes. Previous research-based knowledge, teacher knowledge of students, the pre-assessment instrument, as well as the APS curriculum also drove instruction. The conjectured learning process began with research-based knowledge of students' reasoning and beliefs, the teacher's prior knowledge of the students' mathematical knowledge, student reasoning gained by the teacher-researcher from the preassessment, and elements of the APS curriculum. The conjectured learning process evolved as the teacher-researcher and witness refined the hypothetical learning trajectory following each session of the whole-class teaching experiment.

WCTE1 “Counting Successes” Students are given a newspaper article about eliminating the need for pennies. Using a random number table, students design a simulation to determine the “fairness” of the proposal. What is “fairness”? Is this proposal fair? Why or why not?	**WCTE2** “True/False History Test” There are five questions on the test. Design a simulation to determine the probability of getting 3 correct answers on the test. After 50 simulations, what is the average number of correct responses?
WCTE3 “Free Throw Shooter” The star free-throw shooter on the girls’ basketball team makes 80% of her free throws. She gets about 10 such shots each game. Design a simulation. What does a trial look like? What is the approximate probability that she makes more than 80% of her free throws in one game? How many free throws should she expect to make in a <u>typical</u> game? Explain.	**WCTE4** “Randomly-Generated Outcomes” You will randomly be assigned to one of the following groups. <u>Group 1</u>: Using your calculator, simulate the outcomes of 100 coin tosses. Record your results as a string of H’s and T’s. <u>Group 2</u>: Make up the results of flipping a coin 100 times. Record your results as a string of H’s and T’s. Which results are calculator generated and which are student generated? How can you tell?
WCTE5 “Designing Simulations on the TI-83” Using your calculator, design a simulation for the following situation. The percentage of women in the labor force of a certain country is 30 percent. A company employs ten workers, two of whom are women. Estimate the probability that a company of ten workers would employ two or fewer women by chance. On the basis of your simulation, do you think that women are underrepresented in the company? Why or why not?	**WCTE10** “What is Random Behavior?” Students are given a bag of Jolly Rancher candies. Without looking, groups are to develop a rule for predicting the next Jolly Rancher to be chosen from the bag. Do the data from the past selections help in making a prediction for the next selection? Does the particular pattern of the sequential selections help in making a prediction for the next selection? What is the best you can hope to do, in terms of correct predictions, with any decision rule? Discuss what was learned about predicting future events in a random sequence.
WCTE11 “What’s the Chance?” Students are given 10 tacks and asked to complete 2 experiments: 1) Toss 1 tack ten times and determine empirical probability the tack point will land down; 2) Toss all 10 tacks simultaneously and determine empirical probability. Was there a difference in the results? Explain what may account for the difference.	**WCTE12** “Are these simulation designs valid?” Students are given two simulation designs and asked to determine the validity of each design. Why are these designs valid or invalid? How would you change the design to make it valid?

<u>Figure 5</u>. Sample Tasks Used During the Whole-Class Teaching Experiment (WCTE)

Classroom-Based Analysis Phase

During each instructional session, the witness acted as an observer to the activities and whole-class discussions within the classroom. More specifically, her primary role was to pay particular attention to the case study group. The witness took notes that focused on identifying the emergence and evolution of classroom sociomathematical norms and mathematical practices. At the end of the day, the teacher-researcher and witness discussed and modified the hypothetical learning trajectory relying on the hypotheses of students' development of the learning goals.

As part of each instructional session, students were involved in small group and whole-class discussions. The purpose of the discussions was to highlight the reasoning and beliefs students used during probability simulation tasks. Throughout the whole-class teaching experiment, the teacher-researcher posed questions to students to both assess student reasoning and understanding and to encourage reflective thinking by the students. Student reasoning was assessed according to the steps of the simulation process (Yates et al., 1999) as previously outlined in Chapter 1. Previous research (Zimmermann & Jones, 2002) was used to help identify emerging and evolving beliefs related to probability simulation.

Data Sources

Data sources for this research study included the following: (a) pre-, post-, and retention assessments; (b) audio and videotapes of classroom events; (c) student written work; (d) audio and videotapes of individual student interviews; (e) field notes of the witness; (f) audio tapes of instructional sessions of teacher-research and witness; and (g) teacher-researcher reflections. The pre-, post- and retention assessments, described previously, were one source of data that provided a picture of student understanding of probability simulation over time. Audio and videotapes of student group work and whole-class discussions that were recorded during the classroom instructional sessions provided two sources of data that complemented each other. The tapes helped to document changes in students' reasoning and beliefs about probability simulation as well as provided the means to capture the emergence and evolution of classroom sociomathematical norms and mathematical practices. Student artifacts, such as homework collected from the students, provided further evidence of student reasoning. Additional data on students' reasoning and beliefs about probability simulation were collected through the audio and videotaped interviews of the case-study students. The field notes of both the teacher-researcher and witness, the audio taped instructional sessions, along with the reflections of the teacher-researcher provided data from the perspective of the teacher-researcher. The intent of the variety of data sources was to allow for triangulation of the data; that is, it provided multiple perspectives and interpretations (Eisenhart, 1988) of the data through the rich, descriptive pictures and discourse of student's reasoning in the classroom environment.

Student-Generated data involved a variety of sources. The written assessment instruments provided both quantitative and qualitative data. Each day of the whole-class teaching experiment, one audiotape recorder per group was provided to record each group's discourse providing qualitative data of students' reasoning and beliefs about probability simulation. Other than the assessments, student written work was periodically collected. The students' written work provided a variety of sources that enabled the teacher-researcher to track and record student reasoning and beliefs about probability simulation beyond that of group and classroom discussions. This work was also useful in helping the teacher generate conjectured learning processes. Another source of student-generated data was the clinical interviews of the 4 case-study students. The data collected from both the assessments and the interviews enabled the teacher-researcher to document changes in students' reasoning and beliefs about probability simulation that occurred over the course of the whole-class teaching experiment.

Similar to the student-generated data, teacher-researcher and witness generated data also came from a multitude of sources. The witness took notes as she observed each instructional session. She focused on the case study group but was also an observer during whole-class discussions. The instructional planning sessions between the teacher-researcher and witness were audio taped in order to provide a more complete record of the development and modification of the hypothetical learning trajectories. Finally, after each instructional session, the teacher-researcher recorded journaling notes to further capture the hypothesized learning processes of the students as they were directly related to that day's instructional activities.

Data Analysis

This study utilized a mix-method design (Tashakkori & Teddlie, 1998), containing both a quantitative and a qualitative component. Quantitative analysis was carried out to assess changes in students' reasoning in relation to probability simulation. A repeated measures analysis of variance (Kirk, 1982) was used to assess these changes with the dependent variable being student reasoning across the pre-, post-, and retention assessments.

Analysis of the many data sources began with the quantitative analysis of the assessments. Secondly, Miles and Huberman's (1994) "three part analysis" (pp. 10-11) was used to carryout a qualitative analysis of the whole-class teaching experiment (refer to Figure 1). This analysis focused on students' individual and collective reasoning and beliefs related to probability simulation and the classroom sociomathematical norms and classroom practices that emerged. In the first part of this process, <u>data reduction</u>, codes from all the data sources generated images and impressions of individual and collective student reasoning and beliefs.

Summary

This chapter provided details of the methodology of whole-class teaching experiment used in this study. Outlined were the specifics of the students, the instruments used for assessment purposes, as well as a summary of the instructional activities. Finally, a summary of the data analysis was also provided.

CHAPTER IV

ANALYSIS OF THE DATA AND RESULTS

The goal of this study was three-fold. The first goal was to trace students' individual and collective reasoning about probability simulations during a teaching experiment, and at the same time, determine the extent to which technology played a part in student reasoning. A second goal was to determine beliefs students held about probability simulation and track changes in these beliefs as they occurred throughout the teaching experiment. Finally, it was the intent of this research study to determine what sociomathematical norms and classroom mathematical practices evolved during a whole-class teaching experiment that focused on probability simulation.

Quantitative and qualitative analyses of the pre-, post-, and retention assessments were used to examine changes in students' reasoning before and after the whole-class teaching experiment. In addition, qualitative analysis was used to analyze data collected from the student interviews and whole-class teaching experiment. These data were used to trace the evolution and change in students' reasoning and beliefs as well as to identify the emergence of sociomathematical norms and classroom mathematical practices.

Quantitative Analysis of Pre-, Post-, and Retention Assessments

Of the 23 students who participated in the teaching experiment, 21 students completed all three assessment instruments. Each assessment contained parallel questions that were designed to determine the extent to which a student was able to (a) recognize and evaluate a valid simulation process, and (b) construct a valid probability simulation.

The means and standard deviations of each of the three assessments are presented in Table 2.

Table 2

Means and Standard Deviations for the Pre-, Post-, and Retention Assessment Scores

Assessment	M	SD	N
Pre-	10.2	5.7	21
Post-	17.9	2.9	21
Retention	18.9	2.1	21

Note. The maximum score possible on each of the assessments was 21.

A multivariate test, using Wilks's Lambda, revealed significant differences between the mean scores on the three assessments ($\Lambda = 0.274$; $F(2, 19) = 25.219$, $p < .001$). In view of these significant differences, pairwise comparisons were made to examine changes between the three assessments. These pairwise comparisons, shown in Table 3, indicate that the post-assessment scores were significantly higher than the preassessment scores ($p < .001$), and similarly, the retention assessment scores were also significantly higher than the preassessment scores ($p < .001$). There was not a significant difference between the post- and retention assessments ($p = .116$). The Wilks' Lambda test was performed to determine if the assumption of normality was met (see Figure D1 and Table D1 in Appendix D), and the test revealed no significant departure from normality for any of the three sets of different scores ($p = .499$, $p = .249$, $p = .956$).

Comparing the post- with the preassessment scores revealed that the mean score increased by over 75% while the standard deviation decreased by 3.6 points (see Table 2). Not only do these figures represent a significant increase in student performance from the preassessment to the post-, they also indicate a much more consistent performance for the students at the time of the post-assessment. The mean score on the retention assessment actually increased from that of the post-assessment, albeit not significantly, and the decrease in standard deviation from the post- to the retention assessment is further indication of the consistency in student performance over time. In other words, students' reasoning about probability simulation increased significantly after the whole-class teaching experiment and students maintained the higher level of reasoning six weeks after the teaching experiment.

Table 3

Mean Differences of Pairwise Comparisons Between Assessment Scores

Pairwise Comparison	Mean Difference	p
Pre- and Post-	-7.7	.000
Pre- and Retention	-8.7	.000
Post- and Retention	-1.0	.116

The students' growth in probability reasoning is further illustrated by examining the frequency of valid responses on each of the 10 major components that constitute a probability simulation. Table 4 shows the frequency of valid responses for the pre-, post-, and retention assessments on these simulation components. In all but one of the components (assumptions in the construction task), students demonstrated an increased ability to reason about probability simulation. More specifically, significant progress was made by students in both their ability to use simulated outcomes to determine the probability of a situation and in recognizing the effect of repeated trials on the empirical probability.

Table 4

Valid Student Reasoning in Components of a Simulation by Assessment

	Number of Students (n = 21)		
Process in a Simulation Problem	Pre-	Post-	Retention
Evaluated a Probability Simulation			
Assumptions	1	5	3
Evaluated probability generator	18	21	21
Recognized need for 2-D trial	14	21	21
Accepted randomness of outcomes	7	18	18
Calculated empirical probability given the outcomes	7	18	18
Constructed a Probability Simulation			
Assumptions	5	1	2
Constructed probability generator	12	19	19
Constructed 2-D Trial	11	21	19
Calculated empirical probability	13	19	21
Repetition of trial	6	16	14

Qualitative Analysis of Students' Reasoning

Multiple sources provided data of students' reasoning about probability simulation. One major source of data was the pre-, post-, and retention assessments, and a second major source was the audio and videotapes of the whole-class teaching experiment instructional episodes. The qualitative analysis of each source has been separated into two sections that focus on reasoning.

Students' Reasoning on Pre-, Post-, and Retention Assessments

The written responses of the students as well as the target interviews provided rich and informative qualitative data to support the findings summarized earlier in the quantitative analysis. This section contains the qualitative analysis of student responses on both the evaluation and construction tasks. As the simulation components required for both the evaluation and construction of a probability simulation are similar, tables have been organized to examine student reasoning by process across tasks and assessments. In this manner, student reasoning can be traced as it evolved before and after the whole-class teaching experiment.

For the preassessment evaluation task, students were asked to evaluate a simulation design for a pizza place that had 60 percent of their phone orders for pizza with meat and 40 percent of their orders for veggie pizza. The objective of the simulation was to determine the probability that the next two phone orders were both for meat pizzas. The evaluation task on the post assessment involved a 70% free throw basketball player. In this problem, the goal of the simulation was to determine the probability that the basketball player misses the next two free throws. The retention assessment evaluation task asked students to consider a simulation to determine the probability of getting stopped at a train track both to and from school if the chance of being stopped by a train was 60%. (See Appendix A for each assessment.) In order to identify a valid simulation, students needed to be able to recognize and evaluate the required components or steps involved in the simulation process.

The second part of the assessments was the construction task that concentrated on the students' ability to construct a valid probability simulation and use the results to calculate a probability for the problem. Similar to the evaluation task, the components required to construct a simulation were examined and analyzed. Essentially the components needed to construct a simulation are the same as those needed to evaluate a simulation (see Table 4 for simulation components). In addition to assessing the students' ability to construct a valid simulation, students were also questioned on what impact repeating a simulation many times would have on the empirical probability.

The assessment construction problems, like those for the evaluation problems, were also contextual. In the preassessments, students were asked to construct a probability simulation to determine the probability that the songs played by a radio station at two specified times would both be hip-hop if the probability that the station plays hip-hop is .4, the probability it plays alternative music is .4, and the probability it plays country music is .2. On the post-assessment, students were to design a simulation to determine the probability of both engines failing on a space shuttle given that the probability of the first engine (S1) failing was .2 and the probability of the second engine (S2) failing was .3. Tennis was the context on the retention assessment. Students were to determine the probability of double faulting given the probability of a fault on the first serve was .8 and the probability of faulting on the second serve was .1.

To identify a valid simulation, students needed to be able to recognize and evaluate the required components in the simulation process (Yates et al., 1999). These components include the following: stating assumptions, identifying a probability generator, defining a valid trial, repeat the trial many times, and use the outcomes to determine the empirical probability. Taken together, these simulation process components provide the frame for analyzing students' responses to the two simulation tasks.

Assumptions

Part of the simulation process, according to Yates et al. (1999) is to list the assumptions for the probability simulation design. Table 5 contains different assumptions students stated as they reasoned through both the evaluation and construction tasks for the pre-, post-, and retention assessments. The students made more assumptions on the construction task in the post- and retention assessments than on the preassessment, and in the retention assessment (evaluation task), the majority of students made an assumption about replacing the first drawn chip so as to maintain the target probability at its given value. This was not only a valid assumption; it was a vital assumption for the correct processing of the simulation.

Students who stated assumptions of this kind added more detail to the problem than might be considered necessary. However, it should be noted that these assumptions were not inappropriate. Lacey's reasoning provides an illustrative example. In explaining how she would change the given simulation for the pizza problem, Lacey wrote, "I would consider the day and season it is and if any special orders are made that day." During the interview, Lacey elaborated on her written response. "If there was a holiday or a get-together, then I don't know if that could matter. I just think that if there is a holiday or a get-together then you might order more pizza." When probed further, Lacey explained that she thought the simulation design had "too much fault in it" possibly indicating the simulation did not account for these “special situations.” The type of reasoning exhibited by Lacey was typical of the other students counted in this category. Assumptions involving unaccounted for extraneous variables, noted by students on the preassessment construction tasks included such issues as song length, the amount of time music was played, and the total number of songs played within the targeted time slot.

Table 5

Explicitly Stated Assumptions by Assessment

Assumption	Number of Students (n = 21)					
	Evaluation			Construction		
	Pre-	Post-	Ret.	Pre-	Post-	Ret.
Unaccounted for extraneous variables	1	3	0	7	0	0
Independent events	0	4	3	0	1	0
Did not replace first drawn chip	0	2	0	1	0	0
Replaced first drawn chip	0	0	12	0	1	1
Did not make any assumptions	20	13	6	13	19	20

Note. Student strategies may have been classified under more than one category.
The first assumption listed in Table 5 is “unaccounted for extraneous variables.”

Unlike Lacey and other students who made assumptions about the extraneous variables in the context of the problem, Thor reasoned about extraneous variables associated with actually conducting the simulation. In discussing the validity of the simulation design, Thor responded, “the balls may not completely randomize her selection. Yes, she does have 3/10 blue balls but she may have different sized balls or different textured balls that would cause bias.” Thor recognized the fundamental importance of having the probability generator produce completely random outcomes, even though he may have been somewhat fastidious in his demands for randomization. By way of contrast, Ethan's focus on extraneous variables was more idiosyncratic. "There are more variables to the problem . . . . the best situation that I could think of is that he probably plays country late at night when nobody is listening since nobody like[s] it." Ethan's reasoning is considered idiosyncratic because he used the fact that he did not like country music to assume no one did. Furthermore, Ethan seemed unaware that his assumptions had changed the problem.

The assumption regarding <u>independence</u> of events did not appear until after the teaching experiment. Students, who stated that the events in the simulation were independent of each other, were assuming the outcome of the first event did not influence the outcome of the second event in the trial. During his post-assessment interview, Cade articulated the effect independence of outcomes had on the process, "I thought that each at bat is separate so her probability to get on base would be 30% each time, and if she left the ball out (during the simulation process) then it wouldn't be like separate batting. Then it wouldn't be 30%." Although more detailed than most, Cade's response typified the reasoning of the other students who discussed independence of events.

The third and fourth assumptions that appear in Table 5 are related. Prior to the teaching experiment, none of the students referred to the <u>replacement or non-replacement</u> of the drawn chip in the pizza problem. Yet after the teaching experiment, 2 students specifically stated that replacement should <u>not</u> occur. On the softball problem, Kacy and Ingrid reasoned that "when there is a ball taken out, leave it out to have a better chance of grabbing a blue ball." Both seemed unaware that by not replacing the first drawn ball, the probabilities for the second at-bat were not the same as those for the first at-bat. On the retention assessment, 12 students explicitly wrote that the chip should be replaced in the train-crossing problem on the retention assessment, even though none of the students had made this claim previously.

Only one student on each of the post- and retention assessment construction tasks discussed the replacement of a drawn chip. This may be partly due to the type of probability generator selected, that is, using the random number generator on the calculator, as described by the students, assumed independence of outcomes. (Use of the calculator is explained in more detail later.) Students may have also assumed that if the assumption of replacement or non-replacement was addressed in the evaluation task, it need not be addressed again in the construction task.

Table 5 indicates that prior to the whole-class teaching experiment, more students made assumptions related to the construction task than on the evaluation task. However, following the whole-class teaching experiment, this situation was reversed. This can be traced largely to two sources. One was the decreased use by students of extraneous-type assumptions. The decrease in this type of reasoning caused the number of assumptions made on the construction problem to decrease. The other source was related to the assumption of replacement of a drawn chip or ball. In the evaluation problems, the probability generators provided in the problem involved balls or chips, which required replacement to maintain the original probabilities. In the construction problem, many students used spinners or random-numbers generated by a calculator which have replacement built into the probability generator. Thus, the characteristics of the simulation models used may account for the fact that more students noted the need for replacement on the evaluation problems than on the construction problems in later assessments.

The Probability Generator

The next step in the simulation design process is to model the contextual setting. In other words, in the evaluation task students needed to be able to evaluate the validity of the probability generator used in the simulation problem, and in the construction task students needed to be able to construct a valid probability generator to simulate the problem. Since the reasoning exhibited by the students was different for each of the tasks, student reasoning on the evaluation and construction tasks has been separated. Table 6 contains the reasoning demonstrated by students as they evaluated a probability generator, and Table 7 contains students' reasoning as it was related to constructing a probability generator.

Table 6

Student Reasoning on Evaluating the Probability Generator by Assessment

	Number of Students (n = 21)		
Reasoning	Pre-	Post-	Retention
Number of chips should equal number of orders	1	0	0
Equal amount of colored chips	2	0	0
Probability generator was valid	18	21	21

Evaluation task. Overall, students had little difficulty recognizing a valid probability generator, although three students expressed difficulty with the simulation process on the preassessment. On both the post- and retention assessments, all 21 students demonstrated valid reasoning related to the probability generator for the given simulation problem. With respect to the 3 students who had difficulty on the preassessment, all of them seemed to struggle with the concept of proportion. Dayton's reasoning was typical of these students. When probed about her thinking during the pre-assessment interview, Dayton responded, "I said to put an equal amount [of chips] . . . who's to say that he will get 60% next time. I said that to do this out of a monthly basis you would need an even amount of chips." Dayton seemed to believe that the past could not necessarily be used to predict future outcomes, and since we had no control over the future, she believed that orders for meat and veggie pizza should be equally likely.

Construction task. In the preassessment, 14 of the students were able to explain how they would construct a valid probability generator to simulate the problem. Prior to the whole-class teaching experiment, students primarily referred to manipulative devices, such as chips, balls, and spinners, to construct a probability generator. However, on the post- and retention assessments, about half of the students constructed a probability generator using a randomly generated number on a graphing calculator. Also noteworthy was the increase in the number of students who chose to use a two-bag or two-spinner probability generator on the post- and retention assessments. Table 7 contains a summary of the various strategies students used to construct a probability generator for each of the radio, space shuttle, and tennis problems. The numbers within the parentheses represent those students whose reasoning for that particular strategy was either invalid or incomplete. Following is a more detailed analysis of the students' ability to construct a probability generator for a probability simulation problem.

Table 7

Student Strategies on Constructing the Probability Generator by Assessment

	Number of Students (n = 21)		
Strategy	Pre-	Post-	Retention
None	6	0	0
Other device	1	1	1
One bag of chips or balls	9 (2)	0	0
Two bags of chips or balls	1	3	4
One spinner	4	3 (1)	4
Two spinners	0	4	2
Calculator	0	10 (1)	10 (2)

Note. Student strategies may have been classified under more than one category. Numbers in ( ) denote number of students using invalid strategies.

As noted in Table 7, 6 students did not use a probability generator on the preassessment. Of these 6 students, 2 students provided responses that did not contain any details of a probability generator, and 3 students replied that for the radio problem, they would just turn on the radio itself. Breanna belongs to this second group of students. She wrote, “My simulation would be to turn on the radio at 10:00 a.m. and then again at 2:30 p.m. for every day for 4 weeks straight. At each day, record the type of music played at 10:00 a.m. and 2:30 p.m. as a set.” Tavi was another student who did not construct a probability-generating device. He seemed to understand the proportion necessary for the simulation, but did not relate these numbers to a generator device. Tavi replied, “For every hour, 24 minutes of it would be hip-hop, 24 minutes would be alternative, and 12 minutes would be country. Your chances of hearing a hip-hop song at 10:00 a.m. is 40%

and your chances of hearing another hip-hop song at 2:30 p.m. is also 40%." Tavi did not extend this to a device that would simulate the problem, nor did he define a trial. All of these students were able to construct a probability generator for each of the post- and retention assessments.

Few students used probability generators other than chips/balls, spinners, or the calculator. However, on the radio problem, Edison constructed a valid probability generator using a 10-sided die: "I would take a die with 10 sides (they do exist) and mark off numbers 1-4 as hip-hop, 5-8 as alternative, and 9, 10 as country." Cade used a combination of a calculator and a spinner to construct a probability generator for the space shuttle problem. He said he would "take a calculator and have it choose a random integer from 1-10: (for S1) 1-2 was failure, 3-10 [engine] works. If you get a 1-2, use a spinner divided in 10 spaces, 1-3 failure, 4-10 works."

Using a single bag of chips or balls was the preferred probability generator on the preassessment, but this strategy was not used in the post- or retention assessments. All but 2 of the 9 students who used this strategy did so in a valid way. Kacy was one of those who responded validly using the colored chips strategy, "You can get 10 chips, 4 red (alternative), 4 green (hip-hop), 2 yellow (country)" and put them in a bag. Ingrid and Macy's incomplete responses provided some insight into why students had shown a preference for using chips and one bag. Ingrid explained, "I would do exactly what Stan [the hypothetical student in the problem] did with the chips." On the one hand, neither student had provided further explanation that would reveal their understanding of how to construct a valid probability generator for the radio problem. On the other hand, their responses indicated that many of the students might have chosen chips to simulate the problem because chips were used in the previous pizza problem.

An alternative strategy to using one bag of chips was to use two bags of chips. Oliver's strategy was to use two bags with the same proportion of chips in each bag to represent the radio problem. Oliver wrote, "take 2 bags each with 4 chips marked with an 'H' [for hip-hop], 4 with an 'A' [for alternative] and 2 with a 'C' [for country]. Mark one bag 10:00 and the other 2:30." Oliver was also the only student to employ this strategy involving a combination of two probability generators on the pizza problem. The number of students using the strategy of two bags of chips increased to 3 and 4 on the post- and retention assessments, respectively, and in all instances students demonstrated valid reasoning typified earlier by Oliver.

The structure of the space shuttle and tennis problems might have contributed to the increased use of either a two-bag or two-spinner strategy. In the preassessment radio problem, there were three choices of music that accounted for 100% of all music played in that problem (alternative 40%, hip-hop 40%, and country 20%). Thus, it was likely easier for the students to divide a single bag of chips or a single spinner into equal proportions that would total 100%. However, on the post- and retention assessments, the probabilities in the problems did not total to 100%. For example, recall that in the space shuttle problem it was stated that the probability engine 1 (S1) would fail was 0.2 and the probability that engine 2 (S2) would fail was 0.3. (See Appendix A2: Post-assessment for the detailed problem). Because the total of the probabilities provided in the problem was not 100%, students might have found it easier to reason using two bags or two spinners rather than one. In this way, students treated each outcome separately and then combined the outcomes for one trial. Lacey's post-assessment interview indicated how she approached the complexity of the probabilities in order to construct a valid probability generator. Lacey explained her thinking, "Well I sorta got confused . . . because you're not going to need S2 if S1 doesn't fail. So that's why I just looked at it independently so it would be easier." Lacey was aware that she needed to look at both outcomes for her trial, but decided to essentially run two sets of one-dimensional trials, one for each engine, and then pair the results.

The student strategy of using one or two spinners was similar to that of using one or two bags of chips. On the preassessment, 4 students constructed a valid probability generator using a single spinner. Tasha's design was typical of these students. She explained that she would use one spinner that was labeled "40% hip-hop, 20% country, 40% alternative." As discussed earlier, the probability generator required for the construction problems on the post- and retention assessments required students to reason somewhat differently than on the preassessment. Mai was 1 of only 2 students who provided a valid probability generator using one spinner on the post-assessment construction problem. "You have a spinner with 10 equal spaces. And because there are two power systems, you need to spin twice. First time, 1-8 means it [engine 1] didn't' fail, 9-10 means it failed. The second time 1-7 means it [engine 2] didn't fail, 8-10 means it failed." Mai was able to construct a single spinner that would generate the required probabilities for each outcome of the two-dimensional trial. Thor's approach to the probability generator was unique. Thor created a spinner containing three regions, A, B, and C. According to Thor's diagram, region C accounted for 80%, region A accounted for 20%. Thus, landing in region C was equivalent to getting a fault on the first serve. To simulate the second serve, Thor marked a 10% sector of region C and labeled this region B. Thor wrote that region B accounted for "10% of the 80% [of region C]." Region B represented the probability of getting a fault on the second serve. In essence, Thor's probability generator produced valid one-dimensional outcomes for what was a two-dimensional problem. "If you land in A you didn't fault on the first. B, faulted both. C, missed first made second." Thor's reasoning reflected the increased sophistication with which students constructed probability generators following the whole-class teaching experiment.

Although none of the students used a two-spinner strategy on the preassessment problem, 4 students constructed a valid two-spinner probability generator for the post-assessment problem. Ondrea's explanation typified this strategy; "I'd take 2 spinners with numbers 1-10. Spinner 1 represents S1 [engine 1], so therefore any digit 1-2 would be S1 failing. Spinner 2 represents S2, so any digit 1-3 will be S2 failing."

One of the most interesting trends revealed in Table 7 was the increased use of calculators as probability generators. In the preassessment there was no evidence of calculators being used to simulate the problems. However, this number increased dramatically to 10 students on both the post- and retention assessments. All but one of the students owned a TI-83 or TI-83+ graphing calculator. One of the many functions of this calculator is its ability to generate random numbers and, in particular, random integers. The function on the calculator is known as "randInt." RandInt can be used one of two ways: (a) to generate a random integer n between two designated integers, a and b, such that $a \leq n \leq b$, or (b) to generate n random integers inclusive of the two designated integers. For example, a student could enter into their TI-83 "randInt(1, 20)." This command would generate a random integer from 1 to 20. If a student were to add an additional parameter to the instruction, “randInt(1, 20, 5),” then the calculator would generate five random numbers from 1 to 20.

Out of the 10 students who used the calculator to construct the probability generator, 9 of the responses illustrated a valid strategy. Ethan’s response was typical: “ I would simulate it by using the random number feature on my calculator. I would use 0-9 for both serves. If he got an 8 or a 9 on the first one means he didn’t fault, but any number less than that means he did and then you look at the second number. This one you have to get a 0 to fault on. So for both trials you have to get a number less than 8 and 0 for him to get a double fault.” Unlike Ethan’s model, Kacy’s calculator-generated simulation for the shuttle problem was not valid. Kacy described her simulation design as using “randInt on my calculator and use the variables 1 for S1 and 2 and 3 for S2. If I got a 1 that would mean S1 failed, and if I got a 2 or a 3 that would mean S2 failed.” Kacy was unable to translate the probabilities in the problem into a valid generator. Again, this may have been attributed to the nature of the probabilities within the problem as discussed earlier. The 2 students who were unable to construct a valid probability generator on the retention assessment transformed the two-dimensional problem to a one-dimensional situation similar to what Kacy had done earlier.

No trends were evident in a comparison of the probability-generating device used by each student across assessments. One student used the same device across all three instruments. Five students used the calculator for both the post- and retention assessments, and 8 students used a different device for each assessment. This analysis, coupled with the fact so many students were able to construct a valid probability generator, demonstrates the willingness with which students considered various devices for a simulation design.

Recognized the Need for Two-Dimensional Trial

Both the evaluation and construction simulation problems on each assessment instrument involved a two-dimensional trial. The problems required that students look at a pair of outcomes. As shown in Table 8, 15 of the 21 students recognized that the evaluation problem required a two-dimensional trial in the preassessment. By the post- and retention test, all students were able to demonstrate valid reasoning, using one of two strategies, when evaluating a two-dimensional trial. Furthermore, the students explicitly described a two-dimensional trial in the construction problems on the post- and retention assessments more often than in the preassessment. Following the whole-class teaching experiment, students were less likely to record the simulation outcomes as one-dimensional trials. Table 8 contains a detailed summary by assessment of how students reasoned when asked to evaluate and construct a two-dimensional trial.

In analyzing student reasoning of a two-dimensional trial, student strategies were classified as either valid or invalid. With respect to the valid category, students used two substrategies. The first substrategy, “explicitly stated a two-dimensional trial,” increased substantially after the teaching experiment. Students who exhibited this strategy explicitly stated that the question was about the next two outcomes or they provided specific details about the trial. Oliver’s response provided an example of this substrategy. He wrote that in order to conduct the simulation accurately one would need to “record how many times that he got red twice in a row.” It should be noted that students who explicitly stated a two-dimensional trial also recorded their outcomes as two-dimensional.

Table 8

Student Reasoning on a Two-Dimensional Trial by Assessment

Reasoning	Number of Students (n = 21)					
	Evaluation			Construction		
	Pre-	Post-	Ret.	Pre-	Post-	Ret.
Valid						
Explicitly stated 2-D trial	12	17	11	9[a]	18[a]	16[a]
Calculated/recorded 2-D trial	3	4	10	3	3	3
Invalid						
Incomplete answer/ not enough info	4	0	0	1	0	0
Calculated/recorded 1-D trial	3	0	0	8	0	2

Note. Student strategies may have been classified under more than one category.
[a] Students also recorded their outcomes as 2-D.

Students revealed the second substrategy when they calculated and/or recorded their outcomes as a two-dimensional trial, but had not clearly defined their trial. The strategy these students used to calculate the empirical probability on the evaluation problem for each of the assessments was basically the same as Ugo's strategy. Ugo divided the outcomes into pairs illustrating that he knew he needed a two-dimensional trial. Students were more likely to explicitly define a two-dimensional trial on the construction problems than in the evaluation problems.

Two subcategories of invalid strategies were evident when students reasoned about two-dimensional trials. The first subcategory contained students who provided insufficient detail or justification in which to categorize their response. Students in the second subcategory, "calculated/recorded 1-D trial," reasoned incorrectly when they calculated a one-dimensional empirical probability for the problem, as Bernard did. Given the outcomes, Bernard calculated an empirical probability of getting meat as 25/50 when the problem required the probability that the next two pizzas were meat. More specifically, students recorded outcomes as a string of numbers or letters and in each case a one-dimensional probability was calculated.

Accepted Randomness of Outcomes

Randomness of outcomes is inherent in a simulation process. Therefore, students' willingness to accept the apparent randomness of the outcomes was of interest. Outcomes for each of the assessment instruments were intentionally distributed equally between each possible outcome of the two-dimensional trial. For example, in the preassessment pizza problem the probability of an order for pizza with meat was given as 60% and the probability of an order for a veggie pizza was given as 40%. However, out of 50 outcomes provided to students for this problem, 25 were for meat pizza and 25 were for veggie pizza. (See Appendix A1). Table 9 provides a summary of the various ways students approached the randomness of the outcomes of the simulation process. Due to the structure of the assessment questions, this simulation process was only observed on the evaluation task of the assessments.

Table 9

Student Reasoning on Randomness of Outcomes by Assessment

Reasoning	Number of Students (n = 21)		
	Pre-	Post-	Retention
1) Accepted randomness			
Implicitly accepted randomness	9	18	20
Explicitly accepted randomness	1	0	0
2) Did not accept randomness			
Exhibited representativeness	10	3	0
Relied on theoretical	1	0	0
Too few trials	0	2	1

Note. Student strategies may have been classified under more than one category.

Student strategies related to randomness were separated into two categories: accepted randomness or did not accept randomness. The students who accepted the randomness of the outcomes in a simulation did so in either an implicit or explicit manner (see Table 9). Student acceptance of randomness was considered implicit if no mention was made of the outcomes. In other words when these students were asked if they could determine the probability given the outcomes, they were not disturbed by the equal distribution of the chips. By the time of the post-assessment, 86% of the students accepted the randomness of the outcomes and 95% of the students accepted it on the retention assessment. The number of students who implicitly accepted the randomness of the outcomes grew sharply by the post- and retention assessments.

Evan was the only student to explicitly accept the randomness of the outcomes. When asked if he could use the pizza outcomes to calculate the probability, Evan wrote

"Yes. I still could predict because the ratio is still 6:4 . . . because they came out equal does not mean that it is the true probability." Evan recognized that these outcomes might not be a long-term reflection of the pizza situation.

Students who <u>did not accept the randomness</u> of the outcomes used one or more of three subcategories. The most prevalent subcategory was that of representativeness. That is, the students believed that the sample of simulated outcomes should match or be representative of the population probabilities of the original problem. They did not accept the equal distribution of outcomes in the pizza problem because they believed the outcomes did not accurately reflect the problem situation. Mai, whose response was typical of these students, wrote, "In actuality their percentages of orders were for 60:40, not 50:50 like the outcome." Mai persisted in her reasoning even after the whole-class teaching experiment. She was one of 3 students who was still bothered by representativeness. Nevertheless, her reasoning became somewhat more sophisticated in the post-assessment. Mai explained her reasoning about the softball player problem, "I don't think you can determine the probability because the results of her experiment look very inaccurate, and don't show a true example of her 30% base average." Mai further wrote, "I don't think this is good enough data to determine the probability, and she should do the experiment quite a few more times." Mai, like the other two students, seemed cognizant of the fact that limited trials produce results that contain more variability.

Ondrea was the only student to exhibit the second subcategory whereby she discounted the outcomes and relied on using theoretical probability. She reasoned that probability was determined theoretically, not through simulated results. Ondrea wrote that the outcomes could not be used since "you can approach probability mathematically using equations, not trials." This was the only time this reasoning substrategy was used throughout all three assessments.

Students who used the strategy of too few trials to discount the outcomes of the problem were in the third subcategory. First appearing on the post-assessment, Mai and Thor used this strategy in the softball player problem. They reasoned that the outcomes did not reflect the probabilities originally given in the task. However, they reasoned further that there were not enough trials to provide accurate data. Sadie's reasoning in the retention assessment problem was similar to that of Mai and Thor's. She wrote, "you could try to [determine the probability] but the number of trials is so small that it doesn't get close to the theoretical probability." Sadie, like Mai and Thor, seemed to realize that although these outcomes were possible, they seemed problematic. Therefore, Sadie reasoned that more trials should be done.

Calculated Empirical Probability Given the Outcomes

The objective of a simulation process is to collect data or outcomes representative of a given situation. These outcomes are then used to determine the empirical probability. Thus, a student's ability to determine this probability is integral to the simulation process. In each evaluation assessment task, simulated outcomes were presented in two forms (see Appendix A): (a) as a single count of each color drawn, and (b) as a string of 50 letters representing the color of each chip drawn. The second form enabled students to separate outcomes into pairs in order to determine a valid empirical probability for the two-dimensional problem. Table 10 summarizes strategies students revealed when asked to use the outcomes of the simulation to determine the requested probability.

Table 10

Student Strategies in Calculating the Empirical Probability by Assessment

Response	Number of Students (n = 21)					
	Evaluation			Construction[b]		
	Pre-	Post-	Ret.	Pre-	Post-	Ret.
Did not calculate probability						
Outcomes were invalid[a]	10	2	0	--	--	--
Outcomes were not 2-D	4	6	2	--	--	--
Argued theoretical is better	1	0	0	--	--	--
Did calculate probability						
Calculated valid 2-D probability	3	14	17	9	19	20
Calculated 1-D probability	1	0	0	5	0	1
Calculated using theoretical prob or a mixed method	1	0	1	2	1	0
No pattern	0	0	0	5	1	0

Note. Student strategies may have been classified under more than one category.
[a] This was evidence of representativeness heuristic.
[b] No numeric entry indicates response was not applicable for this problem.

More students were able to calculate a valid empirical probability in each of the post- and retention assessments than in the preassessment. Furthermore, reasoning that initially prevented students from determining the probability of the problem decreased on the post- and retention assessments. Certain types of student reasoning only appeared on the evaluation problems. This was attributable to the fact that students had to reason about outcomes that had been provided to them, rather than trying to make sense of a simulation they designed and executed.

Students' strategies fell into three major categories: those that did not involve the calculation of a probability, those that did involve the calculation of a probability, and those where there was no prevailing pattern. With respect to the category that did not involve the calculation of a probability, students used three substrategies (see Table 10). The first substrategy was particularly dominant in the Pizza Problem and arose from the fact that students did not accept the outcomes as valid because of their reliance on representativeness. As discussed in the previous section on randomness of outcomes, these students were distracted by the equal distribution of outcomes that did not match the probabilities given in the problem. Consequently they were unwilling to use the data or to calculate a probability. This substrategy decreased sharply following the teaching experiment with only 2 students using it in the post-assessment and no student using it in the retention assessment.

Students using the second substrategy observed that the probability of an event like, meat both times, could not be calculated because the outcomes were not two-dimensional. Sadie's response was typical of these students, "No, because this experiment is only for one customer, not two." Although their reasoning was essentially valid, students like Sadie failed to note that they could have paired the outcomes and transformed the original data into two-dimensional data. This substrategy continued following the teaching experiment but students' reasoning was more sophisticated as is reflected in Makaila's response in the post-assessment. She said, "You can't determine the probability Beth gets on base both times with the data and results [in the problem]." She added that it would be "necessary to split the data into groups of two to simulate two-at-bats and take into account of the fact that you are looking to see if she gets on base both times." Even though Makaila did not actually calculate the probability, it was clear that she and the other students who provided this kind of explanation understood what was needed in order to calculate the probability of a two-dimensional trial.

The third substrategy, under the category “did not calculate a probability,” resulted from the fact that one student, Ondrea, believed that probability could only be determined theoretically. There was no further evidence of this substrategy in later assessments by Ondrea or any other student. In summing up the “did not calculate strategy,” it is worth noting that it appeared only on evaluation tasks. By way of contrast “the calculated a probability,” and the “no pattern” strategies appeared on both evaluation and construction tasks.

The strategy where <u>students made a probability calculation</u> revealed itself through three substrategies: a valid two-dimensional probability was calculated, a one-dimensional probability was calculated, or a probability based on both empirical and theoretical considerations was calculated (see Table 10). Students using the first subcategory demonstrated their ability to calculate a two-dimensional probability in two different ways. In the evaluation problems, students used the same strategy that Lacey demonstrated. Lacey explained her process,

> I broke up the count into two sections [pairs of numbers] because the way I’d interpret the simulation and data would be to see if two chips were consecutively pulled out as meat pizzas. This happened 5 times out of the 25 ‘pulls.’

In other words, the students using this first substrategy on the evaluation problems divided the string of outcomes into pairs, counted the number of targeted pairs, and divided their answer by the total number of outcomes. Students using the first substrategy on the construction problems merely recorded their outcomes as two-dimensional trials, and then determined how many successes they encountered compared to the total number of trials. On the post-assessment Bernard shared his reasoning about calculating the probability of both shuttle engines failing; “I did this [found the probability] by taking the number of times I simulated both the engines failing and divided by the number of times I did it.” Not all the students were as articulate as Bernard and simply wrote the probability as a ratio of targeted successes over total attempts. This substrategy dramatically increased following the teaching experiment with well over half the students using this substrategy.

Students using the second substrategy calculated a one-dimensional probability for a problem that required a two-dimensional trial. Interestingly, Bernard used this strategy in the preassessment. Bernard explained, “since the times [number of outcomes for each chip] are equal then the probability is 1/2.” Essentially, Bernard took the recorded 25 meat outcomes and divided it by the 50 total outcomes. Bernard’s response indicated that he did not realize his answer was not the probability for the next <u>two</u> orders rather he was calculating the probability for a one-dimensional problem. This substrategy appeared predominantly on the preassessment and only appeared once more after the whole-class teaching experiment.

Under the category “did calculate probability,” the third subcategory that involved theoretical probability, was not used by many students. However, it did appear consistently before and after the teaching experiment. Students who calculated a probability using a mixture of theoretical and empirical probability demonstrated reasoning similar to Thor’s. According to Thor, “if you only went off of his data, you could predict the wrong probability by multiplying the chance you would get meat for each time (.50 x .50) and come up with 25%, but Stan needs to do a few more tests before he uses his data to make calculations." On the post-assessment, Thor argued that "a 50% chance of getting on base is not close to the theoretical value so it may not be very accurate. She may also need to perform the experiment more times." On the retention assessment, Thor again used the same reasoning. Clearly, Thor consistently believed the outcomes were not representative of what he expected, and this influenced his reasoning when dealing with empirical probability. Furthermore, Thor relied on his knowledge of theoretical probability to use the one-dimensional simulated outcomes to calculate a probability for both the pizza and train-crossing problems, thus combining empirical and theoretical probabilities. Three other students exhibited this mixed reasoning.

The third category, no pattern, reflected those students who reasoned in a more idiosyncratic manner when calculating the empirical probability for the problem. For example, recall that Ethan had argued no one listens to country anyway. Ethan reasoned that since there are only two kinds of music to listen to, the songs would be played evenly, thus the probability is 50%. After the whole-class teaching experiment, this type of reasoning decreased significantly to one student on the post-assessment, and on the retention assessment there was no evidence of idiosyncratic reasoning.

Repetition of Trial

In the simulation process, a trial is repeated numerous times before the empirical probability is determined. The question becomes how many times should a trial be conducted and what are the effects on the probability when the number of simulation trials is increased. Table 11 summarizes students' responses when questioned about the effect of increasing the number of trials on the empirical probability. In general, students exhibited more valid reasoning on the post- and retention assessments. Furthermore,

more students in these later assessments reasoned that increasing the number of trials caused the empirical probability to approach the theoretical value. Their reasoning seemed to be connected to the concept of less fluctuation in the empirical probability over more trials. Because of the structure of the tasks, this simulation process appeared only on the construction tasks.

Table 11

Student Reasoning about the Repetition of Trials by Assessment

	Number of Students (n = 21)		
Response	Pre-	Post-	Retention
Valid			
More accurate results	6	3	8
Less fluctuation of probability	0	10	2
Approaches theoretical, actual, predicted	1	10	7
Invalid			
Little or no effect	8	4	3
Other	2	1	3

Note. Student strategies may have been classified under more than one category.

Basically, student strategies were judged to be either valid or invalid. Student strategies considered to be valid fell into one or more of three substrategies (see Table 11). Students using the first valid substrategy reasoned that increasing the number of trials would produce more accurate results. Thor provided the most articulate response.

> If you only do the experiment 50 times, you may not get a proper ratio that is accurate because there were only 50 trials. Each trial has less of an influence as you do more tests. If you do 100,000 tests, each test will have little change in the

> overall outcome but all the tests together give you a much more accurate representation of your true value.

Thor clearly understood the implications of many trials and was able to explain why a simulation should be repeated many times. Although the other 5 students also agreed that the results would become more accurate, their explanations were less detailed. On the post- and retention assessments, 3 and 8 students, respectively, used similar reasoning. Although these numbers represent a decrease in this strategy, this decrease is more than offset by the increase in the number of students who reasoned that the empirical probability would eventually approach the theoretical, explained in more detail later.

Students who used the second substrategy, “less fluctuation of probability,” were referring to the behavior of the empirical probability over repeated trials. Breanna’s response was typical of students using this substrategy, “The more trials, the probabilities will fluctuate less.” These students understood that initially the probability could vary dramatically and that eventually the empirical probability would settle towards a particular value, namely the theoretical probability. Although this strategy did not appear in the preassessment, 10 students on the post-assessment and 2 on the retention assessment made specific mention of this strategy in their answers.

The third valid substrategy occurred when students referred to the empirical probability approaching a theoretical or predicted value after many trials. Cade’s response was indicative of students who used this substrategy:

> Like flipping a coin and you could do it 50 times and it could be .47. You could do it a hundred times more, and it would still be in the area like we know it’s .5. It’s probably gonna approach .5 so you don’t necessarily need to do like 3000 trials for that because we kinda know it will approach there. With this, if like I were say to do it a hundred more times but after a while say the probabilities that I was getting were like pretty close to the same, then I would say that’s enough

> trials. But like after discussing some stuff in class, like 50 or 100, I don't think you can really put a number on it and get close to the same probabilities over and over again. Like there's not really a chance for it to change drastically.

Cade seemed to be able to recognize that the long run behavior of repeating a simulation would provide the empirical probability that would approach the theoretical probability, and at the same time Cade, like the other students who reasoned similarly, seemed cognizant of the variation inherent in the process. As can be noted in Table 11, the use of this strategy increased considerably after the whole-class teaching experiment.

The invalid strategies students used were classified into two subcategories: little or no effect and other. Students who reasoned that increasing the number of trials would have little or no effect relied on a variety of justifications. Of these students, 6 argued that since the probability generator did not change, the results would not change significantly. Seeming to understand short-term variation, Edison reasoned, "it's all probability, anything could happen within the simulation." Yet he did not realize the contradictory nature of his responses as he further reasoned that there would be little effect because anything could happen. He recognized the variability of simulated results in the short-term but did not seem cognizant of the effect repeated trials would have on the long-term stability of the probability for the problem. Edison's reasoning did not change by the time of the post-assessment.

"Other" invalid reasoning about repetition of trials was generally more idiosyncratic. One such example was Ondrea who explained that the results would change because song popularity and other radio station routines would change. Recall that Ondrea would simulate the radio problem by actually listening to the radio. On the space shuttle problem, Tavi explained that he thought more simulations would be "more realistic.” Although Tavi did not elaborate, he may have believed that the results should be more “accurate” and thus closer to the theoretical. This was supported when his response on the retention assessment was analyzed. When asked on the retention assessment if increasing the number of trials would change the results, Tavi wrote, “yes, because as the number of times increases, the probability gets more precise . . . to a certain point.” Tavi’s earlier reasoning typified that of the other students who also demonstrated an idiosyncratic reasoning strategy.

Summary of Students’ Reasoning on Pre-, Post-, and Retention Assessments

Students were administered three assessments that were used to identify trends in student reasoning as related to probability simulation. A quantitative analysis was completed to examine the change in student performance across the assessments and a qualitative analysis was conducted to identify more detailed trends in student strategies in the various simulation components. The quantitative analysis revealed a significant increase in student performance following the whole-class teaching experiment. Furthermore, student performance was more consistent on both the post- and retention assessments. The qualitative analysis of the assessments uncovered trends in student reasoning that supported the quantitative findings. On the preassessments, students used a vast array of invalid and valid reasoning strategies on each of the simulation components. By way of contrast, student reasoning on the post- and retention assessments was not only more consistent but typically more valid than on the preassessments.

Students' Reasoning During the Whole-Class Teaching Experiment

The primary focus over the 12 instructional sessions was to develop students' reasoning about probability simulation. During the whole-class teaching experiment data collected by both the teacher-researcher and the witness drove the development and modification of the hypothetical learning trajectory. Thus, it was the modified hypothetical learning trajectory that was used to determine daily learning goals and activities for the subsequent instructional session. The intent of this section is to follow the evolution of the students' reasoning about probability simulation, in particular their reasoning about each of the simulation components: assumptions, construction of a probability generator, construction of a trial, randomness of outcomes, calculating empirical probability, and repetition of trial. All referenced activities can be found in Appendix C.

Assumptions

It was not until Session 4 that students were asked specifically to consider the assumptions of a probability simulation problem. Prior to this, learning goals had focused on using and constructing probability generators and defining trials. To help students begin to focus on assumptions, students were explicitly asked to list the assumptions for a given simulation problem. Most students responded using reasoning similar to that found on the preassessment; that is students tended to focus on "unaccounted for extraneous variables" that although valid, were not considered significant to the problem. To illustrate, consider student responses during group work on the "Labor Force" problem (see Appendix C2). Dayton told her group that they were assuming "just 10 workers" even though this was an explicit parameter for the problem. For the same problem, Kane explained to his group that they were assuming each plant does not have to have exactly two women.

Occasionally students referred to assumptions that might be considered more significant in interpreting the simulation results. For example, while working through a problem on having children, Cade made the assumption there would be no twins and Bernard stated that the group was assuming independence of births. The class had discussed the concept of independence and its role in simulation, but few students explicitly stated this as an assumption. However, independence was implied by their choice of probability generator. When students generated random numbers on the calculator, the numbers are generated independently of each other. It is unclear whether any students made this connection. Additionally, there was little evidence of student discussion about the role of independence except when explicitly asked.

Students seldom referred to assumptions in a simulation process unless specifically asked. When assumptions were noted, they were often of a more trivial nature and part of the problem parameters. Yet, by the design of their probability generator, students implicitly assumed independence of outcomes, albeit unknowingly.

Construction of the Probability Generator

In Session 1 of the whole-class teaching experiment, students were given the "Counting Successes" activity (see Appendix C1). In this activity students used a simulation to determine whether one pays $0 or $1 for a cola worth $0.75. The objective of the problem was to eliminate the need for change. Students were given the assignment of digits for the probability generator, and they used a random number table, which had been used prior to the whole-class teaching experiment, to conduct the simulation.

Following the activity the teacher-researcher discussed with the students how the calculator could be used to generate random numbers.

The students quickly adopted the calculator's random number generator as the preferred probability generator for simulation problems. By Session 2 of the whole-class teaching experiment, most students were able to construct a valid probability generator using the calculator. Bernard's strategy was typical of most students. When asked to simulate randomly answering a five-question true/false test, Bernard told his group "1 is true and 2 is false" indicating there was an equal chance of answering either true or false for each question, and that he would use his calculator to randomly generate 1's and 2's.

The concept of equivalent probability generators proved to be slightly problematic for some students. A few students struggled with making the one-to-one correspondence between the probabilities in the problem and the probability generator. The following dialog, taken from Session 3, illustrates the difficulties that some of these students had. Bernard and his group members explained to Dayton why two probability generators were the same. Students were working on simulating an 80% free throw basketball shooter. Cade began by explaining he would use the integers from 1 to 8 to represent the probability of making a shot.

Dayton:	No, 7 [meaning use the integers 1 to 7].
Cade:	No, 8.
Dayton:	But then we should use 0 to 7.
Bernard:	You can do 0 to 7 or 1 to 8.
Cade:	You need ten numbers so . . . [Cade writes down the numbers]. On the calculator you can do 1 to 10, which is the same thing as 0 to 9 on the [random number] table.

Dayton: Yea, it's the same thing.

This group had a similar discussion the next day when Cade chose to use the integers 1 and 2 to represent the probability of having a boy or a girl. Lacey argued that the book had assigned the numbers 0 to 4 for a boy and 5 to 9 for a girl. Bernard and Cade explained to her these were "the same thing" meaning they were equivalent probability generators. Lacey's questioning of her peers' choice of random digits for the probability generator may have been an indicator of a fragile understanding of proportions. In other words, Lacey may not have immediately recognized the equivalence of the two probability generators because she did not see they were proportionally the same. As the following excerpt shows, Dayton's earlier problem with equivalent probability generators may also have resulted from problems with proportions. In designing a simulation for the "Women Working" problem, Dayton questioned Bernard's assignment of integers.

Bernard: You need to have 0 to 6 be men and 7, 8, 9 be women.

Dayton: Why?

Bernard: Because 30% are women.

Dayton: Why wouldn't you just do 8, 9 because they're only taking 2 women?

Lacey: Because we're dealing with 30%. That's the proportion we want.

These exchanges between the students support the analysis of the assessments. Recall that some of the students had struggled on the preassessment with proportions, and Dayton was one of these students. However, as the post- and retention assessments indicate, invalid reasoning related to proportions was not found in later instructional sessions. Although these exchanges between students may have helped some students, like Dayton, reflect on their reasoning about proportions, it is unclear what the exact nature of any change in proportional reasoning may have been.

Up until the end of the whole-class teaching experiment, the simulation problems involved independent outcomes. However, during Session 11, students were given the “Pay Your Bill” problem (see Appendix C3) in which the outcomes were not independent. In this problem, “Tom” is collecting $5 for delivery of the paper for a week from Mr. Bernoulli. Mr. Bernoulli offers Tom a choice of either paying Tom his $5 or letting Tom reach into a bag containing five $1 bills and one $10 bill and drawing two bills. Students were asked to design a simulation for this problem. Initially, students struggled with creating the probability generator for this situation. The following group’s conversation typified the difficulty students had with this particular problem.

Ugo: 1 to 5 equals $1, 6 equals $10. Randint(1, 6, 2). We’ll do this 20 times. Got $2, $2, $11. [Recall that Randint(1, 6, 2) will generate two random integers from 1 to 6, inclusive.]

[Ike begins to question the outcomes and rereads the problem for clarification.]

Ugo: Do you put the dollar back in?

Ike: It doesn’t say.

Sadie: If you put it back in, they’d be drawing the same thing.

Ike: The highest amount is $11.

Ugo: I say put it back in.

[They ask the teacher if they should put the bill back in, and she discusses the problem with them.]

Sadie: One person do randint (1, 6) and another person do randint (1, 5).

Ugo: If he picks a $10, he gets $11 no matter what. If I get 6 on my calculator, then he gets $11.

[Ugo uses his calculator to generate the first number and explains to Mai her role in generating the second number.]

Mai: Why am I doing the calculator? I'll always get $1.

Ugo: No. If you get a 5, then it's $11 because a 5 is $10 [For the second calculator, numbers 1 through 4 would represent drawing $1 and the number 5 would represent drawing $10.]

Most groups displayed similar type reasoning in that they used either two calculators to generate a valid trial or they used one calculator twice to generate two outcomes that combined to form a valid trial.

From the beginning, students had little difficulty constructing valid probability generators. The preferred probability generator proved to be the random number generator on the graphing calculator, which students quickly became adept at using. Furthermore, students had little difficulty in constructing a valid probability generator for simulation problems that involved dependent events even though all problems prior to that point had involved only independent events.

Construction of a Valid Trial

When asked to construct a probability simulation on the preassessment, about half of the students were able to construct a valid two-dimensional probability trial. Of those students who did not construct a valid trial, most had constructed a one-dimensional trial. Therefore, the activity used on the first session of the whole-class teaching experiment was purposefully chosen to have a one-dimensional trial. Once students had worked through a couple of simulation problems involving one-dimensional trials, students were given multi-dimensional trial simulation problems, starting with Session 2.

During Session 2, students were asked to simulate randomly answering a five-question true/false test. Overall, students made an easy transition from a one-dimensional problem to one that was multi-dimensional. The exchange between Bernard and Cade illustrates how easily most students made this modification. Bernard told the group, "1 is true, 2 false." Cade responded, ""We'll do 5 times." Working in tandem, Bernard assigned the random numbers for the probability generator (1 for true and 2 for false) and Cade determined what the trial would look like (5 outcomes representing the 5 questions on the test). One group, however, recorded their outcomes as one-dimensional outcomes.

Ondrea: 1 is true and 2 is false and you record it.

Ethan: It's as easy as that.

Thor: 1 correct and 2 incorrect?

Ondrea: 1 true, 2 false.

[Teacher approaches and asks group about their simulation.]

Ondrea: 1 is wrong, 2 is right.

Ethan: We both got 25.

Teacher: I'm confused. What did you do?

[Ondrea shows teacher her calculator screen.]

Teacher: What does your list represent?

Ondrea: The test.

Tasha: Your test has only 5 questions, not 50.

Ondrea: Thor and I did this yesterday. We combined everything instead of 10 sets of 5.

Ethan: We have to do this 10 times [meaning 10 trials].

[Teacher discusses with the group how the calculator can be used to generate the desired trial.]

In their attempt to simplify the recording of outcomes, Ondrea and Thor seemed unaware that their recorded outcomes did not accurately reflect the problem situation of simulating a five-question test; that is, a situation that involved ordered quintuples. After further discussion, the students changed their outcomes to replicate the outcomes of a five-question test. Students were not seen to combine outcomes in a manner similar to that of Thor and Ondrea's in any of the remaining teaching episodes.

Accepted Randomness of Outcomes

The concept of randomness is central to probability and by natural extension to the study of statistics. Therefore, how students accepted the random outcomes of a probability simulation was relevant to this study. The preassessment revealed that 10 students had exhibited the representativeness heuristic in their reasoning about the outcomes of the simulation. In spite of indications of representativeness on the preassessment, no evidence of this was found during the whole-class teaching experiment. Most students seemed accepting of the outcomes of their simulations. When simulating the true/false test, students would joke with each other that they had received a "perfect score" or that they had missed every question. They seemed to understand that all outcomes were possible, as some students would repeat the simulation to see how many perfect scores they could get.

This apparent discrepancy between the preassessment and instructional episodes may be partially attributed to the group structure of the classroom. In other words, students who exhibited representativeness on the preassessment may have taken fewer opportunities to share their thinking in a group setting. Furthermore, the representativeness heuristic appeared on the evaluation task of the preassessment where students were provided with contrived outcomes intended to help reveal misconceptions. In the classroom, students generated their own outcomes. Thus, students may have given more credence to their outcomes than those given to them in the assessments or, student-generated outcomes may have not produced “unexpected” results. That is, student-generated outcomes may have reflected the probabilities given in the problem, whereby students were not faced with a situation where the representativeness heuristic would be applied.

In anticipation of students struggling with the concept of randomness, the “Randomly Generated Outcomes” activity (see Appendix C4) was done during Session 5 of the teaching experiment. In this activity, students were presented with different strings of outcomes resulting from tossing a coin 50 times. The outcomes were generated either by students in an earlier class who were trying to simulate random outcomes or the string of outcomes was actually generated using the random-number generator on the calculator. A string of outcomes was presented and students were asked whether the string was student or calculator-generated. Thor began by arguing that if the string were random (calculator-generated) then HHHHH would not appear. While some of the students agreed with Thor, others disagreed. Breanna reasoned that because of randomness, calculator-generated strings could be generated by students and visa-a-versa so she did not believe it was a worthwhile activity. At this point, students had not resolved the issue of what might or might not be a string of random outcomes.

During a whole-class discussion later that day, some of the students began to make a distinction between randomness as it related to a single event and the extrapolation of patterns over the long run. In a problem assigned later that period, students were asked to record their longest runs when simulating 20 shots by an 80% free throw shooter. During a class discussion, the teacher-researcher noted that in 14 sets of recorded trials the basketball player never shot exactly 2 baskets. The following is an excerpt from that discussion.

Breanna:	No, he’s an 80% shooter.
Teacher:	Would it be surprising if he shot 20 for 20?
Class:	No.
Breanna:	It would be more surprising if he shot only 2.
Teacher:	Do you think it is more likely for him in one game to shoot 2 out of 20 if he’s an 80% shooter or more likely for him over 20 games to average 5 out of 20.
Kane:	It’s more likely for the one [game of] 2.
Class:	Yea.
Mai:	His average is 80% so they’re looking at a long period of time, and if he has a bad game, it’s going to average out.
Makaila:	If you shot 5 out of 20 for 20 games, you have to have an average of 80%. [Therefore, it is unlikely for an 80% free throw shooter to shoot 5 baskets out of 20 for 20 games.]
Teacher:	So if I change this to 5 games . . . [an 80% shooter to average 5 out of 20 shots over 5 games.]
Mai:	Yea, that’s different. Five games is not that many.

In this conversation, students seemed aware that single events or short runs were more likely to contain unusual or unexpected results than long runs. Furthermore, students were able to recognize the uncertainness in the short term, yet the long-term behavior could be predicted. In particular, students were generally not surprised by the outcomes of simulation problems. In this case Mai expressed a more developed understanding about randomness and its relation to long-term behavior. It is interesting to observe that Mai was a basketball player and the context of the problem may have contributed to her reasoning.

Calculated Empirical Probability

The preassessment demonstrated that although some students experienced difficulty with the two-dimensional nature of the problems, they were generally able to use the outcomes to calculate an empirical probability. During Session 3, students were asked to simulate a problem involving an 80% free throw basketball player and then use the results to determine the probability that she missed at least 3 shots in 10 throws. Although, most students had little difficulty calculating the probability, in a whole-class discussion, Breanna specifically asked, “How do you use your results to determine the probability?” Cade explained how one counts the number of three or more missed shots out of total trials. Up until this point, students had been asked general questions about what results they would “expect,” but had not been asked to use their results to determine the probability of an event. Breanna’s question indicated that some students were unaware of how to transform their simulated results into an empirical probability. Cade’s explanation may have helped students who were unsure of how to calculate this probability. Following this incident, students did not exhibit any difficulty in determining empirical probabilities.

One difficulty a student had in calculating the probability was related to her ability to make sense of the target outcome. The discussion from Cade's group provided some insight into Dayton's struggle. Referring back to the basketball free throw shooter during Session 3, students were asked to determine the probability that the basketball player makes at least 8 free throws out of 10 attempts if she is an 80% free throw shooter. Cade's group had been conducting the simulation and recording their outcomes. Recall from an earlier discussion that Cade had convinced Dayton that a trial consisted of 10 numbers representing the 10 shots.

Dayton: [Reading the problem to the group]. "What's the probability she makes more than 8?"

Bernard: That's where we take our 20 [trials].

Dayton: I don't understand how you do this.

Cade: You add up the number of times you make 8 or more . . .

Dayton: And take it out of 20.

As Cade interpreted it, Dayton's difficulty seemed to be connected to her ability to understand what the target outcomes were for the problem. Once the target outcomes had been identified and counted, Dayton knew the probability was determined by taking the number of target outcomes out of the total number of simulated trials.

In general, students did not have difficulty in taking their simulation outcomes and using them to determine the empirical probability, at least from the analysis of whole-class and individual group discussions.

Repetition of Trial

An integral part of a simulation process is to repeat the trial many times (Yates et al., 1999). Therefore, it was of particular interest to determine how students reasoned about the number of trials they considered sufficient, and how this might have changed during the whole-class teaching experiment. Prior to the whole-class teaching experiment, almost 50% of the students said that increasing the number of trials in a simulation would do little if anything to change the empirical probability.

There was generally little discussion about how many trials should be done for a given simulation, and in the end, students seemed to agree that around 20 trials was sufficient. This was in spite of evidence to suggest that in later instructional sessions students believed that more trials would provide more accurate data. In most cases, students were not told how many trials to conduct for a particular simulation. Therefore, it was up to them to determine how many trials would be appropriate or sufficient. For much of the whole-class teaching experiment, students tended to do between 10 and 20 trials per person in the group. They would then often combine their outcomes before determining the probability for the simulation problem.

During the third instructional session as students were determining the number of trials to conduct, Mai said to her group members “We should do a bunch, maybe 10 times. The group members then decided to each do 10 trials and combine their results. Mai and her group members did not discuss why 10 trials were satisfactory. In fact, when left to work in their groups, students seldom discussed why a particular number of trials were appropriate. However, Cade’s group was an exception. During the same session, Cade suggested to his group that they do 50 sets of data because “I think the more sets you use, the more accurate it is until a certain point. Then it just gets to be the same thing." Cade was able to articulate how a large number of trials affected the probability of a simulation problem. Cade was one of the target students who had been interviewed prior to the whole-class teaching experiment, and his suggestion to the group to do more trials may be attributed to the interview questions that focused on this particular simulation process.

On the following day while working in groups, the teacher asked Thor how many trials he should do. Thor responded, "About 4 million. It depends, until it [the empirical probability] stops changing." Although some students understood the implications of repeated trials, few verbalized this type of reasoning. Informed by student responses, the teacher-researcher and witness modified the hypothetical learning trajectory to include theoretical probability.

Theoretical probability provides students with concepts that help them develop an understanding of the purpose of repeated trials in a simulation process. Therefore, the focus of instruction for Session 6 was theoretical probability. After students had worked through some problems involving theoretical probability, they were asked to work on simulation problems. The teacher-researcher facilitated a discussion on how theoretical probability related to simulation. At the same time, she used the overhead graphing calculator to reconstruct the simulation problem students had just completed. In this particular problem, students were asked to design a simulation to determine the probability of two out of three traffic lights being green. Students had determined that the theoretical probability was 37.5% of getting two green lights.

Teacher: [Simulating on overhead.] How many trials should I do?

Ugo: 30 something.

[Teacher did 30 trials and empirical probability for problem was 52%.]

Teacher: Would it [the results] change if I did more simulations?

Kane: I don’t think it will affect the data. Each day you have the same situation.

Teacher: What if I’m looking over the long run?

[Students did not respond, so teacher did more trials on overhead calculator.]

Teacher: Any observations? Will I get to 20%? [Will the empirical probability approach 20%.]

Sadie: If you do it a long, long time then you’ll probably get close to our theoretical value.

As the teacher-researcher and witness conjectured about the hypothetical learning trajectory for Session 9 of the whole-class teaching experiment, they felt that many students did not yet understand how many trials was “enough” and why many trials were desirable. As a result, a lesson was designed to demonstrate graphically what happened to the empirical probability as the number of trials was increased. The graphing calculator was used to simulate and collect data. The data were then used to graph the empirical probability against the number of trials. A whole-class discussion followed the teacher-researcher demonstration.

Teacher: Tell me what you can interpret about the graph.

Ugo: The more trials that you run, the more I think accurate is the word I’m looking for, either accurate or precise. The graph fluctuates less because you have more data, and it’s harder to fluctuate your data points.

Kane: The more data you gather, the closer you get to your predicted value.

Teacher: How many trials should I do to simulate something?

Lacey: The more trials you do, the closer you'll be to the estimated probability.

Teacher: What happens when I do 500 trials? This [referring to the graph] is 202 [trials].

Dayton: Flatten out.

Teacher: You don't think it's going to peak any more?

Dayton: Nope

Breanna: I don't think 20 trials would be wrong as long as you say you did 20 trials.

Teacher: So if you were trying to simulate the spread of a disease, 20 trials would be enough?

Breanna: I think doing 20 trials isn't necessarily wrong, but of course more trials would be better. It'll give you a more accurate number [empirical probability].

According to the above dialog, the students had developed a rich, conceptual understanding of how the number of trials in a simulation affects the empirical probability. Students also seemed to make a connection between the behavior of the empirical probability and the increased number of trials. The strength of this connection was borne out in the post- and retention assessments where students referred to the "target" probability value by many different terms, such as the "predicted value," the "estimated probability," or a "more accurate number."

Even though the discussion of repeated trials seemed to have been resolved by Session 9, students continued to struggle with how many trials were enough. As Breanna argued, "I think doing 20 trials isn't necessarily wrong, but of course more trials would be better." Students ultimately decided that while many trials was ideal, for classroom purposes 20 trials should be enough.

Summary of Students' Reasoning During the Whole-Class Teaching Experiment

The reasoning exhibited by students in the classroom during the whole-class teaching experiment supported the findings of the assessments in that students were able to construct valid probability generators and trials for simulation problems. Student difficulties related to the construction of the probability generator were connected to the concept of proportionality. After some discussion, students were able to use their simulated results to determine an empirical probability. Finally, student reasoning during the whole-class teaching experiment concerning the number of trials mirrored student reasoning found on the assessments. That is, students would complete approximately 10 to 20 trials but would reason trials should be repeated until the empirical probability approaches the theoretical probability. It should be noted that the judgment heuristic of representativeness did not appear during the whole-class teaching experiment. The analysis of the classroom dialogs provided insight into the evolution of some of the reasoning strategies, and the role that group and whole-class discussion played in affecting change in students' reasoning strategies.

The Role of Technology

In its Advanced Placement Statistics curriculum, the College Board (2000) mandates that students have access to technology. Moreover, technology lends itself to simulation in that it enables students to conduct many trials in a relatively short period of time. Thus, one of the research questions this study sought to address was how technology impacted students' reasoning about probability simulation.

Because of time constraints and limited availability to computer labs, technology was confined to the graphing calculator. As mentioned previously, students were required to own a graphing calculator and most students had a Texas Instrument's TI-83 graphing calculator. However, one student in the class owned a TI-86 graphing calculator that had significantly different menu options than the TI-83. It is also noteworthy that prior to the whole-class teaching experiment, students had been using a graphing calculator for at least two or three years in other mathematics and science classes. Therefore, students were comfortable with the technology. Nonetheless, it became evident that some students knew more about the technology than others and certain students were quicker to pick up the newly learned features.

Table 12 presents a summary of the ways in which technology impacted the learning of the students during the whole-class teaching experiment. An elaboration of each category follows the table.

Table 12

The Impact of Technology

Impact
In class, students preferred the calculator as a probability generator to other devices
Calculator syntax focused students on simulation components
Calculator provided a transparent medium for dealing with dependent events
Programming capabilities of the calculator had limited value
Calculator provided a common language to discuss simulation

Preferred Calculator as Probability Generator Over Other Devices

During classroom instruction, students exhibited a preference for the calculator as the probability generator over other devices, such as spinners, chips, etc. In Session 1 of the whole-class teaching experiment, students completed a simulation activity using a random number table. After discussing the activity and comparing results, the teacher-researcher introduced students to the random number generating feature of the calculator. By all indications, most students had not seen or used this feature before this time. Students seemed to immediately adopt the random number generating feature of the calculator as their preferred probability generator. When spinners were brought in the following day, the students discussed valid ways to use the spinners for simulation problems, but preferred the efficiency of the calculator.

The post- and retention assessments further support the willingness of the students to use the calculator to simulate problems. On the preassessment none of the 21 students used a calculator to design their simulation, however, in later assessments about 50% of students described a simulation design that involved using the calculator to randomly generate the outcomes. In spite of the focus on using the calculator to do simulations in the classroom, over half of the students chose a valid non-calculator method of simulation on the post- and retention assessments. These results indicate the mental flexibility the students demonstrated in constructing valid probability generators using both technology and other random generators, such as chips, balls, and spinners.

Calculator Syntax Focused Students on Simulation Components

A second impact of technology was that the calculator syntax helped to focus the students on two of the simulation components, namely the assignment of random digits and the trial. Students became very skilled at using the calculator to simulate problems. Early on in the whole-class teaching experiment, students were rattling off such lingo as randint (1, 2, 5) to simulate a woman having 5 children, and everyone understood that this jargon referred to the assignment of the numbers 1 for boy, 2 for girl (or visa versa), and the five represented the 5 children. The calculator syntax focused the students on two of the components of simulation: assigning random digits to match the probabilities of the problem and defining the trial. Therefore, the graphing calculator helped students to think about the trial. That is, in entering the syntax into their calculator to generate the random numbers, students had to first be able to assign random digits appropriately. Then they had to be able to determine how many outcomes were required for a valid trial. So by using the calculator to simulate problems, students had lots of practice in both assigning the digits to match the probability of the problem and in defining the trial; in essence, their calculator-generated representations made the parameters of simulation explicit.

Calculator Provided a Transparent Medium for Dealing with Dependent Events

Students' proficiency in using the calculator to simulate problems enabled students to view the calculator as a transparent medium for dealing with dependent events. From Session 2 of the whole-class teaching experiment, students used their calculators to run simulations by generating strings of random numbers. Up until the final session, simulation problems involved independent events. In other words, the probability for one outcome was not affected by the previous outcome. During Session 12, students were given the "Pay Your Bill" problem that involved pulling two bills out of a bag (see Appendix C3). The second bill drawn was not independent of the first bill drawn. Thus, the syntax students had been using to simulate problems on the calculator was no longer valid. To enter something like RandInt (1, 6, 2) would assume that the second bill drawn had the same chance of being chosen as the first bill drawn. Even though a couple of groups faltered initially, the groups ultimately designed valid simulations. Ike shared with the class how his group handled the problem.

> The actual amount of money we're going to need is $10 or $11. We, on one calculator, got a random number from 1 to 6, and then, if we got a 6 we automatically knew we had $11 because we had 6 equal to $10 bill. And if we got any other number, we got a second calculator to make this go faster and had a randint 1 through 5, 5 being a $10 bill. So if we got say a 2 on the first calculator and go to a second calculator and got a 5, we got $11. But if we got a 3 on the second calculator then it would just be $2.

Ike's group realized that they would have to alter how they usually designed their simulation process, and explained how they used two calculators to design a valid probability simulation. In other words, the calculator provided a transparent medium for dealing with dependent events. Students in this group were able to make the connection between the context of the problem to the use of two calculators to reflect the dependent events in the problem.

<u>Programming Capabilities of the Calculator Had Limited Value</u>

In an attempt to utilize the programming capabilities of the TI-83 calculator, it was determined that these capabilities had limited value to the students. Because of the programming capabilities of the calculator, during Session 4 of the whole-class teaching experiment, students were given an activity that guided them through a programming exercise. The purpose of the program was to run a defined simulation, calculate the empirical probability for all accumulated trials, and display both the outcomes and the updated probability. Once the initial programming was completed, the students needed only to hit the enter key to update the outcomes and display the new probability. In this manner, as students completed more trials, they could see the probability settle towards a particular value. Although students appreciated the effectiveness of the program, generally they were frustrated with the time-consuming nature of the programming. Additionally, many students had done little if any programming previously and did not try to make sense of what the commands were doing. According to Kane, “By the time you press all the buttons, you could have it done.” Evan was particularly frustrated. He was the only student to own a TI-86, and no one else in the class, including the teacher-researcher, was able to help him translate calculator commands for his particular calculator. Evan was able to use his calculator for generating random numbers, like that for the TI-83. The activity for programming simulations on the TI-83 seemed to have limited value for instructional purposes because of the frustration experienced by students in programming their calculators. Some students, though, did use the program to run many simulations and commented on the results. Oliver ran his simulation “195 times,” and Thor told his group, “I’ve done a lot. It’s giving me the probability overall.” Since students had limited themselves to about 40 trials per simulation, they did not readily see the advantages in automation that programming could offer if, for example, 500 or more trials were completed.

Calculator Provided a Common Language to Discuss Simulation

A final impact of technology on how students reasoned about probability simulation was that the calculator provided students with a common language with which to discuss probability simulation. Recall that one way to simulate a problem is to enter “RandInt (1, 2, 10).” This calculator command will randomly generate 10 integer values of 1 or 2. Using the syntax of the calculator, students would use calculator-speak to discuss their simulation designs with one another. For example, in simulating 20 shots by an 80% free-throw shooter, Breanna told her group that she was doing “randint 0, 9, 20.” Although Kane asked Breanna for an explanation of what her directions meant, Makaila was able to translate Breanna’s abbreviated calculator-speak into directions of how to use the calculator to simulate the problem, and at the same time, Makaila also correctly interpreted Breanna’s numbers for Kane. Other students also referred to this calculator-speak to explain their simulations to their classmates.

In summary, technology influenced how students reasoned about probability simulation in a number of ways. Prior to the whole-class teaching experiment, students had not used the random number generator feature on the calculator. However, once they were shown how to use it, they embraced its use for the remainder of the teaching experiment. Furthermore, students prior familiarity and comfort with the technology extended into this unit. However, the concentrated use of the calculator in the classroom did not hinder students using other valid probability generators on assessments. In general, students found the calculator to be the efficient and easy device to use for probability simulations. In spite of this, students struggled with the programming capabilities of the calculator. However, this struggle could be partially attributed to the lack of classroom time devoted to programming. In an instructional environment, the calculator proved to the preferred probability generator, and the number of students who used the calculator to create simulations increased from the preassessment to the post- and retention assessments. The calculator was also found to focus students on the components of a simulation process as well as provide a transparent medium for simulating dependent events. Finally, the use of the calculator provided students with a common language with which to discuss their simulation designs.

Qualitative Analysis of Students' Beliefs

Beliefs held by students can influence how they reason about probability simulation. In other words, reasoning and beliefs together determine how a student approaches and thinks about problems. Thus, a goal of this study was to examine the beliefs students exhibited as they reasoned about probability simulation. As in the case of students' reasoning, multiple sources provided data of students' beliefs about probability simulation: (a) pre-, post-, and retention assessments; (b) audio and videotaped instructional sessions; and (c) teacher-researcher and witness field notes. Schoenfeld's (1985) definition combined with other research about beliefs (Fischbein & Gazit, 1984; Fischbein et al., 1991; Fischbein & Schnarch, 1997; Garfield & Ahlgren, 1988; Piaget & Inhelder, 1975; Shaughnessy, 1992; Zimmermann & Jones, 2002) guided the identification and analysis of students' belief about probability simulation. More specifically, the analysis of beliefs was guided by codes like the following: a student said they "believed," a student responded to a question that asked what they believed, or a student added an afterthought to their response that reflected a disposition.

The qualitative analysis of students' belief has been separated into two sections. The first section contains a summary of the analysis of student beliefs that were revealed predominately in the assessments with supporting evidence from instructional sessions. The second section summarizes student beliefs that were unique to the classroom and not necessarily found on the written assessments.

Students' Beliefs on Pre-, Post-, and Retention Assessments

Table 13 contains a summary of the beliefs students expressed during the various assessments. It also shows the number of students found to hold each belief. These beliefs have been categorized as either helpful or problematic. Helpful beliefs are considered to be potentially beneficial in learning because they can be linked to a normative view of simulation. Problematic beliefs, on the other hand, may constrain students from learning various aspects of simulation because they incorporate misconceptions that are contrary to a normative view of the simulation process.

Helpful Beliefs on the Pre-, Post-, and Retention Assessments

Students' helpful beliefs were diverse and these beliefs were held to varying degrees. Most of the helpful beliefs identified in the assessments were related to trials in a simulation process. Generally, students believed the probability generator should correspond to the probabilities in the problem. Other helpful beliefs were related to increasing the number of trials and about assumptions in a simulation problem.

Inherent assumptions in a simulation model. Students demonstrated the belief that assumptions were inherent in a simulation model. The number of students who held this belief increased from the preassessment to the retention assessment. Of particular interest, though, was how beliefs about the nature of the assumptions changed across assessments. The beliefs about assumptions initially held by students were more idiosyncratic than those exhibited by students after the whole-class teaching experiment.

Table 13

Student Beliefs by Assessment

Belief	Number of Students Pre-	Post-	Retention
Helpful beliefs			
Inherent assumptions in simulation model	8	8	15
Probability generator should correspond to given probabilities	18	21	21
Number of trials (n) that should be simulated			
• $n \leq 10$	6	1	0
• $10 < n < 100$	8	3	8
• $n \geq 100$	4	15	10
• Other	3	2	3
Context influences number of trials	0	5	0
Increased trials make empirical probability more “accurate”	7	16	15
Problematic belief			
Probability generator should not correspond to the given probabilities	2	0	0
Representativeness	10	3	0
Outcome approach	2	0	0

Of the 8 students who held the belief that assumptions were part of the simulation process on the preassessment, 7 expressed some type of idiosyncratic assumption, similar to Lacey’s assumptions that were discussed previously. Recall that Lacey assumed that the situation would change according to the season, day, and so on. Although the number of students who held the belief that assumptions were part of simulation remained at 8 for the post-assessment, the assumptions students referred to changed in nature. Assumptions were less idiosyncratic and related more to the mechanics of the simulation process, such as the assumption of independence of events and the replacement of a drawn chip.

Breanna's response on the post-assessment was typical of students who noted that independence of events was assumed as part of the simulation: "Lora must take note of the fact that she is assuming one bat is independent towards the 2^{nd} bat of Beth. Therefore, Lora assumes that each time Beth goes up to bat she will always have a 30% chance one will get onto base at each bat." At the time of the retention assessment, 15 of the 21 students held to some degree the belief that assumptions were a part of the simulation process, and all explicit assumptions were related to independence and replacement.

Probability generator should correspond to given probabilities. A strong belief that appeared even before the whole-class teaching experiment was that the probability generator should reflect the probabilities that were provided in the contextual problem. On the pre-assessment 18 of the 21 students held this belief. According to their responses, these students either explicitly or tacitly believed the simulation was valid because the probability generator matched the probabilities of the problem. Makaila's response to the preassessment pizza problem exemplified this belief: "Yes [the simulation is valid], because he took each 'order' with the same probability 6:10 for meat and 4:10 for without meat. This means that the data should accurately show how probable two meat pizza orders are." Makaila determined the probability generator was valid because she made a one-to-one correspondence between the probability generator of chips and the probability stated in the problem.

Number of trials that should be simulated. Student responses to the question "how many times would you do the simulation" provided insight into students' beliefs about the number of trials needed for a simulation process. As Table 13 reveals, the number of trials students believed should be simulated increased substantially from the preassessment to the post-assessment. Before the whole-class teaching experiment, 14 students expressed the belief that less than 100 trials was appropriate and only 4 of the students believed that more than 100 trials should be completed. On the post-assessment, these numbers were reversed. More specifically, 4 students held the belief that less than 100 trials on the post-assessment was satisfactory, whereas 15 students believed they should do more than 100 trials.

Two of the students whose beliefs about the number of trials were classified as "other" on the preassessment had designed their simulation to actually listen to the radio and both said they would do this for a month. The third student when asked how many times he would repeat the simulation process answered "many, many times." Students' "other" responses on the post- and retention assessments referred to non-numeric answers. That is, these students believed that trials should be repeated until the probability settled around a particular value. More specifically, these students seemed to hold the belief that the more trials one did, the more accurate or better the results. This belief was also reflected in the responses of students who said they would do more than 100 trials. Thor's response typified the reasoning of these students, "I would perform the simulation 100's of times to acquire a good estimate of the actual [probability]."

<u>Context influences number of trials.</u> Students in this group students believed the context of the problem directly influenced how many trials one should do. This belief was identified in students who justified the number of trials they would do by referring to the context of the problem. Interestingly, this belief only appeared on the post-assessment. When students were asked how many trials they should do, 5 students on the

post-assessment justified their answer by referring to the context of the problem. The problem involved the probability of two failed engines on a space shuttle. Thus, students seemed to believe that precision was especially critical. As Thor explained when asked how many simulations he would do, “As many as possible. This is a space shuttle. Lives depend on this value. You want it to be accurate to the nth degree.”

Increased trials will make empirical probability more “accurate.” The final helpful belief identified on the assessments was the belief that increasing the number of trials will make the empirical probability more precise, or as many students stated, more “accurate.” Prior to the whole-class teaching experiment, one-third of the students held this belief, and on the post- and retention assessments, the number of students holding this belief more than doubled. Other students believed that eventually the probability would level off at some value or it would fluctuate less. Basically, all of these beliefs reflect the disposition of the students that increasing the number of trials will cause the empirical probability to approach a theoretical value. Kacy stated this belief succinctly, “the greater amount of results will give a more accurate probability.” This belief was also evident during instructional episodes. Kane explained, “The more data you gather, the closer you get to your predicated value.”

Problematic Beliefs on the Pre-, Post-, and Retention Assessments

The problematic beliefs that appeared on the assessments were related to judgment heuristics. Furthermore, these beliefs appeared predominately on the preassessments and did not appear on the retention assessment. The most common heuristic used was that of representativeness, although two students seemed to demonstrate the outcome approach on the preassessment.

Probability generator should not correspond to given probabilities. One problematic belief that was identified on the assessments was the belief that the probability generator given in the evaluation task should not correspond to the probabilities stated in the problem. Two students did not believe the probability generator was valid for the pizza problem. Instead they believed that the chance of either type of pizza should be equal rather than match the probabilities given in the problem. Because of their belief, they suggested making the number of chips of each color the same. Kacy's explanation reflected her belief, "There is a better chance that the next two pizzas will have meat because there are 2 more red chips which stand for meat than blue chips. So Stan has a higher chance pulling out red but he still had a good chance of getting blue." Therefore, Kacy recommended that the probability generator be changed so that the number of chips of each color should be equal. In spite of the probabilities given in the problem, Kacy believed that there should be an equal chance of either meat or veggie pizza. However, on each of the post- and retention assessments all 21 students held the belief that the probability generator should correspond to the probabilities given in the problem.

Representativeness. As shown in Table 13, the representativeness heuristic (Kahneman & Tversky, 1972) appeared in 10 of the 21 preassessments. Moreover, this belief did not appear on the retention assessment. In particular, these students held the belief that the outcomes of the simulation should always approximate the given probabilities in the problem. In the pizza problem, the outcomes of meat and no meat were purposefully made to equal each other. When asked if the outcomes could be used to determine the probability Tavi explained, "No, because your percentage probability has changed and you’re left with a 50/50 chance of having the next two orders containing meat.” Like Tavi, the 10 students who exhibited the representativeness heuristic believed that the outcomes for the pizza problem should reflect the 60% and 40% for each type of pizza as given in the problem.

A more subtle form of representativeness appeared in the reasoning of 3 students. These students demonstrated the belief that repeated trials would not substantially change the empirical probability because it was all related to “proportions.” Kane, Bernard, and Macy exhibited the representativeness heuristic in their belief that given a valid probability generator, simulated trials would always represent the probability of the problem regardless of the number of trials conducted.

In Kane and Bernard’s case, they persisted in their beliefs that increasing the number of trials would have little or no effect as is evident by their responses on the assessments. Kane wrote on each of his three assessments that the probability would not change if the number of trials were increased. Kane’s response on his post-assessment provided a glimpse of his fragile beliefs. When he had simulated the problem, his outcomes resulted in an empirical probability equal to the theoretical probability for the problem. Kane wrote that the probability would not change significantly since his empirical probability was equal to the theoretical. Although Kane seemed to have some understanding that the empirical probability would eventually approach some number, he did not understand the inherent randomness of a small number of trials and its effect on the probability. Said another way, Kane was unaware that with a small number of trials, the empirical probability could fluctuate dramatically.

Bernard’s more detailed assessment responses reflected his belief that since the problem dealt with proportions the empirical probability would not change. When asked if the probability would change, Bernard said “No, because it is dealing with percents.” In his post-assessment interview, Bernard explained that he did not think the probability would change “because in most cases there is a theoretical probability and every time you do it, it gets closer and closer to that. And so it wouldn’t really matter the further you get.” Macy’s beliefs about the empirical probability seemed to vacillate across the three assessments. In the pre- and retention assessments, she believed the probability would become more accurate. On the post-assessment she stated that the probability would “probably only change a couple percents because the data will still be the same for the most part.” Macy, like Bernard, seemed to believe that the probability would stay the same regardless of the number of trials because it was all “proportional.” These students believed that the problem was basically about proportions, and that the empirical probability of a simulation would mirror or be representative of the original problem regardless of the number of trials.

<u>Outcome approach.</u> A final problematic belief identified was that of the outcome approach. People who reason using the outcome approach believe they are asked to predict the outcome rather than determine the probability of an event (Konold et al., 1993). In the preassessment, Ingrid and Kacy exhibited the outcome approach heuristic. When asked to design a simulation for the radio problem on the preassessment, Kacy provided a valid simulation design and Ingrid’s explanation was incomplete. However, both students proceeded to predict what they believed would happen. Kacy wrote, “I predict a 30% chance of hearing another hip-hop song.” Ingrid responded, “My guess

would be one hip-hop and one alternative." The unsolicited responses of these girls may be an indication they held the belief that in the process of designing a simulation, they were also being asked to predict the outcomes of the simulation. The outcome approach never appeared again.

Students' Beliefs During and After the Whole-Class Teaching Experiment

As described above, a variety of both helpful and problematic beliefs appeared on the written assessments. A number of additional beliefs became apparent during the analysis of the whole-class teaching experiment as well as during the target students' interviews that were conducted immediately after the conclusion of the teaching experiment. Again, these beliefs have been categorized as either helpful or problematic.

Helpful Beliefs During the Whole-Class Teaching Experiment

Three helpful beliefs appeared during the whole-class teaching experiment: (a) empirical probability was more reliable than theoretical; (b) number of trials was different for practical classroom use versus the "ideal"; and (c) the purposes of simulations were useful but limited. The first helpful belief discussed in this section appeared during the target interviews when students were asked whether empirical or theoretical probability was more reliable. The other two helpful beliefs appeared during the whole-class teaching experiment, and each of them generated intense classroom discussions. These beliefs were related to the role of simulations and how many trials were required for a simulated process. Because of the scope of questions asked on the assessments, none of these beliefs had appeared in the analysis of the assessments.

Empirical probability versus theoretical probability. One helpful belief that was held by students was the belief that the empirical probability was actually more reliable than the theoretical probability. During the post-assessment interviews, the interviewer asked 3 of the 4 students a follow-up question; “Let’s say that you have theoretical probability and then the one from the simulation. Which one is more dependable in terms of describing the situation?” Of the 3 students, 2 students responded that the empirical probability was more reliable than the theoretical. According to Lacey, “I would say experimental [empirical] probability just because that is using an actual simulation as long as there is no bias. Just because theoretical, you know, it’s how it’s suppose to be but things don’t always happen. You don’t always have the ideal probability.” Bernard expressed similar beliefs and reasoning. This belief also appeared during the whole-class teaching experiment when the teacher-researcher asked the class a similar question. Breanna answered, “I just like simulated better. It seems more real.”

In his response, Cade seemed to believe that collecting the actual data rather than simulating it would serve two purposes. One was to provide the theoretical probability, and the other was to provide more reliable data. Cade reasoned “. . . if you’re actually doing it and getting the data then it’s a lot better than trying to simulate this. I mean, for all I know, I could be way off than actually getting the data.”

Lacey, Bernard, and Breanna believed that the simulation process inherently contained random fluctuation that would reflect a more realistic situation, unlike a theoretical probability that was void of random fluctuation. Cade, however, held the belief that there was too much randomness inherent in the simulation process. Therefore, a theoretical probability was more reliable.

<u>The number of trials: Ideal versus practical.</u> Students who exhibited this belief felt there was a difference in the number of trials that should be done “ideally” if time and

effort were not a factor and the number of trials that could "practically" be done, say for a class problem. Throughout much of the whole-class teaching experiment, students discussed how many trials should be done for a simulation process. Toward the beginning of the instructional sessions, most students held the belief that between 10 and 20 trials was sufficient. Kane, in particular, believed that many trials were unnecessary. During Session 4 Kane told Dayton that he had done 10 trials, and when Dayton said she had done 15, Kane informed her, "it's a waste to do 5 more" [than he had done]. This comment supported Kane's beliefs on his assessments when he had stated that completing more trials would have little or no effect on the probabilities.

Breanna seemed especially concerned with how many trials were needed for a simulation process. She held a strong belief that more trials were more accurate, but she seemed overly concerned about exactly how many trials was enough. During the final session, Breanna reasoned with the teacher-researcher:

> How can you set a number? If I were to do this and say did 50, you'd say 'oh no, you didn't do enough to actually get that number.' I would disagree with you. I'd say I know I did 50 trials and this is what I got from it. You're like saying like if you did 100 trials, you're wrong from the person who did 200 trials. Not necessarily wrong, just not as accurate. You're saying if you want to find a statistical answer you have to do as many [trials] as possible, but I don't understand how someone can set that number.

Breanna understood the implications of repeated trials, but her belief about the number of trials needed for a simulation seemed connected to how her answer would be marked. Breanna was concerned that her answered would be considered as incorrect because she completed an insufficient number of trials, in spite of the fact that a person could not know exactly how many trials was sufficient. After further discussion, the students agreed there was an ideal number of trials and a practical number of trials for classroom purposes. Recall that a similar belief was found on the assessments. After the teaching experiment, students generally recorded between 20 and 50 trials, but when asked how many total trials they would do, 15 students demonstrated the belief that more than 100 trials should be completed.

Beliefs about the purpose of simulations. In general, students believed that simulation was a useful tool. That is, they believed that simulations were considered helpful in providing information. However, when asked why someone would want to simulate an event, most student responses were limited to examples used in class; however one student generated an idiosyncratic response.

In his post-assessment interview, Cade suggested that simulations could be used in marketing to determine what "percentage of people are gonna use this product." Later, Cade also said simulation could be used to test products. Bernard and Dayton's answers were limited to examples that had appeared in class or on the assessments, such as engine failure and determining free throw percentages. Lacey's interview response revealed a belief that simulations were used to determine a probability, however, her suggested uses for simulation were idiosyncratic. When asked, "If it's everyday life when would you ever see yourself using simulation?" Lacey answered, "When would I use probability? Um, I would probably see the probability that I fall asleep before I get my homework done. The probability that I'll see my friends on a weekend, or I'll have to work on a Saturday." Although Lacey's suggestions were unlike class examples, they were idiosyncratic in that they contained little practical value.

Approximately halfway through the whole-class teaching experiment, the teacher-researcher asked students "why do we do simulations? What's the point?" Lacey said it was “to be more accurate.” Edison suggested, “To test a theory. To see the difference between simulated and theoretical.” And Ondrea responded, “To make it at least as random as possible.” During the last session of the whole-class teaching experiment, Sadie said that the point of simulating an event was “to predict.” Taken together, these responses indicated that students were beginning to develop rudimentary beliefs about the purpose of simulation. That is, students were beginning to believe that a randomized simulation could be used to obtain data to test a theory.

Problematic Beliefs During the Whole-Class Teaching Experiment

Beyond the written assessments, only one problematic belief was found during the instructional episodes. This was the belief that a theoretical probability could always be determined. The students did not seem to have any conception that determining the theoretical probability in some complex situations could be intractable. This belief proved to be very resistant to change.

When students discussed the number of trials required for a simulation process or the purpose of conducting simulations, students would often refer to the “theoretical” probability. Students talked about the empirical probability approaching a theoretical probability or that a simulation was used to verify if the results were the same as the theoretical probability. As Edison was quoted earlier, the purpose of doing a simulation was “to see the difference between simulated and theoretical.” Thus, the teacher-researcher posed the following question to the students; “Can we always determine the theoretical probability?” Breanna responded, “Yes. I think so.” Many others in the class echoed this response. The general belief of the students was that a theoretical probability could always be determined. As a result of this dialog, the teacher-research and witness chose to have the students engage in the “Thumbtack Activity.” Even though this activity was not a simulation, but rather an actual experiment, it was chosen to help students see that a theoretical probability was not always possible to calculate. When asked how they would determine the theoretical probability of a tack landing point-touching the floor if 10 thumbtacks in a cup were tossed onto a flat surface, Kane replied “use physics.” On the following day in an attempt to create a perturbation in the students’ beliefs that a theoretical probability could always be determined, students were presented with a wooden cube in which a corner had been haphazardly cut off. The exposed surface had been painted black. Students were again asked how the theoretical probability could be determined and again they said physics could be used. What then transpired was a whole-class discussion about the number of trials required for a simulation (discussed earlier) and theoretical versus simulated results. Following is an excerpt of that discussion:

Ondrea: You’re shooting for doing trials to get to that 50% [theoretical] but you know it’s 50% so what’s the point of proving you’ll get it?

[Some discussion has been omitted for brevity and flow of discussion.]

Teacher: [To Ondrea] Since we can figure out the probability, why do we do the simulation?

Ondrea: Exactly!

Sadie: Sometimes you’re not going to know the theoretical probability.

Breanna: Sometimes you can get the theoretical probability in 2 trials. Why would you need 1000?

Teacher: Is that the goal, to get the theoretical probability?

Cade:	I think the point of simulation is to prove the theoretical probability. Like I can walk up to him and say you have a 1 in 50 chance of dying.
Teacher:	Is there always a theoretical probability?
Students:	[Some respond “yes” and others “no.”]
Teacher:	Give me an example where there isn’t a theoretical probability. [Students do not offer any suggestions.] So how do you figure out the probability that someone dies from cancer?
Cade:	You can get the entire population.
Teacher:	That didn’t come from theoretical mathematics.
Ugo:	[To the class.] You have to look at the data, that’s her [the teacher’s] point!

At the end of this last instructional session, many students continued to hang on to the belief that a theoretical probability could always be determined for any event. However, a few students were beginning to demonstrate a belief that simulated data could provide useful information for prediction purposes.

Summary of Students’ Beliefs During the Whole-Class Teaching Experiment

Unlike student reasoning, student beliefs revealed during the whole-class teaching experiment differed from those beliefs exhibited on the assessments. The analysis of group and classroom dialog revealed that some of the beliefs evolved and transformed in a helpful manner. That is, students’ beliefs about empirical probability, the number of trials for a simulation, and the purpose of simulations changed in a manner that would be considered a more normative point of view. At the same time, a problematic belief was also revealed: a theoretical probability could always be determined. For some students, this belief was deep seeded and difficult to change.

Sociomathematical Norms and Classroom Mathematical Practices

Throughout the whole-class teaching experiment, students became quite adept at constructing and conducting simulation problems. In the process of learning about the simulation process, four sociomathematical norms arose during both the small group discussions and the whole-class discussions. The sociomathematical norms that emerged were general in nature and refer to what constituted a valid (normative) simulation or a valid approach to some part of the simulation process. The classroom mathematical practices that were identified indicate the link between sociomathematical norms and mathematical practices; that is for each sociomathematical norm there are a number of corresponding mathematical practices (Cobb, 1999). The sociomathematical norms and their corresponding classroom mathematical practices are summarized in Table 14.

Table 14

<u>Sociomathematical Norms and Classroom Practices</u>

Sociomathematical Norms	Classroom Mathematical Practices
What counts as a valid simulation	• Must use a valid probability generator • Need to ensure a trial is valid • Must conduct at least 30 trials per group
What counts as a valid probability generator	• Must have 1-1 correspondence to probabilities given in problem • Will accept equivalent probability generators • Will use the calculator as the mechanism for generating a trial
What counts as a valid trial	• Number of outcomes in a trial must correspond to target question
What counts as a sufficient number of trials	• Will combine results of 10 trials per person in group • Number of trials depends on whether we are dealing with an “ideal” or a “practical” situation

As indicated by Table 14, a process can on one level be considered a sociomathematical norm and on another level be considered a classroom mathematical practice. For example, as students evaluated the validity of a simulation, they developed classroom mathematical practices about the probability generator, the trial, and conducting the simulation to collect data. These led to the need for new sociomathematical norms. For example, as soon as students adopted the practice of using a valid probability generator, they needed to establish a new sociomathematical norm, “what counts as a valid probability generator" This generative process reveals the reflexivity between sociomathematical norms and classroom mathematical practices.

What Counts as a Valid Simulation

Beginning with Session 1, students began to ask the question "What counts as a valid simulation" In the "Counting Successes" activity (see Appendix C1), students were given detailed information to carry out the simulation process. Throughout the activity, students repeatedly questioned and discussed with each other the specifics of the simulation. For example, Dayton clarified with her group that the probability generator consisted of "75 or less you pay $1, greater than 75 you pay nothing." Dayton further explained that one should "look at the first two of every one." That is, using the random number table, use two digits to simulate the random number. Lastly, Dayton also focused her group on the number of trials to conduct when she asked them "How many times do we do it [the simulation]?" From the sociomathematical norm of "what counts as a valid simulation" three classroom mathematical practices emerged: (a) must use a valid probability generator, (b) need to ensure the trial is valid, and (c) must conduct at least 30 trials per group.

Classroom mathematical practices related to a valid simulation process. The first classroom mathematical practice that emerged was that a valid probability generator must be used. The practice began to appear in Session 1 as students reasoned about the validity of the probability generator specified in the "Counting Successes" activity. During Session 2, the teacher-researcher gave students two examples of probability generators for the "Counting Successes" problem. One probability generator was equivalent to the one students had been given in the original problem, but used different digits. The second probability generator was incorrect. When asked if these were valid probability generators for the problem, students recognized the one-to-one correspondence of the alternative generator that was valid, and they also noted that the second probability generator was invalid because there was not a correspondence between the probabilities in the problem and the second probability generator. At this point, students were beginning to develop a taken-as-shared approach (Cobb, 2000) to requiring the use of a valid probability generator for a simulation design. During Sessions 3 and 4, isolated incidents appeared where a student questioned the probability generator, such as the time Dayton required help from her group in making sense of how to construct the probability generator for the basketball free throw problem. However, by Session 5, it appeared that the classroom mathematical practice of using a valid probability generator had become taken-as-shared by the students. More specifically, in the process of constructing a probability generator for a particular problem, students discussed the exact assignment of random digits, but no longer required their peers to justify why the probability generator was valid.

The second classroom mathematical practice that emerged was that it was necessary to ensure a trial was valid. Like the previous mathematical practice, this practice began to appear during Session 1. Although students had been given what constituted a valid trial for the "Counting Successes" activity, students discussed why it was appropriate for this problem. Students discussed how one 2-digit number on the random-number table represented buying one can of cola. Thus, the trial was one 2-digit number from the random-number table. The following day, students began to construct simulation problems that also required them to recognize a valid trial. By the end of this session, they had developed the classroom mathematical practice that a valid trial was required for a simulation, and that a trial was valid if it corresponded to the target outcome or outcomes. For example, students simulated randomly answering a five-question true/false test. While most groups determined that a trial was 5 numbers representing the 5 questions on the test, recall that Ondrea's group had run their outcomes together in one long string. Following a whole-class discussion facilitated by the teacher-researcher, Ondrea and her group determined that their outcomes were invalid for this problem and proceeded to correctly identify a valid trial. After this episode, the classroom mathematical practice of ensuring a valid trial for a simulation appeared to be taken-as-shared by the students in that they did not require each other to explain their choice of a trial.

The third classroom mathematical practice identified was that at least 30 trials per group must be conducted for a simulation. In the "Counting Successes" activity, students were asked to "purchase 60 colas." After that first day, students were not generally told how many trials they should do. During the first few sessions, students would ask one another how many trials the other person or group was doing, and the responses varied from as low as 10 to as many as 30. At the same time, student groups decided to "share the work" by each member conducting 10 to 20 trials and combining their results as a group. Thus, by approximately Session 3, the classroom mathematical practice of a group conducting at least 30 trials for an in-class activity had emerged. It appears that this classroom mathematical practice became taken-as-shared without a whole-class discussion, but rather was determined within individual groups.

What Counts as a Valid Probability Generator

Like that of the previous sociomathematical norm, the sociomathematical norm of "what counts as a valid probability generator" appeared during Session 1. As discussed in the previous section, Dayton had discussed with her group that the probability generator consisted of "75 or less you pay $1, greater than 75 you pay nothing." In another group dialog, Ingrid explained to Edison and Tavi how the numbers were assigned for a valid probability generator. In spite of limited discussion about what actually constituted a valid probability simulation, students frequently discussed the need to use a valid probability generator. It seems that the class quickly recognized that there was an unstated norm that you must construct a valid probability generator. Three classroom mathematical practices emerged related to validating a probability generator. One mathematical practice was related to the assignment of random digits; another mathematical practice referred to the acceptance of equivalent probability generators; and a third mathematical practice involved the use of the calculator as a probability generator.

Classroom mathematical practices related to a valid probability generator. One classroom mathematical practice that emerged was that a valid probability generator would reflect a one-to-one correspondence to the probabilities stated in the problem. A probability generator was considered valid if random digits had been assigned to model the outcomes in the problem. This classroom mathematical practice began to emerge during Session 1 when reasoning through the "Counting Successes" activity. As discussed earlier, students helped other group members make sense of the probability generator by appealing to a one-to-one correspondence between random numbers from a random number table to the probability of paying $1 for a cola. As early as Session 2, it became standard practice for the students to immediately construct the probability generator and share this with other group members as if to ensure the accuracy of their generator. If all members did not agree on a particular probability generator, a negotiating process took place until everyone was in agreement. This negotiation relied upon a matching of numbers from a random number table to corresponding numbers that matched the probability in the problem. The classroom mathematical practice did not become taken-as-shared until after the emergence of another classroom mathematical practice, that of acceptance of equivalent probability generators which is discussed below.

The mathematical practice that equivalent probability generators were acceptable emerged during Sessions 3 and 4. At this time, students were still negotiating with each other about what constituted a valid probability generator. On an assigned homework problem students were to assign random digits to simulate 50% democratic and 50% republication. Ondrea shared that she had assigned the digits 01 to 50 democratic and 51 to 100 republican. Breanna responded that she had used 0 to 9, assigning odd digits to represent democratic and even digits to represent republican. After a short discussion, students agreed both methods were not only valid, but their probability generators were also equivalent. After this point, students seemed to have taken-as-shared that equivalent probability generators existed and were acceptable, and at the same time, the classroom mathematical practice that valid probability generators required a one-to-one correspondence to the probabilities in the problem became taken-as-shared.

A third classroom mathematical practice that appeared in relation to a valid probability generator was the use of the graphing calculator as a tool for generating random numbers. In Session 1 of the whole-class teaching experiment, the teacher-researcher demonstrated how to use the random-number generator on the graphing calculator. Students immediately adopted the calculator as their preferred method of generating random digits. Later when spinners were discussed, students were able to discuss how to use them to construct valid probability generators, but given the choice students preferred the graphing calculators for simulation problems done in class. Thus, the classroom mathematical practice of using the calculator to generate trials became taken-as-shared as early as Session 2.

What Counts as a Valid Trial

The sociomathematical norm of "what counts as a valid trial" emerged during the first session. Even though students were given the entire simulation process from assignment of random numbers to what counts as a trial for the situation, students clarified and discussed with each other what a valid trial should look like. This sociomathematical norm was better illustrated in later sessions when it became necessary for students to construct their own trial. During Session 2, Dayton's group clarified what counts as a valid trial. In this situation, the group was creating a simulation for a five-question true/false test.

Bernard: 1 is true, 2 is false.

Cade: We'll do 5 times [for the 5 questions].

Dayton: We're not doing true/false. We're doing right or wrong.

In this excerpt, Dayton was clarifying that the trial did not represent what the answers to the questions were but rather it represented whether the outcomes were right or wrong.

Thus, this group was discussing the sociomathematical norm of what constituted a valid trial.

Classroom mathematical practice related to a valid trial. Students developed the classroom mathematical practice that the number of outcomes in a trial must correspond to the target question. This practice emerged in two different forms. One form of the mathematical practice emerged when students referred specifically to the context of the problem and matched the trial to the required outcomes. For example, during Session 2 Mai explained to Oliver “each trial represents taking a test.” Similarly, Breanna defined a trial to her group as “consists of 10 shots.” In each case, the student was connecting the trial directly to the context of the problem.

The second form of the classroom mathematical practice of matching the outcomes in a trial to the target question emerged when students talked about a trial in “calculator-speak.” Calculator-Speak occurred when students referred to a simulation design as randInt(1, 2, 5), meaning they would generate a string of 5 randomly generated ones and twos to simulate a particular problem. Some students developed this mathematical practice of calculator-speak approximately midway through the instructional sessions. Thus, for some of the students calculator-speak became a taken-as-shared form of the classroom mathematical practice requiring students to match the number of outcomes on the calculator with the n-tuple required for a valid trial. However, even if a student used calculator-speak to refer to a trial, that student was still required to explain the assignment of random digits even though the number of outcomes in the trial was assumed by the student to be understood by other group members. This was probably attributable to the fact that there were equivalent ways to assign numbers for a probability generator so it was necessary for students to clarify how they had assigned integers for their particular probability generator. Consequently, although the classroom mathematical practice of requiring that the number of outcomes in a trial must correspond to the target question emerged, how numbers were assigned for a probability generator could not be taken-as-shared by all students in the class.

What Counts as a Sufficient Number of Trials

The sociomathematical norm of “what counts as a sufficient number of trials” actually appeared after the classroom mathematical practice of combining individual results for a “group” simulation. There were some individual discussions about how many trials should be done. For example, during Session 3 Cade commented to his group that “I think the more sets you use, the more accurate it is.” Generally, students told their group members they would do 10 or 20 trials without justifying their decision. During Session 6, the teacher-researcher asked the class how many trials they should do for a simulation. Students did not agree how many trials were sufficient. Some argued that 30 trials were enough. Others reasoned that enough trials should be conducted to get close to a “theoretical” value even though few students or groups conducted more than 50 trials. Another class discussion took place during Session 9 at which time another classroom mathematical practice emerged, the number of trials was dependent on an “ideal” versus a “practical” situation. A detailed description of the related classroom mathematical practices follows.

Classroom mathematical practices related to number of trials. Two classroom mathematical practices emerged related to this sociomathematical norm. The first classroom mathematical practice was that each person in the group would complete approximately 10 trials and then combine their results. For much of the whole-class teaching experiment, students generally followed the practice that each student would do between 10 and 20 and then combine their results. Thus, the number of trials for a particular simulation was usually around 40 to 80. However, within individual students and different groups, these numbers would vary. Dayton informed her group, during Session 3, that she would complete 4 trials. No one addressed this, although moments later, Cade said "Let's only do 20. This is going to take forever." This practice turned out to be the average number of trials this group would simulate, and in fact, it was the average for all the groups with episodic exceptions. In one case, Oliver completed 134 trials and on another problem he completed 115.

Another classroom mathematical practice that emerged among many of the students was to <u>discuss</u> that many trials were needed; yet students would usually complete less than 100 trials. Early in the teaching experiment, Lacey told her group "I think the more sets you use, the more accurate it is until a certain point. Then it just gets to be to be the same thing." In spite of Lacey's insight, the group still limited the number of trials to around 50 sets for a group total. The classroom mathematical practice was found in other instances as well. When the teacher-researcher asked Thor and his group how many trials they thought they needed, Thor responded "about 4 trillion. It depends. Until it stops changing." The group actually did about 20 trials per person. The teacher-researcher asked the class a similar question about the number of trials. Breanna explained that more trials would be better and provide a more accurate number. However, Breanna's group regularly completed about 20 trials per person. Toward the end of the whole-class teaching experiment, a classroom mathematical practice that emerged and was taken-as-shared by the students was that there was an ideal number of trials and a practical number of trials. An ideal number of trials was considered to be enough for the empirical probability to approach the theoretical probability. A practical number of trials was considered sufficient for class activities and homework. The students resolved this mathematical practice during a whole-class discussion about what counts as a sufficient number of trials, and this practice became taken-as-shared by the students in spite of repeated attempts by the teacher-researcher to facilitate discussions that what students practiced should reflect their beliefs. That is, if they believed that trials should be repeated until the empirical probability becomes more precise or ceases to fluctuate, then this is what they should practice on their class problems and homework.

Summary of Analyses of the Data and Results

The analyses of the quantitative and qualitative data provided evidence to support significant progress by students in reasoning about probability simulation. Student scores on the post- and retention assessments were significantly higher than the scores on the preassessment. Student reasoning on the components of the simulation process following the whole-class teaching experiment generally became more valid and at the same time, their reasoning became more sophisticated and normative.

Technology was found to play an integral part in the instruction of probability simulation. In some ways it supported and encouraged reasoning, and in other ways it discouraged or prevented it. The familiarity and efficiency of the calculator made it the preferred probability-generating device. Furthermore, the syntax required for generating random numbers proved beneficial in focusing students on integral components of the simulation process, and it also provided a common language with which students could communicate about probability simulations. Furthermore, the calculator was found to be a transparent medium for students in dealing with dependent events. Finally, the programming capabilities of the calculator, although a potentially powerful tool, proved to be more frustrating than helpful.

This study also revealed that students hold certain beliefs related the probability simulation. Some of these beliefs were considered helpful to instruction and others were considered more problematic. Furthermore, some of these beliefs changed as the teaching experiment progressed, while others proved resistant to alteration. An analysis of helpful beliefs revealed many beliefs, including the belief that assumptions were an inherent part of the simulation process. Students also exhibited the belief that the probability generator should correspond to the probabilities stated in the problem. Although students believed that the more trials completed, the closer the empirical probability would get to the theoretical probability, students differentiated between what they believed to be an ideal number of trials and a practical number of trials. Moreover, some students exhibited the belief that the context of the problem directly affected the number of trials that one should do. Other helpful beliefs that were revealed included the belief that the empirical probability was more reliable than the theoretical probability and the use of simulation had potential benefits.

Although a number of problematic beliefs were revealed, two of these beliefs turned out to be more resistant to change than others. The representativeness heuristic appeared prior to the whole-class teaching experiment, and in spite of a decreased occurrence on later assessments, there was still evidence of its use on the post-assessment. Finally, another problematic belief identified was the belief by students that a theoretical probability could always be determined. This belief, in particular, was held very strongly by many students.

During the teaching experiment four sociomathematical norms were identified from which emerged a number of classroom mathematical practices. The first sociomathematical norm that appeared was “what counts as a valid simulation.” Connected to this norm were three classroom mathematical practices: (a) a valid probability generator was required, (b) a valid trial is needed, and (c) at least 30 trials per group should be done. The second sociomathematical norm that appeared was related to the justification of a valid probability generator. The practices that emerged within this norm included the following: (a) the probability generator must have a one-to-one correspondence to the probabilities in the problem, (b) equivalent probability generators are acceptable, and (c) the calculator will serve as the mechanism for generating a trial. The concept of a valid trial encapsulated the third sociomathematical norm. From this norm developed the classroom mathematical practice that the number of outcomes in a trial must correspond to the target question. Finally, the fourth sociomathematical norm required the justification of what was a sufficient number of trials. Two practices were identified related to this norm: (a) a group’s simulation results would be the combined outcomes of 10 trials per person, and (b) the number of trials depended on whether the students were dealing with an ideal or a practical situation. Some of the classroom mathematical practices became taken-as-shared by the students, such as the need to ensure a valid trial and the requirement of having a one-to-one correspondence to probabilities given in a problem. Other practices were observed but not necessarily shared by all students. For example, one such mathematical practice was that one must conduct at least 30 trials for a simulation problem.

CHAPTER V

SUMMARY, DISCUSSION, AND CONCLUSIONS

The purpose of this study was to examine the role instruction played in changing students' reasoning and beliefs about probability simulation. At the same time, this study sought to examine what impact technology had on students' reasoning and beliefs. Additionally, it was the goal of this study to investigate the sociomathematical norms and classroom mathematical practices that evolved during a whole-class teaching experiment (Cobb, 2000) that focused on probability simulation. This chapter provides a summary of the study and a discussion of the findings as they are related to the current research base. The limitations of this study, the implications for curriculum and instruction, and the recommendations for future research are also addressed.

Summary of the Study and its Findings

Probability simulation is a critical part in the study of high school level statistics (NCTM, 2000; College Board, 2000). Simulations provide students with the opportunity to look at long run behavior patterns and develop an understanding of probability. When students construct and analyze probability simulations, they are, in a sense, formulating and analyzing mathematical models, and thus developing skills necessary to solve more open-ended type problems (Lesh et al., 1997; Lesh & Clarke, 2000; Lesh & Lehrer, 2000).

What little research is available about probability simulation indicates that students struggle with the two-dimensional nature of probability problems (Benson, 2000; Benson & Jones, 1999; Zimmermann & Jones, 2002). Furthermore, students carry a myriad of beliefs, some helpful and some problematic, that affect how they reason about probability simulation (Zimmermann & Jones, 2002). Finally, it is has been suggested that appropriate use of technology can help students develop a better conceptual understanding of simulations (Balacheff & Kaput, 1996; Ben-Zvi ; Biehler, 1991; Konold, 1991b).

Despite NCTM's (2000) recommendations for increased focus on statistical thinking and the fact that more students are enrolling in the Advanced Placement Statistics courses (Straf, 2002), there is little research that provides teachers with instructional guidance. This study sought to bridge the gap between teaching practice and research as it related to probability simulation. Moreover, it was the intent of this study to help fill the void in research regarding students' individual and collective reasoning and beliefs about probability simulation.

Purpose of the Study

Using a whole-class teaching experiment (Cobb, 2000) as the setting for the research, this study investigated the development of students' reasoning and beliefs as they worked through the components of probability simulation. In particular, this study addressed the following research questions:

1. How does high school students' individual and collective reasoning about probability simulation change during a whole-class teaching experiment? What role does technology play in this change?
2. How do high school students' beliefs about probability simulation change during a whole-class teaching experiment?

3. What kind of sociomathematical norms and classroom mathematical practices evolve during a whole-class teaching experiment that focuses on probability simulation?

Methodology

A class of 23 high school students enrolled in an Advanced Placement Statistics course participated in a 12-day whole-class teaching experiment. Twenty-one of the students were administered a pre-, post-, and retention assessment to provide quantitative data that could be used to examine changes in student reasoning. These assessments along with the audio and videotapes of the instructional episodes provided qualitative data related to students' reasoning and their beliefs about probability simulation. The whole-class teaching experiment was grounded in Cobb's (1999) developmental research cycle, consisting of an instructional development phase and a classroom-based analysis phase. During the instructional development phase, a review of content-specific research was conducted which in turn informed the development and modification of the hypothetical learning trajectory (Simon, 1995) and guided the selection of classroom tasks and activities. The tasks and activities were implemented during the classroom-based analysis phase where data were gathered about students' reasoning and beliefs. These data were used to modify and refine the hypothetical learning trajectory and to trace changes in students' individual and collective reasoning as well as their beliefs.

Results of the Study

Quantitative and qualitative analyses provided numerous results related to students' reasoning and beliefs about probability simulation. The analyses also revealed the impact of technology on students' reasoning as well as sociomathematical norms and classroom mathematical practices that emerged during the whole-class teaching experiment.

Students' Individual Reasoning about Probability Simulation

A Wilks's Lambda multivariate test revealed that there was a significant difference ($p < .001$) between the mean scores on the pre-, post-, and retention assessments. Furthermore, a pairwise comparison indicated that the post-assessment scores were significantly higher than the preassessment scores ($p < .001$), and similarly, the retention assessments scores were significantly higher than the preassessment scores, ($p < .001$). However, there was no significant difference between post- and retention scores even though mean scores were slightly higher on the retention assessment. Growth in students' ability to reason about probability simulation was further substantiated by an analysis of student reasoning across six different components as they both evaluated and constructed a probability simulation: assumptions, probability generator, two-dimensional trial, accept random outcomes, calculate empirical probability, and repetition of trial. In 5 of the 6 major simulation components, the frequency of valid responses increased from the preassessment to the post-assessment. More specifically, students made significant progress in their ability to use simulated outcomes to determine the probability of an event and to recognize the effect of repeated trials on the empirical probability. The key patterns of students' reasoning on each of the six different components of simulation will be described in turn.

Not only did more students explicitly state assumptions after instruction, such assumptions became more normative. These normative assumptions related mainly to maintaining a valid probability generator by replacing a drawn chip and to the independence of events.

Students became more versatile in their ability to recognize and construct a valid probability generator. In spite of initially demonstrating some difficulty in constructing a valid two-dimensional probability generator, student reasoning progressed to a point where students were able to construct valid probability generators for multi-dimensional trials and trials involving dependent events. During instruction students tended to rely on their graphing calculator to generate random numbers. However, they showed versatility in using other devices, such as spinners, balls, and chips, in the post- and retention assessment problems. What invalid reasoning occurred resulted from students' inability to use proportions correctly and to recognize a valid sample space, especially in the case of two-dimensional trials.

Over 85% of the students accepted the randomness of simulated outcomes. For those students who had difficulty with randomness, the representative heuristic was a barrier to accepting "unexpected" results. Interestingly, representativeness only occurred on written assessments and not during class instruction.

Because students overcame their difficulties in calculating two-dimensional trials, all but one student were able to calculate an empirical probability. The students' ability to calculate the probability was connected to their ability to define a valid two-dimensional trial and then use their definition to calculate the empirical probability.

With respect to repetition of trials, students showed increased awareness that as the number of trials increased the empirical probability approached the theoretical value.

While they did not always want to simulate a substantial number of trials, they were aware that the empirical probability showed less fluctuation and more stability as the number of trials grew.

Impact of Technology on Students' Reasoning

The calculator was found to have impacted students' individual and collective reasoning in five different ways. First, students showed a preference for the calculator as a probability-generating device, especially during instruction. Next, the syntax of the calculator seemed to help students to focus on the components of the simulation process. It also provided students with a transparent medium for dealing with more complex problems that involved dependent rather than independent events. The exclusive use of the graphing calculator provided a common language for students to discuss probability simulation. The fifth impact of technology contrasted with the others as it was not beneficial in helping students reason about simulation. In spite of the powerful programming capabilities of the calculator, it was found that the difficulty of learning how to program the calculator outweighed its potential benefits.

Students' Beliefs About Probability Simulation

Students revealed a number of beliefs considered either helpful or problematic as related to instruction. Many of the helpful beliefs were directly related to the simulation components; such as assumptions are necessary in carrying out a simulation, and the probability generator should correspond to the probabilities in the problem. Other helpful beliefs appeared that were related to simulation trials. Namely, students held the belief that although many trials provide a more precise empirical probability, they also believed there exists a difference between an ideal number of trials and a practical number of trials. Even more specifically, some students believed that the context of the problem has an influence on the number of trials. Three problematic beliefs were discerned: (a) representative heuristic, (b) outcome approach, and (c) the theoretical probability can always be determined. The first two problematic beliefs focused students' attention on irrelevant aspects of the simulation while the third belief tended to devalue simulation as a viable strategy. The three beliefs are considered problematic because they rival not only a normative view of probability, but more specifically, of probability simulation.

Students' Collective Reasoning About Probability Simulation

In examining students' collective reasoning, four sociomathematical norms emerged. Each of these sociomathematical norms spawned one or more classroom mathematical practices. The following classroom mathematical practices were related to the sociomathematical norm "what counts as a valid simulation": (a) you must use a valid probability generator, (b) you need to ensure the trial corresponds to the sample space of the contextual situation, and (c) you must conduct at least 30 trials per group. Three practices emerged within the sociomathematical norm of "what counts as a valid probability generator": (a) you must have 1-1 correspondence to probabilities given in problem, (b) equivalent probability generators are acceptable, and (c) the calculator will be used to generate trials. The third norm "what counts as a valid trial" revealed just one mathematical practice: the number of outcomes in a trial must correspond to target question. The last sociomathematical norm "what counts as a sufficient number of trials" produced three classroom mathematical practices: (a) combine results of 10 trials per person in a group, and (b) recognize the limitations on the number of trials created by time constraints.

Discussion of the Findings

The first research focus of this study was to examine students' reasoning as they worked through problems involving probability simulation. This focus also included an investigation into the role and impact of technology. A second research focus was to identify and track students' beliefs as they reasoned about simulation. Finally, this research set out to identify sociomathematical norms and classroom mathematical practices that emerged during a whole-class teaching experiment that focused on probability simulation.

Students' Reasoning About Probability Simulation

Across all six simulation components, students made significant progress in their reasoning throughout the whole-class teaching experiment. The following provides an interpretation of how these changes occurred as well as how the findings of this study relate to current literature.

Assumptions. From the beginning of the whole-class teaching experiment, students noted implicit assumptions in a simulation design. What is significant is how students in this study moved from a more idiosyncratic way of reasoning about assumptions to one that was more normative. In particular, these normative assumptions related primarily to maintaining a valid probability generator, and in the case of a small number of students, to assumptions relating to independent events. Yates et al. (1999) note that independence is often a critical assumption. It is possible that the number of students assuming independence may actually have been even higher than indicated by the post- and retention assessments since the syntax of the calculator assumed independence of events, and thus students may have reasoned they did not need to make this explicit. Students' limited attention to assumptions may also be attributable to a limited instructional focus on assumptions by the teacher-researcher. Very little instructional time was spent exploring or discussing the role of assumptions in the simulation process, and it was not until Session 4 of a 12-session teaching experiment that the teacher-researcher focused students' attention to assumptions.

Probability generator. In spite of almost exclusive use of the graphing calculator during instruction, students displayed tremendous versatility in their choice and explanation of valid probability generators on the post- and retention assessments. One possible explanation for students' increased ability to construct valid multi-dimensional trials is the use of calculator technology. The syntax required by the calculator focused students' thinking on assigning appropriate valid random numbers and recognizing the number of outcomes needed for a trial. With that said, the calculator may have worked against students who struggled with proportionality. According to Borovcnik and Peard (1996), some students may benefit from a more kinesthetic approach that utilizes manipulative devices, such as spinners and chips before relying on the calculator. Such hands-on devices may help students develop more concrete cognitive models before progressing to more abstract simulation models generated by technology.

The fact that students did not generally struggle with constructing a valid probability generator confirms Benson and Jones' (1999) research that suggests constructing probability generators is not generally problematic for students, at least in the case of one-dimensional trials. Analysis of the preassessment found that an incomplete understanding of proportionality contributed to the difficulty some students had in constructing a valid probability generator. This finding is consistent with earlier research (Garfield & Ahlgren, 1988; Green, 1983) that suggests the mathematical concept of proportionality is a hindrance to students' ability to reason probabilistically.

Two-dimensional trials. After instruction, students appeared to have overcome their difficulties in defining a two-dimensional trial. Furthermore, students were able to construct valid trials of complex, multi-dimensional problems during instruction. The use of the troublesome representative heuristic that appeared on the preassessment did not appear on later assessments or during instruction. Because a whole-class teaching experiment focuses more on collective reasoning, changes in individual reasoning with respect to representativeness were difficult to trace. Thus, it is unclear how changes may have occurred in the reasoning of individual students to modify misconceptions, like representativeness. Of the approximately one third who had transformed a two-dimensional trial into a one-dimensional trial on the preassessment, almost all of these students were able to define valid two-dimensional trials after the teaching experiment. The numerous simulations done in class may have provided the necessary opportunities to reason about multi-dimensional trials. In addition, the use of the graphing calculator may have played a role in helping students develop valid reasoning about trials since students needed to use the correct syntax for the calculator to generate appropriate random numbers.

Benson & Jones (1999) reported that subjects were able to construct a valid one-dimensional trial; however, two-dimensional trials proved difficult for even the college students in their sample. The results of the preassessment bear this out, and students who were unable to recognize or construct a valid trial typically transformed the problem to a one-dimensional trial supporting the findings of Zimmermann and Jones (2002).

However, instruction and technology seem to have played a role in helping students develop valid reasoning about multi-dimensional trials because 90% were able to construct valid trials for two-dimensional problems during the retention assessment.

Randomness and empirical probability. Whether a student calculated an empirical probability was often related to the student's willingness to accept the random outcomes of a simulation. Of the students who did not accept the randomness of the outcomes, all 10 did not calculate a probability. Thus, it is not surprising that as students' acceptance of the random outcomes increased, so did students' ability to calculate a valid probability.

It turns out that student misconceptions for the two simulation components are the same. That is, they are both related to representativeness. Although evidence of the representativeness heuristic was found on each of the written assessments, albeit in decreasing amounts, it was never identified during class instruction. A number of issues may be connected to the apparent resolution of the representative heuristic. A combination of instructional strategies, including having students simulate various problems, compare the results obtained by different groups, and having the whole-class discuss randomness, may have helped students develop an understanding of randomness in simulated problems. On the one hand, integrity of the belief in the random integer feature of the graphing calculator may be partially responsible as students relied solely on it for generating random numbers. On the other hand, the group structure of the class may have contributed to the apparent lack of the representative heuristic during instructional sessions. That is, students who reasoned using representativeness may not have been mathematically confident, and, therefore, may have contributed less during group dialogs.

The fact that representativeness only appeared on the preassessment evaluation task and not on any of the construction tasks may also help to explain the inability to identify use of the representativeness heuristic during instruction. Representativeness was found when students were provided with outcomes they believed did not represent the given problem. During instruction, students always generated their own outcomes. Thus, the students may have placed a greater faith in student-generated outcomes or "unexpected" imbalances across outcomes did not occur; thereby limiting the opportunity for them to fall victim to representativeness. Notwithstanding these hypotheses, representativeness did not appear on the retention assessment even when students faced the previously troublesome evaluation task. Hence, there is evidence that most students seemed to have overcome representativeness even across the time gap between post-assessment and retention assessment. If students had resolved the misconception of representativeness, it is unclear how student reasoning changed.

The findings of this study support the conclusions of Batanero & Serrano (1999) that students struggle with the concept of randomness. The findings also mirror those of Zimmermann and Jones (2002) who found similar evidence to suggest that some students struggle with the concept of randomness, and the difficulties are related to reasoning based on representativeness (Kahneman & Tversky, 1972). Although Batanero and Serrano concluded that instruction failed to help students make sense of randomness, there is more evidence in this study, vis-à-vis the retention assessment, that students did overcome problems with representativeness and as a consequence gained a better appreciation of randomness.

<u>Repetition of trial.</u> Students became more aware that as the number of trials increased the empirical probability would approach a theoretical value. In particular, students noted that the empirical probability would show less fluctuation and more stability as the number of trials grew. Activities had been designed to focus students' thinking on tracing the empirical probability as the number of trials was increased. A dual instructional approach of having students repeat numerous trials as well as analyze a graphical display to track the change in the empirical probability as the number of trials was increased helped students to develop an understanding of the long-term effects to the empirical probability. The effects of this dual approach were evident as students reasoned during their post- and retention assessments and during class discussions that focused on how many trials should be conducted. Students often made comments such as "until it stops fluctuating" or "until it approaches a theoretical value." They recognized that with enough trials, the probability would fluctuate less and settle towards a theoretical value. Aspinwall and Tarr (2001) found that a student's understanding of experimental probability was related to his or her understanding of the law of large numbers. Thus, the increase in the ability of students in this study to calculate the empirical probability may have helped students to understand the concept that the empirical probability will eventually settle towards some theoretical value.

<u>The Role of Technology</u>

Technology was found to play an integral part in students' reasoning about probability simulation. It was probably most influential in focusing students' thinking on constructing a probability generator and in defining a two-dimensional trial. The calculator also helped students to understand the effect of many trials on the empirical probability. The syntax for the graphing calculator required that the students understand the assignment of random digits in order to be able to define a valid trial. The many simulation problems students worked helped them to become proficient at using the calculator for simulations, and therefore they also became more adept at assigning random digits and defining valid trials.

Another positive impact of the graphing calculator was that it enabled students to complete many trials quickly, which in turn led to a deeper understanding of the long-run effects to empirical probability. As Biehler (1991) had suggested, technology proved to be beneficial in helping students explore how empirical probability changed over numerous trials. The positive impact of technology on students' reasoning in probability simulation supports Biehler's (1991) and Konold's (1991b) claims that technology can be used to both enhance learning and to develop a richer understanding of simulation.

Students' Beliefs about Probability Simulation

Students' beliefs about assumptions became more helpful during the teaching experiment, and they focused specifically on beliefs about maintaining the probability generator. It is likely that as students developed the classroom mathematical practice of ensuring that a valid probability generator had a one-to-one correspondence with the problem, their beliefs about assumptions changed. Their belief in a one-to-one correspondence appeared to be intuitive and by the post- and retention assessments they were explicitly stating their belief that a drawn chip must be replaced to guarantee that correspondence was maintained.

Even though students would only do 30 or 40 trials during a classroom simulation, 81% developed the belief that more than 100 trials were desirable. This was tied to their belief that with enough trials the empirical probability would eventually approach the theoretical probability. Through the course of selected activities and problems, students were able to see how the empirical probability changed as more trials were added. The graphical demonstration of how the empirical probability changed as the number of trials increased seemed to have a major impact on students' beliefs as evident in their responses both during instruction and on written assessments. Students frequently would explain that additional trials would make the empirical probability more precise or the empirical probability would fluctuate less and become more stable. The repeated experiences of examining these changes appeared to have a major influence on students' beliefs about the long-run behavior of the empirical probability.

Most of the students' problematic beliefs were related to misconceptions about probability. Representativeness and the outcome approach were two types of misconceptions uncovered in the qualitative analysis of this study. Although the representativeness heuristic (Kahneman & Tversky, 1972) appeared throughout all three written assessments, the frequency with which it was found decreased substantially. In the case of the outcome approach (Konold et al., 1993) there was only one recorded instance and this appeared prior to instruction. As mentioned earlier, the methodology of this study limited the teacher-researcher's ability to trace how changes occurred in individual thinking. However, it can be conjectured that technology may have played a part in changing students' beliefs, especially those related to representativeness. Because of the opportunities students had to compare and discuss simulation outcomes with group peers and classmates, they were able to see a variety of sampling distributions. Furthermore, students' increased knowledge of empirical probability and repeated trials may have also contributed to a more complete understanding of the random nature of probability simulation.

A number of beliefs were identified in this study, and in accord with the literature it seems that students' beliefs were a major contributor to students' reasoning about probability simulation (Fischbein & Gazit, 1984; Fischbein et al., 1991; Fischbein & Schnarch, 1997; Garfield & Ahlgren, 1988; Piaget & Inhelder, 1975; Shaughnessy, 1992). Some of the beliefs identified in this study were also found in earlier studies, specifically those related to assumptions, the probability generator, and long-run effect of the empirical probability (Zimmermann & Jones, 2002). The results of this study also support Zimmermann & Jones' (2002) conclusion that students exhibited the problematic belief that simulated outcomes should always represent the probabilities in the problem.

Sociomathematical Norms and Classroom Practices

During the whole-class teaching experiment, four sociomathematical norms arose from the students' discourse: what counts as a valid simulation, what counts as a valid probability generator, what counts as a valid trial, and what counts as a sufficient number of trials. Each of these four norms led to the development of a number of classroom mathematical practices. Taken together, the sociomathematical norms and classroom mathematical practices provide insight into the collective reasoning of the students in a classroom environment (Cobb, 2000).

What counts as a valid simulation. During the first session, students were grappling with the notion of what was required for a valid simulation. The group structure of the classroom lent itself to the development of classroom mathematical practices as a social outgrowth of interactions about individual students' thinking.

Without being provided with the simulation steps outlined by Yates et al. (1999), it became necessary for the students to communally develop classroom mathematical practices that essentially replicated these steps so that they would have a common process to follow and discuss with other class members. Thus, it was necessary that the classroom mathematical practices associated with the sociomathematical norm of justifying a valid simulation become taken-as-shared so that students could discuss and make sense of each other's simulations. The ease with which students developed a multi-step process for designing a simulation suggests that the mechanics of simulation are within easy grasp for many students in the upper grades of high school.

What counts as a valid probability generator. Because a valid probability generator is a critical component of the simulation process, classroom mathematical practices quickly emerged from the sociomathematical norm of justifying a valid probability generator. Students immediately and rather easily adopted the practice of ensuring a one-to-one correspondence between the probability generator and the context of the problem. The ease with which students developed such a correspondence is not surprising as similar results were found in earlier studies (Benson, 2000; Benson & Jones, 1999; Zimmermann & Jones, 2002), some of which involved much younger children. Another practice emerged when students discussed the equivalence of generators. Some of the problems presented to students were intended to promote multiple ways of assigning random digits for a valid probability generator. Students compared generators, and then through a negotiating process, they determined that equivalent probability generators were acceptable. The last classroom mathematical practice of using the graphing calculator to act as the probability generator, like the previous two practices, became taken-as-shared early in the teaching experiment. Students' ready access to a calculator and their familiarity with using it most likely helped to encourage the development of this technology-oriented mathematical practice. Biehler (1991) and Konold (1991b) advocate the use of technology as a learning tool suggesting that it can be used to construct simulation models and thus help students to develop a better understanding of simulation. The findings of this study give merit to Biehler's and Konold's suggestion given that the calculator's use as a probability generator seemed to be beneficial in helping students make sense of the simulation process, in particular the modeling of contextual situation.

What counts as a valid trial. Very early during student discourse, the sociomathematical norm concerning a valid trial emerged. This was not surprising because preassessment test results corroborated research that students struggled with two-dimensional probability problems (Benson, 2000; Benson & Jones, 1999; Zimmermann & Jones, 2002). Hence, it was necessary for students to grapple early with the notion of what constituted a valid two-dimensional trial. The choice of activities, group work, and whole class discussions seemed to facilitate a more complete understanding of what constituted a valid trial. In addition, the increasing difficulty of the multi-dimensional trials seemed to provide students with contextual and mathematical skills that they could discuss and build upon. The choice of activities also seemed to allow for the emergence of the classroom mathematical practice of requiring that the number of outcomes in a trial must correspond to the number of outcomes in the target question. While the calculator seemed to help facilitate student understanding of what was required for a valid trial, its use also precipitated the emergence of calculator-speak. As students relied upon communicating about simulation via calculator speak, it became necessary that the mathematical practice of requiring the number of outcomes in a trial correspond to the number of outcome in the target question become taken-as-shared so that students could communicate with one another

What counts as a sufficient number of trials. Although it took until the end of the teaching experiment for the classroom mathematical practice of recognizing the time constraints of conducting an ideal number of trials to appear, the practice of completing about 40 trials per group emerged on the second day. There was little discussion by students to discern how they determined that an average of 30 to 40 trials was sufficient for a problem. However, the experiences of comparing simulation results combined with classroom discussions about the long-term consequences of repeated trials seemed to influence the emergence of the practice related to an ideal number of trials. Even though students would agree that more trials would provide a better answer, they limited themselves to a “practical” number of trials for what they termed “classroom purposes” in spite of the fact that they believed that an “ideal” number of trials would provide better results.

In retrospect, other instructional decisions could have been made to help students see that 40 trials are not sufficient for determining a more precise empirical probability (K. Berk, personal communication, June 29, 2002). Specifically, 40 trials do not provide a picture of the variability inherent in so few trials. One instructional suggestion is to relate the number of trials to opinion polls. It is likely that students would readily agree that polls of only 40 people would not provide a very accurate picture of people’s opinions. Another suggestion would be

to create a graphical representation of the simulation outcomes generated by each student so as to view the variability of these outcomes and see how this distribution changes as more trials are conducted. Finally, students could be asked to explore the standard deviation of simulation problems and how it changes as the number of trials is increased. These instructional suggestions may help students to overcome the classroom mathematical practice that 40 trials are sufficient for a simulation problem.

Limitations of the Study

Three limitations related to the findings of this study have been identified. The limitations concern the level of reasoning on probability simulation before instruction, data on students' individual reasoning, and the dual role of teacher –researcher.

One limitation of this study was the large number of students who possessed a higher level of reasoning about probability simulation prior to the whole-class teaching experiment. This was higher than would have been predicted from studies by Zimmermann and Jones (2002) and Benson and Jones (1999). Although this class contained students with varied mathematical backgrounds and abilities, there was still a relatively large group of students that demonstrated a high level of understanding of probability simulation as revealed on the preassessment. Hence, there may be limitations as to the extent to which this class represents high school students' thinking on probability simulation.

A second limitation was the restricted amount of data available on how individual reasoning and beliefs changed. As a result of the methodology of the whole-class teaching experiment, changes in individual reasoning were difficult to trace. For example, the representative heuristic was used by a number of students prior to the teaching experiment, but no evidence of this type of reasoning was found during instruction. Additionally, the combination of a focus on collective reasoning and the exclusive use of the calculator by students as a probability-generator device, may have hidden individual misconceptions or fragile beliefs related to probability simulation.

A third limitation of this study was the dual role of the teacher and researcher. As Ball (2000) recognizes, there are distinct advantages to being both the teacher and the researcher. As the teacher, one can draw from her own personal knowledge and experience as a teacher. A teacher also possesses first-hand knowledge of how her students interact with each other and the teacher, as well as insight into the knowledge that students bring to the classroom. Additionally, the teacher and students have a shared history. Yet this intimate knowledge of how students reason and interact can also act as a limitation in this type of study. The complex nature inherent in the dual role of teacher and researcher can make it difficult to provide the objectivity required for analysis. For example, the teacher must make instructional decisions for the benefit of her students that may not be beneficial in studying change in students' reasoning. Therefore, this dual responsibility can be both advantageous as well as a limitation.

Implications for Curriculum and Instruction

The findings of this study can help curriculum leaders and mathematics teachers by providing valuable information on how students reason about probability simulation. The implications for curriculum contain recommendations for changes to curriculum to help facilitate development of student reasoning about probability simulation. The implications for instruction are recommendations targeted specifically to classroom teachers.

Implications for Curriculum

Probability simulation is a requisite of the Advanced Placement Statistics curriculum (College Board, 2000), and NCTM (2000) recommends its inclusion in any high school curriculum containing probability and statistics. Therefore, the results of this study are particularly useful to curriculum planners and mathematics teachers.

As indicated by the findings of this study, students were able to overcome most of the difficulties that prevented valid reasoning about probability simulation. In the process, students developed a stronger understanding of random behavior, empirical probability, and the connections between empirical probability and theoretical probability. Thus, the implementation of probability simulation in the high school curriculum is not only appropriate, but also encouraged. Probability simulation could also be incorporated at the middle school limiting the complexity of simulations students are asked to do. For example, problems would focus primarily on one-dimensional trials. The topic of probability simulation integrated early in the curriculum may help some students develop a better conceptual understanding of probability and random behavior, which are considered critical concepts for all students according to NCTM (2000).

Implications for Instruction

As a teacher and researcher, it was important that this research be able to provide classroom teachers with recommendations that could be implemented during the course of instruction. The findings of this study support the following recommendations for instruction.

Given the impact technology had on helping students develop a richer understanding of probability simulation and its components, it is essential that technology play a major part in simulation instruction. Technology had a positive influence on how students reasoned about randomness, trials, and empirical probability, all of which are critical components of simulation. The technology used in this study was limited to the graphing calculator. However, computers may prove to be even more beneficial considering their speed and computing power, and hence, their ability to run a larger number of trials quickly. Software, such as Fathom, that is specifically designed to focus on statistical processes, graphical displays, and various data distributions could be used to look at larger and more complex probability simulations.

Analyses revealed that students in this study developed a higher level of understanding of probability simulation that involved multi-dimensional trials and both independent and dependent events. Thus, activities should be created that capitalize on these strengths and extend the reasoning of students to even more challenging and complex probability simulation problems. In particular, teachers should present problems that require more thought provoking assumptions, more demanding insights when defining a trial, and decision-making related to important concepts, like independence, dependence, and conditional probability.

The more normative view of assumptions students developed during the teaching experiment was encouraging but also limiting. Students failed to recognize that other implicit assumptions may exist in a simulation, such as there are only two kinds of pizza or no commercials will be on the radio station, and limited themselves to independence and maintaining a valid probability generator. While these assumptions are important, they are rarely the only implicit assumptions. Thus, instruction should include a larger and more explicit focus on assumptions. Students should develop the understanding that inference is based on both explicit and implicit assumptions. Likewise, given the resistance students had in actually doing many trials, instruction should incorporate activities or problems that help students to understand the variability in a simulation of only 40 or 50 trials. Looking back, there are a number of activities and whole-class discussions that could be incorporated that specifically target the concept and role of assumptions and the practice of doing many trials.

Recommendations for Future Research

This study utilized a whole-class teaching experiment methodology. Therefore, it was difficult to trace changes in individual reasoning. Future research needs to examine how individual reasoning and beliefs about probability simulation change during instruction. The research also needs to identify what catalysts promote or impede such change. Lesh and Kelly (2000) describe a three-tiered approach to teaching experiments. Tier 3 is the researcher level where researchers construct models so as to analyze and interpret the reasoning and actions of students and teachers in a learning environment. The research literature for this study exemplifies what has been accomplished at Tier 3. Tier 2 is the teacher level. At this level the teacher develop ways to collect data about what and how students learn in an effort to influence student learning. Lastly, Tier 1 is the student level. The intent of this tier is to focus on how the individual students develop ideas or models. Such a model would be helpful in tracing how individual students resolve misconceptions, like representativeness. Similarly, research is also needed to explore how students' concept of proportionality changes as they reason about probability simulation. In particular, more Tier 1 level research is needed that looks at how students build concepts of theoretical and empirical probability and how they make connections between these two concepts.

A major focus of this study was to investigate the students' individual and collective reasoning and beliefs about probability simulation. Although students made tremendous strides in reasoning, the findings of this study provide a limited view of the role of the teacher and how her reasoning and beliefs influenced instructional decisions about teaching probability simulation. Additionally, how did modifications to the hypothetical learning trajectory change the way the teacher would have otherwise approached instruction, and how do the hypothetical learning trajectories of the teacher lead to breakthroughs in learning by the students? Research at the Tier 2 level is needed to study the cognitive processes and pedagogical practices employed by teachers that support and facilitate learning about probability simulation.

This study demonstrated that probability simulation is accessible to high school students and that instruction can be used to help students develop a richer understanding of simulation and its related mathematical concepts, randomness, empirical probability, and probability. Furthermore, a simple graphing calculator can be a powerful technological tool that can be used to enhance concepts like constructing probability generators and trials, and act as a tool to provide a substantial number of trials. Finally, by its very nature, probability simulation encourages students to share and debate key concepts that promoted understanding.

REFERENCES

Aspinwall, L., & Tarr, J. E. (2001). Middle school students' understanding of the role sample size plays in experimental probability. Journal of Mathematical Behavior, 20, 229-245.

Balacheff, N., & Kaput, J. J. (1996). Computer-based learning environments in mathematics. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick & C. Laborde (Eds.), International Handbook of Mathematics Education (pp. 469-501). Dordrecht, The Netherlands: Kluwer.

Ball, D. (2000). Working on the inside: Using one's own practice as a site for studying teaching and learning. In A. Kelly & R. Lesh (Eds.), Handbook of research design in mathematics and science education (pp. 365-402). Mahwah, NJ: Lawrence Erlbaum Associates.

Batanero, C., & Serrano, L. (1999). The meaning of randomness for secondary school mathematics. Journal for Research in Mathematics Education, 30, 558-567.

Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. Mathematical Thinking and Learning 2, 127-155.

Benson, C. T., & Jones, G. A. (1999). Assessing students' thinking in modeling probability contexts. The Mathematics Educator, 4(2), 1-21.

Benson, C. T. (2000). Assessing students' thinking in modeling probability contexts. Unpublished doctoral dissertation, Illinois State University, Normal.

Biehler, R. (1991). Computers in probability education. In R. Kapadia & M. Borovcnik (Eds.), Chance Encounters: Probability in Education (pp. 169-212). Dordrecht: Kluwer.

Biehler, R. (1993). Software tools and mathematics education. In C. Keitel & K. Ruthven (Eds.), Learning Through Computers: Mathematics and Educational Technology (pp. 68-100). Berlin: Springer Verlag.

Biehler, R. (1994). Probabilistic thinking, statistical reasoning and the search for causes: Do we need a probabilistic revolution after we have taught data analysis? In J. Garfield (Ed.), Proceedings of the Fourth International Conference on Teaching Statistics (pp. 34-52). Minneapolis, MN: The International Study Group for Research on Learning Probability and Statistics.

Blumer, H. (1969). Symbolic interactionism: Perspective and method. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Borovcnik, M., & Peard, R. (1996). Probability. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick & C. Laborde (Eds.), International Handbook of Mathematics Education (pp. 239-287). Dordrecht: Kluwer.

Carpenter, T. P., & Fennema, E. (1988). Research and cognitively guided instruction. In E. Fennema, T. P. Carpenter & S. J. Lamon (Eds.), Integrating research on teaching and learning mathematics (pp. 2-17). Madison, WI: University of Wisconsin, Wisconsin Center for Education Research.

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. American Educational Research Journal, 26(4), 499-531.

Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. Mathematical Thinking and Learning, 1, 5-43.

Cobb, P. (2000). Conducting teaching experiments in collaboration with teachers. In A. Kelly & R. Lesh (Eds.), Handbook of research design in mathematics and science education (pp. 307-335). Mahwah, NJ: Lawrence Erlbaum Associates.

Cobb, P. & Bauersfield, H. (Eds.) (1995). The emergence of mathematical meaning: Interaction in classroom cultures. Hillsdale, NJ: Lawrence Earlbaum Associates.

Cobb, P., & Whitenack, J. W. (1996). A method for conducting longitudinal analyses of classroom videorecordings and transcripts. Educational Studies in Mathematics, 30, 213-228.

College Board. (2000). A Guide to the Advanced Placement Program. Princeton, NJ: AP Order Services.

delMas, R. C., & Bart, W. M. (1989). The role of an evaluation exercise in the resolution of misconceptions of probability. Focus on Learning Problems in Mathematics, 11, 39-53.

Doerr, H. M. (1998). A modeling approach to non-routine problem situations. In S. Berenson & K. Dawkins & M. Blanton & W. Coulombe & J. Kolb & K. Norwood & L. V. Stiff (Eds.), Proceedings of the nineteenth annual meeting, North American Chapter of International Group for the Psychology of Mathematics Education (Vol. 1, pp. 441-446). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Doerr, H. M., & Tripp, J. S. (1999). Understand how students develop mathematical models. Mathematical Thinking and Learning, 1, 231-254.

Dörfler, W. (1993). Computer use and views of the mind. In C. Keitel & K. Ruthven (Eds.), Learning from Computers: Mathematics Education and Technology (pp. 159-186). Berlin, Germany: Springer.

Eisenhart, M. A. (1988). The ethnographic research tradition and mathematics education research. Journal for Research in Mathematics Education, 19, 99-114.

English, L. (1993). Children's strategies for solving two- and three-dimensional combinatorial problems. Journal for Research in Mathematics Education, 24, 255-273.

English, L. D., Jones, G. A., Lesh, R., Bussi, M. B., & Tirosh, D. (2002). Future issues and directions in international mathematics education research. In L. English (Ed.), Handbook of international research in mathematics education (pp. 787-812). Mahwah, NJ: Lawrence Erlbaum Associates.

Falk, R. (1983). Children's choice behavior in probabilistic situations. In D. R. Grey & P. Holmes & V. Barnett & G. M. Constable (Eds.), Proceedings of the First International Conference on Teaching Statistics (pp. 714-726). Sheffield, UK: Teaching Statistics Trust.

Fast, G. R. (1999). Analogies and reconstruction of probability knowledge. School Science and Mathematics, 99(5), 230-240.

Fennema, E., Franke, M. L., Carpenter, T. P., & Carey, D. A. (1993). Using students' mathematical knowledge in instruction. American Educational Research Journal, 30, 555-583.

Fischbein, E., & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? Educational Studies in Mathematics, 15, 1-24.

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. Journal for Research in Mathematics Education, 28, 96-105.

Fischbein, E., Nello, M. S., & Marino, M. S. (1991). Factors affecting probabilistic judgments in children and adolescents. Educational Studies in Mathematics, 22, 523-549.

Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. Journal for Research in Mathematics Education, 19, 44-63.

Gnanadesikan, M., Scheaffer, R. L., & Swift, J. (1987). The art and techniques of simulation. Palo Alto, CA: Dale Seymour Publications.

Green, D. R. (1983). A survey of probability concepts in 3,000 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, UK: Teaching Statistics Trust.

Hawkins, A. S., & Kapadia, R. (1984). Children's conceptions of probability -- a psychological and pedagogical review. Educational Studies in Mathematics, 15, 349-377.

Hogg, R. V., & Tanis, E. A. (1997). Probability and statistical inference (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Jaworski, B. (1997). The centrality of the researcher: Rigor in a constructivist inquiry into mathematics teaching. In A. R. Teppo (Ed.), Qualitative research methods in mathematics education. Journal for Research in Mathematics Education Monograph Series, Number 9 (pp. 112-127). Reston, VA: Narional Council of Teachers of Mathematics.

Jiang, Z., & Potter, W. D. (1994). A computer microworld to introduce students to probability. Journal of Computers in Mathematics and Science Teaching, 13(2), 197-222.

Jones, G. A., Langrall, C. W., Thornton, C. A., & 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., Thornton, C. A., Langrall, C. W., & Tarr, J. E. (1999). Understanding students' probabilistic reasoning. In L. V. Stiff & F. R. Curcio (Eds.), Developing Mathematical Reasoning in Grades K-12: 1999 Yearbook (pp. 146-155). Reston, VA: National Council of Teachers of Mathematics.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. Cognitive Psychology, 3, 430-454.

Kirk, R. E. (1982). Experimental Design (2nd ed.). Pacific Grove, CA: Brooks/Cole Publishing Co.

Kolata, G. (1997). Understanding the news. In L. A. Steen (Ed.), Why numbers count: quantitative literacy for tomorrow's America (pp. 23-29). New York: College Entrance Examination Board.

Konold, C. (1991a). Understanding students' beliefs about probability. In E. von Glaserfeld (Ed.), Radical Constructivism in Mathematics Education (pp. 139-156). The Netherlands: Kluwer Academic Publishers.

Konold, C. (1991b). *ChancePlus: A Computer Based Curriculum for Probability and Statistics* (Second Year Report), Amherst, MA: University of Massachusetts Scientific Reasoning Research Institute.

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. Journal for Research in Mathematics Education, 24(5), 392-414.

Lesh, R., Amit, M., & Schorr, R. Y. (1997). Using "real-life" problems to prompt students to construct conceptual models for statistical reasoning. In I. Gal & J. Garfield (Eds.), The assessment challenge in statistics education (pp. 123-138). Amsterdam, The Netherlands: IOS Press.

Lesh, R., & Clarke, D. (2000). Formulating operational definitions of desired outcomes of instruction in mathematics and science education. In A. Kelly & R. Lesh (Eds.), Handbook of research design in mathematics and science education (pp. 113-149). Mahwah, NJ: Lawrence Erlbaum Associates.

Lesh, R. & Kelly, A. (2000). Multitiered teaching experiments. In A. Kelly & R. Lesh (Eds.), Handbook of research design in mathematics and science education (pp. 197-230). Mahwah, NJ: Lawrence Erlbaum Associates.

Lesh, R., & Lehrer, R. (2000). Iterative refinement cycles for videotape analysis of conceptual change. In A. Kelly & R. Lesh (Eds.), Handbook of research design in mathematics and science education (pp. 665- 708). Mahwah, NJ: Lawrence Erlbaum Associates.

Miles, M. B., & Huberman, A. M. (1994). Qualitative Data Analysis. Thousand Oaks, CA: Sage.

Munisamy, S., & Doraisamy, L. (1998). Levels of understanding of probability concepts among secondary school pupils. International Journal of Mathematical Education in Science & Technology, 29(1), 39-45.

National Council of Teachers of Mathematics. (2000). Principles and Standards for School Mathematics. Reston, VA: Author.

Piaget, J., & Inhelder, B. (1975). The origin of the idea of chance in children (J. Leake, L. & P. Burrell & H. D. Fischbein, Trans.). New York: W. W. Norton (Original work published 1951).

Polaki, M. V. (2000). Using instruction to trace Basotho elementary students' growth in probabilistic thinking. Unpublished doctoral dissertation, Illinois State University, Normal.

Romberg, T. A. (1992). Perpspectives on scholarship and research methods. In D. A. Grouws (Ed.), Handbook of Research on Mathematics Teaching and Learning (pp. 49-64). New York: Macmillan.

Ruthven, K. (1996). Calculators in the mathematics curriculum: The scope of personal computational technology. In A. J. Bishop & K. Clements & C. Keitel & J. Kilpatrick & C. Laborde (Eds.), International Handbook of Mathematics Education (pp. 435-468). The Netherlands: Kluwer.

Scheaffer, R. L., Gnanadesikan, M., Watkins, A. E., & Witmer, J. (1996). Activity-based statistics. New York: Springer-Verlag.

Schoenfeld, A. H. (1985). Mathematical problem solving. Orlando, FL: Academic Press.

Schrage, G. (1983). (Mis-) Interpretation of stochastic models. In R. Scholz (Ed.), Decision making under uncertainty (pp. 351-361). Amersterdam: North-Holland.

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, 285-316.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), Handbook of Research on Mathematics Teaching and Learning (pp. 465-494). New York: Macmillan.

Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. Journal for Research in Mathematics Education, 26, 114-145.

Steffe, L. P., & Thompson, P. W. (2000). Teaching experiment methodology: Underlying principles and essential elements. In A. Kelly & R. Lesh (Eds.), Handbook of research design in mathematics and science education (pp. 267-306). Mahwah, NJ: Lawrence Erlbaum Associates.

Straf, M. (2002). A quiet tide. Amstat News, *296*, 2-3.

Tarr, J. E., & Jones, G. A. (1997). A framework for assessing middle school students' thinking in conditional probability and independence. Mathematics Education Research Journal, 9, 39-59.

Tashakkori, A., & Teddlie, C. (1998). Mixed methodology: Combining qualitative and quantitative approaches. Thousand Oaks, CA: Sage.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 5, 207-232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.

Tzur, R. (1999). An integrated study of children's construction of improper fractions and the teacher's role in promoting that learning. Journal for Research in Mathematics Education, 30, 390-416.

Wares, A. (2001). Middle school students' construction of mathematical models. Unpublished doctoral dissertation, Illinois State University, Normal.

Walton, K. D. & Walton, J. D. (1987). Probability: Actual trials, computer simulations, and mathematical solutions. Reston, VA: Mathematics Education Trust. (ERIC Document Reproduction Service No. ED295843)

Watkins, A. E., Roberts, R. A., Olsen, C. R., & Scheaffer, R. L. (1997). Teacher's guide: AP statistics. New York: College Board.

Wilder, P. (1994). Models and metaphors in reasoning about probability. In H. Mellar & J. Bliss & R. Boohan & J. Ogborn & C. Tompsett (Eds.), Learning with artificial worlds: Computer-based modelling in the curriculum (pp. 85-93). London: Falmer.

Yackel, E., Cobb, P., & Wood, T. (1990). The importance of social interaction in children's construction of mathematical knowledge. In T. J. Cooney & C. R. Hirsch (Eds.), Teaching and learning mathematics in the 1990s: 1990 Yearbook (pp. 12-21). Reston, VA: National Council of Teachers of Mathematics.

Yates, D., Moore, D., & McCabe, G. (1999). The practice of statistics. New York: W. H. Freeman and Company.

Zimmermann, G., & Jones, G. A. (2002). Probability simulation: What meaning does it have for high school students? Canadian Journal Science, Mathematics, and Technology Education, 2, 221-236.

APPENDIX A

ASSESSMENT INSTRUMENTS

Preassessment

Task 1: Pizza Problem

A student, Stan, was given the following problem.

The Pizza Wagon has determined that 60 percent of their phone orders for pizza contain meat (sausage, pepperoni, etc.) and the remaining 40 percent of their phone orders are for pizzas with no meat (cheese, veggie, etc.). What is the probability that the next two phone orders for pizza are each with meat?

To simulate the Pizza Wagon's situation, Stan used colored chips. Stan chose 6 red chips to each represent an order for pizza with meat, and he chose 4 green chips to each represent an order for pizza without meat. To simulate the actual order, Stan put all 10 chips into a bag, shook the bag, and drew out one chip. He recorded the color, put the chip back, and then repeated this action a total of 50 times.

1) Remembering that the Pizza Wagon is trying to determine the probability that the next 2 pizzas have meat, do you think that Stan's simulation would enable him to determine the probability that the next two pizzas have meat? **Justify your response.**

2) If you don't think Stan's simulation will work, how would you change the simulation to determine the probability that the next two pizzas have meat?

3. **Suppose Stan conducted his experiment 50 times and his results were as follows:**

RRRGRRRGGGGRGRRGGGGRRRGRGRRGGGRRGGGRRGGRRRGGRGRGGR

Red chip	**25 times**
Green chip	**25 times**

Using the outcomes or the results from Stan's experiment, could you determine the probability that the next two phone orders for pizza have meat? If your response is "yes," calculate the required probability and explain your reasoning. If your response is "no," explain why not.

Task 2: Radio Problem

The school radio station plays three types of music: hip-hop, alternative, and country. The DJ uses a format such that the probability he plays hip-hop is 0.4, the probability he plays alternative is 0.4 and the probability that he plays country is 0.2. If you turn your radio on at 10:00 am and then again at 2:30 pm, what is the probability that both times you hear a hip-hop song?

4. How would you simulate this situation to determine the probability that both times you hear hip-hop? Assume you would have access to such things as spinners, chips, dice, calculator, or anything else you think may help. Describe your simulation precisely and make sure you provide me with enough detail that I could go and repeat the simulation. (Use the back of the paper to write your answer.)

5. Make up some data that you think your simulation would produce and write it below.

6a. Using your data in #5, determine the probability that both times you turn on the radio you hear a hip-hop song. Explain how you did this.

6b. How many times would you do the simulation? Why?

6c. Would your solution change if you did the experiment 50 times, 1000 times, or 100,000 times? Explain why or why not.

Post-assessment

Task 1: Free Throw Shooter Problem

Lora was asked to design a simulation for the following problem.

Beth's basketball statistics show that historically, when she is at the free throw line, Beth makes about 70% of her free throws. What is the probability that Beth misses both free throws?

To simulate the Free Throw Shooter situation, Lora used colored balls. Lora let 7 red balls represent making the shot, and she let 3 blue balls represent missing the shot. To simulate the actual problem, Lora put all 10 colored balls into a bag, shook the bag, and drew out one ball. She recorded the color, put the ball back, and then repeated a number of times.

1) Remembering that you are trying to determine probability that Beth misses both free throws, do you think that Lora's simulation would enable her to determine the probability that Beth misses both free throws? **Justify your response.**

2) If you don't think Lora's simulation will work, how would you change the simulation to determine the probability that Beth misses both free throws?

3. **Suppose Lora conducted her experiment 50 times and her results were as follows:**

RBBBBRRBRBBBRRRRBBRBRRRRRBBRBRRRBBRRRRRBBBRBRRBRR

27 Red (made)
23 Blue (missed)

Using the results from Lora's experiment, could you determine the probability that Beth missed both free throw shots? If your response is "yes," calculate the required probability and explain your reasoning. If your response is "no," explain why not.

Task 2: Space Shuttle Problem *

A primary power system, S_1, on a space shuttle has a backup system, S_2. If S_1 fails during a mission, S_2 automatically takes over. Suppose the probability that S_1 fails during a mission is 0.2 and the probability that S_2 fails is 0.3. What is the probability that both power systems fail?

4. How would you simulate this situation to determine the probability that both power systems on the space shuttle fail? Assume you would have access to such things as spinners, chips, dice, calculator, or anything else you think may help. Describe your simulation precisely and make sure you provide me with enough detail that I could go and repeat the simulation. (Use the back of the paper to write your answer.)

5. Make up some data that you think your simulation would produce and write it below.

6a. Using your data in #5, determine the probability that both power systems fail. Explain how you did this.

6b. How many times would you do the simulation? Why?

6c. Would your solution change if you did the experiment 50 times, 1000 times, or 100,000 times? Explain why or why not.

Retention-assessment

Task 1: Train Crossing Problem

Tim was asked to design a simulation for the following problem.

> **Alana has to drive over a set of railroad tracks on her way to and from school everyday. Every time she crosses, Alana has a 60% chance of being stopped by a train. What is the probability that Alana gets stopped by a train both on her way to AND from school on any given day?**

To simulate the Train Crossing situation, Tim used colored balls. Tim let 6 red balls represent getting stopped by the train, and he let 4 blue balls represent not getting stopped by the train. To simulate the actual problem, Tim put all 10 colored balls into a bag, shook the bag, and drew out one ball. He recorded the color, put the ball back, and then repeated a number of times.

1) Remembering that you are trying to determine probability that Alana is stopped by the train both times she crosses, do you think that Tim's simulation would enable him to determine the probability that Alana is stopped by a train both times? **Justify your response.**

2) If you don't think Tim's simulation will work, how would you change the simulation to determine the probability that Alana is stopped by a train both times?

3. **Suppose Tim conducted his experiment 50 times and his results were as follows:**

   RBBBBRRBRBBBRRRRBBRBBBRRRRBBRBRRRBBRRRRRBBBRBRRBBB

   25 Red (stopped by the train)
   25 Blue (not stopped by the train)

   Using the results from Tim's experiment, could you determine the probability that Alana was stopped by a train both times she crossed? If your response is "yes," calculate the required probability and explain your reasoning. If your response is "no," explain why not.

<u>Task 2: Kyle's Serve Problem</u>

> **When playing tennis, a player gets a second chance to serve if his first serve is out of bounds (considered a "fault"). Kyle's tennis statistics show that he has a 0.8 probability of getting a fault on his first serve and a 0.1 probability of getting a fault on his second serve (assuming his first serve is a fault). What is the probability Kyle faults on both serves?**

4. How would you simulate this situation to determine the probability that Kyle faults on both serves? Assume you would have access to such things as spinners, chips, dice, calculator, or anything else you think may help. Describe your simulation precisely and make sure you provide me with enough detail that I could go and repeat the simulation. (Use the back of the paper to write your answer.)

5. Make up some data that you think your simulation would produce and write it below.

6a. Using your data in #5, determine the probability that Kyle faults on both serves. Explain how you did this.

6b. How many times would you do the simulation? Why?

6c. Would your solution change if you did the experiment 50 times, 1000 times, or 100,000 times? Explain why or why not.

APPENDIX B

INSTRUCTIONAL PROGRAM: HYPOTHETICAL LEARNING TRAJECTORIES

Session 1

Learning Goals

- To construct a valid probability generator.
- To determine and explain an event or trial for the simulation.
- To use the probability generator to simulate a 1-dimensional probability situation.
- To use the simulation results to compare with “expected” results.

Instructional Activities

Activity – “Counting Successes”
(Scheaffer, Gnanadesikan, Watkins & Witmer, 1996).

Conjectured Learning Process

- Students will be able to construct a valid probability generator using a 1-1 correspondence between the probability given in the problem and numbers found in the random number table.
- Students will be able to use the random number table to determine the outcomes as required in the problems.
- Some students will attempt to solve the problem using theoretical probability, thus questioning the need to simulate the problem
- Some students may question the randomness of their results if the outcomes of the simulation do not seem to match the stated probabilities of the problem.

Session 2

Learning Goals

- To construct a valid probability generator.
- To determine and explain an event and trial for the simulation.
- To use the probability generator to simulate a multi-dimensional probability situation.
- To use the simulation results to compare with “expected” results.

Instructional Activities

Activity – “True/False History Test”
(Scheaffer et al., 1996).

Conjectured Learning Process

- Students will be able to construct a valid probability generator using a 1-1 correspondence between the probability given in the problem and numbers found in the random number table.
- Students will most likely struggle with the multi-dimensional nature of the outcome (i.e. 5 answers on a test).
- Students may struggle with being able to recognize the sample space.

Session 3

Learning Goals

- To construct a valid probability generator using other devices than random number generator on calculator (dice, spinners).
- To determine and explain an event and trial for the simulation.
- To use the probability generator to simulate a multi-dimensional probability situation.
- To be able to recognize the sample space.
- To use the simulation results to compare with "expected" results.
- To address (list) assumptions inherent in the simulation problem.

Instructional Activities

Activity – "Free Throw Shooter" &
"Blood Bank" (Scheaffer et al., 1996)

Conjectured Learning Process

- Students will be able to construct a valid probability generator using a 1-1 correspondence between the probability given in the problem and various devices.
- Some students will continue to with the multi-dimensional nature of the outcome (i.e. 10 shots in a game).
- Some students may struggle with being able to recognize the sample space.
- Most students (if not all) should be able to recognize how to represent the 33% in the "Blood Bank" problem on a random number generator (other than the calculator).
- While students will be able to recognize some assumptions inherent in the problems, they may not recognize the most relevant ones (i.e. assuming the free throw shooter shoots 10 shots every game and each shot is independent of the previous shot).

Session 4

Learning Goals

- To recognize and understand the concept of randomness
- To use the probability generator to simulate a multi-dimensional probability situation.
- To use the simulation results to compare with "expected" results.
- To address (list) assumptions inherent in the simulation problem.
- To use technology (calculators) to design & conduct simulations.
- To look at "long run" behavior in simulated situations

Instructional Activities

Activity – "Designing Simulations on the TI-83" (Rossman, 1996)

Conjectured Learning Process

- Students will struggle with the concept of randomness.
- Students will have difficulty recognizing randomly vs. man-made outcomes.
- Will not recognize that runs are likely in long sequence of randomly generated outcomes.
- Using the technology, students may struggle with "logic" of programming.

Session 5

Learning Goals

- To recognize and understand the concept of randomness
- To address (list) assumptions inherent in the simulation problem.
- To use technology (calculators) to design & conduct simulations.
- To look at "long run" behavior in simulated situations

Instructional Activities

Activity – "Randomly-generated outcomes"
2nd Session on "Designing Simulations on the TI-83"
Homework: "Women Working" (Gnanadesikan, Scheaffer, & Swift, 1987)

Conjectured Learning Process

- Students will struggle with the concept of randomness.
- Students will have difficulty recognizing randomly vs. man-made outcomes.
- Will not recognize that runs are likely in long sequence of randomly generated outcomes.
- Using the technology, students may struggle with "logic" of programming.

Session 6

Learning Goals

- To use tree diagrams to determine the sample space of a situation
- To recognize and use the Multiplication Principle to determine the number of outcomes in a given situation.
- To be able to determine the theoretical probability of simple situations.
- To use technology (calculators) to design & conduct simulations.
- To compare long run behavior of simulations with theoretical probability

Instructional Activities

Activity – "Tree Diagrams"

Conjectured Learning Process

- Students will be able to draw tree diagrams.
- Students will be able to use the Multiplication Principle to determine the number of outcomes in a sample space.
- Some students may still not recognize the significance of long run behavior of simulations. However, now that we are comparing to theoretical, it may help more students to realize that a simulation must be done "many" times.
- Students will become increasingly comfortable with using the calculator to design and run simulations.

Session 7

Learning Goals

- To recognize mutually exclusive events and be able to determine the probability of such events.
- Be able to recognize non-disjoint events and be able to determine the probability of such events.
- Be able to construct Venn diagrams to illustrate sample space and use to determine the probability of an event.

Instructional Activities

Activity – "Probability Workshop" (Rossman, 1996) and teacher examples

Conjectured Learning Process

- Students may not be familiar with Venn diagrams. However, once shown, students will be able to draw and use to determine probability of events.
- Students may struggle with the difference between mutually exclusive events and those that are not.
- In familiar situations (such as cards), students will be able to determine the probability of mutually exclusive events and those that are not. However, students may have difficulty transferring this knowledge to unfamiliar situations.

Session 8

Learning Goals

- Be able to construct Venn diagrams to illustrate sample space and use to determine the probability of an event.
- To be able to determine the conditional probability of an event.

Instructional Activities

Activity – "Tree Diagrams," teacher examples and "Probability Workshop" (Rossman, 1996)

Conjectured Learning Process

- Given a familiar contextual example, students will be able to make sense of Venn diagrams and use them to determine the probability of an event.
- Some of the students should recognize how to determine the conditional probability of an event. For a few students, this is almost entirely new.
- Students will be able to determine P(A), however, they may struggle with determining P(A and B).
- Likewise, students will be able to determine P(A), but may have difficulty with determining P(A | B) in order to calculate P(A & B).

Session 9

Learning Goals

- Be able to determine the theoretical probability of independent events.
- Be able to compare empirical and theoretical probabilities.
- To recognize the long run behavior of simulations
- To recognize that one cannot predict the next event or predict in the short run.

Instructional Activities

Theoretical problems involving coin tosses –
demo simulation with TI-83 ProbSim.
Problem involving "false positive" of Aids test

Conjectured Learning Process

- By looking at a graph of toss # vs. probability of heads will understand the long-run behavior of simulations.
- Students have difficulty applying probability concepts. They will have difficulty interpreting the Aids problem and then determining how to solve. Most students will also have difficulty recognizing that P(at least one) is the same as 1-P(none)
- Students will apply various heuristics in trying to "predict" the outcome of the Jolly Rancher activity. They will struggle with the limitations of predicting the next or in the short run and how this connects to examining long run behavior to draw conclusions.

Session 10

Learning Goals

- To recognize, interpret and use complements of events.
- To recognize that not all simulated events can be determined by theoretical probability.
- To recognize that one cannot predict the next event or predict in the short run.
- To recognize the role of independence in determining probabilities.

Instructional Activities

Warm-up worksheet on sets involving complements.

Activity – "What is Random Behavior – Jolly Ranchers" (Rossman, 1996)

Conjectured Learning Process

- By using a smaller set of outcomes, students will have an easier time recognizing what a complement is and how to use it to simplify problems.
- Students will apply various heuristics in trying to "predict" the outcome of the Jolly Rancher activity. They will struggle with the limitations of predicting the next or in the short run and how this connects to examining long run behavior to draw conclusions.
- Students will be surprised the role independence plays in the tack experiment.

Session 11

Learning Goals

- To be able to simulate an event with an unspecified number of trials.
- To recognize that not all simulated events can be determined by theoretical probability.
- To recognize the role of independence in determining probabilities.

Instructional Activities

Activity – "What's the Chance" – focus is on independence (Scheaffer et al., 1996)

Conjectured Learning Process

- Students will be surprised the role independence plays in the tack experiment.
- Students may struggle with the fact that the number of trials is unspecified.
- Students will continue to not do "enough" trials to provide a "confident" probability of a situation.

Session 12

Learning Goals

- To determine if a simulation is valid.
- To recognize that not all simulated events can be determined by theoretical probability.

Instructional Activities

Problem – "Are these simulation designs valid?"

<u>Conjectured Learning Process</u>

- Students will begin to do more trials to determine empirical probability.
- Related to the valid simulation designs, in the second case, students will intuitively believe that it is invalid. With some probing (and coaxing) they will determine the actually outcomes of the game to determine if the probability generator for the simulation is valid.

APPENDIX C

INSTRUCTIONAL ACTIVITIES

C1. Counting Successes (Scheaffer, Gnanadesikan, Watkins, & Witmer, 1996)

Read the article from the *Milwaukee Journal* (May 1992) entitled "Non-cents: Laws of probability could end need for change," which follows:

a. Does this seem like a reasonable proposal to eliminate carrying change in your pocket?
b. Do you think the proposal is fair? Explain your reasoning.

Non-cents: Laws of probability could end need for change

Chicago, Ill.—AP—Michael Rossides has a simple goal: to get rid of that change weighing down pockets and cluttering up purses.

And, he says, his scheme could help the economy.

"The change thing is the cutest aspect of it, but it's not the whole enchilada by any means," Rossides said.

His system, tested Thursday and Friday at Northwestern University in the north Chicago suburb of Evanston, uses the law of probability to round purchase amounts to the nearest dollar.

"I think it's rather ingenious," said John Deighton, an associate professor of marketing at the University of Chicago.

"It certainly simplifies the life of a businessperson and as long as there's no perceived cost to the consumer it's going to be adopted with relish," Deighton said.

Rossides' basic concept works like this:

A customer plunks down a jug of milk at the cash register and agrees to gamble on having the $1.89 price rounded down to $1 or up to $2.

Rossides' system weighs the odds so that over many transactions, the customer would end up paying an average $1.89 for the jug of milk but would not be inconvenienced by change.

That's where a random number generator comes in. With 89 cents the amount to be rounded, the amount is rounded up if the computerized generator produced a number from 1 to 89; from 90 to 100 the amount is rounded down.

Rossides, 29, says his system would cut out small transactions, reducing the cost of individual goods and using resources more efficiently.

The real question is whether people will accept it.

Rossides was delighted when more than 60% of the customers at a Northwestern business school coffee shop tried it Thursday.

Leo Hermacinski, a graduate student at Northwestern's Kellogg School of Management, gambled and won. He paid $1 for a cup of coffee and a muffin that normally would have cost $1.30.

Rossides is seeking financial backing and wants to test his patented system in convenience stores.

But a coffee shop manager said the system might not fare as well there.

"Virtually all of the clientelle at Kellogg are educated in statistics, so the theories are readily grasped," said Craig Witt, also a graduate student. "If it were just to be applied cold to average convenience store customers, I don't know how it would be received."

Source: Milwaukee Journal, May 1992

Activity

2. Investigate a single random outcome per trial.
   Suppose the soft drink machine you use charges $.75 per can. The scheme proposed by Mr. Rossides requires you to pay either $0 or $1, depending on your selection of random number. You select a two-digit number between 01 and 00 (with 00 representing 100). If the number you select is 75 or less, you pay $1. If the number you select is greater than 75, you pay nothing.

   a. From a random number table choose a random number between 01 and 100. If this represents your selection at the drink machine, how much do you pay for your drink?
   b. The article suggests that things will even out in the long run. Suppose that over a period of time you purchase 60 drinks from the machine and use the random mechanism for payment each time. This can be simulated by choosing 60 random numbers between 01 and 100. Make such a selection of 60 random numbers.

      i. How many drinks did you pay $1? What is the total amount you paid for 60 drinks?

      ii. If you had paid the $.75 for each drink, how much would you have paid for 60 drinks? Does the scheme of random payment seem fair?

iii. C2. Labor Force Problem*

Use a separate sheet of paper to record your answers!! Using your calculator, design a simulation for the following situations.

1. The percentage of women in the labor force of a certain country is 30 percent. A company employs ten workers, two of whom are women.

   a. Design a simulation to answer the questions below:

      i. What assumptions are you making?

      ii. Define a trial for this simulation

      iii. How are you assigning the outcomes to random digits?

      iv. What commands did you put into your calculator?

      v. How many trials did you do to answer each question?

   Questions:

   A. Using your simulation results, determine the probability that this would occur by chance.

   B. Estimate the probability that a company of ten workers would employ two or fewer women by chance.

   C. Estimate the expected number of women that a company of ten workers would employ.

   D. In simulating the number of women among the ten workers, what number occurs most frequently?

   E. On the basis of your simulation, do you think that women are underrepresented in the company? Why or why not?

C3. Pay Your Bill (Walton & Walton, 1987)

Pay Your Bill

When Tom tries to collect $5 a week for delivering the daily paper, Mr. Bernoulli offers Tom a choice. He proposes that he either pay Tom $5 weekly or each week let Tom reach into a bag containing five $1 bills and one $10 bill and draw two bills. Being a doubter, Tom is suspicious that Mr. Bernoulli's scheme is a scam.

1. Design a simulation for this problem. Describe your design in detail.

2. Record the outcomes of your simulation.

3. According to your simulation results, which method of payment should Tom choose? Justify your answer.

4. What is the expected value for this problem?

5. How does this compare with your expected value?

C4. Randomly Generated Outcomes (Scheaffer et al., 1996)

The activity used in the whole-class teaching experiment was an adaptation of the following activity. The adaptation concentrated on turning an individual or group activity into a whole-class activity.

Provide students with a piece of overhead transparency and overhead marker. Randomly assign each student either a “C” for calculator-generated or a “S” for student-generated. Make sure only that student knows whether he or she is a “C” or an “S”. Have them mark the upper right corner of their overhead transparency with their designated letter.

Then have each student consider tossing a coin 50 to 100 times. They are to generate these outcomes (H or T) according to their earlier designation.

The results are collected and displayed one at a time on the overhead for students to judge whether the outcomes are calculator-generated or student-generated. The “C” or “S” is covered and revealed after students have made their guesses. A discussion about randomness follows the activity.

APPENDIX D

TEST FOR NORMALITY OF ASSESSMENT SCORE DIFFERENCES

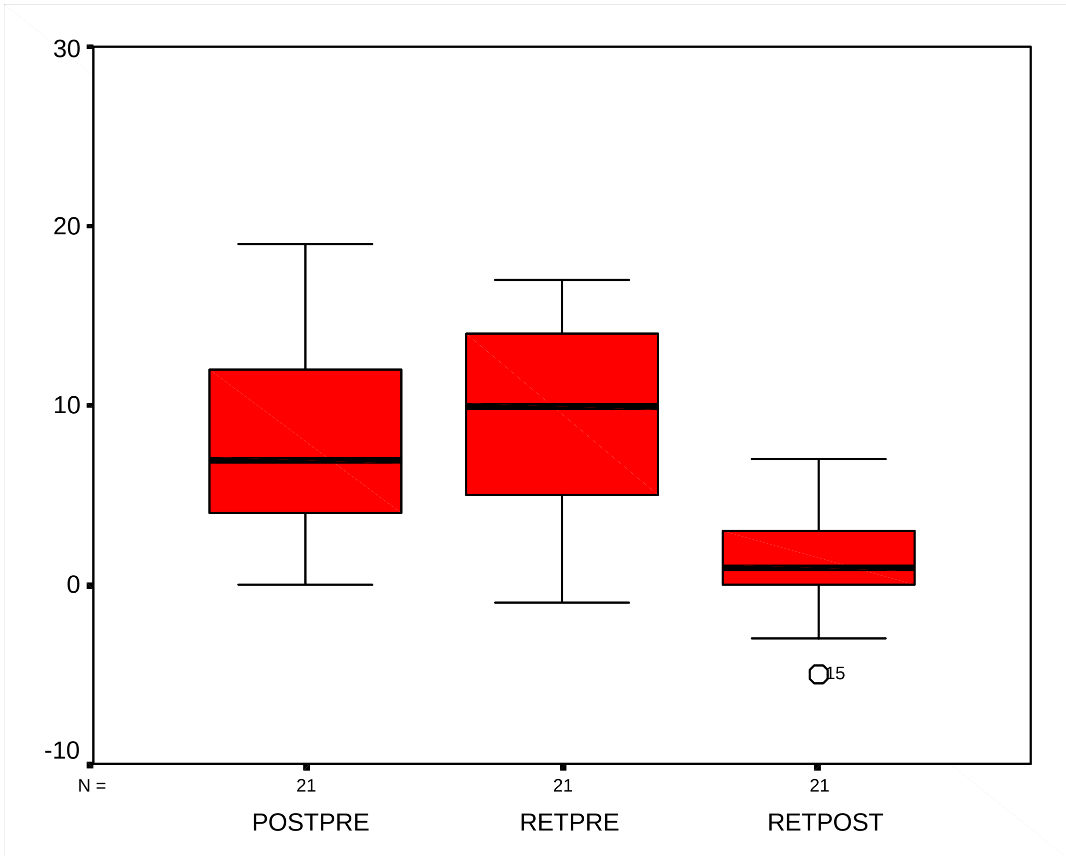

Figure D1. Boxplots for Assessing Normality

Table D1

Shapiro-Wilk's Test of Normality of Assessment Score Differences

Difference	Statistic	df	p
Pre- and Post-	.960	21	.499
Pre- and Retention	.937	21	.249
Post- and Retention	.984	21	.956

Probabilistic Simulation Pre-assessment

TASK 1: PIZZA PROBLEM

A student, Stan, was given the following problem.

The Pizza Wagon has determined that 60 percent of their phone orders for pizza contain meat (sausage, pepperoni, etc.) and the remaining 40 percent of their phone orders are for pizzas with no meat (cheese, veggie, etc.). What is the probability that the next two phone orders for pizza are each with meat?

To simulate the Pizza Wagon's situation, Stan used colored chips. Stan chose 6 red chips to each represent an order for pizza with meat, and he chose 4 green chips to each represent an order for pizza without meat. To simulate the actual order, Stan put all 10 chips into a bag, shook the bag, and drew out one chip. He recorded the color, put the chip back, and then repeated this action a total of 50 times.

1) Remembering that the Pizza Wagon is trying to determine the probability that the next 2 pizzas have meat, do you think that Stan's simulation would enable him to determine the probability that the next two pizzas have meat? **Justify your response.**

2) If you don't think Stan's simulation will work, how would you change the simulation to determine the probability that the next two pizzas have meat?

3. **Suppose Stan conducted his experiment 50 times and his results were as follows:**

RRRGRRRGGGGRGRRGGGGRRRGRGRRGGGRRGGGRRGGRRRGGRGRGGR

Red chip	**25 times**
Green chip	**25 times**

Using the outcomes or the results from Stan's experiment, could you determine the probability that the next two phone orders for pizza have meat? If your response is "yes," calculate the required probability and explain your reasoning. If your response is "no," explain why not.

TASK 2: RADIO PROBLEM

The school radio station plays three types of music: hip-hop, alternative, and country. The DJ uses a format such that the probability he plays hip-hop is 0.4, the probability he plays alternative is 0.4 and the probability that he plays country is 0.2. If you turn your radio on at 10:00 am and then again at 2:30 pm, what is the probability that both times you hear a hip-hop song?

4. How would you simulate this situation to determine the probability that both times you hear hip-hop? Assume you would have access to such things as spinners, chips, dice, calculator, or anything else you think may help. Describe your simulation precisely and make sure you provide me with enough detail that I could go and repeat the simulation. (Use the back of the paper to write your answer.)

5. Make up some data that you think your simulation would produce and write it below.

6a. Using your data in #5, determine the probability that both times you turn on the radio you hear a hip-hop song. Explain how you did this.

6b. How many times would you do the simulation? Why?

6c. Would your solution change if you did the experiment 50 times, 1000 times, or 100,000 times? Explain why or why not.

Probabilistic Simulation Post-assessment

TASK 1: FREE THROW SHOOTER PROBLEM

Lora was asked to design a simulation for the following problem.

Beth's basketball statistics show that historically, when she is at the free throw line, Beth makes about 70% of her free throws. What is the probability that Beth misses both free throws?

To simulate the Free Throw Shooter situation, Lora used colored balls. Lora let 7 red balls represent making the shot, and she let 3 blue balls represent missing the shot. To simulate the actual problem, Lora put all 10 colored balls into a bag, shook the bag, and drew out one ball. She recorded the color, put the ball back, and then repeated a number of times.

1) Remembering that you are trying to determine probability that Beth misses both free throws, do you think that Lora's simulation would enable her to determine the probability that Beth misses both free throws? **Justify your response.**

2) If you don't think Lora's simulation will work, how would you change the simulation to determine the probability that Beth misses both free throws?

3. **Suppose Lora conducted her experiment 50 times and her results were as follows:**

   RBBBBRRBRBBBRRRRBBRBRRRRRRBBRBRRRBBRRRRRBBBRBRRBRR

   27 Red (made)
   23 Blue (missed)

Using the results from Lora's experiment, could you determine the probability that Beth missed both free throw shots ? If your response is "yes," calculate the required probability and explain your reasoning. If your response is "no," explain why not.

TASK 2: SPACE SHUTTLE PROBLEM

A primary power system, S_1, on a space shuttle has a backup system, S_2. If S_1 fails during a mission, S_s automatically takes over. Suppose the probability that S_1 fails during a mission is 0.2 and the probability that S_2 fails is 0.3. What is the probability that both power systems fail?

4. How would you simulate this situation to determine the probability that both power systems on the space shuttle fail? Assume you would have access to such things as spinners, chips, dice, calculator, or anything else you think may help. Describe your simulation precisely and make sure you provide me with enough detail that I could go and repeat the simulation. (Use the back of the paper to write your answer.)

5. Make up some data that you think your simulation would produce and write it below.

6a. Using your data in #5, determine the probability that both power systems fail. Explain how you did this.

6b. How many times would you do the simulation? Why?

6c. Would your solution change if you did the experiment 50 times, 1000 times, or 100,000 times? Explain why or why not.