TEACHERS' UNDERSTANDING OF STUDENTS' CONCEPTIONS ABOUT CHANCE:

AN EXPERT-NOVICE CONTRAST

by

OLGA LUCIA ZAPATA CARDONA

(Under the Direction of Jeremy Kilpatrick)

ABSTRACT

This study investigated teachers' perception of students' thinking about chance. In particular, the study explored how teachers anticipated and explained students' difficulties with the ideas of chance, and the strategies teachers claimed to use in order to help students reorganize their thinking. The study was motivated by the literature on reasoning about chance in which researchers present detailed descriptions of the difficulties students have in thinking about uncertainty. Yet little research has focused on teachers' practices associated with students' difficulties. Little is known about how teachers perceive students' difficulties, how they anticipate students' struggles, and what kinds of strategies are useful for teachers to deal with those difficulties.

Two teachers, one expert and one novice, members of an AP Statistics learning community, participated in this study during fall 2007 and the beginning of winter 2008. They were observed five times in the learning community meetings, and interviewed three times for about 1 hour each. The interviews explored four core ideas in statistics that have been associated with the sources of students' difficulties about chance: sample space, randomness, independence, and the law of large numbers. The interview protocol included 12 episodes whose tasks had been previously used in research. Data were collected in the form of observations, interviews, and artifacts and then analyzed using grounded theory, interpretativism, and an expert-novice contrast.

The results of this study highlighted that the expert and novice teachers exhibited differences in the way they perceived students' difficulties and in the way they dealt with them. The expert was able to identify students' difficulties at an early stage in the discussion; the novice, however, underestimated students' difficulties and needed a longer exposure to the episodes to recognize the difficulties. Data also showed that although the teachers used similar strategies, the way they integrated them differed.

The characteristics of expertise in teaching are still murky, but results from this study suggest that novice teachers are more open to considering students' difficulties when they reflect deeper on students' struggles. Consequently, teacher preparation and teacher professional development programs should be designed in ways that challenge teachers to reflect on students' difficulties.

INDEX WORDS: AP Statistics, learning community, chance reasoning, expert-novice

TEACHERS' UNDERSTANDING OF STUDENTS' CONCEPTIONS ABOUT CHANCE:

AN EXPERT-NOVICE CONTRAST

by

OLGA LUCIA ZAPATA CARDONA

B.Ed., Universidad de Antioquia, Colombia, 1996

M.S., Universidad Nacional, Colombia, 1999

M.Ed., Universidad de Manizales, Colombia, 2002

A Dissertation Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2008

TEACHERS' UNDERSTANDING OF STUDENTS' CONCEPTIONS ABOUT CHANCE:

AN EXPERT-NOVICE CONTRAST

by

OLGA LUCIA ZAPATA CARDONA

Major Professor: Jeremy Kilpatrick

Committee: Denise S. Mewborn
Andrew Izsák

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2008

DEDICATION

To my very first teacher: my mother.

ACKNOWLEDGEMENTS

The completion of this dissertation would not have been possible without the support and encouragement from so many people. I gratefully acknowledge Natasha and Eric for allowing me to enter to their world, for the time they gave to participate in this study in spite of all their other occupations, and for their openness about their thoughts.

I express deep gratitude to Mrs. Sanders and members of the AP Statistics learning community who were so generous in allowing this research to take place at their meetings. Thanks for their willingness to open the doors to me, for letting me in, and for sharing their thoughts and experience with me and for their support.

Words might be not enough to express my gratitude to my major professor Dr. Kilpatrick, who in spite of his busy agenda being in several advisory committees and a mathematics education ambassador around the world, always had time for a detailed reading, patient editing, and thoughtful comments and suggestions throughout my work. His comments and suggestions always challenged my thinking, pushed me to think differently, and showed me better ways to express my thoughts. He always was a source of inspiration, encouragement, and motivation, and has been a wonderful mentor.

I express my gratitude to the members of my advisory committee Dr. Mewborn and Dr. Izsák. I thank each of them for their time and expertise. Their dedicated reading, thoughtful comments, and thought-provoking questions helped me reflect upon and shape my thoughts and my writing.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS v

LIST OF TABLES ix

LIST OF FIGURES x

CHAPTER

1 INTRODUCTION AND BACKGROUND 1

Background 1

AP Statistics 2

The Four Ideas 4

The Expert-Novice Contrast 8

Problem Statement and Research Question 9

2 A REVIEW OF RELEVANT LITERATURE 11

Research on Thinking about Chance 11

The Expert-Novice Paradigm 13

3 METHOD 17

The AP Statistics Learning Community 18

My Position in the Study 21

Participants 21

Observations 23

Interview Instrument 25

Interviews 26

Interpretativism 27
Grounded Theory 29
Data Analysis 29
Evolution of the Framework 31
Principles for Teaching and Learning Ideas of Chance 35
4 ANALYSIS 37
Sample Space 37
Randomness 53
Independence 69
Law of Large Numbers 84
Conclusion 96
5 REVISITING THE DATA 98
Students Learn by Doing 100
Take into Account Students' Previous Knowledge 106
Use Cooperative Learning 107
Students Learn by Confronting Misconceptions 108
Use Technology to Visualize and Explore 109
Point Out Common Misuses of Chance 110
Conclusion 111
6 SUMMARY AND CONCLUSIONS 113
Relating the Findings to the Research Questions 113
Benefits from Participation 118
Limitations 119

Significance of the Study 120

Implications for Teacher Education Programs 121

Recommendations for Future Research.. 124

REFERENCES 126

LIST OF TABLES

Page

Table 1: AP Statistics Learning Community Members' Professional Backgrounds 22

Table 2: Dates of Data Collection...24

Table 3: Proportion of Attributes in the Fake Sequences Task 65

Table 4: Frequency of Reference to Principles of Learning Statistics in Teachers' Reflections..99

LIST OF FIGURES

Page

Figure 1: Natasha’s solution to the Bags and Marbles Task 39

Figure 2: Probabilities for the Bags and Marbles Task as interpreted by Natasha........................ 40

Figure 3: Two spinners.. 47

Figure 4: Natasha’s student failing to list the sample space 49

Figure 5: Natasha’s sketch for the Gumball Machine Task. 54

Figure 6: Natasha’s picture to solve the Balls in Urns Task 75

CHAPTER 1

INTRODUCTION AND BACKGROUND

This report tells the story of two Advanced Placement (AP) Statistics teachers, one expert and one novice, and their reflections on understanding students' difficulties in reasoning about chance and designing strategies to help students overcome such difficulties. The teachers were participants in an AP Statistics learning community that supported them in improving their teaching and their students' performance. Interviewing and observing these teachers showed me how they viewed students' struggles with chance and the strategies they used to support students' reasoning. I used the expert-novice contrast to analyze the teachers' responses.

Background

Chance and data have an important place in everyday life. The news media present reports of economic and social statistics, results of medical studies, odds in sports, and results of opinion polls. Many citizens need to be familiar with statistics to deal successfully with their jobs. Farmers use the results of agricultural research and crop forecasts to improve production. Insurance companies use statistical records to set the rates of their policies. Engineers monitor the quality, performance, and reliability of their products using statistical tools. There are many practical reasons to consider statistical literacy as needed by every citizen today.

Although participation in modern society requires an increasing familiarity with data and chance, research shows that people have difficulty reasoning in situations of uncertainty (Garfield & Ahlgren, 1988). The research literature is full of studies that describe the difficulties students (and even teachers) have in dealing with uncertainty, but little has been revealed about the pedagogical implications of those difficulties and how teachers might help students overcome them. Research has also shown that teachers' struggles in dealing with uncertainty situations are similar to those of students; little is known, however, about how teachers deal successfully with students' difficulties, and almost nothing is known about exemplary practices in teaching AP Statistics. This study connected what is known about students' difficulties in reasoning about uncertainty with how two teachers, one expert and one novice, dealt with such difficulties.

This study examined four core ideas underlying statistics and probability: sample space, randomness, independence, and the law of large numbers. These ideas have been considered by some researchers to be foundational and yet very problematic for students studying statistics and probability (Gal, 2005; Garfield & Ahlgren, 1988; Garfield & Ben-Zvi, in press; Shaughnessy, 2003). It explored, using an expert-novice contrast, how AP Statistics teachers anticipate and explain students' difficulties in uncertainty situations and what strategies teachers claim to use to assist students in reorganizing their intuitive reasoning into more formal thought.

AP Statistics

Advanced Placement Statistics (AP Statistics or AP Stats) is a college-level course in high school mathematics offered by the College Board. Students who take the AP Statistics course are expected to take an examination at the end of the school year. If they successfully complete the course and the examination, they may receive college credit, advanced placement, or both for a one-semester introductory college statistics course (College Board AP, 2007). The purpose of the AP Statistics course is to introduce students to the major concepts and tools for collecting, analyzing, and drawing conclusions from data. Students are exposed to four broad conceptual themes:

1. Exploring Data: Describing patterns and departures from patterns

2. Sampling and Experimentation: Planning and conducting a study

3. Anticipating Patterns: Exploring random phenomena using probability and simulation

4. Statistical Inference: Estimating population parameters and testing hypotheses

Students who enroll in AP Statistics should have completed a second-year course in algebra and, therefore, are usually in their junior or senior year when they take the course. This course is one of the College Board's more recent offerings. The first AP Statistics examination was administered in May 1997, and nearly 7,600 students took the test. In 2007, approximately 98,000 students took the test. Since the first examination in 1997 the number of students taking the test has been increasing, but the students' average scores have been going down. This phenomenon could have several explanations. First, the increase in the number of students taking the class implies an increase in the number of teachers teaching AP Statistics, which means that there may be many new teachers without enough preparation trying to figure out how to get students ready for the test. Second, since the first examination in 1997, the exam has had some additions in each section, which means that the exam given in 2007 was longer than the one given in 1997, but the time given to the students to complete the test did not increase. This aspect may also have affected students' performance (C. Franklin, personal communication, April 15, 2008).

There is a great complexity in the teaching of probability and statistics. Teachers say that as students progress through the AP Statistics curriculum, they do not seem to get enough preparation to deal efficiently with uncertainty situations. In fact, the literature has shown that even after instruction, students continue to have difficulties in reasoning about chance (Garfield & Ahlgren, 1988; Garfield & Ben-Zvi, in press).

The College Board, which designs and controls all AP courses, does not require that teachers of the courses have special training. The Board recognizes that there are many paths toward becoming an effective teacher of an AP course. However, some school districts require the teacher to attend a College Board training session before the teacher is allowed to teach an AP course.

The Four Core Ideas

The AP Statistics syllabus covers a variety of topics; most of them are typically taught in an introductory statistics course at college. These topics are organized into four categories: data exploration, design of studies, probability distributions with emphasis on sampling distributions, and basic methods of statistical inference. Behind the categories there are four underlying ideas that many researchers have considered not only as problematic but also as crucial for understanding uncertainty (Gal, 2005; Garfield & Ahlgren, 1988; Garfield & Ben-Zvi, in press; Shaughnessy, 2003). These core ideas, as I said above, are sample space, randomness, independence, and the law of large numbers. Gal calls the ideas of variation, randomness, and independence the “big ideas,” whereas Cobb and Moore (1997) consider the role of randomness to be the most relevant. The four core ideas should be considered not as specific topics in the AP Statistics curriculum but rather as underlying ideas across the curriculum. For example, randomness is not a topic for Monday’s lesson; it is a fundamental topic for understanding of data analysis, production of data, probability, and inference—the four categories of the AP Statistics curriculum.

The core ideas are neither exhaustive nor exclusive. They are not exhaustive, because there are other difficulties in reasoning about uncertainty (see, e.g., the research by Kahneman & Tversky, 1972, 1996; Tversky & Kahneman, 1973, 1974, 1983). They are not exclusive, because they often intertwine and overlap. For example, a gambler in a coin-tossing game who decides to bet on tails because the previous three results were heads reveals a misunderstanding of two fundamental ideas of chance: the randomness of tossing a fair coin and the independence of each of the four events. The following section describes each of the core ideas that were explored in this study.

Sample Space

To list all the possible outcomes after tossing a coin is without doubt a simple event, and many researchers agree that school children are able to establish the sample space before doing an actual experiment. When the event, however, is more complex than a simple event—like tossing two coins—students begin to have difficulties. Sample space is a dense but foundational concept for chance reasoning. Many researchers have stated that the misunderstanding of this concept is the beginning of learners' difficulties in calculating probabilities because to list the entire sample space the learner needs to be systematic and exhaustive. Part of the difficulty in listing the entire sample space has been attributed to weaknesses in combinatorial reasoning. For example, Shaughnessy (2003) indicated, by analyzing data from the 1996 National Assessment of Educational Progress (NAEP), that students were weak in the concept of sample space and that even when students were able to list the entire sample space, they were not always able to use it to make further predictions. Jones, Langrall, Thornton, and Mogill (1999) also mentioned the difficulty of sample space; in fact, they introduced the expression *sample space misconception* in which students eliminate some outcomes because they have occurred in a previous trial or do not consider some others because they look very unlikely.

Randomness

The concepts of randomness and chance variation are closely related. For example, if we are interested in sampling, we need to consider randomness in the selection of the sample. This means that all the members of the population have the same probability of being chosen, but if we are interested in having several samples, we would expect that the sample would be slightly different each time we do the sampling. Thus, the difference among the samples is due to chance variation.

Randomness is not an isolated concept or a topic in statistics curricula. On the contrary, it is a fundamental idea in the development of probabilistic and statistical thinking, and it is found in the contexts of reasoning about data, reasoning about probability experiments, and reasoning about information from the media. People need to understand chance variation of random processes to make sense of statements in the media about group differences, sampling error, margin of error, significance of differences, likelihood of results, and other assertions related to statistical inferences (Jones, Langrall, & Mooney, 2007). The understanding of chance variation is central to the study of statistics, and a consideration of it is precisely what differentiates a deterministic phenomenon from a stochastic one.[1] It is a very complex concept in statistics; however, students at all grades have weak conceptions of this idea and are not aware of its importance in thinking about chance (Reading & Shaughnessy, 2004). This weakness may arise because schooling teaches students processes and algorithms that reinforce that idea of a

[1] Determinism is a philosophical proposition that states that every event in the world is causally determined by an unbroken chain of prior occurrences and that there is at any instant exactly one physically possible answer; whereas a stochastic model states that the next state of a process cannot be fully determined in advance and considers a range of possible answers. A stochastic phenomenon has a nondeterministic nature; it is random.

deterministic world. Students quickly learn to expect right or wrong answers in a world where variation in answers is uncomfortable.

Some researchers have explored students' ideas of randomness (Batanero & Serrano, 1999) by creating fake and true chains of outcomes of experiments and having students identify the fake sequence. These researchers found that some students conceive of randomness as associated with regularity, balance, or unpredictability of a sequence. A secondary analysis (Zawojewski & Shaughnessy, 2000) of the probability items from the 1996 NAEP showed that students were deterministic in their answers. In a "Gum Ball Machine" problem (this task is discussed later in chapter 4), students were asked to predict how many red balls would appear in a group of 10 gumballs obtained from a gumball machine containing 100 gumballs: 50 red, 30 blue, and 20 yellow. Most students gave exact an answer (say, 6), and only 1 student out of a sample of 232 students considered a range of possible numbers for the answer.

Independence

The difficulties students have with the concept of independent events have been explored by a number of mathematics and statistics educators (e.g., Borovcnik & Bentz, 1991; Falk & Bar-Hillel, 1983). "Independence is one of the key concepts of probability theory" (Borovcnik & Bentz, p. 89), but it is an assumption easily violated by researchers. This violation may occur because independent probabilities are easier to calculate than dependent ones. This is, if A and B are independent events then P (A/B) = P (A); but if A and B are not independent P (A/B) = P (AB) / P (B), which is more difficult to calculate than the previous expression. A well-known example of the violation happens, for instance, when in experimental studies the assumption of independence in the allocation of treatments is not valid because a "without-replacement" condition makes the event dependent. In theory, if a population is implicitly infinite, the sampling is assumed to be independent. But in practice that is not usually the case; the population in most experiments is finite. Thus, this core idea of independence is essential, together with the idea of randomness, for understanding conditional probabilities, one of the hardest topics in AP Statistics (Shaughnessy, 1992). Several research studies, however, have shown that students in their reasoning violate the assumption of independence (Falk & Bar-Hillel). This violation generates other errors in students' reasoning.

Law of Large Numbers

Typically, people think that any size sample should reflect the characteristics of the parent population. The law of large numbers says that if an experiment with independent outcomes is repeated many times, in the long run the observed probability is close to the expected probability. For example, in tossing a fair coin 1000 times an observer could expect to get close to 500 heads, whereas in tossing the coin 4 times the outcomes should have more variability. A misunderstanding of the law of the large numbers is due to a failure to recognize that the experimental probability is close to the theoretical probability only when the number of trials is large. This core idea was included in this study because research has shown that it is a hard idea to grasp (J. F. Wagner, 2006). The phenomenon has been studied by some researchers under the topics of representativeness and sample size.

The Expert-Novice Contrast

The expert-novice contrast has been used in many fields of research to study what expertise is like, its effects, and how it is exhibited. Such studies have motivated much research on expertise in education and have shown that the characteristics of expert and novice teachers' performance resemble those of experts and novices in other fields (Carter, Sabers, Cushing, Pinnegar, & Berliner, 1987). There are qualitative differences between experts and novices in perceptions, thinking, knowledge, processing and using information, problem solving, and decision making.

The expert-novice research paradigm has some limits, but the nature of that contrast has been fruitful in other research fields (e.g., Benner, 1984; Chase & Simon, 1973; Hidi & Klaiman, 1983). The expert-novice contrast has been criticized by many scholars because it is a misleading comparison; however, it was useful to me in highlighting the different ways in which different teachers approached students' difficulties. Understanding expert and novice teachers' approaches may have several advantages: orient novice teacher education programs, inform decisions of policy makers, guide professional development, and inform research on teaching and teacher education.

Much research in mathematics education with an expert-novice contrast has used as novices the student teachers in their last year of preparation (e.g., Borko et al., 1992; Livingston & Borko, 1990) and as experts the student teachers' cooperating teachers. This characteristic has received much criticism because student teachers do not have full status as teachers; they are still in the process of training. Also, choosing participants at the extremes of the continuum of expertise has not been very useful for understanding the nature of that expertise. I used two practicing teachers to draw an expert-novice contrast. The use of a practicing teacher who was nonetheless a novice in teaching an AP Statistics course was designed to contribute to an understanding of the nature of expertise and how novices become experts.

Problem Statement and Research Questions

This study was focused on high school AP Statistics teachers who were in a volunteer AP Statistics learning community to discuss topics related to their teaching. I wanted to find out how they perceived students' struggles in chance situations and the strategies they used to help their students understand the situations and overcome difficulty.

In this qualitative study framed by interpretivism and grounded theory, I observed 8 high school teachers in an AP Statistics learning community and interviewed 2 of those teachers in depth. My goal was to get a description of (1) the teachers' notions of students' difficulties and (2) the strategies they used in instruction to help students overcome difficulties. Specifically, in this study, I addressed each of the following questions:

1. How do an expert and a novice teacher of AP Statistics anticipate students' difficulties in chance situations?
2. How do an expert and a novice teacher of AP Statistics explain students' difficulties in chance situations?
3. What strategies do an expert and a novice teacher of AP Statistics claim to use to assist students in the reorganization of their intuitive thinking into more formal thinking about chance?

To address these questions, in the chapters that follow I provide an overview of the relevant literature (chapter 2), detail the theory and the methodology employed (chapter 3), and present my findings (chapters 4 and 5). Finally, I discuss the implications of the study and make suggestions for future research (chapter 6).

CHAPTER 2

A REVIEW OF RELEVANT LITERATURE

Research on Thinking about Chance

Studies in thinking about chance abound not only in statistics education but also in mathematics education and educational psychology. Most of the studies on this topic have focused exclusively on how students think, and studies on how teachers assist students to overcome their intuitive thinking about chance are scarce.

To a large extent, the research on thinking about chance has been characterized by studies that explore how students think in uncertainty situations. This line of research has focused mainly on describing what students think when they are exposed to chance, what strategies of thought they use, and what the most common errors in their reasoning are. Researchers have typically used the expressions *judgment heuristics*[2] and *biases* to refer to strategies of thinking that lead people to errors and the word *misconceptions* to refer to errors in thinking that happen when people believe in concepts that are objectively false.

Research on thinking about chance may be classified into two approaches. The first has looked at data collected exclusively from forced-choice responses. That is, multiple-choice tasks have been given, generally, to college students and then analyzed through statistical tools. Researchers have sought to identify whether the students reasoned properly and to identify the

---

[2] Heuristics refer to strategies of thinking that act as aids for learning. The majority of the heuristics people use in problem solving are the result of personal experience or deterministic models of thinking. That is, people tend to think about right and wrong answers instead of considering a range of possibilities.

more common errors made by students in their solutions (Kahneman & Tversky, 1972, 1996; Lecoutre, 1992; Moutier & Houdé, 2003; Ross & DeGroot, 1982; Tentori, Bonini, & Osherson, 2004; Tversky & Kahneman, 1973, 1974, 1983). These studies have been helpful in illuminating the common errors students make in reasoning about chance, but the studies have not been successful in providing a detailed description of students' thinking or in offering explanations of the reasons that students may favor one option over others.

The second approach to research on thinking about chance has also employed multiple-choice items in some cases, but this approach has further explored students' reasoning more deeply using clinical interviews (Jones et al., 1999; Konold, 1989, 1995; Piaget & Inhelder, 1975; Polaki, 2002; J. F. Wagner, 2006). These studies have become great resources for understanding students' thinking about chance and have served as good complements to the studies using the first approach.

Findings from research on thinking about chance have been varied and sometimes contradictory. Some studies, for example, have found that thinking about chance improves with age (Piaget & Inhelder, 1975; Ross & DeGroot, 1982), whereas others have found that age has little influence (Batanero & Serrano, 1999). Some studies have shown that instruction has a positive impact on students' growth in thinking about chance (Jones et al., 1999; Polaki, 2002), and others have shown that inappropriate reasoning about chance is widespread and difficult to overcome, despite good instruction (Garfield & Ben-Zvi, in press).

The main message from research is that ideas of probability and statistics are very difficult for students to learn and that such ideas generally conflict with students' primary intuitions about data and chance (Garfield & Ahlgren, 1988; Shaughnessy, 1992). Overall, the literature that describes students' difficulties in reasoning about chance is much more abundant than the literature that treats how to overcome such difficulties (Garfield & Ahlgren). There are many references to students' misconceptions about chance but limited references to pedagogical insights into the teaching of this complex topic. The present study, therefore, may be useful in making connections between students' difficulties and what teachers might do to help students overcome those difficulties.

Shaughnessy (1992) suggested that research on teachers' knowledge of probability was needed, but there has been only a limited response to his suggestion. Stohl (2005) and Garfield and Ben-Zvi (in press) list the studies done in this area, but those studies have been focused on content knowledge and not on pedagogical content knowledge. They have not explored, for example, what tools and instructional strategies teachers use to help students transform their primary intuitions into formal thinking. Additionally, this line of research has centered most of its attention on preservice teachers. Thus, there is a need to explore what practicing teachers do with their knowledge of chance to assist their students in transforming intuitive knowledge into normative knowledge.[3]

The Expert-Novice Paradigm

> "Experts and novices apparently perceive information differently, remember information differently, and use different criteria to judge the utility of the information that they perceive and remember." (Carter et al., 1987, p. 148)

Researchers from different fields have been interested in how experts and novices process and use information for making decisions. A number of consistent conclusions have been found

[3] Intuitive knowledge refers to understanding constructed from the senses that is sometimes immediate and without reasoning; it is the foundation upon which other genuine knowledge must be established. In contrast, normative knowledge is usually associated with the knowledge about theories, beliefs, and prepositions that help people to describe the reality and see how things should be and how to value them, which things are good or bad, or which actions are right or wrong.

in domains of chess playing (Chase & Simon, 1973), problem solving in physics (Chi, Feltovich, & Glaser, 1981), nursing (Benner, 1984), note taking (Hidi & Klaiman, 1983), and sports among others. These studies reveal that there are qualitative differences in the thinking and actions of experts and novices. Experts and novices perceive information differently, remember information differently, and use information differently.

Results from other areas of research have inspired the field of education. Without doubt, some seminal work in expertise in teaching is attributed to Leinhardt and colleagues (Leinhardt, 1986, 1989; Leinhardt & Greeno, 1986; Leinhardt & Smith, 1985; Leinhardt, Weidman, & Hammond, 1987), and to Borko and colleagues (Borko et al., 1992; Borko & Livingston, 1989; Livingston & Borko, 1990). Many other researchers have found in these results the foundations for their work (Sternberg & Horvath, 1995; Swanson, O'Connor, & Cooney, 1990).

Research in expertise in mathematics teaching has been carried out mainly in the field of cognitive and developmental psychology by using a variety of approaches. One approach has been to study closely and in depth the performance of extraordinary teachers to identify their routines and common characteristics of their teaching (Leinhardt, 1986; Leinhardt et al., 1987). These studies have provided good examples of what the practice of expert teachers looks like. The other approach, which has been the most common tradition in the study of expertise, has been to compare the performance of expert and novice teachers to identify what the experts do differently from the novices. Within this approach two methods have been identified: one by direct observation of teachers teaching and the other by creating hypothetical situations in which teachers are asked to make decisions. Leinhardt (1989), following the former method, investigated expert and novice elementary mathematics teachers by looking at three important elements in their mathematics lessons: agendas, lesson structure, and explanations. Leinhardt found that expert teachers (1) constructed lessons characterized by fluid movements from one type of activity to another, (2) had transparent goals, (3) showed sophistication in the subject matter, (4) were flexible but coherent in the lesson structure, (5) demonstrated rich connections between topics, (6) had effective management of time and students' engagement, and (7) had rich examples and representations. These characteristics, in contrast, were weakly present in the novice teachers' lessons.

Borko and colleagues' studies (Borko et al., 1992; Borko & Livingston, 1989; Livingston & Borko, 1990) also followed the methodology of observing teachers. Their results revealed that novice teachers exhibited limited pedagogical content knowledge about students' learning. Novice teachers' knowledge structures for mathematics and for the teaching of mathematics were insufficiently developed, insufficiently interconnected, and insufficiently accessible to enable flexible, responsive teaching.

Other researchers have used a different methodology: using simulated situations to explore teachers' thinking and decision making. Carter et al. (1987), for example, interviewed expert, novice, and postulant[4] teachers based on a hypothetical task where teachers have to take over another teacher's classroom. Information for planning that included students' information cards, a grade book, homework, tests, and teachers' notes were offered to teachers, and then they were questioned about their plans for instruction. Results showed that expert teachers considered as relevant information that for novice and postulants was not important. Expert teachers brought rich schemata to the interpretation of phenomena, and such schemata appeared to provide them

[4] The researchers call *postulant* those professionals who were interested in teaching but had not had any training in pedagogy. Postulants were different from novices because they had experience in their specific fields such as chemistry, engineering, or computer technology but not in teaching.

with a framework for meaningfully interpreting information. More recently, Sternberg and Horvath (1995) used the prototypical performance of experts in a variety of domains to construct a prototype of the expert teacher. They concluded that there are three main characteristics in which experts differed from novice teachers: knowledge, efficiency, and insight.

The existing research on the expert-novice paradigm offers interesting contrasts. Researchers have compared the characteristics of expert and novice teachers near the extremes of the continuum of expertise. However, there is not a clear understanding of the development of expertise. As a result, research needs to focus on individuals in the continuum of expertise, not necessarily at the extremes or at a point in time, to be able to study the development of expertise.

As indicated in chapter 1 (p. 9), much of the research done using the expert-novice paradigm has considered student teachers as novices and their cooperating teachers as experts. From the field of supervision, however, we know that student teachers do not have full responsibility for the classroom, do not have full autonomy in the classroom, and cannot make many decisions without consulting with their cooperating teachers. Many of the norms they follow in the classroom have been previously established by the cooperating teachers. Student teachers have not fully completed their preparation; they still have the status of students. Student teachers and beginning practicing teachers have different realities. So, better candidates to be considered novice teachers would be those practicing teachers in their first years of teaching as it was done in this study. Another problematic aspect in the tradition of expert-novice research is that those cooperating teachers have been usually experienced teachers but not necessarily experts. In this study, I combined three criteria for expertise to select the expert teacher.

CHAPTER 3

METHOD

I selected a qualitative research method to address the research questions for this study because it seemed to be the most reasonable approach to explore expertise in teaching. Interviews and observations of teachers in an AP Statistics learning community were the main sources of data. In addition, I collected the AP Statistics learning community meeting agendas, laboratories, slides, and other artifacts. Interpretivism as a theoretical perspective and a grounded theory approach were used to generate descriptions of teachers' understanding of students' conceptions of chance.

Two teachers were interviewed and observed in the setting of an AP Statistics learning community where they meet with other high school teachers interested in AP Statistics. First, I visited a meeting of the learning community, and after the first visit I asked all members to participate. All of them agreed to be observed during the meetings. Additionally, I asked two teachers for extra participation that would involve in-depth interviews. These teachers were selected for two reasons. First, they fit the well-defined characteristics of expert and novice I was looking for, and second, they volunteered to participate and to communicate their thoughts about their understanding of students' conceptions about chance. I determined the expertise of the participants by the numbers of years they had taught AP Statistics, the percent of their students passing the College Board AP Statistics exam, and their reputation based on references by key informants from the community.

AP Statistics has been in the high school curriculum for only about 10 years, so I considered more than 5 years of teaching AP Statistics as a criterion of expertise. AP Statistics students are required to take the College Board examination at the end of the year if they are to receive AP credit. I considered 75% or more of one's students passing the examination in the previous year to be a second criterion for expertise. The test is graded in a scale of 5 and obtaining a score of 3 or higher is considered a passing score. Additionally, I asked a university faculty member and some AP Statistics teachers in the community for names of AP Statistics teachers they considered experts. Those names that were repeated from different sources were considered potential AP Statistics expert teachers. An expert teacher had to meet all the criteria. In contrast, I considered teachers in their first or second year of teaching AP Statistics as novices. The criterion of the students passing the examination was taken into account, but not all the participants could be evaluated with this criterion because many of them had not ever had students taking the examination. Out of the eight teachers, only one met the criteria for being considered expert, and three met the criteria for novices. I asked all four to participate and two agreed—one expert and one novice.

The AP Statistics Learning Community

The AP Statistics learning community was a volunteer gathering supported by a partnership project designed to improve science and mathematics education in a Southeastern state. The project developed four partnerships across the state to work collaboratively to increase achievement in mathematics and science in grades from preschool to 16, each partnership consisting of a university and several partner schools. Funded by a National Science Foundation grant over 5 years, the project sought to increase achievement by providing challenging science and mathematics curricula; increasing and sustaining the number, quality, and diversity of P–12 science and mathematics teachers; and increasing the responsiveness of higher education to the needs of schools. Activities of the project included the formation of P—16 learning communities, the support of postsecondary instructional innovations in undergraduate science and mathematics classes, and the coordination of individualized professional development for P—12 science and mathematics teachers.

For this study, the focus was on the work of a learning community associated with one of the four partnerships. This partnership coordinated several learning communities at the elementary, middle, and high school levels. The learning communities had different formats: Some supported teachers within the schools, and others supported teachers in specific subject matter across the schools. This partnership had 7 learning communities at the high school level. Five learning communities supported teachers in mathematics and science within the school, and 2 supported teachers in a specific subject matter: AP Calculus and AP Statistics. These communities brought together high school teachers and higher education faculty who collaborated to develop and share professional knowledge in science and mathematics. The main goal of each learning community was to improve students' achievement in science and mathematics. To accomplish the goal, the learning communities promoted primarily: (1) the replication of teaching practices that have a positive impact on students' learning in science and mathematics; (2) collaboration between P—12 and higher education faculty; (3) making the work of learning community participants public by sharing effective and authenticated practices across schools, districts, regions, the state, and nation; (4) results-oriented work that led to improved student achievement in the teaching and learning of science and mathematics; and (5) collaborative inquiry that allowed reflection on teaching, methods, materials, and assessment for an effective implementation of plans for improvement.

Although all the learning communities in the partnership project were intended to contribute to the same general goal, the specificity of the subject matter of the AP Statistics learning community generated additional goals. In the first meeting, at the beginning of the school year, the team made a list of the teachers' needs and the aspects of their teaching practice for which they needed support. That list guided the learning community meetings throughout the year. Specific goals from the AP Statistics learning community were as follows: (1) create a pacing guide for the year; (2) develop a deep knowledge of statistics content, concepts, and pedagogy; (3) collect resources such as activities, projects, and assessments; (4) explore the use of technology in teaching; (5) practice scoring student work using College Board standards; (6) develop review strategies for the College Board examination; (7) determine course grades that are appropriate for AP students; (8) discover enrichment activities beyond the standard College Board curriculum; (9) share successful lessons; (10) share the learning community experience with other AP Statistics teachers; and (11) develop strategies to motivate students.

The AP Statistics learning community was started by Mrs. Sanders,[5] one of the project staff members associated with a university who coordinated the partnership. She advertised the learning community throughout the school districts in the project. She called each school to get the name and email address of each AP Statistics teacher and sent each teacher an individual invitation to join the community. Fourteen teachers were invited to participate, 4 never attended, and 10 became members. By the time data collection for this study began, the learning community was in its second year and had 8 members. Although Mrs. Sanders had not taught AP Statistics, she acted as a facilitator at the meetings and made sure the goals and needs expressed by the members were being met. The AP Statistics learning community received additional

[5] All the participants' names in this report are pseudonyms.

support from a university faculty member in statistics with considerable experience with AP Statistics and statistics education.

The meetings were held in a neutral place (a local chamber of commerce) close to every member's home where they did not feel they were in a work environment. The meetings took place once a month during the school year for about 2 hours each.

My Position in the Study

I knew through one of my professors about the AP Statistics learning community, and I asked Mrs. Sanders, the coordinator, for an interview and permission to visit the learning community and see what they did there. She invited me to the first meeting of the year, and I was an observer. My role changed a little bit at the second meeting because Mrs. Sanders asked me to participate by sharing some of the activities we did in a college statistics class in which I was the teaching assistant. I agreed to share the activities, and my role changed from observer to participant observer.

Participants

The participants in this study were 2 teachers in the AP Statistics learning community. The participants in the learning community were 8 teachers from 7 different high schools in a Southeastern state, and they were expected to be teaching AP Statistics. Two members, however, were teaching not statistics but algebra and geometry, and one was teaching statistics online (and not in a regular classroom). Only five members were teaching AP Statistics. Those teaching algebra and geometry had joined the AP Statistics learning community because they planned to teach AP Statistics in the near future. The participants had diverse experience in teaching AP Statistics, ranging from 0 to 8 years, and diverse experience teaching mathematics as described in Table 1.

Table 1
AP Statistics Learning Community Members' Professional Backgrounds

Teacher	No. years teaching	No. years teaching AP Statistics	Highest degree attained
Eric – The Expert	17	8	Specialist Math. Ed.
Natasha – The Novice	4	1	Master's Math. Ed.
Mrs. Sanders	22	0	Specialist Math. Ed.
Brandon	2	0	Master's Math. Ed.
Henry	2	0	B.S Math. Ed.
Hanna	10	5	Master's Math. Ed.
Melinda	25	1	Specialist Math. Ed.
Trevor	6	2	Master's of Business
Ophelia	23	3	Master's Math. Ed.

It is important to mention that this learning community was particularly small because the population of AP Statistics teachers was also small. Most high schools in the area had at most one section of AP Statistics per year. The AP Statistics learning community was a voluntary gathering, which suggests that the teachers who participated in the meetings were concerned about improving the teaching and learning of statistics. A brief description of the participants, Eric and Natasha, is provided next.

Eric, the Expert Teacher

Eric was the expert teacher in this study. He was a high school teacher with 17 years of experience; he had taught AP Statistics for 8 years. For 7 consecutive years, his students had taken the AP Statistics examination, and 80% of his students passed it in the year previous to this study. He was recommended by a university professor of statistics and by Mrs. Sanders and was well known in the community of AP Statistics teachers as an experienced teacher. In the AP Statistics learning community, he determined the course of the activities through his suggestions and questioning. He had bachelor's, master's, and specialist degrees in mathematics education. At the time of the study, he had National Board Certification, had been the mathematics department chair at his high school for 3 years, and had worked for the College Board at a Summer Institute on AP Statistics for 4 years. The summer following the year of the study would be his first time serving as a reader for the AP Statistics examination. His interest was in the school mathematics curriculum, and he hoped to become a mathematics curriculum director.

Natasha, the Novice Teacher

Natasha was a high school teacher who had taught algebra in middle school for 4 years and then had taught AP Statistics for 1 year. She had started teaching AP Statistics because when she moved to high school that was the only class she was offered, and she accepted the challenge. The year previous to the study had been her first year of teaching AP Statistics and the first year that she had students take the AP Statistics examination. Around 29% of her students had passed the exam (6 out of 21 students). She had a bachelor's and a master's degree in mathematics education, and her interest was in teaching statistics from an activities-based point of view. Before starting teaching AP Statistics, she attended a week-long workshop from the College Board. This training session was recommended for her school district but it was not mandatory.

Observations

I observed 5 learning community meetings from September 2007 to January 2008. I videotaped 4 of the meetings; Table 2 contains a calendar for the data collection. The first meeting was not videotaped because it was the meeting in which the members of the AP Statistics learning community were asked to participate in the study; I, however, took field notes that I expanded later. I used a video camera in the back of the room for recording the meetings. I took field notes that I expanded on as soon as possible. I also collected artifacts such as agendas from the meetings, laboratories, slides, hand-outs and sheets of paper that revealed the teachers' work. I used the artifacts and the videotapes as aids for expanding the field notes.

The observations provided a useful source to understand what teachers said in the interviews even if in the learning community meetings they did not speak directly about students' difficulties with the core chance ideas of the study: sample space, randomness, independence, and the law of large numbers. Observations might not have provided direct answers to the research questions, but watching teachers talk about their experience in teaching AP Statistics helped me understand their perceptions about teaching statistics. The video captured the complexities of the discussions among the teachers and allowed for in-depth analyses of the complex interactions that occurred in the meetings. Videotaping produced a rich source of data that allowed me to validate findings from other sources of data and was crucial for the retrospective analysis that occurred when observations had ceased.

Table 2
Dates of Data Collection

			Teacher interview date	
Year	Month	Observation date	Eric	Natasha
2007	September	17		
	October	15		24
	November	5	14, 27	7, 28
	December	10	7	
2008	January	14		

Interview Instrument

To explore the teachers' understanding of students' reasoning about chance, I developed an instrument consisting of 12 episodes for the interviews. An episode consisted of two parts: a task and either a hypothetical student's incorrect reasoning or specific results from research that used the task. All the tasks had been previously used in research. Episodes 1, 2, and 3 explored the idea of sample space. In these episodes, the majority of the students described failed to make a precise estimation of the probability because they did not list the entire sample space. Episodes 4, 5, and 6 explored the idea of randomness (and variation) in different scenarios: sampling, surveys, and random sequences. Even though the episodes were selected to investigate the idea of randomness, the idea of independence also came up. Results from studies where those items were used reveal that some students were misled by the form of the questions asked and that many students had a poor understanding of randomness. Episodes 7, 8, and 9 explored the law of large numbers. Previous studies where these items were used indicate that it is a hard idea to understand (Kahneman & Tversky, 1972; Tversky & Kahneman, 1974; J. F. Wagner, 2006). Even if a student is familiar with the law in a specific context, say coin tossing, presenting the idea in a different context might lead the student to a wrong answer. Finally, Episodes 10, 11, and 12 explored the idea of independence. Studies where these items were used showed that students' responses were highly influenced by previous outcomes of the same experiment.

The instrument was not designed to require the teachers to work on the tasks to express their thinking about it. In fact, that aspect was not relevant. I was interested in how the teachers anticipated students' difficulties and what kinds of strategies the teachers might use to help students to readdress their thinking about chance. From time to time, however, the teachers worked on some tasks and only then expressed their perceptions.

Interviews

I interviewed Eric and Natasha in depth during the fall of 2007 (see Table 2). Interviews were held at the teacher's school, usually at the end of the last period. This aspect could be a concern because teachers are usually tired after teaching all day, which may influence what they say in an interview. However, after the last period was when the teachers had the most flexibility in their schedules.

The two teachers were interviewed three times for about one hour each. There was at least one week between interviews to allow me enough time for transcription and reflection. The teachers were asked general questions about their teaching background, school, and curriculum. Additionally, the teachers were asked for their interpretations of students' difficulties in the episodes that concerned the four core ideas. Each core idea was studied using three episodes. The rationale for using three episodes for each idea was that more information could be obtained about a phenomenon if it were explored by using different settings. Sometimes if the wording in one episode was not clear, a teacher might feel stuck and not say much. But because the idea was explored using different scenarios, there were several opportunities to get an accurate sense of the teacher's thoughts.

In each interview, the teacher was presented with four episodes (three exploring the same core idea and one exploring a different one). The first interview explored the three episodes related to sample space; the second, the episodes on randomness; and the third, the episodes on independence. The law of large numbers was explored across the interviews with one episode at the end of each one. Each task of the episode was shown to the teachers, who were then asked for their anticipation of students' approaches to solving the problems. Once the teachers had offered their opinions and explanations about students' approaches, either a student's incorrect reasoning or results from research were shown to the teachers. The teachers were asked again for their explanations of students' difficulties and for the strategies they would use to address students' difficulties. The interviews were audiorecorded and transcribed as soon as possible so that I could ask follow-up questions and seek clarifications of the teachers' comments in a subsequent interview.

Interpretativism

The theoretical perspective for this study was interpretivism. Interpretativism emerges in contradistinction to positivism, which proposes that the social sciences should follow the method of research of natural sciences. Crotty (2004) describes interpretivism as an attempt to understand and explain human and social reality. It has branched into three historically important divisions: hermeneutics, phenomenology, and symbolic interactionism. The present study was based on a phenomenological and a symbolic interactionism approach. A phenomenological approach is connected to a philosophical movement dedicated to describing the structures of experience as they present themselves to consciousness. Phenomenology is the study of the structures of consciousness that enable people to refer to objects outside of themselves. As a result, the structures give meaning to the abstract content of acts of the mind such as remembering, desiring, and perceiving. Thus, phenomenology may be defined as the study of how meanings are built up in the course of experience and as a study of a subject's everyday experiences described from the subject's point of view and as the subject understands them. Hence, "phenomenology births in the effort to identify, understand, describe and maintain the subjective experiences of the respondents. … Phenomenological research … emerges as an exploration, via personal experiences, of prevailing cultural understanding" (p. 83).

From a phenomenological perspective, the researcher tries to gain access to individuals' life worlds, that is, each person's world of experience, where consciousness exists. Conducting interviews is a common method for gaining access to an individual's life worlds. The objective of interviewing is to reveal a participant's standpoint regarding either the experience or the phenomenon of study. A phenomenological theoretical perspective enables researchers to examine human experience in detailed ways.

A symbolic interactionism approach considers human beings as purposive agents that engage in self-reflective behavior. "They confront a world that they must interpret in order to act rather than a set of environmental stimuli to which they are forces to respond" (Schwandt, 1994, p. 124). Symbolic interactionism requires that the researcher "actively enter the worlds of people being studied in order to see the situation as it is seen by the actor" (p. 124). Conducting observations is a common method to enter to participants' world. In this study, the participant observations were a good tool for understanding the participants' interpretations of the phenomenon studied. Observations add rigor when combined with other methods and are valuable as an alternative source of data for enhancing cross-checking. Interviews are limited sources of data because the participants can report only what they perceive. The participants' perceptions are distorted because of personal bias, and data from interviews can easily be affected by the interviewees' emotions. In this study, the observations provided a check on what was reported in the interviews, and the interviews provided an internal view of what was externally observed.

These approaches were appropriate for the present study because the focus of attention was on the perceptions and experiences of the participants: what individuals said they believe, the feelings they expressed, and explanations they gave. The goal was the reconstruction of the experience as it had occurred in a natural setting and the exploration of how the teachers "naturally" reacted to certain students' responses without any deliberate intervention to alter the teachers' thinking.

Grounded Theory

A grounded theory approach guided the analysis of the data. This approach looks systematically at the qualitative data to generate theory and to explain *why* the world is the way it is. It is a simplification of the world that helps clarify and explain how it works. It assumes that the processes of data collection, coding, analysis, and theorizing are simultaneous, interactive, and progressive. The theory is grounded in the data gathered, but the researcher adds his or her own insight into why those experiences exist and in that way creates the theory (Maxwell, 2005).

Grounded theory refers not to any particular theory but to the theory that is inductively developed during a study and in constant interaction with the data gathered (Glaser & Strauss, 1967). Inductive analysis consists of a series of categories, codes, and coding that are used to develop theories by meticulously inspecting data to disclose communalities and ideas. In this study, I invested many hours in coding the data, categorizing those codes, and reflecting on their implications. This process was important for identifying concepts and locating supportive excerpts from the data. Using inductive methods allowed me to identify concepts, describe them, and compare them across participants.

Data Analysis

The data were analyzed using grounded theory (Glaser & Strauss, 1967) and interpretativism (Crotty, 2004; Schwandt, 1994). The grounded theory and interpretativism are both constant comparative methods of analysis in which the analysis is often explained as a process of organizing data, breaking data into manageable units through codes, synthesizing the data through categories, and finding patterns and regularities among all data collected to generate theory. This process occurs both during and after data collection. Consistent with these descriptions, the data analysis in this study was not a step in a linear process but was instead embedded in all aspects of the study, from data collection to writing. Analysis was based on (a) transcribing and reading transcripts during the data collection to shape subsequent data gathering in the direction of the study, (b) expanding field notes and reading field notes to shape the subsequent data gathering, (c) coding using a constant comparative method of inductive analyses, (d) watching and rewatching videotapes of the learning community to validate or invalidate findings, and (e) writing.

I transcribed the interviews as soon as possible after they occurred. The process of transcribing helped me reflect on and identify specific aspects that needed clarification. Once I had gathered data from both teachers related to the same episodes, I made matrices to identify similarities and differences between the teachers' responses. These matrices were useful tools to establish hypotheses to be validated when all of the data had been collected and to identify patterns, trends, and paradoxes. I went over the recordings and the transcripts again to highlight ideas that seemed to be important. I used the constant comparative method of inductive analysis to generate categories that might help to provide an understanding of the data. This part of the process involved coding the data and comparing those codes across the teachers' responses. I named those codes generated from the data *the principles for teaching and learning ideas of chance* which are described later in this chapter. As I coded, I compared the codes across the teachers; the similarities and differences among them became apparent.

The expansion of the field notes and the watching and rewatching of the videotapes were important in the analysis because they allowed me to confirm or disconfirm the findings from the interviews. If the teachers' interactions and participation in the learning community did not offer discrepant evidence to reject the findings previously established, then findings from the interviews were considered valid. Finally, I went back over the interviews to identify the teachers' statements that supported the codes previously identified, and I quoted those statements in this report.

Evolution of the Framework

There have been several attempts to list the ideas that should be taken into account in the teaching of statistics and probability. More than a decade ago, Garfield (1995) reviewed research related to teaching and learning of statistics and proposed 10 principles of learning statistics. The principles focused on promoting learning for understanding in student-centered learning environments. The principles were as follows:

1. Students learn by constructing knowledge.
2. Students learn by active involvement in learning activities.
3. Students learn to do well what they practice doing.
4. It is easy to underestimate the difficulty students have in understanding basic concepts of probability and statistics.
5. It is easy to overestimate how well students understand basic concepts.
6. Learning is enhanced by having students become aware of and confront their misconceptions.
7. Calculators and computers should be used to help students visualize and explore data, not just to follow algorithms to predetermined ends.
8. Students learn better if they receive consistent and useful feedback on their performance.

9. Students learn to value what they know will be assessed.

10. Use of the suggested methods of teaching will not ensure that all students will learn the material.

These principles are general aspects of the teaching and learning of statistics. Their generality limited my initial idea of using them as a framework for the analysis. In this study, I was interested not in the entire AP Statistics curriculum but in how the four core ideas (sample space, randomness, independence, and law of large numbers) were integrated in the curriculum. Additionally, the principles are not exhaustive, and some of them overlap. For example, Garfield (1995) explained that the third principle, *students learn to do well what they practice doing,* involved those actions that teachers design to engage students in learning. According to her, such practice involves hands-on activities, activities using cooperative small groups, and activities that require working on computers. However, the principle *students learn by active involvement in learning activities* also involves students working in groups. In her classification, therefore, there is overlap among the principles, including overlap among Principles 2, 3, and 7. Because Garfield's principles were general and non-exhaustive, I decided not to use them in the analysis. The data were explained to some degree by the principles, but some aspects were not explained. Garfield's principles were, however, very helpful in illuminating a possible direction for the analysis. I realized that if I wanted to use her principles, I would have to make some modifications.

In my effort to revise Garfield's principles, I looked at Shaughnessy's (2003) recommendations for teaching probability. These recommendations were inspired by the results of the 1996 National Assessment of Educational Progress (NAEP), in which students performed poorly on tasks related to chance and probability. Shaughnessy's recommendations to teach probability were the following:

1. Begin to teach probability at a young age, and continue throughout the school years.
2. Emphasize the value and importance of building the sample space for a probability experiment.
3. Make connections between probability and statistics. In particular, connect the notion of sample space in probability with the concept of variation in statistics.
4. Introduce probability *through* data; start with statistics to get to probability.
5. Adopt a problem-solving approach to probability. Give students opportunities to investigate probability problems or chance situations on their own and to conduct their own stochastic projects.

Note that Shaughnessy's recommendations emphasize the connections between probability and statistics. Except for the second recommendation about emphasizing the listing of the sample space, however, his recommendations are as general as the principles suggested by Garfield. These recommendations also failed to help me analyze my data completely.

Finally, I looked at a more recent work, the GAISE recommendations for teaching an introductory statistics course (included AP Statistics) (Franklin & Garfield, 2006). This set of recommendations was developed by the college group of the GAISE project, and the intention was to help students attain the goals for the introductory statistics course. The recommendations were as follows:

1. Emphasize statistical literacy and develop statistical thinking.
2. Use real data.
3. Stress conceptual understanding rather than mere knowledge of procedures.

4. Foster active learning in the classroom.

5. Use technology for developing concepts and analyzing data.

6. Use assessment to improve and evaluate students' learning.

These recommendations have been generally accepted as foundational for teaching an introductory college statistics course (included AP Statistics). There are several similarities between the GAISE recommendations and the principles for learning statistics suggested by Garfield (1995). As with Garfield's principles, however, these recommendations address general features of an introductory statistics course and did not necessarily help me analyze data that focused specifically on the four core ideas. Although some recommendations were relevant, others were far from the scope of this study. After examining these three frameworks, I decided to develop a set of principles that originated from the data and that were more helpful in the analysis for this study. There were categories in the data that were not explained by any of the three frameworks. For example, Eric brought up the need to discuss in class those misuses of chance found in advertisements and news articles. This category was not considered in any of the frameworks. The GAISE project recommends the use of real data, which remotely resembles Eric's idea of bringing misuses of ideas of chance into the classroom. However, the main concern of the GAISE recommendation is related to the authenticity of the data so as to engage students in explorations in contexts that allow them to ask and answer meaningful questions. It does not directly address Eric's concern. Consequently, the data suggested the inclusion of the principle *point out common misuses of chance*. In the next section, I described the six principles for teaching and learning ideas of chance that I used as the framework for the analysis.

Principles for Teaching and Learning Ideas of Chance

1. Students learn by doing.

This principle states that students are not able to master concepts, think critically, communicate ideas, or make arguments when they are not encouraged to do so repeatedly in different contexts. In this study, I looked at instances in which the teachers suggested some type of practice in their descriptions to help students in their thinking about chance. Some examples of activities that incorporate this principle are simulating with physical counters, simulating with technology, drawing diagrams, and making graphical representations.

2. Take into account students' prior knowledge.

This principle states that students come into the classroom with a lot of prior knowledge that teachers should take into account when designing learning environments. In this study, I was interested in seeing whether the teachers took into account students' prior ideas about chance when the teachers designed their interventions.

3. Use cooperative learning.

This principle implies that students working together have a fine opportunity to discuss their reasoning and ideas with others, which allows them to become more involved in their learning than when working alone. In this study, I was interested in whether the teachers proposed some sort of teamwork or small-group discussions to encourage cooperative learning.

4. Students learn by confronting misconceptions.

Learning is enhanced by having students acknowledge and confront their own misconceptions. When students are encouraged to make guesses and predictions, they are more likely to care about the results. Discrepancies between predictions and actual results are useful in checking students' understanding as well as addressing their misconceptions about chance. In this study, I was interested in those moments in which the teachers suggested ways for students to confront their misconceptions.

5. Use technology to visualize and explore.

This principle states that calculators and computers should be used to help students visualize and explore data and not merely to carry out algorithms. The use of computers in instruction can help students to see the data represented in different ways, which can improve students' understanding of abstract ideas about chance. In this study, I paid attention to those occurrences in which the teachers took advantage of the available technology to suggest explorations that might help students in visualizing and understanding ideas, properties, concepts, and theories.

6. Point out common misuses of chance.

This principle implies that students learn when they are exposed to situations in which they have to confront the validity of information from news stories and advertisements. When students are confronted with these kinds of situations, they are likely to be more critical next time they find conclusions based on chance in the mass media. Also, this exposure provides a fine opportunity to acquaint students with data from real-world situations. In this study, I paid special attention to those instances in which the teachers suggested bringing examples from the mass media in which chance information was misused to be discussed and criticized in the classroom.

CHAPTER 4

ANALYSIS

This chapter is devoted to results concerning the three research questions: (1) How do an expert and a novice teacher of AP Statistics anticipate students' difficulties in chance situations? (2) How do an expert and a novice teacher of AP Statistics explain students' difficulties in chance situations? (3) What strategies do an expert and a novice teacher of AP Statistics claim to use to assist students in the reorganization of their intuitive thinking into more formal thinking about chance? I present a description of each episode, how Natasha and Eric anticipated the students' difficulties, how they explained those difficulties, and what pedagogical strategies they suggested for dealing with students' difficulties. The discussion is presented according to the four core chance ideas: sample space, randomness, independence, and the law of large numbers.

Sample Space

Difficulties with listing the whole sample space in an experiment can be attributed to a weakness in combinatorial reasoning that later affects the estimation of probability. The sample space idea was explored with three episodes: Bags and Marbles, Tossing a Coin, and Spinners. Researchers report that when students are given these tasks, they may have problems calculating the probability because they do not list the entire sample space.

These tasks have the property that students can be misled by the data in the problem. For example, the fact that the coin has a 50% chance of landing heads or that the spinner has a 50% chance of landing on black could mislead students; they might transfer the "fifty-fifty" property of each single event to the whole sample space. In situations like these, Tversky and Kahneman

(1983) identified psychological difficulties that can confound students' reasoning. Shaughnessy (2003) added that the difficulties are more than just psychological. They are associated with the students' lack of the mathematical skills needed to analyze or list the outcomes for the problem and the kind of training students are exposed to. Thus, Shaughnessy suggested that "students need more experience listing the set of all possible outcomes in probability experiments" (p. 222).

Episode 1 – Bags and Marbles

Consider the hypothetical game: There are two bags with red and blue marbles in equal proportions. Player B holds the bags. Player A, with closed eyes, picks a marble from the first bag and a marble from the second bag. The marbles are compared and returned to their respective bags. If the marbles are of the same color, Player B wins one dollar; otherwise, Player A wins one dollar. Is the game fair?

Student reasoning: Player B has a better chance of winning the game because there are three outcomes: Either both marbles are blue, both are red, or the colors are different. Each outcome has a probability of 1/3, and since two outcomes favor Player B, the probability that Player B will win is 2/3. (Taken with modifications from C. H. Wagner, 1981)

In discussing students' possible approaches to this task before I showed her the student's reasoning, Natasha mentioned that "students might start by drawing a picture." Immediately she started drawing a picture of the bags and the marbles. She drew a diagram of the bags first. Then, when she started to draw the marbles, she realized that the task did not state how many marbles there were. The task stated only that the marbles were in equal proportions. To deal with that problem, she decided to simplify the task by assigning one red and two blue marbles to the first bag, and two red and four blue marbles to the second bag, as shown in Figure 1. Then she listed the sample space. Natasha predicted that students could start solving the task by listing all the possibilities or by simulating the game, trying out the game with bags and marbles or some other material. It is interesting that Natasha's predictions about students' approaches resembled the strategies she had used to solve the task.

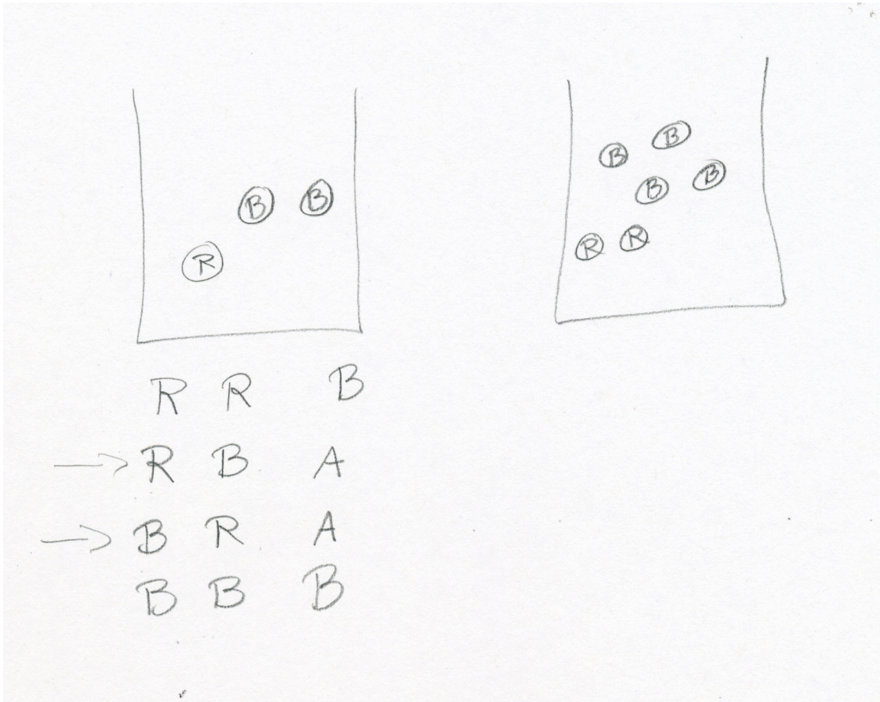

Figure 1. Natasha's solution to the Bags and Marbles Task.

It should be noted that the wording of the task is ambiguous. The expression "there are two bags with blue and red marbles in equal proportions" allows different interpretations. One might conclude that both bags have the same number of blue marbles as red marbles, which was probably the intended interpretation, but one might also conclude, for example, that the first bag has x marbles, with half red and half blue, and the second bag has $2x$ marbles, with half red and half blue. Natasha interpreted the task in a way that the fist bag has the same proportion of marbles as the second bag, but the proportion of red and blue marbles within each bag was different. Her interpretation meant, however, that the game was not fair anymore. In her version, the probability of drawing the same colored marbles from both bags is greater than the probability of drawing different colored marbles, as illustrated in Figure 2.

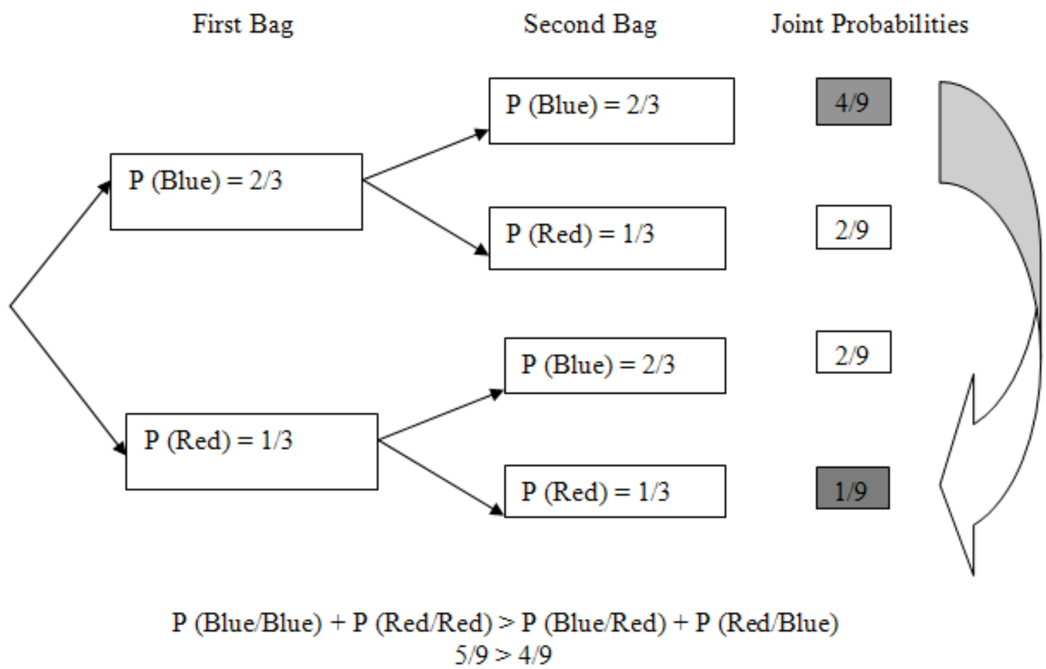

Figure 2. Probabilities for the Bags and Marbles Task as interpreted by Natasha.

In discussing the anticipation of students' difficulties, Natasha said, "Sometimes students have a hard time distinguishing between these two outcomes. Getting a red [ball] first and a blue second, and then a blue first and a red second." To expand on her explanation of this difficulty, she related the Bags and Marbles Task to the classical problem in elementary probability of finding a three and a four in rolling two dice. There are two different ways to get a three and a four, but students do not see it at first glance. Natasha was very confident in making this prediction; she was familiar with the difficulty in other contexts and was able to transfer that difficulty to this task.

Later, when Natasha saw the particular student's reasoning in which the student miscalculated the probability, she said that the student's error might be that she or he did not consider all the possibilities for the sample space. She stated the difficulty clearly, but she did not offer further explanation of the reasons for it.

Natasha said that if she had a student struggling as this particular student was, she would suggest playing the game to find the experimental probability and compare that with the theoretical probability. She said:

> I would actually have them play the game, and then find the experimental probability. Once they have found that by doing maybe a hundred trials, … [they could] compare … to see if the experimental probability actually matches theoretically what they're claiming. (Natasha interview; October 24, 2007)

Natasha also suggested having students write out the sample space for all the outcomes, but this suggestion was made only after she saw the student's incorrect reasoning in which some elements of the sample space were omitted. Initially, she had stated that the listing of the sample space would be a student's first approach.

For the same episode, Bags and Marbles, Eric read the task and anticipated that students would have different difficulties according to their level of thinking. He said that one of the main difficulties for students, however, would come from the lack of a systematic approach to listing the sample space:

> I have different level students, so some of them would actually want the bags and the marbles and to try it, and see how it works. Other students would say, "Well, how many [marbles]?" … "Tell me about the marbles, …there's red ones and blue ones right? …They're in equal proportions." So, some of them would start with the factoring tree, a tree diagram. Others would wanna do it empirically. ... I would say most of the mistakes would come from trying to do [the listing of the sample space] without some kind of systematic approach whether [it] be a tree diagram, or contingency table, or something along those lines. (Eric interview; November 14, 2007)

An interesting aspect of Eric's reflections is that he suggested that students would have different approaches according to their level of thinking. When Eric saw the students' incorrect reasoning, it only confirmed what he had predicted. He said:

> When you talk about in this problem not knowing the number of total outcomes, that causes more probability errors than just about anything else. They forget [to] count one, or … sometimes they double count a success—those types of things. I say it's pretty frequent. (Eric interview; November 14, 2007)

Eric was very explicit about the fact that listing the entire sample space was crucial for the correct estimation of the probability; he mentioned other factors, however, that might be associated with this difficulty. He stated that the difficulty students have with number facts, fractions, multiplication, and proportions might contribute to their difficulties in reasoning about chance. He said another factor might be that students struggle a great deal with word problems because an immense amount of content (English prose) in the task "turns them off." When the language of the task is abstract, students tend to have difficulties in interpretation. Eric said that the expression "There are two bags with red and blue marbles in equal proportions" could be difficult for students to understand because the wording seems abstract, and students usually prefer something more concrete: "It says that they are equal proportions, but that is not enough for them, that is too abstract." An additional explanation of students' difficulties that Eric pointed out was the students' deterministic view of the world; some just want to find the right answer. He said that this view is very difficult to change because in AP Statistics students have to deal with investigation tasks, and they often lack the patience to do it. They want an answer. Sometimes that feeling of wanting the right answer and not finding it generates apathy, and students start asking questions like, "When am I ever going to use this?"

Eric also said that the fact that the student did not have all the outcomes listed might have contributed to the difficulty, but he liked the idea that the student attempted to list the possibilities. Eric mentioned that although the student was wrong in his calculation, he demonstrated interesting reasoning. Eric said, "Well, you know, this is a quick approach. I like the fact that he is actually [wondering] 'What are the possible outcomes?' That is a good thing for kids to get used to doing."

Eric mentioned different pedagogical strategies that he could use in instruction to help students overcome this difficulty. First, he suggested doing a simulation with physical material—acting out the game but adding shading to the properties of the marbles. That is, a light shade for the marbles in the first bag and a dark shade for the marbles in the second bag. By using the shading, students would understand that two marbles of the same color could happen in two different ways as in rolling dice; it is possible to get a five and a three in two different ways. Second, he suggested enumerating the entire sample space because if students do not know how many different elements they are looking for before they get started, they might miss one or two. To help students with listing the sample space, Eric mentioned that it would be helpful to use the fundamental counting principle, the multiplication principle, or a tree diagram. Third, he suggested doing simulation with other physical counters such as coins and dice. Fourth, Eric mentioned that this task gives a perfect scenario to introduce students to the concept of expected value to calculate the theoretical probability and to compare that with the experimental probability. However, he was aware that in the AP Statistic curriculum the concept of expected value is at the end of the probability unit.

Episode 2 – Tossing a Coin

A coin is tossed four times. What is the probability of getting two heads?

Student reasoning: Since a single coin has ½ chance of landing heads, in four tosses we would expect to get two heads. So the probability of getting two heads is ½.

In discussing this task, Natasha expressed pretty much the same ideas as she did in the Bags and Marbles Task. First, she worked out the problem; she made a tree diagram and partially listed the sample space. Then she predicted the students' approaches using the following description:

> Well, they could start by listing the possibilities that could happen when they flip the coin. They're tossing it four times. They could possibly get all four heads, and they would … keep on listing all of the outcomes by making a systematic list. (Natasha interview; October 24, 2007)

It is notable that Natasha did not consider that students could have difficulties in situations in which they have to make estimates based on an exhaustive list of the sample space. She was very confident that students were going to list all the possibilities in their first encounter with the situation. Students usually do not do that, however, because they do not regularly apply a systematic strategy to list all the possible outcomes (Shaughnessy, 2003).

I showed Natasha a student's incorrect reasoning in which the student concluded that the probability of getting two heads in tossing a coin four times was half. When Natasha saw the student's incorrect reasoning, she was able to recognize and talk about some of the students' difficulties more clearly. She said:

> He [the student] is just thinking about the coin having a fifty-fifty chance [to land either heads or tails], and he is not thinking about all of the different arrangements, all of the different ways that you can get two heads in the outcomes. (Natasha interview, October 24, 2007)

In this excerpt, it is clear that Natasha considered not making the list of the possible outcomes as a potential difficulty in the student's reasoning. She was much more precise about students' difficulties after she saw the second part of the episode, the results from research in which a student struggled calculating the probability. Her explanations were better developed when she was exposed to the student's reasoning. This aspect is revealing because it shows that sometimes the perception of students' difficulties is not immediate for teachers that do not have a broad repertoire of examples. With some help or hints, however, teachers can detect where students make common mistakes and could take advantage of that knowledge to design interventions in which they make sure that at least some students will make such mistakes.

The strategies Natasha suggested to help students with their difficulties in listing the sample space were similar to the strategies described in the Bags and Marbles Episode. First, she suggested doing 50 times the simulation of flipping the coin 4 times; recording the results of the simulation and calculating how many times students get two heads. Second, she suggested making an organized list of the outcomes by using tree diagrams.

The discussion with Eric of the Tossing a Coin Episode was richer. Immediately after reading the task, he noted that it might have several answers if it does not specify that there are "exactly two heads." Eric's predictions of students' difficulties in this task were very similar to the predictions he expressed in the Bags and Marbles Task. He identified that the listing of the sample space was fundamental for an accurate calculation of the probability. An additional feature Eric noted was that students' performance would depend on the students' year in school. The AP Statistics course is generally offered to students in their junior and senior years (sometimes sophomore year), and Eric predicted differential performance in students across years. He said that young students (sophomore) would be more inclined to solve the problem from a concrete point of view (e.g., tossing the coins) while the older ones (senior) would probably do it by using some theoretical rules (e.g., multiplication rule).

When he was shown the student's specific reasoning, he confirmed his predictions and was very interested in figuring out the student's reasoning. He said:

> I like the first part, "Half a chance of landing heads." So okay, he is just multiplying the probability times the number of tosses … so we get two heads. So the probability of getting two heads is a half. Well … it's really interesting, … coming up with the half is a pretty good idea. (Eric interview, November 14, 2007)

This description reveals that Eric was interested in understanding the student's way of reasoning. He also mentioned that the fact that the student had an argument could be taken as a starting point to confront students with their own thinking. In his effort to understand the student's reasoning, he listed the 16 different possibilities in the sample space, an activity in which he spent some time. This was the only task Eric worked out completely. Note that in the Bags and Marbles Episode, he paid also close attention to students' reasoning and noted that students would have different approaches according to their level of thinking. This characteristic reveals that Eric was concerned about the students' way of thinking.

Eric described several pedagogical strategies to help students in formalizing their reasoning about chance. Some were similar to the strategies he had suggested in the previous episode, but some had new features. He suggested, first, asking students for some conjectures and having them write down their thinking. Second, have students do the experiment of flipping the coin many times and combine the class's results to get an empirical distribution. Then, have students find the probability of getting two heads from the empirical distribution they generated. Third, have students write down the entire sample space. Fourth, have students compare their initial predictions with the empirical distribution and ask those who were close to explain their thinking. Then compare the empirical distribution with the theoretical.

Eric's suggestions, however, were not as isolated as I described them above. The following excerpt portrays better how he integrated these actions:

> We talk a lot about … empirical versus theoretical. So, I would have … each kid do this [toss a coin] maybe ten times, and it takes twenty-four kids and putting together, so we have two hundred and forty [data points]. And we would look then at the empirical probability, and we'd say, "Well, that's counter to what he [the student] is thinking. …" We'd ask for some conjectures, … this is a good idea to get started. …Have people write down what they're thinking. … [Then,] take out the coins and … count the number of times we get two [heads], and then see where that probability is compared to what some of the people were thinking. And then, if somebody were right, then I would ask them, "So could you explain how you were thinking about this kind of problem?" And then from there we would go to enumerating the sample space, counting the successes, and going back again, reinforcing the counting principle. (Eric interview; November 14, 2007)

Note that one big difference between Natasha and Eric was in the way they thought about students' thinking. For Eric, students' thinking was the starting point for his interventions. But Natasha did not make any reference to students' thinking.

Episode 3 – Spinners

A student claims that if he spins two spinners simultaneously (as in Figure 3), there is a 50% chance that both of them will end up on black. Do you agree or disagree?

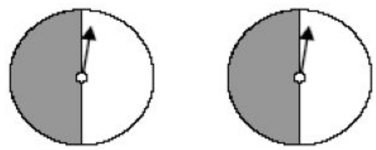

Figure 3. Two spinners.

Results: Researchers analyzed the responses of more than 1000 12th-grade students in the 1996 NAEP, and only 8% disagreed with the statement and correctly reasoned that the chance of both spinners landing on black was only ¼. (Taken from Zawojewski and Shaughnessy, 2000).

Natasha read the task and started listing the sample space and calculating the probability using the multiplication rule. Then she predicted, as in the two previous episodes, that students would start by listing all the possible outcomes to be able to calculate the probability. That, however, is not what students do. Zawojewski and Shaughnessy (2000) found that students had difficulty in this task because they were not exhaustive in listing the sample space.

It is interesting that Natasha made the same prediction in the three tasks. Later, she mentioned that if students have difficulty in this task it would be because they would transfer the idea of each spinner being half black to the set of two spinners. She said: "Since it is half black and half white, many students might perceive that there was a fifty percent chance that both of them would end up on black. They would agree with this problem."

Natasha read the results from research, and then she noted that this problem was similar to the Tossing a Coin Task. She immediately stated that the students' difficulty was because they did not list all the possible outcomes in the sample space. This time she was more accurate at stating students' difficulties, but she did not offer additional reasons to explain them.

The strategies Natasha described to help students in the formalization of their thinking were similar to the strategies expressed in discussing the previous tasks. First, have students do the experiment with the spinners several times to get the experimental probability. Second, have students list the entire sample space. The following gives her description:

> We can simulate it; we have spinners that we use in our classroom. So, they could do an experiment where they actually do, maybe, fifty trials and find the experimental probability that both land on black. And then, maybe they would begin to see that maybe something else is going on here in the problem. …You can either go into the multiplication rule … or listing the outcomes in the sample space to help them understand what it is one fourth and not fifty percent. (Natasha interview; October 24, 2007)

Later, Natasha described the connections between the experimental and the theoretical probability:

> When you are doing simulation a large number of times, your experimental probability would get closer and closer to theoretically what it should be. And that is one of the basic concepts that I teach at the beginning of the probability unit. So that when we do problems like this, that is something that they should already understand. And when they end up getting their experimental probability through the simulation, they should realize that that should be close to theoretically what it should be. And when it's not close to their fifty percent chance, then they know that there is something more to this problem than they had originally thought about. (Natasha interview; October 24, 2007)

Although Natasha was explicit in making the connections between the experimental and theoretical probability, she underestimated the difficulty that students might have with the concept. She said that she teaches that relationship at the beginning of the probability unit to prevent students from having difficulties in situations that require contrasting results from experiments with theory. The fact that she introduced the concept previously does not ensure that the students would recall it without help. In fact, there are many examples from research where students seem to be familiar with a concept (say, the law of large numbers), but they fail to apply it in a real world situation (e.g., J. F. Wagner, 2006).

At the end of this interview, Natasha told me that she had assessed her students using this task in a test she gave the previous year. I inquired about the students' performance on the test, and she said that in general the students did well. I asked her if I could take a look at those test papers, and she agreed. We reviewed the test papers one by one and found that many students had missed that question because they agreed with the statement. Natasha was surprised by the number of students that missed the question. We did not count exactly how many students there were, but I would estimate that about one third missed the question. Figure 4 shows an example of one of Natasha's students failing to calculate the probability because he did not list the entire sample space; the student listed only WW, WB, and BB. Natasha added the option BW when she was grading the test.

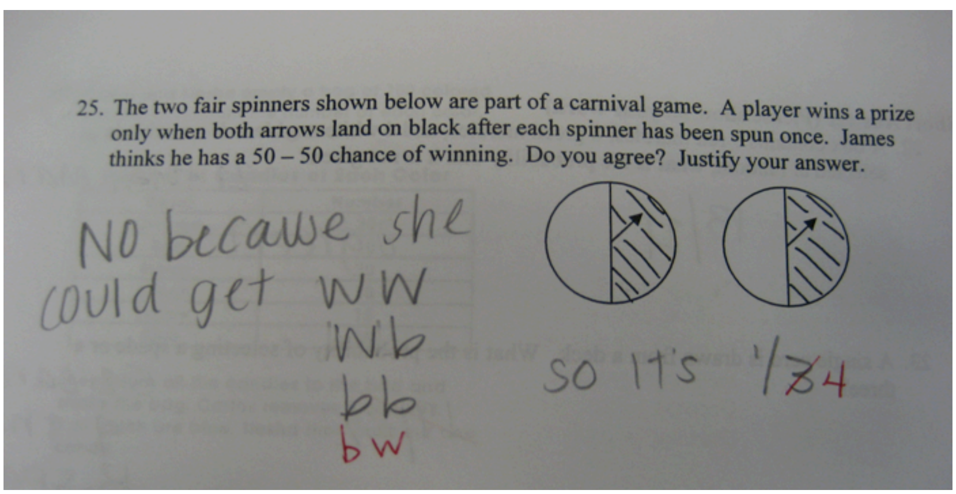

Figure 4. Natasha's student failing to list the complete sample space.

The fascinating part of going over the tests with Natasha was the interest she showed in doing that. She was very engaged in reviewing each student's written explanations on the test, and she realized that she had underestimated the difficulty of the task and had not analyzed the students' performance task by task. She had looked exclusively at their overall performance. In other words, Natasha was developing her pedagogical content knowledge by reviewing her students' work.

The Spinners Task was discussed with Eric. As soon as he read the task, he said that he would have to assume that the spinners have to be independent to be able to apply the multiplication rule. He recognized that this task was very similar to the Tossing a Coin Task and that students might have difficulties comparable to those expressed in the previous tasks.

Although Eric successfully identified the students' difficulties in this task, he was surprised to see the results, in which only a small percentage (8%) of students correctly disagreed with the statement. He said, "We are not doing a very good job with teaching probability." He looked at the task more carefully and stated that besides the student's difficulties other factors might have influenced the results. He said that the first AP Statistics examination took place after the 1996 NAEP, and he hoped that if the same question were asked today, the percentage of students getting a right answer would be higher. In doing so, Eric revealed that he was aware of the students' typical thinking as well as the more global issues in the United States. Eric explained that some additional reasons for students' difficulties might be that students are not trained in statistics and that there is not a lot of probability in the high school curriculum unless students take statistics. He also said that not long ago, around 2000, when he got the curriculum guide and the pacing guide for his mathematics classes (Algebra I and Algebra II), the recommendations were "skip the statistics part." He said that with those recommendations he did not find support in the high school curriculum for students' development of ideas related with chance.

Some pedagogical strategies that Eric described were similar to those described in the two previous tasks: Simulate with the spinners enough times to construct an empirical distribution, and estimate the empirical probability. Some different strategies Eric suggested were the introduction of technology for simulations and the use of the empirical distributions generated throughout the simulations to derive the rules like the multiplication rule. He said:

> You could even do it [simulation] in your calculator. …We randomly select two binary digits [in two different columns], two binary integers [0 and 1], and sum them [the digits in the columns]. And if the sum was two, you win. So, you could get ten thousand repetitions in five minutes in your classroom. And then put them all together and have a distribution. And then you can see that ... twenty-five hundred are wins. …So, now it's a rule. There is something that we can prove and then go from the empirical and derive the formula ourselves. (Eric interview; November 14, 2007)

When expressing the strategies, he related them to the state performance standards' expectations, which were going to start the following year. He mentioned that the new standards suggest engaging students in experimentation before introducing them to the formulas.

An important characteristic of Natasha's reflections throughout the three episodes was that she worked out the tasks before giving her opinions. One might infer several things. First, these tasks were not familiar to her, and she therefore needed to work them out to be able to express an opinion about students' thinking. However, there is disconfirming evidence from the Spinner Task. She was familiar with the task because she had used it to assess her students in the previous year. She worked out the task even though she was familiar with it. Natasha might have thought that identifying her own struggles by working out the tasks might help her to have an indication of students' possible difficulties. Second, her approach to these tasks was based on concrete representations. She made diagrams and suggested simulations with concrete material for the three tasks. If the language was abstract, she reduced the task to a simpler one. I was not interested in how the teachers worked a problem, but I did not stop them from doing it, and I gave them the time they needed. Eric, in contrast, did not work out a problem to be able to predict the students' difficulties. He worked out only the Tossing a Coin Task, and he did that in the middle of the discussion, right after seeing the student's reasoning and probably in an effort to understand the student's thinking. I think that Eric was probably familiar with the tasks.

Natasha did not consider students' previous ideas in designing the strategies. Her strategies seemed appropriate but not necessarily connected to the difficulties students showed. Eric, in contrast, paid close attention to students' reasoning, and he showed an interest in trying to understand the causes of students' mistakes. The data reveal that he took into account the students' mistakes to build upon his strategies and that he was interested in confronting students with their own thinking.

Eric brought into the discussion several factors that might contribute to students' difficulties. He stated that students' sense of fractions, multiplication, and proportions, the abstraction in the language, their deterministic view of the world, the lack of probability in the curriculum, and other global issues in the United States could also contribute to students' difficulties. Natasha, in contrast, did not mention any of these factors to explain students' difficulties.

The simulations recommended by the two teachers were different. Natasha was very interested in the use of physical material, from coins to other physical counters. Doing the simulation using dice instead of coins, for example, does not change the essence of the simulation. Eric, however, besides proposing physical material for simulations, suggested including shading the counters to help students understand that, in listing the sample space, the order matters (see the Bags and Marbles Episode). He also recommended doing simulations using technology and deriving the theoretical rules and properties from the empirical data.

Randomness

The understanding of randomness is central to the study of statistics, and a consideration of randomness is basic to understanding chance variation. It is common, however, to find students having difficulties understanding this foundational concept. According to Falk (cited in Sahaunessy, 2003), people tend to infer randomness when it is not really present and are not reliable at recognizing random outcomes or generating them. In this study, the idea of randomness was investigated using three episodes: Gumball Machine, MovieWorld, and Fake Sequences. The episodes explored the ideas of randomness in different settings: random selection, sampling methods, and random sequences. In these tasks, students were asked to predict the results of a random selection in a mixture of colored gumballs, design an appropriate technique of sampling to estimate the responses to a raffle, decide the accuracy of different sampling techniques, and determine if a given sequence was randomly generated or made up. I present the discussion task by task.

Episode 4 – Gumball Machine

A gumball machine has 100 gumballs: 20 are yellow, 30 are blue, and 50 are red. The gumballs are well mixed inside the machine. Jenny gets 10 gumballs from this machine. What is your best prediction of the number that will be red? (Taken from the 1996 NAEP, Zawojewski & Shaughnessy, 2000).

Results: Only 7% of the student from a sample of 232 students gave an extended response, and 14% a satisfactory response. The majority of the students give exact answers like 5, and only one gave a range for the answer.

We started by discussing the difficulties students might encounter in solving the task. Natasha indicated that students might have difficulty if they assume equal probability in the outcomes. She said, “I think that a difficulty that students might have with this is that they read that there are three colors: yellow, blue, and red. And they might assume that each one has one third chance of being selected.” This was not the prediction I was expecting from this question, because I was exploring the idea of randomness. That is a reasonable prediction, however, and a common difficulty in students' reasoning about chance. In the 1996 NAEP study, the researchers reported that they found the equiprobability bias in students' answers. They reported that "students believe that well-mixed populations should produce uniformly distributed samples of color, regardless of the percents of colored balls in the machine" (Zawojewski & Shaughnessy, 2000, p. 262). The researchers, however, did not report how often they found the equiprobability bias in the students' reasoning.

Natasha also predicted that students could solve the task by drawing pictures. I think that she mentioned the picture approach without being aware that the gumball machine had 100 gum balls. It was not clear what type of picture she was referring to, but in any case, it would have been time consuming to draw. This task did not need to be worked out to be able to come up with a prediction; Natasha, however, made a sketch (Figure 5) to reflect her prediction. It was interesting that her prediction was similar to the students' prediction in the 1996 NAEP. She did not consider a range of possible values.

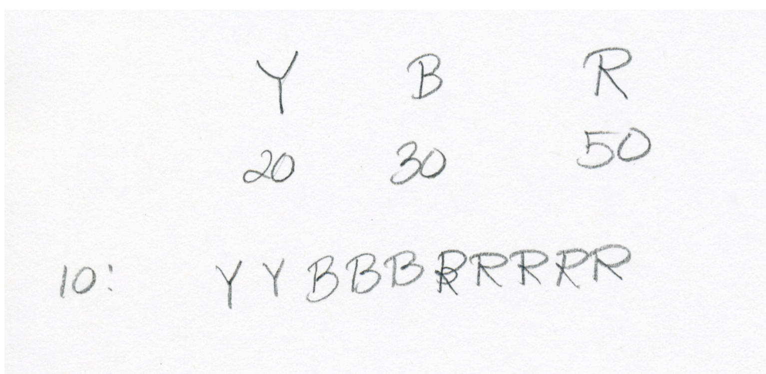

Figure 5. Natasha's sketch for the Gumball Machine Task.

Natasha mentioned that a good approach could be to reduce the task to 10 gumballs. I was concerned about that suggestion. Sampling with 10 gumballs increases the variability in the number of read gumballs and violates the law of large numbers (discussed later in this chapter).

Her goal might have been to simplify the situation for students' better understanding, but her suggestion indicated that she did not consider the implications of simplifying the task.

When Natasha saw the second part of the episode in which the students' difficulties were described, she said: "Okay, I see. They just wrote down the number instead of explaining what could possibility happen in that situation." Her statement suggests that she focused exclusively on the fact that students did not give explanations, but she ignored the aspect that almost all the students gave exact answers and did not consider a range of possible answers. Because Natasha ignored the second part of the results, her explanations were oriented mainly to justifying why students do not give extensive explanations. To explain this phenomenon, she mentioned that the tradition in mathematics classes is that students are not asked to explain their thinking. They are asked to solve for *x* or to find exact answers, but they are not trained to explain their thinking. She also mentioned that a hard part was having students write down what they think; they can explain it to you but not write it down. In her reflections there was no reference to reasons for the students' tendency to give exact answers, which is not surprising, given that she did not give a range of possible answers.

Natasha suggested several pedagogical strategies to help students with this task. First, she suggested having students reread the problem so that they would understand that there are different numbers of yellow, blue, and red gumballs. This strategy was oriented to helping overcome the equiprobability bias she had predicted earlier, which suggested that she proposed strategies according to the difficulties she was able to predict. Second, she suggested reducing the task to a simpler task. These two strategies were her own—the strategies she used to work out the task. Third, she mentioned drawing a picture, but for this task that strategy was unrealistic, given that there were 100 elements. Fourth, she suggested doing the simulation with physical counters (replace the gumballs by little colored blocks), taking samples of ten blocks, and recording the number of reds.

Note that these strategies did not specifically address the two problems revealed in the results from research. They did not address the students' difficulties in explaining their thinking or their tendency to give exact answers. Although Natasha predicted the equiprobability bias, I wanted her to talk about the fact that students preferred exact answers, but I was not successful.

In his discussion of the Gumball Machine Task, Eric mentioned that the idea of independence could be conflicting here and that the task was not specific about that aspect. Another difficulty he mentioned concern the students' poor understanding of probability and their poor strategies for problem solving. He mentioned that most of the time students just want an equation so they can get an answer, and when they have to do an application they fail because they do not have strategies for problem solving. His prediction was associated with the students' deterministic view of the world, where students favor exact answers and chance variation is not comfortable for them. Note that Eric also mentioned this difficulty in the discussion of the idea of sample space. He did not predict the problem with giving an exact answer instead of a range.

After Eric saw the results from the research, he was surprised by the low percentage of students that offered an extended answer but said that students' thinking is not the only factor contributing to their difficulties. He said that there were other factors outside of students' thinking that might increase their difficulties. He said that the NAEP is not a mandatory examination, which might have affected the results. He also said that the way the question was asked might have encouraged students to give an exact answer, but if students had been asked for an interval, the results would have been different. This suggestion is consistent with the recommendations made by the researchers who explored students' performance on this task.

They found that the wording of the item may have contributed to the poor performance; "likely range" and "likely interval" would have been more appropriate wording than asking for a numerical prediction (Zawojewski & Shaughnessy, 2000).

Eric also said that the mathematics curriculum promoted the notion of expected value but left the notion of confidence intervals exclusively to statistics. He said that under those conditions, the mathematics curriculum was not contributing to overcoming students' difficulties and that it was not strange that students were not familiar with these ideas unless they had had some exposure to statistics.

Eric suggested several pedagogical strategies. Some were similar to those he suggested in previous tasks, but there were some new features. First, he suggested doing a physical simulation, but he knew that the simulation would take time. Second, he said that a much better strategy than a physical simulation would be a simulation using the TI-83 and exploring different quantities. He described it:

> Assign digits to outcomes, and then randomly select numbers on their calculator. And then … they get a sample of ten, and then they would count the number of red ones. And then do that again and again and again and again. And so, they can get some empirical … data collection. And then …get a pretty close estimation [of the experimental probability]…. Once you've established this idea, …you can change all these numbers around, or get different values, and … see that the same strategy still works regardless of the numbers. …That is what we always do. (Eric interview; November 27, 2007)

Third, Eric suggested having the students explain their reasoning. He said, "You have to explain your reasoning, or … tell me about what you were thinking about, or something like that, because one answer of course isn't enough in our class."

The main message from the use of this episode is that it was not useful to explore the idea of randomness with Natasha. I was prepared for such trouble, and for that reason, I prepared three episodes for each idea. In case I had difficulties with one episode, I would have two others to look at. With Eric, the situation was different. I was able to obtain a more detailed picture, although he did not predict the students' difficulty of considering an exact answer instead of a range of possible answers. He predicted some other difficulties; he was able to talk about that specific students' difficulty only when he saw the results where students' struggled to give a range of possible values. The fact that neither of the teachers identified the students' tendency to give exact answers suggested several things. It might suggests, for example, that such a difficulty was more complicated to identify than I had initially thought, or that the item has serious difficulties in its wording, as suggested by Zawojewski and Shaughnessy (2000).

Eric's strategies to help students in their reasoning were richer than Natasha's. His strategies included using simulations with technology, exploring the empirical distributions with different quantities, and listening to students' reasoning. Natasha's strategies did not address the idea of randomness directly, which might be because she did not identify that idea in the discussion.

Episode 5 – MovieWorld

A class wanted to raise money for their school trip to MovieWorld on the Gold Coast. They could raise money by selling raffle tickets for a Nintendo Game system. But before they decided to have a raffle, they wanted to estimate how many students in their whole school would buy a ticket. So, they decided to do a survey to find out first. The school has 600 students in Grades 1–6 with 100 students in each grade. How many students would you survey, and how would you choose them? Why?

Suppose that five students in the school conducted surveys.

a) Shannon got the names of all 600 children in the school and put them in a hat, and then pulled out 60 of them. What do you think of Shannon's survey?
b) Jake asked 10 children at an after-school meeting of the computer games club. What do you think of Jake's survey?
c) Adam asked all of the 100 children in Grade 1. What do you think of Adam's survey?
d) Raffi surveyed 60 of his friends. What do you think of Raffi's survey?
e) Claire set up a booth outside of the snack shop. Anyone who wanted to stop and fill out a survey could. She stopped collecting surveys when she got 60 kids to complete them. What do you think of Claire's survey?

Who do you think has the best survey method? Why?

Results: The majority of the students in Grades 3, 5, 7, and 9 gave answers like "Ask everyone." Only 7% of the students suggested appropriate randomization, and only 22% evaluated [chose] the preferred technique of Shannon's survey (task taken from Watson, Kelly, Callingham, & Shaughnessy, 2003)

In Natasha's reflections about this episode, I observed some interesting aspects. The first part of the task asked students to propose a technique of sampling. In discussing the first part, Natasha said that students might select a stratified sampling: "They would wanna get variety in their sample, so maybe they would choose an equal number of students from each of the grade levels to include in their survey." The second part asked students to choose the best sampling technique from among a list of five techniques. Natasha observed that students would be able to notice the errors in some of the sampling techniques described in the situation. She stated:

> I think that my students would quickly pick up on [the fact] that several of these ideas have error in them, something that would lead to bias and maybe skewed results. For example, in (b), when Jake asks students at the computer games club, where they're gonna be more willing to buy a ticket for Nintendo than other students in the school most likely, so that would weight heavily in favor of buying tickets. Same thing with Adam: First graders are gonna have a different opinion than sixth graders. (Natasha interview; November 7, 2007)

Natasha had a nice conjecture in describing students' approaches to solving the problem but ignored the possibility that students might have difficulties. Her predictions were very far from the results obtained when these tasks were used in research, where only a small percentage of the students proposed either a randomized or a stratified randomized technique. It is clear that Natasha underestimated the difficulties that students might have with sampling.

When Natasha saw the results of studies where these tasks had been used, and realized that only 7% of the students suggested appropriate randomization, she was more open to considering students' difficulties. She explained:

> They think, the ones that gave answers like "Ask everyone," they think that everybody's opinion is important. And so they don't wanna leave any one single person out of their survey. So, they think that they should ask everyone. They feel [that] if they're not asking

> everyone, then it's not fair. So, they are not realizing that you can still get a representative sample using only a portion of your population and not making that inference connection. (Natasha interview; November 7, 2007)

One of Natasha's attempts to explain students' difficulties was related to the students' lack of experience. She stated that students might have those difficulties because they had not been exposed to similar situations:

> Because they feel like everybody's opinions are different, and maybe those are the experiences they've had in school. …They're used to just being within a small classroom. So, if the teacher wants to know what the class's favorite color is, it's easy to ask everybody in the whole class and compile the results. So, they just haven't had experience, I don't think, in comparing sample results to population results and seeing the similarity between the two. (Natasha interview; November 7, 2007)

It is interesting the way Natasha described the motivations to suggest data collection that happens in the classroom. The expression "if teachers want to know what the class's favorite color is" could suggest that in her vision, teachers' motivations come before students' motivations. Confirming evidence for this statement was found in one of the learning community sessions (December 10, 2007). The teachers in the learning community were concerned with pacing, especially because they had different schedules in their high schools. Most of the teachers were using the book *The Practice of Statistics* (Yates, Moore, & Starnes, 2003). They were reviewing the pacing guide, and Natasha realized she had spent 23 days on a chapter for which the pacing guide had suggested 12 days. For the same chapter, Eric spent 13 sessions. Natasha said, "I was looking at this pacing guide here, and I realized I spent way too much time on chapter three. That is maybe because I like chapter three too much." The fact that Natasha referred to herself as the reason for extending the time spent on the chapter made me think that students were not her priority in teaching.

It is worth noting that Natasha expressed different opinions about students' difficulties before and after knowing the results of research. In her first encounter with the first part of the

MovieWorld episode, she almost ignored the possibility that students might have difficulties. In the second encounter, she was more open to considering that possibility. One might conjecture that, at some point, showing teachers the results from research might help them think about students' approaches that they had not previously considered and might also favor teachers' reflection on students' thinking. Another conjecture is that Natasha was not familiar enough with the tasks to allow her to identify students' difficulties, but it also might reflect her inexperience as teacher.

Natasha recommended two pedagogical strategies to help students with their difficulties in sampling. First, she suggested showing students different sampling techniques and talking about the advantages and disadvantages of each technique. She said:

> I show them different scenarios … for how to select the sample. For example, … ask the people in their bus … whether they would buy a ticket for the … game system. …Ask every fourth person in the cafeteria lunch line if they would buy a ticket. …Make an announcement on the morning news and ask for volunteers to come out to the office to take part in …[the] survey. … Use a random method where they have a ten-sided number cube and … role it three times to come out with a three-digit ID number … [to] select the students in … [the] sample. ... So, I have them look at those four different situations. And I have them list the advantages and disadvantages they see in each situation, and we talk about those. From there, I go on to give them a definition for the meaning of the sampling method that goes with each situation. And we talk about how bias could be a part of each one of these situations, and which one do you think is gonna give us the most representative sample. (Natasha interview; November 7, 2007)

Second, Natasha suggested asking students to come up with their own techniques of sampling and comparing them with those described in the task.

Natasha's expression "I go on to give them a definition for the meaning of the sampling method" might reflect her perception of teachers as providers of information. However, it could simply mean that she was reporting what she had done in the past and not necessarily thinking about the scenario at hand. To contrast this aspect with Eric's view, recall the Spinners Task. He suggested having students do the simulations to get empirical distributions for that task. Later he would use that resource to derive the multiplication rule. Consequently, Eric perceived his role of teacher as a facilitator.

The discussion with Eric about the MovieWorld Task was a little different. In responding to the first part of the episode, he was able to anticipate some students' difficulties. He suspected that some students might not recall randomness in their first attempt to suggest a sampling method. He said:

> Well, they wanna do it. … It seems silly, but whatever grade they are in, that is what they want to, that is what they want to survey first or the most. Seniors don't care what freshman think. So, … if we wanna survey some kids to get an idea, then they would probably just pick the people around them. "What do you think? Who in class would buy one?" … I know my students would choose their friends before they even begin to think of random sampling … because that is the way they think the world revolves around. (Eric interview; November 27, 2007)

It is clear that Eric anticipated the difficulties students would have with the idea of randomness. He knew that randomness is not the first idea students take into account when they are asked to design a sampling technique. In his attempt to explain the reasons for students' difficulties, he mentioned, as he had in previous tasks, the students' conception of the world as an invariable entity where chance variation is barely taken into account.

Later, when Eric saw the results of previous studies in which students failed to provide a sampling technique, he agreed with the results and explained that students might not have a clear understanding of what sampling really means:

> Maybe they are thinking they have to have a right answer. … You know, sampling only gives you the attitude of so many. And if you really wanna be right, you gotta ask everybody. And not thinking that, you know, it might be too costly or too time consuming. … So, just trying to get one right answer could be what they are thinking. (Eric interview; November 27, 2007)

Eric suggested some pedagogical strategies to help students deal with the idea of randomness. First, he suggested having a discussion in class in which he would ask leading questions like: Do we need boys and girls in the sample? Do we need people from every grade? What do Jake's friends do? Why are Jake's friends in that club? He said that the idea of this discussion was to help students to identify flaws in the sampling techniques. He said:

> Once they express how they feel, … they need kind of leading toward the vocabulary of the subject and starting identifying some different flaws and different sampling techniques. And of course, you always wanna steer them right into the randomization. (Eric interview; November 27, 2007)

Second, he suggested having discussions in groups. He said, "Have those prompts [different sampling techniques], and let the kids, you know, talk it over and [argue] those things out. … If they have this conversation in class, then they wouldn't just say 'Ask everybody.'"

These excerpts illustrate that Eric was able to anticipate with some accuracy the students' difficulties in thinking about the idea of randomness. Natasha, in contrast, underestimated the students' difficulties in the first encounter with the episode, but she was able to predict students' difficulties more clearly after she saw the results in which students failed to suggest random sampling. The exercise where the teachers predicted students' difficulties and then contrasted them with results from research seemed to be a great stimulus to engage the teachers in reflections about students' struggles with the ideas of randomness. It also seemed to help the teachers look deeper at students' thinking. Thus, sharing students' difficulties with novice teachers could be valuable help for them to become aware of the complexities of students' thinking.

The teachers' explanations for the difficulty with the conception of randomness in sampling were different. Natasha suggested that a reason would be the students' lack of exposure to similar situations, whereas Eric attributed students' difficulties to a variety of factors such a limited vision of a world that students see as invariable, a poor understanding of variability and a weak interpretation of sampling.

The teachers’ pedagogical strategies involved talking to students, but Natasha’s intervention seemed more directly instructional; she would explain and provide definitions for students. Although Eric’s strategies involved leading questions, they also engaged students in team work.

Episode 6 – Fake Sequences

Some children were each told to toss a coin 40 times. Some did it properly. Others just made it up. They put H for heads and T for tails.
Maria: T T T H T H H T T T H T H H H H H T H T H T T H T H H T T T T H H H T H H T H H
Daniel: H T H T T H H T H T H H T T H T T H H T T H T H H T T H T H T H T H T H T T H T
Martin: H T T T H T T H H H T H T T T T T H T H T H H T H T T H H H H T T T H T T H H H
Diana: H T T T H T T H T H T T T H T T T T H H T T T H T T H T T H T T T T H T T T H T
ltem 1: Did Maria make it up? How can you tell?
ltem 2: Did Daniel make it up? How can you tell?
ltem 3: Did Martin make it up? How can you tell?
ltem 4: Did Diana make it up? How can you tell?

Results: From a normative point of view Items 1 and 3 are correct. Items 2 and 4 were made up. Most students consider all sequences to be random except for Diana’s. (Situation taken from Batanero & Serrano, 1999)

This episode was included because people have a general tendency to reject sequences with long runs of the same result. I was interested in seeing whether the teachers were aware of how difficult this task can be for students. According to the researchers that used this situation (Batanero & Serrano, 1999), two attributes of the sequences were manipulated for the task: the proportion of heads (H) and the length of runs and, consequently, the proportion of alternations (A) (changes in the type of outcomes from head to tail or from tail to head). Table 3 shows the proportions; a value of .5 means that the proportion is close to the theoretical value. Items 1 and 3 were intended to appear random, and Items 2 and 4 were intended to appear made up. In Diana’s sequence, for example, the runs of heads were quite short and the proportion of heads was smaller than the proportion of tails.

Table 3
Proportion of Attributes in the Fake Sequences Task

Item	Child	Heads (H)	Alternations (A)
1	Maria	.53	.54
2	Daniel	.48	.74
3	Martin	.48	.51
4	Diana	.30	.54

In Natasha's discussion, she predicted that one of the main problems in students' reasoning would be an incorrect understanding of randomness. Students would think that alternation would be a property of randomness but long runs would not. She said:

> They [students] think that the one that would be the most random, in their opinion, would be the one that probably be most like heads, tails, heads, tails. ... It would mainly alternate each time with some variety, just a little bit of variety in there. So the entire outcome should have approximately half heads and half tails. I think that that would be the one that students would most likely choose, as the one that was done properly [random]; the one that was not made up. ... They think that is very unlikely to get a streak of heads. And so they think that if you get heads ten times in a row, that is very unlikely to happen. … I think that would be a difficulty students would have. (Natasha interview; November 7, 2007)

Natasha was successful in predicting some of the students' difficulties. It was interesting that she was able to integrate the two attributes (the proportion of heads and the proportion of alternations) to explain students' difficulties in recognizing the randomness in the sequences. Natasha, however, at this point had not offered any indication that she was able to recognize which sequences were probably random and which were probably made up.

When Natasha saw the results from research and realized that most students recognized Item 4 as made up but not Item 2, she worked out the task. She counted the numbers of tails and heads and paid close attention to the length of the longest runs in each sequence.

Natasha suggested two pedagogical strategies to help students to understand the situation. First, she suggested doing simulations. Have students flip a coin 40 times, write down the longest streak of heads, and share the class's results. Second, use the binomial probability formula to calculate the probability of a streak of heads (she used 12 as an example, which is a highly unlikely streak) and compare that with the experimental probability obtained from the class's results.

Eric predicted that if students did not have the notion of chance variability, they would have difficulties with this situation. He added that students would be more concerned about counting the number of heads and tails than about the long strings of the sequences. He said:

> They [students] just go through and count up the numbers of heads instead of looking at how many in a row you get. … That might be some of the reasoning that they use, … or if you look at the sequence, the largest run, either head or tails. And if they see it is too many, I think, that intuitively they would think that they had to make that [the sequence] up even though that would be [the] actual flipping. … If someone is making it [the sequence] up, you usually won't get seven or eight heads in a row. … But when you are talking about how a kid thinks about it, he might think, if they saw eight or nine heads in a row, or tails, the person has to be making it [the sequence] up because they know that it should be fifty-fifty. (Eric interview; November 27, 2007)

Eric was accurate in predicting some of the difficulties students might have in this task. He explained that if students are not shown or taught about the idea of variability, then they have difficulty recognizing it.

When Eric saw the results from research in which students recognized Item 4 as made up but not Item 2, he confirmed his predictions and said:

> Well, I didn't study all the other ones [the sequences], but if you look in four [Item 4], it is the only one that doesn't have, well, there is one sequence of four tails. But everything else is pretty short. … This [one], [Item] two, … there is no runs at all. So that has gotta be made up. (Eric interview; November 27, 2007)

Eric also explained that another reason for students' difficulties with the idea of variability is that they do not find support in the mass media. He said that students are exposed to a great deal of misinformation in TV and sport news that counters what they are taught in a statistics course. He said that sport news shows "bombard" students with wrong information that they have to reconcile with what they learn in school. Sometimes it is not easy; many times that information confuses the students. He gave the following example: A baseball star is hitting three-forty for the whole year, and he has had 12 at bats in a row without a hit. Sportscasters conclude that the baseball player is due for a hit. According to Eric, that conclusion shows an incorrect understanding of long runs, and students that listen to that information start developing incorrect ideas about chance.

Eric recommended several strategies to help students to understand the notion of variability in this task. First, he suggested a visual strategy such as printing the sequences in colors, H in blue and T in red (a variation could be using dots and dashes, or zeros and ones). Second, do a simulation using the calculator, using zeros for heads and ones for tails, record the longest strings, and make a histogram with the class's results. Third, he recommended having discussions about variability with the information they find in the mass media like, for example, misuses of probabilities in sport news. Fourth, have this same situation as a game in the classroom. Have some students make up a sequence and have others do the actual flipping of a coin (or rolling a die), then pair the students to guess which sequences are made up. Fifth, he suggested using the binomial distribution to calculate the probability of a specific string and compare that with students' results.

In this episode, both teachers were able to predict with some accuracy the students' possible difficulties. Both teachers agreed that students might look at the number of heads and tails but not at the length of the strings. When they explained the reasons for difficulties, they agreed about the possibly incorrect understanding of randomness.

Natasha's strategies were very similar to those she had suggested in previous discussions: create an empirical distribution and compare it with the theoretical. Eric, in contrast, offered more and richer strategies. He suggested visual representations, simulations with physical counters and also with technology, grouping of the class's results, team work, and discussion of misleading information from the mass media. Eric had a more elaborate and detailed repertoire of strategies to pick from than Natasha, which make me think that Natasha was still building her collection of strategies and in the process of building her pedagogical content knowledge to offer reasonable strategies to her students. She had to spend some time going over the tasks to understand them, whereas Eric was able to express his opinion as soon as he saw the tasks.

Confirming evidence that Natasha invested time in working out the tasks was observed in a learning community session (December 10) in which the teachers explored a Random Babies Activity. I brought the activity to share with the teachers in the meeting (this activity was used in the college statistics class for which I was the teaching assistant. A similar but extended version of this task was suggested by Rossman and Chance, 2003). The task said that four mothers gave birth to four babies, but the hospital staff decided to return the babies to their mothers at random. The teachers were asked to explore the number of times each mother got the right baby and to explore the behavior in the long run. This activity allowed the introduction of concepts such as sample space, random variable, and expected value. The activity had a simulation part and a theoretical part that required exhaustive listing of the sample space and calculations. All the teachers were engaged in the simulation part to combine the group's simulations, but only Natasha did the theoretical part and worked through the whole activity.

While Natasha was doing the activity, the other teachers started discussing the concepts and ideas that could be introduced to students by using this activity. The teachers mentioned aspects like introduction of descriptive variables, construction of probability distribution empirically, comparison between empirical and theoretical probabilities, reinforcing the ideas of probability, counting rules, and sample space. The teachers also mentioned that this activity could be nicely combined with simulation with on-line applets. Natasha did not express her thoughts about the concepts that would be introduced with this activity because she was concentrating on working out the task.

Independence

The notion of independence between two events intuitively means that the occurrence of one event makes it neither more nor less probable than the other event occurs. This idea, although apparently simple, is difficult to understand, and it is responsible for numerous errors in students' reasoning about chance. The idea of independence is closely related to the ideas of randomness and variability, because if one idea is present, the others are likely to be present as well. In this study, the notion of independence was explored with three episodes: H-T Sequences, Balls in Urns, and Predicting the Lottery. The tasks explored the idea of independence in sequences in which each trial does not depend on the previous one. In this section, I present the teachers' discussions task by task.

Episode 10 – H-T Sequences

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

a) H-H-H-H-H-T
b) H-T-H-T-T-H
c) H-H-H-H-T-H
d) H-T-H-T-H-T
e) All four sequences are equally likely.

Results: Many students answer (b) because it is most representative of the population. The majority of students consider that (a) is not representative of the population and that (d) is too regular. Few students choose (e).

This task was used by Kahneman and Tversky (1972) and later by Shaughnessy (1977) and Konold, Pollatsek, Well, Lohmeier, and Lipson (1993).

At first glance this task might seem essentially the same as the Fake Sequences Task discussed in the previous section. These two episodes, however, can be used for different purposes, and the reasoning students use in the tasks might differ. In the Fake Sequences Task, the number of trials is large enough to suggest the behavior of the parent population. In the H-T Sequences Task, in contrast, there are only six trials, and the respondents do not have an opportunity to see the behavior over a long run. Also, the Fake Sequences Task asks the respondents to decide whether the sequences are random. In solving the task, the respondents would likely use the notion of randomness as suggested in the wording. In the H-T Sequence Task, in contrast, the respondents are asked to predict the most likely sequence.

Researchers who have explored the H-T Sequence Task (Kahneman & Tversky, 1972; Konold et al., 1993; Shaughnessy, 1977) agree that there is a wide range of possible approaches that respondents use in their reasoning, such as the law of small numbers, the outcome approach, and the gambler's fallacy.[6] Difficulties appear, however, when the attribute of independence is not taken into account. This task does not suggest that students recall a definition of *randomness*. They might use a definition, but it is not suggested in the wording of the task. The framing or the wording of the task highly influences the type of reasoning evoked. Zawojewski and Shaughnessy (2000) reported a similar issue with the Gumball Machine Task, discussed in the previous section, where students did not consider a range of possible answers because the task did not suggest it explicitly.

[6] I discussed the law of large numbers later in this chapter but not *the outcome approach* or *the gamblers' fallacy*. The outcome approach is the heuristic that influences respondents to give exact answers to probability questions rather than to recognize what is likely to occur. A detailed discussion of the outcome approach is offered by Konold (1989) and Konold et al. (1993). The gambler's fallacy is the false belief that the probability of an event in a random experiment is based on preceding results of the same experiment. Tversky and Kahneman (1974) detail this fallacy.

Natasha's reflections revealed that she recognized some plausible difficulties that students would have with the concept of independence. She said, "They [the students] aren't thinking about how each event is independent from the other events; that each flip of the coin is independent from the next flip of the coin." After she expressed her predictions about students' possible difficulties, I showed her the results when this task was used in research. Such results say that students are inclined to pick the sequences that look most representative of the population. She observed:

> When students flip coins, they think that there should be some type of alternating scheme to their results. They tend to think that getting a string of heads or a string of tails in a row is very unlikely to occur. (Natasha interview; November 28, 2007)

These excerpts show that Natasha was able to predict some students' difficulties before the results were shown to her. Note that the reasons she gave to explain students' difficulties in this task were similar to those she gave to explain students' difficulties in the Fake Sequences Task. Natasha said that she would look at the length of the strings of heads but that students might not look at that specific aspect of the sequences. Natasha also predicted that some of the sequences from the task might mislead students:

> I think that the students would be thinking that when you flip a coin, there is a fifty percent chance of getting heads and a fifty percent chance of getting tails. So, I think that their first inclination would be … selection (d), the one that alternates heads, tails, heads, tails, heads, tails, because they see that as fifty percent for each one. I think that would be the students' first guess or inclination. (Natasha interview; November 28, 2007)

She also said that not considering that each event is independent of the previous one might cause problems in students' answers.

Natasha proposed several strategies to help students deal with their misunderstanding in this task. First, she suggested explaining concepts to students. She said, "I would talk to them about what independence means, and how when you flip the coin one time it is not gonna have any effect on what result you get when you flip it a second time." Second, she suggested doing the simulation with the coins and listing the experimental outcomes. For the simulations, she recommended a possible modification; she said that they could use different materials like dice and spinners instead of coins. Third, she proposed listing the sample space to find the theoretical probability.

Note that the strategies she suggested for this task were similar to those she suggested in previous tasks: simulation to compare with the theoretical distribution. From time to time she proposed exchanging the initial physical counters for other physical counters, but the essence of the simulation was not modified by changing the counters. The consistency in the strategies she proposed to use makes me think that her actions could be predictable. Her limited repertoire of strategies suggests that if a new situation were proposed, she would probably approach it by using simulations with physical counters and comparing the empirical distribution with the theoretical distribution.

When Eric saw the H-T Sequence Task, he said that students would be inclined to pick either Option (b) or (d). He explained that the reason is that students would try to integrate their ideas of "fairness" and the idea of "fifty-fifty" chance to select the choices that give the impression of being fair. He expressed that in the short run everything is possible, but students might not recall that idea to make the decision. Eric provided some predictions for students' reasoning, but he never mentioned the idea of independence.

When Eric read the results, he confirmed his prediction about students' tendency to select the sequences that seem more balanced. However, he did not mention the idea of independence as one of the reasons for students' errors. He related the errors to faulty conceptions of variability, fifty-fifty chance, and fairness. He also mentioned, as he did in the tasks on randomness, that the misuses of chance in sports news, advertisements and television make it difficult for students to develop the correct notions of concepts related to chance. An example of this is the generalized belief that basketball players are confident after making a couple of shots but tense after missing several shots in a row. This phenomenon is also known as the “hot hand,” and it is an illustration of the misunderstanding of chance in sports news by ignoring the high probability of long runs. The examples Eric gave were not directly related to the idea of independence because he tied this task to the notion of randomness. It is worth noticing that sports situations are not good examples to model chance situations because anxiety and other psychological issues play and important role. However, the misuses of probability theory in sports stories contribute to prolonging the students’ difficulties with ideas of chance.

Eric described several pedagogical strategies to help students overcome difficulties with this task. First, he mentioned discussing the concepts of variability, small sample, run, number of trials, and other concepts associated with chance. Second, he suggested doing simulations using graphic calculators to generate empirical distributions. Third, he recommended using the theory of combinatorics to find the theoretical probabilities and determine that each choice has the same probability. He suggested taking advantage of the theoretical distribution to discuss the symmetry of the probability distribution. Fourth, he suggested bringing misuses of probability from the mass media to be discussed in class.

To summarize the discussion of the H-T Sequence Task, both teachers were able to identify the sequences that students would be more inclined to select. However, the explanations the teachers gave for students’ difficulties were slightly different. Natasha stated explicitly that the absence of the concept of independence might cause difficulties. Eric did not make any reference to the concept of independence; he mentioned the ideas of variability and fairness as well as the strong influence of incorrect information from the sports news in the students' construction of the idea of chance.

Both teachers suggested the use of simulations, but Natasha recommended the use of physical counters, whereas Eric proposed the use of technology. Simulations with physical counters are a good starting point for engaging students in thinking about chance, but they are time consuming. There are situations in which simulations with physical counters are impractical because the experiment is complex and has a large sample space. The H-T Sequences Task is a good example: In tossing a coin six times, the sample space is large—64 different possibilities. In experiments with large sample spaces, it is difficult to get a good sense of the random phenomena if the simulations are not run a large number of times. Simulations like this could take the whole class period, and after flipping the coin 20 times students do not get much out of the activity. I am not sure whether the inclusion of simulations with physical counters in Natasha's classroom could be the reason for her difficulties in time management. As noted previously, she said in one of the learning community sessions (December 10) that she spent twice as much time on a chapter as Eric did. If that were the case, time management could also be considered a factor of expertise, which is consistent with the results from research (Leinhardt, 1989) revealing that novice teachers were less efficient in time management than expert teachers.

Episode 11 – Balls in Urns

An urn contains two white balls and two black balls. Shake the urn thoroughly, and blindly draw two balls, one after the other without replacement.

a) What is the probability that the second ball is white, given that the first ball is white?
b) What is the probability of the first ball being white, given that the second is white and the color of the first is not known?

Results: This item was given to university students. Most of them had no difficulty calculating 1/3 for (a), but 50% responded ½ for (b). Most students are not able to recognize that the second draw cannot influence the first draw. Situation suggested by Falk and Bar-Hillel (1983).

In discussing this episode, Natasha said that students could start by drawing a picture, and immediately she started drawing a picture (Figure 6). She sketched the first two urns to represent Part (a) and the second two urns to represent Part (b). However, Part (b) was more abstract than Part (a), and she realized that she was not able to solve the problem just by using a graphical representation. She added that her students always have difficulties with situations that involve conditional probabilities and complicated language. Natasha spent some time working out the first part of the task and spent even more time reading and trying to understand the second part. She said that Part (b) was more complex than Part (a). She again appeared to predict students' difficulties according to her own difficulties.

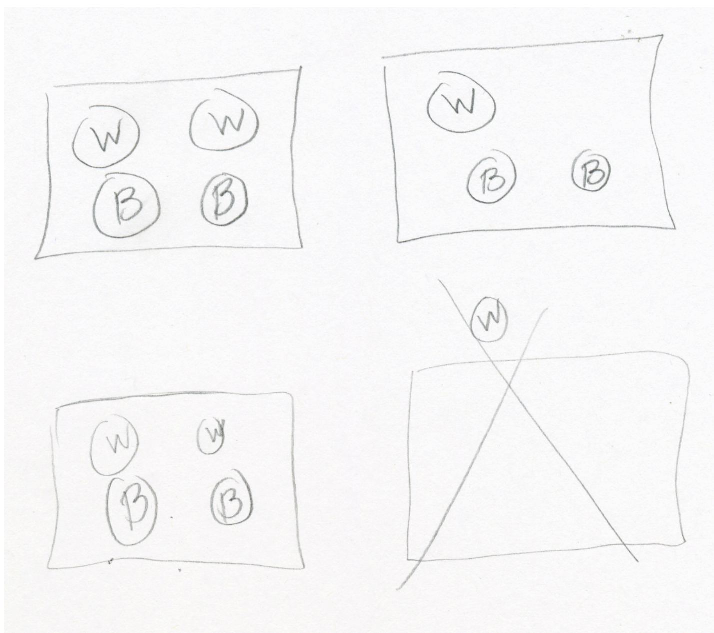

Figure 6. Natasha's picture to solve the Balls in Urns Task.

The pedagogical strategies Natasha suggested to help students with their difficulties were pretty much the same strategies she had suggested for the tasks previously discussed. She suggested doing simulations with physical counters and varying some properties of the experiment. She said:

> I would … have … different blocks or balls, and I would have students maybe take out the first ball. And then, find out how the probabilities have changed when you draw the second one without replacement. And then, I would have them record these probabilities. And then I would do it a second time, but the second time I would have them replace the ball. And I would compare the two results so they could see what happens when you replace them and what happens when you don't replace them. (Natasha interview; November 28, 2007)

In discussing the Balls in Urns Task, Eric predicted that the abstract condition of the second part of the problem would be confusing for students. He also predicted that Part (b) would be harder than Part (a) because it had an abstract extra condition. Additionally, he mentioned that if students do not have a strategy to systematically organize the outcomes of the experiment, they could have trouble. Eric had made this observation previously in the discussion of the idea of sample space, where he explained that not having a systematic strategy to list all the possible outcomes of an experiment was associated with an incorrect estimation of probability.

Eric saw the results from research and explained that students' deterministic ideas of wanting "right answers" instead of considering multiple options could prevent them from exploring the task from different points of view. He further explained that students have difficulties with the ideas of conditional probabilities and independence at the high school level because these ideas are exclusive to the statistics curriculum. He said that there is no other subject in the secondary school curriculum that includes such ideas.

I found several instances in which Eric showed signs of being reflective about the curriculum. Similar kinds of reflections were absent from Natasha's discussions. Eric also expressed his concerns about curriculum in the learning community sessions. In two meetings, when the teachers were reviewing the pacing guides, Eric mentioned the idea of skipping chapter 10 from the textbook (by Yates et al., 2003). He justified skipping it because the chapter introduces significance testing when the standard deviation of the population is known. He said, “That is an unrealistic situation” because the standard deviation of the population is rarely known; it can be estimated with the standard error but not known (December 10). Another reason he gave was that there were many topics in chapter 10 that could be integrated into other chapters (January 14). This observation suggests that reflecting about the curriculum could be a characteristic of expertise.

Eric recommended several strategies to help students with their difficulties in the Balls in Urns Task. First, he suggested having different ways to organize and visualize the possible outcomes. He talked about constructing contingency tables and tree diagrams. Second, he proposed doing a real simulation with the balls and urns; students would close their eyes for the second part of the task. Third, he suggested having a systematic way to list the whole sample space of the experiment. Fourth, he recommended having small group discussions. He said that having students work in collaborative groups might be useful for them: It prevents students from making the same errors over and over again, it helps students with their communication skills, and it helps students explore different strategies. He said:

> Have small groups discussion … would be a good way to do it. Where you put them in groups, and you give some guys this idea: “Why you don’t try the simulation?” And you guys, “Why don’t you see if you can work with the balls, and act it out, and get some kind of strategy [pattern] going there?” … And bring them [students] back together and see what everybody thinks. (Eric interview; December 7, 2007)

Fifth, he proposed doing a simulation with the calculator to get an empirical distribution. Eric’s strategies involved simulations using physical counters and using technology. Simulations with the physical counters were treated as prompts for the group discussions, whereas the simulations using technology were treated as alternatives to visualization or to give an approximate estimation if the theoretical approach was confusing for students.

To summarize the teachers' discussions about the Balls in Urns task, although this task was used to explore the idea of independence, it also brought up the idea of the sample space and variability. Both teachers agreed that Part (b) was more complex than Part (a). In Natasha's discussion, however, I could not identify whether she predicted this difficulty or inferred that her struggles with Part (b) could be translated as students' difficulties.

In Eric's discussion, he did not even attempt to work out the task; he offered his predictions as soon as he finished reading. This aspect of having Natasha working out the situations was very common across the episodes, and it was evident also in the learning community meetings. In contrast, Eric usually read the task, predicted students' difficulties, and figured out pedagogical strategies without working the task. Additional instances from the learning community sections helped to support this conclusion. In one of the meetings (December 10), Mrs. Sanders (the coordinator) brought in some statistical problems for the team to review. Eric was familiar with most of the tasks and made comments on the solutions and on the concepts that could be introduced with the tasks. Natasha, in contrast, had to work out each problem and listened attentively to Eric's comments. Being able to anticipate a solution or possible students' difficulties without working out the tasks could be another characteristic of expertise.

Natasha did not provide reasons to explain students' difficulties with the Balls in Urns Task. Eric, in contrast, offered a variety of reasons, such as students' tendency to prefer exact answers, an absence of ideas of probability in the secondary curriculum, and a lack of a systematic approach to listing the sample space.

Natasha's pedagogical strategies to help students with this task were similar to the strategies she had offered in previous discussions. She suggested doing simulation with physical counters. Eric's suggested pedagogical strategies were richer. He suggested visual representations, simulations with physical counters, simulations using calculators and group discussions.

It is interesting that Natasha did not consider group work in any of her suggested strategies, whereas Eric did. There are many ways to interpret this difference. Beginning teachers are more hesitant to have students work in collaborative groups than experienced teachers are, probably because of the classroom management issue (Raymond, 1997). Eric had been in the classroom 11 years longer than Natasha, and she was in only her second year of teaching at the high school level. A few years in the classroom might not have been enough for Natasha to explore different ways of engaging students in their learning and to discover that many issues in students' intuitive reasoning could be addressed by having students work cooperatively. The aspect of involving students in small cooperative groups could be another characteristic of expertise.

Episode 12 – Predicting the Lottery

If you could choose any one of the following tickets [for a lottery], which would you chose?

a) 1, 2, 3, 4, 5, 6
b) 5, 10, 15, 20, 25, 30
c) 2, 13, 19, 27, 30, 38
d) Use the same numbers I lost last week.
e) Use the same numbers I won last week.
f) No preference.

Results: Respondents judge the outcome (a) less likely than (c). They do not take into account the independence of the events, and they do not seem to believe that order or patterns are likely to be associated with random events. (Situation used by Begg and Edwards,1999.)

In this task, Natasha observed that respondents would be more inclined to pick the sequence that did not have a clear pattern, like being multiple of 5 or being in ascending order. She said that respondents might be inclined to select Choice (c) because it seems random. She also explained that if respondents have difficulties, it would be because they fail to recognize that the sequences were equally likely to happen. She saw the results and confirmed her prediction. However, the idea of independence did not come out in her explanation; the idea of randomness predominated in her justifications.

Natasha's suggested two pedagogical strategies. First, she recommended using the random generator in the calculator to generate numbers from the lottery, doing a large number of trials and checking the experimental probability of one of the sequences. Second, she suggested reducing the situation to a simpler situation. She said, "We could do a simpler situation, … maybe we only have five numbers, and we are trying to find the combinations of drawing out two. And that would be something maybe we could simulate in class." In all the data, this was the only instance in which Natasha suggested the use of technology, which provides revealing evidence of the limited use of technology in her pedagogical strategies. She presumably felt more comfortable with strategies using concrete material, or at least she was more familiar with them. I did not ask specifically about the absence of technology in Natasha's suggested pedagogical strategies, but I confirmed it in one learning community meeting (November 5). She said that she felt the need to explore statistical software and more activities that involved technology.

The second strategy Natasha suggested was a simplification of the original task. Her idea was to have a simpler task to present to students and be able to simulate it in class. I am still not convinced, however, that the new task was able to depict the ideas of independence that the initial situation described. The situation Natasha suggested did not characterize the patterns represented in the initial situation.

Natasha mentioned that she liked the Predicting the Lottery Task and that she could give it to her students. She commented, “I do like this example; I might pose it to my students in class one day and see what the kind of answer they come out with. I think that would be very interesting.” This expression reveals that Natasha was eager to look for and try new things in her classroom. In several instances Natasha made comments about collecting tasks to use in her teaching. There is confirmatory evidence from the learning community sessions that reveals that building up a wide repertoire of activities was one of Natasha’s interests.

At the end of the meeting on October 15, Eric brought up the idea that the teachers in the learning community needed to review their goals for the year. In the following session, on November 5, Mrs. Sanders asked the members to list the issues they wanted to work on for the rest of the year. Several issues came up: assessment, time management, grading, homework checking, pacing guide, review questions, technology use and others. The aspects that interested me were the issues suggested by Eric and Natasha. Eric was concerned with students’ motivation. He wanted to understand why students were not “in tune” with the class. He wanted to explore whether the reduction of hours they had in the schedule the previous year could have been affecting students’ motivation in the class.[7] He wanted to look at other teachers’ pacing guides and explore whether they were seeing the same effect on students’ motivation and performance that he was experiencing. In contrast, the issues Natasha brought into the discussion were the need to compile more activities to work in the class, review questions for the AP Statistics examination, and activities using technology. Although both teachers were concerned

[7] The year of the study Eric was teaching the same AP Statistics class he had taught the previous year; however, the number of days had been reduced. Before 2007 he had 180 days with 90-minute periods to teach the class. After 2007 the number of days was reduced to 90 with 90-minute periods.

about improving their teaching, there was a difference in focus. Natasha's issues were associated with her performance as a teacher and with her own actions. Natasha presumably thought that having a great number of activities to pick from might give her more confidence in her teaching. Eric's issues, in contrast, were related to students. He wanted to find out how to get students motivated in the class. This contrast suggests that expert and novice teachers have different needs and priorities. Expert teachers may be more interested in students, whereas novice teachers may be more interested in developing their expertise by expanding their repertoire of activities.

This finding is consistent with research in teaching education (Moir, 2004) that suggests that teachers go through stages of development. In the survival phase, teachers are busy developing the curriculum, looking for materials and strategies, and testing what things work and what do not. The amount of time teachers devote to the development of the curriculum does not leave much time for reflecting on other important aspects of teaching. A late stage of that framework refers to reflection. At this stage, teachers are able to think deeply about features of past lessons and materials; they reuse those that have worked well and redesign those that might need adaptation. They have time to reflect on other teaching issues like curriculum, student learning, and professional development. Although Moir's framework was designed to explain the development of first-year teachers, it could be extended to explain the development of practicing teachers. Other literature suggests that beginning teachers have egocentric concerns about teaching and that only when they have resolved these concerns do they shift their considerations to the impact that teaching is having on students (Schwab, 1973).

Eric's reflections on this task showed that he too was able to predict some students' difficulties, although he did not explicitly mention the idea of independence. Eric was able to predict that the structure of the lottery tickets would prevent students from seeing that all the tickets have the same probability of coming up. He said that students would think that the sequences were not random and that they would be inclined to select Choice (c). Eric mentioned that a lack of training could influence the students' ability to recognize random phenomena. He identified other students' difficulties but related them to the idea of randomness:

> You start constructing your sequence of numbers based on something else, like you know, the multiples of five. … I don't think that they [the students] see that as the same as some just random numbers. … The numbers having something in common would reduce their random ability—that is not even a word. But … when you have numbers all in a row, … they just see it all over the place, you know, random. When it is random, they [tickets] shouldn't be one through six, you know. That is not random. How can you say that that is random? (Eric interview; December 7, 2007)

When Eric was presented with the results from research, he agreed and confirmed his prediction about students' difficulties: "I can see where, how they would think like that. But, I mean, to stop that from happening is just more training in probability and, you know, the true nature of randomness, you know, random phenomena."

The fact that Eric did not mention the idea of independence could be explained in several ways. First, the ideas of independence and randomness are closely related in thinking about chance, and it is difficult to separate them. In this study, for example, the idea of randomness showed up easily in the discussion of the episodes on independence, and vice versa. Another explanation might be that the instrument for exploring the idea of independence was not powerful enough to produce a detailed description.

Eric's suggested diverse pedagogical strategies to help students in their reasoning about this idea. Some of his suggested strategies were similar to those recommended in previous tasks but some considered new features. He suggested, first, doing simulations with manipulatives and using technology. Second, students might compute the probability of occurrence of the sequences and compare them so as to recognize that they are equally likely. Third, students might collect real data on wining numbers from a local lottery for several consecutive years and analyze the results in class. Fourth, a teacher might offer training in different contexts to allow students to become more familiar with random phenomena.

It is worth noticing that the idea of independence was the most difficult one to analyze. Natasha only associated students' difficulties with the notion of independence in the H-T Sequences Episode, but she did not do it in the discussion of the Balls in Urns nor in Predicting the Lottery Episodes. This result is ironic because the wording of the H-T Sequences Task does not explicitly mention independence and Natasha talked about this idea. The other two episodes explicitly point out that students had difficulties because they were not aware of the independence condition but that was not enough to make Natasha talk about this construct. Eric did not relate students' difficulties to the notion of independence in any of the three episodes designed to explore this idea, which was surprising. That does not mean that Eric was not aware of the difficulties that independence can cause in students' reasoning because he brought it up when discussing the Spinners and the Gumball Machine Episodes. This result is hard to explain but as I mentioned before it might be due to the fact that the ideas of randomness and independence are closely related and that the episodes used may not have been good ones to prompt independence.

Law of Large Numbers

The law of large numbers says that as the number of independent trials (with the same probability) of a random process increases, the observed (experimental) probability gets closer to the expected (theoretical) probability. Many students, however, have difficulty understanding this law and think that what happens in the long run can be extended to the short run. This idea was explored with three episodes: Hospital, Coin Game, and Post Office. All the tasks asked the students to estimate the probability of an event in which repetitions were done both a few times and a large number of times. The results when these episodes were used in research revealed that students do not recognize the influence of the sample size in estimating the probability. In this section, I present Natasha's and Eric's reflections task by task.

Episode 7 – Hospital

> A hospital registers all babies born each day. Which event is the most probable one?
> a) Eight out of ten newborns are girls.
> b) Eighty out of one hundred newborns are girls.
> c) Eight hundred out of one thousand newborns are girls.
> d) a, b, and c have the same probability.

Results: Researchers found that the majority of the college students choose wrongly the Option (d), whereas the correct answer is (a). A similar situation was used by Tversky and Kahneman (1973, 1974) and later used by Begg and Edwards (1999) in research with Australian teachers.

In Natasha's reflections on this task, she noted that because the fractions were the same, students might conclude that the probabilities were the same. She said:

> They [students] would go ahead and maybe start by setting up each one of these probabilities as a fraction. So, you would have eight out of ten, eighty out of one hundred, and eight hundred out of one thousand. And then they would reduce the fractions, and they would say "Oh! They are the same, they have the same probability." So they would be inclined to choose (d). (Natasha interview; November 28, 2007)

A reason Natasha gave for students' difficulties was that they do not take into account the number of trials; they might not notice the difference in estimating the probability. She mentioned it using the following terms: "Not thinking about whether or not the number of trials that you conduct makes a difference in your probability. I am thinking about the differences between experimental and theoretical probabilities."

When Natasha saw the results from research in which the majority of the students selected Option (d) she was not surprised, because the results were close to her predictions. She also said that she agreed with the results. When she was asked for the pedagogical strategies she would use in the classroom in case her students had that difficulty, she related the Hospital Task to a similar activity (the classical task of rolling two dice) she had used in her class some days before. She described the strategies:

> I did something similar in my class the other day. We were looking at rolling two dice and looking at the probability of achieving each sum. And so they found them experimentally by rolling a pair of dice sixty times. And I had them record their experimental probability, and then, we combined the class's results to find the class's experimental probability. So, instead of just having sixty trials, I think, we ended up with seven hundred and twenty trials. And then we found the theoretical probability of achieving each sum, and they could see that the class's experimental probability was closer to the theoretical probability than just their individual experimental probability was. (Natasha interview; November 28, 2007)

These strategies were fairly similar to the ones she had described in the previous discussions of sample space, randomness, and independence: do the simulation with physical counters, keep track of the experimental probability, combine the class's results, and compare with the theoretical probability.

Eric's predictions about the students' difficulties in this task were also focused on the fact that the ratios, not the probability, were the same. He said that the quantities in the task could mislead students. He said:

> Well, if you form the ratio of all those numbers, you get the same thing. So leading to eight out of ten, point eight ratio in each one. And right away they are thinking, "Well if it is the same, you know, eight out of ten, eight out of ten, eight out of ten, even though it's eighty out of a hundred and eight hundred out of a thousand." That [led] them to error (Eric interview; December 7, 2007)

When Eric saw the results of research in which students did not find differences among the choices, he confirmed his predictions and said, "That doesn't surprise me, because they are going to confuse the ratios with the probabilities." He explained that reasons for students' difficulties were a poor understanding of variability and the law of large numbers.

Eric recommended several strategies to help students deal with their difficulties in the Hospital Task. He suggested using technology to explore the binomial coefficient for each choice of the task to illustrate the effect of the sample space. He said:

> Use it [binomial coefficient] in each one of those [choices], and then help them see, you know, what the probability is for each one. And then, if you take the probability for each one and show them how it keeps going down, then you can show the effects of the sample size on the probability. … And then, you know, look at the whole bigger picture, with all three of them [probabilities], all in front, in front of them. (Eric interview; December 7, 2007)

His suggested interventions were rich and integrated. He suggested exploring the binomial distribution with different attributes through simulations, looking at graphic representations, and discussing. The explorations Eric suggested were not proposed exclusively to solve the tasks, but they were intended also to investigate other aspects of the binomial distribution such as skewness, center, and spread. The following excerpt describes in a better way Eric's suggested explorations:

> Well, we use TI-83, and we just start with a sample size, like, or probability of a half and a sample size of maybe a hundred. And you can … probably get no successes or one or two or three. …And it'll graph, it'll do a histogram for you, a real nice one. And you can see that it's symmetric and all that. And you go back and change the probability to, you know, point six, point seven, point eight. And you can just see the whole thing just stacking up. … [Then, we] project it on the chalkboard, and we draw errors with different colors. … We start with a nice bell shape, and we end up with something else. … And that is how we … discuss it. And that gets them thinking about … if the probability is a half, and you are finding that you are getting eight out of ten successes. [But] the probability is supposed to be a half. … That doesn't happen as often with a hundred as it would do with ten, because we would have a class, let's say, twenty, and everybody can randomly generate some numbers from a distribution. … Let's say a binomial distribution where ten or a hundred or twenty. I mean, we are trying to … start at ten and go to twenty or thirty or something like that. Then, they all have a different set of numbers. And then, we take those proportions and do a distribution of the sample proportions, and look at … how it changes as the sample size increases. And then, you look what probability I can get, eighty out of a hundred, if the probability is a half. And you see pretty quickly that it's not very likely at all. (Eric interview; December 7, 2007)

For the Hospital Task, both teachers identified that the property of the choices of having the same ratio could be a difficulty for students. Similarly, both teachers noted that the understanding of the law of large numbers was crucial to solve successfully this type of task. Although both teachers suggested simulations and comparations of the empirical with the theoretical probability, Eric suggested the use of technology to explore different aspects of the binomial distribution. He also emphasized in his strategies the discussions in groups as a good way to make students talk about their ideas and confront their thinking with others. It is interesting that Eric was not focused only on how to help students solve the situation but also on how he could take advantage of a task to introduce other aspects of probability that might be important in the development of the ideas of chance.

Episode 8 – Coin Game

Suppose you were asked to play the following games each involving flipping a coin a number of times. In each game, you have the choice of flipping your coin 10 times, 50 times, or 100 times. For each game, which of these coin-flipping strategies would you choose if you wanted to win, or does the number of flips not make a difference?

a) You will win if more than 70% of your coin flips land HEADS. Would you flip 10 times, 50 times, 100 times, or does the number of flips not make a difference?
10 times _____ 50 times _____ 100 times _____ No difference _____

b) You will win if exactly 70% of your coin flips land HEADS. Would you flip 10 times, 50 times, 100 times, or does the number of flips not make a difference?
10 times _____ 50 times _____ 100 times _____ No difference _____

c) You will win if between 40% and 60% of your coin flips land HEADS. Would you flip 10 times, 50 times, 100 times, or does the number of flips not make a difference?
10 times _____ 50 times _____ 100 times _____ No difference _____

d) You will win if exactly 50% of your coin flips land HEADS. Would you flip 10 times, 50 times, 100 times, or does the number of flips not make a difference?
10 times _____ 50 times _____ 100 times _____ No difference ____

Results: Similar problems to this one have been also suggested by Tversky and Kahneman (1973,1974). Most students are not able to recognize that the number of times the coin is tossed influences the estimation of the probability. Maria (a participant in J. F. Wagner's study), for example, in answering Part (c), simultaneously chose “no difference” and “100 times.” (Situation taken from J. F. Wagner, 2006)

Although Natasha was able to talk about some of the most common difficulties in students’ reasoning about the law of large numbers, there were also some instances indicating that she underestimated students’ difficulties. For example, in the first encounter with the Coin Game Episode, she said that students might recall the idea that “the more times that you flip a coin, the closer your experimental probability is gonna get to your theoretical [probability].” She was very confident that students might base their reasoning on this idea; however, research has shown that having the concept is not enough for students to transfer their understanding to another context (Garfield, 1995; J. F. Wagner, 2006).

When Natasha saw the results from research she was more open to considering some students’ difficulties. She said:

> Some students might not see it. It does not make a difference, because the coin is just random each time, and you cannot control it. So, they might not understand that flipping [the coin] more times is going to actually get you closer and closer to what it should be, the probability [that] should be theoretically. (Natasha interview; October 24, 2004)

Natasha gave an interesting definition of *randomness*; she said it was “something you cannot control.” She used this definition to explain that many students might select the choice “no difference” just because with the randomness of the coin, they could get any combination regardless of the number of flips.

Natasha suggested two pedagogical strategies to help students to deal with their difficulties with the law of large numbers. First, she suggested doing the simulations by flipping a coin 10 times and 50 times, recording the experimental probability, compiling the class’s results to get an extensive empirical distribution, and estimating the probability using the empirical distribution. Second, she suggested doing simulations by rolling dice a small number of times and a large number of times, combining the class's results, and comparing the results. Although Natasha mentioned the second strategy as a different one, the two strategies were much the same. The only feature that changed was the object, die instead of coin; but the essence of the simulation was not altered.

After we had discussed the Coin Game Task, Natasha asked if I would mind sharing it so she could use it in her classroom. She used the Coin Game Task with her students during the following week, and when we met again for discussion, she told me of her experiences with the task. She found that the students used different approaches to solve the task and struggled more than she had thought they would. Some students set up the fractions, some did not find differences among the means, some went straight to the formulas of binomial probability, and only a few recommended simulations.

In discussing this task, Eric knew that students would need to have a clear understanding of the law of large numbers. He said that Choice (c) could be complex and Choice (d) a little unrealistic because of the word *exactly*. He was not sure, however, what kind of difficulties students might have. He said, "I know they might have difficulties with it, but I am not sure how they would respond to the number of times."

When Eric saw the results and realized that Maria picked simultaneously "100 times" and "no difference," in Choice (c), he was very interested in trying to figure out Maria's reasoning. He liked the fact that Maria had some correct ideas. He said:

> She got the right idea that in here ["100 times" choice], right, most of the time you are gonna get, you are forcing something in there. But if you only flip it ten, it just might not happen. … Certainly by fifty and a hundred, but then pick up both. I think she is getting the idea that [there is] no difference. She just needs more practice, I think. (Eric interview; November 14, 2007)

The discussion of Maria's reasoning was interesting because it showed once more that Eric was concerned about students' thinking. As in the discussion of sample space, Eric showed an interest in understanding the students' reasoning and tried to build strategies based on their thinking. Later, he talked about other possible reasons to explain students' difficulties. He said that one difficulty might be a poor understanding of the law of large numbers but that the lack of exposure to similar situations and a poor understanding of variability might also contribute to the difficulties.

Eric suggested several strategies to help students to deal with difficulties in solving the Coin Game Task. First, he recommended simulations using a TI-83 graphing calculator or statistical software, doing a thousand repetitions of the game, drawing a diagram of all possibilities, and exploring the shape of the distribution. Second, he suggested bringing to the class similar situations so that students could become familiar with ideas of chance. Third, he recommended using software to reinforce the concepts (e.g. binomial coefficient, binomial distribution, binomial probability and chance variability) that students need to do these kinds of tasks successfully. Fourth, he suggested keeping folders with problems that students have most difficulties with and have them make the mistakes in class so that they can generate discussions.

The contrast between Natasha and Eric about the Coin Game Task was interesting. Natasha was able to talk about some of the most common difficulties students have in reasoning about the law of large numbers; however, she underestimated students' difficulties. Eric, in contrast, did not identify the precise difficulties students could have with the task but clearly saw that a poor understanding of the law of large numbers could cause difficulties. Both teachers agreed that a clear understanding of chance variability was needed to understand the law of large numbers.

The big difference between the teachers was related to the pedagogical strategies they suggested. Natasha suggested pretty much the same strategies she had proposed in the discussions of sample space, independence, and randomness. I have stated previously that the fact that she recommended the same strategies across the tasks could mean that she had a limited repertoire of strategies to pick from and that she was in the process of building her pedagogical content knowledge. However, it also might mean that she did not find differences across the core ideas needed to treat them differently. Eric, in contrast, recommended a varied set of strategies. Some were similar to the strategies suggested in previous tasks, but he showed an interest in understanding students' reasoning as a starting point to design his strategies and in being systematic about the difficulties students more often have. It is interesting to observe that the data did not reveal a single occurrence in which Natasha mentioned an interest in understanding students' reasoning.

Episode 9 – Post Office

When they turn 18, American males must register for the draft at the local post office. In addition to other information, the height of each male is recorded. The national average height of 18-year-old males is 5 feet 9 inches. Every day for 1 year, 25 men registered at Post Office A, and 100 men registered at Post Office B. At the end of each day, a clerk at each post office computed and recorded the average height of the men who had registered there that day. Which would you expect to be true? (Circle one)

a) The number of days on which the average height was 6 feet or more was greater for Post Office A than for Post Office B.
b) The number of days on which the average height was 6 feet or more was greater for Post Office B than for Post Office A.
c) There is no reason to think that the number of days on which the average height was 6 feet or more was greater for one post office than for the other.
d) It is not possible to answer this question.

Results: Students show no systematic preference for the correct answer (a), and the modal answer is "same." Situation taken from J. F. Wagner (2006) and adapted from Well, Pollatsek, and Boyce (1990), after Kahneman and Tversky (1972).

Natasha read the task and was able to describe some difficulties that students could have in solving this task. She said that students might not recognize the differences among the choices. She said:

> I think that students would probably … choose (c). There is no reason to think that that the number of days was greater … for one post office than for the other. Because they would not think that there would be a difference, no matter how many men … you … have. They would each kind of average each other out. (Natasha interview; November 7, 2007)

Natasha said that students' difficulties with the Post Office Task could be explained because students do not recall the idea of variability and do not recognize the influence of the sample size on the variation of the average height.

When Natasha was asked for the strategies she might suggest for helping students to reorganize their thinking about the idea of law of large numbers, she stated that she would use the "penny lab," and she described it:

> Well, one activity that we do, that you've probably heard of, is one with the penny lab, where the students bring in pennies just from home, and they write down the age [date] of the penny. And we make a dot plot of all of the ages that they have collected, and they see that the dot plot is skewed to the right because most of the pennies that they bring in are younger [newer] pennies. Most of them are brand new, which means that they have an age of zero or they are one or two years old. And then, we collect samples from the population that they brought in. So, first I have them collect samples of sizes five and find the average age of their samples of size five. And then, we make another dot plot of their samples of the mean of their samples. And then, I have them take samples of size ten, and we make another dot plot of the mean of their samples of size ten. And then, I have them take samples of size twenty, and … we compare to see as your sample size increases what happens to the mean of all of your samples.(Natasha interview; November 7, 2007)

The penny lab was pretty much the same as the strategies that Natasha suggested for the other tasks. A new feature of this activity was the graphical representation (dot plot) for the experimental probability, which Natasha had not mentioned in the strategies she suggested previously. When she mentioned graphical representations, it was for a different purpose. She suggested using graphical representations as an attempt to help students understand and visualize the task before solving it, but she seldom used graphical representations to explore the physical characteristics (shape, spread, skewness) of the empirical distribution of probability.

In discussing the Post Office Task, Eric predicted the students might not find differences between the average men's heights registered by the two offices. He said that students might approach this task from an algebraic point of view; the majority would take the averages of the two post offices and would conclude that the means are equal. He said that students have difficulties "recognizing that the variability is reduced with larger sample sizes."

By looking at the results in which the students did not show preference for the correct answer (a) and not recognizing that the sample size influences the estimation of probability, Eric confirmed his predictions. He said that the idea of variability needed for understanding the law of large numbers does not appear in the secondary school curriculum unless the students take statistics:

> I can see that most college or most high school kids don't get this notion of variability based on sample size. Unless, you know, you take statistics, I don't know where you would get it in any of our curriculum. (Eric interview; November 27, 2007)

He also observed that students need to be exposed to these ideas more often to become familiar with them. He said:

> They have never encountered that idea [the law of large numbers]. … I can see, like we say at the beginning, most of them would just pick "there is no difference." … I think their biggest problem is that we don't have a lot of data analysis or inference even in our math curriculum. It's just lack of training or never been exposed. Had they seen an activity, like, … a sampling activity, then they would have some kind of idea, but before that I don't know. (Eric interview; November 27, 2007)

Eric suggested several pedagogical strategies to help students to deal with this difficulty. First, he recommended constructing confidence intervals for the means of the men's heights at each post office to compare the variability between the two. Second, he also suggested the "penny lab" that Natasha described. Third, he suggested doing simulations for each post office using technology and combining the class's results:

> We could take our TI-83 and have them randomly select twenty-five guys from a normal distribution. And we consider the heights of American men normally distributed, that would be an assumption I would let them have. And then, you could take average of those twenty-five, … and then have each person in the class do it. And so then you will see … your scores and find an average. And then, do it again, … randomly select the hundred and have the calculator give the mean for your hundred. And see … once you get all of those means on little Post-it notes, blue ones for Post Office A, and green ones for Post Office B, and see how many of each one was the average … over six feet. And so we have this physical count through our simulated [activity]. (Eric interview; November 27, 2007)

Both teachers accurately recognized the same difficulties that the research literature has identified in students' reasoning about the law of large numbers. In the Coin Game Task, Natasha had underestimated the students' difficulties, and Eric was not sure of the type of difficulty students might have. Natasha's situation might be explained because she was not familiar with the task, but except for that task, both teachers were familiar with tasks exploring the law of large numbers, were confident in talking about students' difficulties, explained reasons for those difficulties, and suggested pedagogical strategies.

The law of large number was the core idea that the teachers felt most confident talking about. Their familiarity with that idea might have occurred for several reasons. First, the textbooks (i.e., Agresti & Franklin, 2006; Yates et al., 2003) they were using make several references to the law. Second, the AP Statistics examination contains many situations in which students need to apply either the law of large numbers or one of its implications, which implies that AP Statistics teachers need to be familiar with such situations so as to coach students. I found confirmatory evidence during the learning community meetings. Some sessions were devoted to reviewing material from previous AP Statistics examinations. The teachers reviewed multiple-choice questions and free-response questions, and several questions required a clear understanding of the law of large numbers.

Although the goal of this study was not to evaluate the impact of the learning community on the teachers' professional development, it is worth noticing that the teachers showed a positive attitude in engaging in the suggested activities and actively contributed to the discussions accordingly to their expertise. In most cases Eric led the discussions, but the rest of the teachers took turns to lead the exchange of ideas of topics in which they felt familiar. Some teachers said they had put off some topics in their classes because they did not want to teach them without having discussed them with the members of the learning community. This feature makes me think that the teachers found authentic support in the learning community for their specifics needs and that it could be considered a promising format for professional development for teachers

Conclusion

The results of the analysis show that both teachers were able to identify students' difficulties with foundational ideas about chance. Eric made more refined predictions than Natasha, who frequently needed some extra help to see the difficulties clearly. Eric, in contrast, was able to identify those difficulties at first glance, even before knowing the results from research reporting those difficulties.

Both teachers gave explanations for students' difficulties. Natasha's explanations were associated with the understanding of concepts related to chance, whereas Eric offered a variety of reasons to explain students' difficulties. Although he also related students' difficulties to poor understanding of statistical concepts, he identified other factors such as students' difficulties with fractions, multiplication, and proportions; a deterministic view of the world that causes students to give exact answers; a lack of inclusion of probabilistic ideas in the secondary curriculum, and a negative influence from the mass media.

The strategies that the teachers used to help students readdress their difficulties in chance situations were different. Natasha used an experimental approach. For almost all the tasks she suggested doing simulations with physical counters to get the empirical distribution, combining the class's results, and comparing with the theoretical distribution. Her suggestions revealed an approach that responded to the new challenges of the state performance standards, which required that students be involved in a great deal of experimentation before they go into the theory. Natasha's experimental approach showed her special characteristics. She was not just a novice teacher but a beginning teacher eager to learn how to be a better teacher.

The pedagogical strategies that Natasha suggested, however, did not differ much from idea to idea and did not make a strong connection with probability theory. She used much the same strategies across the core ideas and tasks. Eric's suggested strategies, however, were rich, integrated, and differentiated across the core ideas and tasks. He went further than constructing an empirical distribution through simulations with physical counters and contrasting it with the theoretical distribution. He was very skillful in using technology not only to calculate and do simulations but also to explore properties of the distributions. He made strong connections among the results from simulations and probability theory. He supported cooperative learning and was interested in students' reasoning. He found real-world situations to be discussed in class. He was skillful in managing the pacing guides as recommended. There is no doubt that these differences in the teachers' reflections can be attributed to their differential experience in teaching AP Statistics.

CHAPTER 5

REVISITING THE DATA

This chapter is dedicated to analyzing the strategies the teachers proposed to use to help students reorganize their thinking to overcome their difficulties. The strategies reviewed in this chapter are the same as those described in chapter 4 but analyzed from a different point of view. For this analysis, I used the six *principles for teaching and learning ideas of chance* presented in chapter 3.

I looked at the data I had, and I coded them according to the strategies the teachers claimed to use in the classroom to help students with the reorganization of their thinking. I ended up with a long list of actions that I tried to organize in a meaningful way. During the organization process, I first considered three other frameworks (described in chapter 3) to see how well the principles and recommendations would work with the data. At some point, each framework was useful in analyzing portions of the data but not in analyzing the data completely. However, there were moments where the frameworks were insufficient, and I had to create a new framework with categories generated from the data as suggested.

Any strategy is a long-term plan of action designed to achieve a particular goal, and its nature is extensively premeditated. Strategies are used to make a problem easier to understand and solve. I classified the strategies that Eric and Natasha claimed to use to help students reorganize their thinking about chance according to the principles for teaching and learning ideas of chance. I went back to the data and counted the number of times the teachers referred to each principle from framework, and I created Table 4, which reveals that Natasha and Eric had some similarities and differences in their strategies. In this section, I describe principle by principle those similarities and differences.

Table 4

Frequency of Reference to Principles of Learning Statistics in Teachers' Reflections

Principle	Action	Natasha	Eric
Students learn by doing	Make diagrams and graphical representations	5	6
	Do simulations with physical counters to construct empirical distributions	12	9
	Do simulations with technology to construct empirical distributions	1	8
	Construct theoretical probability distribution	5	6
	Compare empirical with theoretical distribution	3	3
Take into account students' prior knowledge	Take into account students' prior knowledge as starting point in the intervention	0	4
	Make students express their ideas and then lead them from there	0	1
Use cooperative learning	Use small-group discussions	1	2
	Have students argue about the flaws in sampling techniques	0	2
Students learn by confronting misconceptions	Ask students for conjectures and contrast with results	0	3
	Have a folder with difficult problems and create situations where students make mistakes that reveal misconceptions	0	1
Use technology to visualize and explore	Use technology to reinforce concepts	0	3
	Explore the properties of distributions by varying parameters	0	4
	Visualize the influence of sample size in sampling	0	2
Point out common misuses of chance	Bring to the classroom situations from the real world where incorrect reasoning about chance has been used	0	2

Students Learn by Doing

The greatest similarities between Natasha and Eric involved the principle *students learn by doing*. The principle is defined as those actions that teachers design to engage students in learning. It involves hands-on activities such as graphical representations, simulations using physical counters and technology to create empirical distributions, tree diagrams (or other approaches) to find the theoretical distribution, and explicit comparisons between empirical distributions and the theoretical distribution.

Students need to practice what they learn in different contexts, and they learn better when they apply ideas in new situations. Having students learn a concept in only one context does not ensure that they will recall the concept in other situations. Research has shown, for example, that students may be familiar with the law of large numbers in contexts of tossing coins, but when the context is changed to another real-world situation, they might have problems identifying the concept and might not think of the law (Garfield, 1995). The more students are exposed to certain situations, the more likely they are to recognize the concepts across different contexts. In coding this principle, I noted several categories of actions in the teachers' descriptions. The categories with the most occurrences were the following: diagrams and graphical representations, simulations, find theoretical probability and comparison between experimental and theoretical probabilities.

Diagrams and Graphical Representations

Eric and Natasha both mentioned several times the convenience of having graphical representations to help students visualize in concrete ways what is abstract. As Table 4 shows, Natasha mentioned this action 5 times and Eric 6 times. In discussing the Tossing a Coin Task, where students were asked for the probability of getting two heads in tossing a coin four times, Natasha said that organizing the information by means of tree diagrams could be useful for students:

> Making a list by using tree diagrams. So, on the first toss we can get heads or tails, and on the second toss heads and tails, and so on. So that might be another visual strategy to help them see the different paths for finding the outcomes. (Natasha interview; October 24, 2007)

In discussing the Gumball Machine Task (the task asked for the best prediction of getting 5 red balls in 10 trials if the machine has 20 yellow, 30 blue, and 50 red balls), Eric also mentioned that graphical representations might be useful for students to understand the task. He mentioned an additional detail; however, he said that for large samples the use of such representations might be inefficient:

> Well, … you can use … diagrams or … a tree diagram. If they are totally stuck, you go back … and when you choose ten of them [colored balls], you can get a pretty extensive tree. … Gosh, there's gotta be a better way for us to look at this. … But certainly a tree diagram. (Eric interview; November 27, 2007)

I conjectured that the teachers' consideration of graphical representations as a resource to use in the classroom to support students' understanding suggested that they had previously reflected about students' difficulties with such abstractions. I double-checked the episodes in which the teachers proposed the use of graphical representations, and for those same situations they had predicted that the students might have problems making an organized list of the sample space. Hence, the use of graphical representations seems to be an action these teachers thought suitable to support students' reasoning when they appear to struggle. Proposing this action, however, could also reflect the teacher's own way of approaching such situations. Evidence to support this conjecture comes from the first interview with Natasha. In discussing the Bags and Marbles Task, in which two players pick red and blue marbles from two bags, she was asked to predict some of the students' approaches to solving the problem. She immediately started making a drawing of the bags and the marbles (see Figure 1). Only after the drawing was done did she mention that making pictures might be one of the students' approaches to solving the problem. Although the two teachers suggested the use of graphical representations, Eric went further than Natasha; he noted that graphical representations might have limitations when the sample space is large.

Simulations

The use of simulations is a nice tool to engage students in hands-on activities and a great resource to help them to understand random phenomena, especially if connections with the theory are explicitly made. Both teachers saw the value of involving students in making simulations; however, the way they conceived and used simulations differed. Natasha, for example, mentioned several times that simulations were a means for students to see concretely abstract aspects of statistics. One characteristic of simulations that she noted was that they should be done with physical counters such as coins, dice, and marbles. Eric, in contrast, expressed his perceptions and use of simulations in a more holistic way. In addition to proposing the use of physical counters, he added the use of technology and integrated other elements into his suggested actions. The following example illustrates this contrast. In discussing strategies for solving the Bags and Marbles Task (Episode 1), Natasha said, "Maybe they [the students] could simulate this. They could actually try out the game themselves with bags and marbles or something similar to marbles, and they could start keeping track of their trials and seeing what the outcomes are."

Eric, in contrast, gave a more detailed description when simulations were going to be used. In discussing the Tossing a Coin Task, he described his strategy. I noted that his strategy integrated several actions; it was not just a simulation using concrete material. He integrated the contrast between the empirical and the theoretical probability, an estimation of probability from the empirical distribution, an outline of the theoretical probability, a synthesis of the class's results, a sketch of students' conjectures, a contrast of students' previous beliefs with empirical evidence, an exhibition of students' thinking by having them talk about their reasoning, and the reinforcement of statistical concepts (see excerpt of Eric's description of suggested strategies for the Tossing a Coin Task in Chapter 4, p. 46).

Besides his holistic view of simulations, Eric used technology to support students' learning. Table 4 shows that Natasha mentioned the action "use simulation with physical counters" 12 times, whereas Eric did it 9 times. Eric, however, suggested simulations using technology 8 additional times, a suggestion that was almost nonexistent in Natasha's reflections. This difference might suggest that Natasha operated from a concrete point of view or it could reflect Eric's familiarity with technology.

I infer that the teachers' use of simulations was informed by their reflections on engaging students in meaningful activities and by their knowledge about the power of simulations. I also infer that simulation was a common practice in their teaching because there were several instances in which the teachers mentioned the use of simulation. I suspect that this practice, however, might have been a response to the College Board instructions that the AP Statistics course be taught with an active approach in which students are not faced with formulas without being engaged in experiences that support the construction of their knowledge (College Board AP, 2007).

It is important to mention that all the episodes used in the interview protocol could have been simulated using either technology or physical counters and that Natasha had a strong preference for simulations with physical material over those with technology. This preference might suggest that she was comfortable using the simulations with the physical counters and that she was aware of the positive influence of simulations in the classroom. However, her strong preference for simulations with physical counters, in spite of having the technological resources at her school, might also imply that she had not had effective training in the use of technology in instruction. One might think that with training she would have been able to integrate the simulations with the use of technology. There is nothing wrong with doing a simulation with physical material, but it should be confined to the first stages in exploring data. If the exploration requires a large number of repetitions, it is appropriate to use technology.

Theoretical Probability

Finding the theoretical probability of a specific distribution was a common action suggested for both teachers, as Table 4 shows. Natasha suggested this action 5 times whereas Eric suggested it 6 times. For example, Eric mentioned that applying theoretical rules such as the multiplication rule and combining it with tree diagrams and contingency tables could be helpful for students to find the size of the sample space and consequently the theoretical probability. He also mentioned that finding the theoretical probability would be a good way to help students understand the reasoning behind some tasks. In the discussion of the Post Office Task, for example, Eric suggested having students "build a confidence interval for each one [post office] and see which one has the biggest confidence interval."

Similarly, for the Bags and Marbles Task, Natasha suggested having students find the expected value. She said "find the expected value for the game and compare each player probabilities". It is not surprising that this action appeared in the teachers' suggested pedagogical strategies because the tradition in teaching statistics, for a long time, has been approaching problems from a theoretical point of view. This practice is, plugging numbers into formulas to find the theoretical probabilities. It is interesting, however, that although the teachers proposed finding the theoretical probabilities; this was not an isolated action. Usually, it was suggested in combination with other actions such as finding the experimental probability to support previous findings. This integration could reflect the teachers' awareness of having multiple representations of the same situation for better students' understanding.

Comparing Experimental with Theoretical Probability

Another category that showed the similarities between Natasha and Eric was the explicit statement that both teachers made about linking the results from empirical evidence with theory. This is, they wanted to establish connections and make comparisons between the experimental probabilities obtained from simulations and the theoretical probabilities obtained from applying multiplication rules, counting principles, contingency tables, and tree diagrams. Table 4 shows that both teachers suggested the comparison the same number of times. In the Spinners Task, for example, Natasha suggested that the distributions of results from simulations could be compared with the theoretical distribution when the number of trials was large.

Eric also mentioned the comparison between an empirical and the theoretical distribution when discussing the Tossing a Coin Task (see excerpt of Eric's description of suggested strategies for the Tossing a Coin Task in chapter 4, p. 46). I infer that the teachers recalled this action on a regular basis because the curriculum materials they used in their teaching (Agresti & Franklin, 2006; Yates et al., 2003) suggested using this approach. Nonetheless, it could also have come from their reflection about teaching practice.

Take into Account Students' Previous Knowledge

Theories of learning argue shown that students approach new learning situations with extensive prior knowledge. Theories such as constructivism (Von Glasersfeld, 1984, 1987), transfer in pieces (DeSessa & Wagner, 2005), and conceptual change (Posner, Strike, Hewson, & Gertzog, 1982) agree that the learner constructs new knowledge in contexts where prior learning has taken place and uses prior ideas to make sense of the world. Thus, evidence that the teachers considered students' previous ideas about chance in designing their interventions was important to understanding their perceptions about learning. Table 4 shows that Eric, on 5 occasions, mentioned using students' prior knowledge as essential in designing his interventions. For example, the Tossing a Coin Episode showed the teachers a student's incorrect reasoning. The student said that since a single coin has a 50% chance of landing heads, in four tosses we would expect to get two heads. Confronted with this situation, Eric considered that a good starting point could be the student's argument:

> The fact that he has an argument, … there is something to be said for that. And you … try to get him to think more [about the student's initial argument] … There is nothing wrong with this [student's argument] as a starter. …And then don't just say "You're wrong," …or … tell the kid that this is wrong instead of just showing him how it should be done. It's better to bring in the coins, and let him experiment with why it's not right, and then have him see that it's not you. It's not the teacher against the student. … It's just not the right logic that he is applying. … It's not me that's telling this is wrong. It's the situation they can actually see. (Eric interview; November 14)

In contrast, the principle of considering students' prior knowledge was absent from Natasha's reflections. One can infer that if a teacher takes into account students' prior knowledge about chance when designing strategies, that teacher is aware of students' previous ideas in the construction of knowledge. It might also mean that the teacher has reflected on the importance of the principle for students' learning. If there is no evidence of the principle in the data, one can conclude that the teacher ignored the principle. Also it could mean that the teacher was not aware of the principle, or that the specificity of the tasks does not reveal the principle in the teacher's practice.

Use Cooperative Learning

The principle of cooperative learning is related to the concepts of active learning and cooperative work, which according to some researchers contains many advantages for teaching and learning (Magel, 1998). Although the implementation of this principle in statistics education has advantages, it was mentioned only a few times in the teachers' reflections, and there was no detailed description. From the evidence in the teachers' talk, however, one might make some interpretations of these teachers' perceptions about having students work in cooperative groups. Eric mentioned, in an interview, how having students work in collaborative groups might be useful for them: It prevents students from making the same errors over and over again, and it helps them with their communication skills. He also said that such an approach might help students explore different strategies (see Eric's suggestion of teamwork for the Balls in Urns Task in chapter 4, p. 77).

In contrast, evidence of this principle in Natasha's reflections was rare. There was only one instance in which she brought up cooperative learning as useful for helping students in their reasoning about chance. The idea of students working in cooperative groups was slightly different for the two teachers. The kind of examples the teachers gave suggested that they had different conceptions of this practice. For example, Eric's conception of cooperative learning was associated with the idea of social construction of concepts, arguing, and exploring the variety of strategies that students come up with. Natasha, in contrast, said that students talking with other students may help them to get other ideas to expand their answers. It seems that her idea of having students work in cooperative groups did not necessarily include students confronting or reassessing their thinking. She described the use of group work in these terms:

> Have them [students] say it [idea] out loud, or even sharing their ideas with another student [and] not just [with] me. And sometimes talking to another student gives them more ideas. And they're able to use those to help justify and explain their answer as well. (Natasha interview; November 7, 2007)

As I mentioned in chapter 4, this difference in teachers' strategies could also be due to a difference in the teachers' experience, which would be consistent with teacher education research about classroom management. Experienced teachers are more open to having students working in groups because they have overcome the difficulties of classroom management. Beginning teachers, instead, prioritize discipline and hesitate to explore nontraditional teaching methods that risk disrupting the apparent order of the class (Raymond, 1997).

Students Learn by Confronting Misconceptions

Researchers in statistics education have found that students' ideas of chance improve when they make predictions before gathering data and then compare the empirical results with their original predictions (Castro, 1998; Garfield & delMas, 1991; Shaughnessy, 2003). I looked for evidence of this principle in the data and found that on several occasions (see Table 4), Eric talked about having students confront their misconceptions. He mentioned actions like having students write down their initial predictions and having them compare those predictions with empirical evidence. He also mentioned having a folder of the most common students' errors so that if students made those errors, they could talk about them in class. Eric spoke about confronting students' misconceptions in the following way:

> We ask for some conjectures like this. This is a good idea to get started with, [to] have people write down what they're thinking. … In my folders I have the problems … the kids mostly have troubles with. And … I'm thinking ahead when I am teaching, when I'm introducing it. … Be sure, that … [I] have them make this mistake in class so we can

> talk about it. And so it's kind of like, what would you call it? Proactive intervention strategy? (Eric interview; November 14, 2007)

Eric's verbal descriptions suggest that he mentioned confronting students' misconceptions several times because he had previously thought about that. The fact that he had a collection of problems on which students usually make mistakes suggests that he had been systematic enough in noticing those mistakes to collect them so as to use his experience with future students. In contrasting Eric's reflections with Natasha's, I founded it surprising that she did not mention confronting students' misconceptions in her reflections. The absence of a concern with misconceptions might also reveal Natasha's conception of learning or that she was not aware of the power of confronting students with their previous ideas. Another alternative could be that Natasha did not have enough experience in AP Statistics to construct the pedagogical content knowledge required to anticipate students' misconceptions.

Use Technology to Visualize and Explore

Many technological devices are available today, and many of them are highly relevant for applications and modeling. These technologies include calculators, computers, the Internet, computational and graphical software, instruments for measuring, instruments for performing experiments, and so on. The use of technology in statistics classrooms has been increasing not only as a tool for calculating precise answers but also for exploring and manipulating data in versatile ways. Technology has proven itself to be a powerful tool in exploring properties of the distributions of probability that if done by hands would typically take a lot of time. Research has also shown the positive influence of technology in changing and expanding students' thinking about chance (Drier, 2000). One of the biggest differences between Eric and Natasha was the way they integrated this principle into their teaching practice.

Several of the Eric's reflections concerned the use of technology to support students' reasoning about chance. He mentioned using technology not only to do simulations, as expressed previously, but also to explore and visualize properties of the distributions. He described several situations in which he would use technology in his teaching practice for different purposes. He mentioned the use of technology to explore the characteristics of probability distributions by varying attributes like sample size, population size, probability, center, and variability. Further, he talked about the power of technology in producing graphical representations and in showing the effect of changes in the sample size.

The variety of his examples suggests that he had extensively explored the use of technology in the classroom and that he was aware of some of its advantages. The fact that Eric talked about the use of technology in very different ways showed that he had a wide collection of actions that he was able to use depending on the situation. In contrast, the principle of using technology in the classroom to explore data was nonexistent in Natasha's reflections (see Table 4). She did not offer options for exploring data or visualizing the characteristics of the probability distributions. One explanation for Natasha's failure to refer to the use of technology to help students explore and visualize data might be her lack of training in the use of technology and her preference for exploration with more familiar procedures, like combining the class's results on the board.

Point Out Common Misuses of Chance

Television sports broadcasts, newspaper advertisements, and other mass media display conclusions and interpretations based on data and ideas about chance. Students are faced with conclusions that are often wrong but widespread. Because the media are authoritative, students start to integrate such misuses into their reasoning. This principle indicates that students should be exposed to, analyze, discuss, and criticize misuses of chance from normative points of view as well as contrasting the misuses with their knowledge of chance. I found that the use of this principle did not appear very often in the teachers' reflections, but it is relevant to note it. Eric mentioned twice in his strategies that he would bring into the classroom situations from mass media where chance ideas had been used incorrectly so that he could illustrate the errors and make students think critically. He said that those situations were great ways to get students to talk about the accurate use of chance ideas. He said that having students talk about such situations in class made them more critical and prepared them for a future in which they would have similar information in front of them.

Natasha, in contrast, did not mention misuses of chance in the media. One reason might be that she had other priorities. She was still exploring how to handle the curriculum and prepare students for the AP Statistics examination. But another reason might be that she was still developing the pedagogical content knowledge necessary to recognize situations that could be relevant and appealing to students.

Conclusion

An analysis of the teachers' strategies according to the principles for teaching and learning ideas of chance reveals differences and similarities in the teachers' reflections. Both teachers suggested strategies that engaged students in hands-on activities. Diagrams and graphical representations, simulations with physical counters to construct empirical distributions, construction of the theoretical distribution of probability, and comparison between experimental and theoretical distributions were strategies suggested by both teachers. One big difference was the absence of simulations using technology in Natasha's reflections, whereas Eric offered a wide variety of simulations, including simulations with physical counters and simulations with technology.

Eric's suggested strategies had some references to principles such as students' prior knowledge, cooperative learning, confronting misconceptions, using technology to visualize and explore, and point out common misuses of chance. Evidence of the use of these principles was almost nonexistent in Natasha's suggested strategies. This finding makes it clear that Natasha's repertoire was limited to hands-on activities, and she did not consider other principles widely suggested in the statistics education literature for the design of her strategies.

CHAPTER 6

SUMMARY AND CONCLUSIONS

The purpose of the study was to investigate teachers' perceptions of students' thinking about chance. In particular, I was interested in how teachers understand students' struggles with the ideas of chance, how teachers anticipate and explain students' difficulties, and what strategies teachers purpose to use to help students reorganize their thinking. Two teachers, an expert and a novice, and members of an AP Statistics learning community, participated in this study during the fall of 2007. They were observed in meetings of the learning community and were interviewed in depth. Data were collected in the form of observations, interviews, and artifacts and were then analyzed using grounded theory (Glaser & Strauss, 1967), interpretativism (Crotty, 2004; Schwandt, 1994) and an expert-novice contrast (Borko & Livingston, 1989; Leinhardt, 1989; Livingston & Borko, 1990).

The teachers exhibited differences in the way they perceived students' difficulties and in the way they dealt with them. These findings are consistent with the findings from other research studies on expertise in teaching (Borko & Livingston, 1989; Leinhardt, 1989; Livingston & Borko, 1990). The more experienced the teachers, the more complex and integrated are their interventions for students.

Relating the Findings to the Research Questions

The three research questions were addressed in chapter 4, where I described how the novice and the expert teacher predicted and explained students' difficulties in chance reasoning and the pedagogical strategies the teachers suggested using to help students with such difficulties. The results showed that the expert teacher was more accurate in predicting students' difficulties than the novice teacher was. The expert teacher was able to identify students' difficulties in the first encounter with the tasks, whereas the novice teacher needed more time and hints to be able to identify such difficulties. Several times the novice teacher identified students' difficulties only after being shown some examples of the struggles students tend to have with the situations. In addition, the level of the difficulties described by each teacher was different. The novice teacher frequently tended to underestimate students' difficulties and explained them in terms of students' lack of understanding of concepts related to chance. The expert not only associated students' difficulties with students' lack of understanding of ideas of chance but also described the influence of external factors like the organization of curriculum, incorrect uses of chance in the mass media, and a tradition of thinking in which exact answers are privileged over answers that require exploration and exploration.

The results revealed too that there were differences in the teachers' perception of students' difficulties across the four core ideas. The novice teacher's descriptions of pedagogical strategies did not reveal differences across either the tasks or the core ideas. The expert teacher's descriptions of pedagogical strategies were oriented according to the difficulty identified and integrated with real-world situations when possible.

The pedagogical strategies that the novice teacher suggested to help students in the reorganization of their thinking were centered in teachers' actions and were not specifically tied to students' thinking. The expert, in contrast, suggested strategies that integrated the teacher's actions with the students' actions and thinking. The expert teacher's descriptions of pedagogical strategies integrated the teachers' actions with students working in teams, students revealing their ways of thinking, and students confronting their own misconceptions.

Research reports indicate that in different fields experts and novices perform differently (Benner, 1984; Chi et al., 1981; Hidi & Klaiman, 1983; Swanson et al., 1990). Results from the present study concur with those from previous research. Several characteristics of the expert teacher's reflections were absent in the novice's reflections and might constitute a potential framework to explain expertise. First, the teachers' conceptions of teaching and the teacher's role were different. The novice teacher's reflections revealed that her conception of teaching was associated with the teacher's actions, teaching was centered on the teacher, teaching responded to the teacher's motivations instead of to the students' needs, and the teacher was a provider of information. The expert teacher's conception of teaching, in contrast, was highly tied to students' thinking, and he saw the teacher as a facilitator.

Second, the teachers managed time differently. The novice teacher claimed to have difficulties in managing time efficiently, whereas the expert teacher was confident with time management and was able to integrate several elements in his teaching without spending extra time. Third, the teachers' familiarity with the material differed. The novice teacher had to work the tasks to be able to express her opinion about students' difficulties, whereas the expert teacher's familiarity with the tasks allowed him to give reasonable predictions of students' difficulties just by reading the task statement. This finding is consistent with Swanson and colleagues' (1990) findings that, in solving problems, novice teachers tended to represent problems in terms of the solutions whereas expert teachers tended to place priority on defining and representing the problems as well as evaluating possible strategies.

Fourth, the expert teacher showed a reflective attitude toward the curriculum. The expert teacher had several references to the appropriateness of certain topics in the curriculum. He also mentioned that to strengthen students' understanding of chance ideas at the high school level, students should be exposed to statistical concepts early in their schooling by means curricula in other subjects. The novice teacher, in contrast, did not show any indication of being reflective about the topics that should be included in the AP Statistics curriculum or even considering the curriculum as a potential source of students' difficulties.

Fifth, the novice and the expert elucidated different priorities in improving their teaching. The novice teacher thought that expanding her repertoire of activities might have a positive impact on her teaching. The expert teacher was concerned about understanding and promoting students' motivation. The novice teacher's priorities in professional development were focused on the teacher, whereas the expert teacher's were focused on the students. This characteristic is consistent with the "teachers' conceptions of teaching" mentioned previously, which reveals from another angle the egocentric concerns of the novice teacher.

Sixth, the way the teachers perceived students' difficulties differed. The expert teacher was more accurate in predicting students' difficulties and offered a wider range of explanations than the novice teacher. Seventh, the variety of suggested strategies differed. The expert teacher offered a variety of strategies that were integrated and that responded to the students' difficulties. The novice teacher's suggested strategies were similar throughout the core ideas and were less integrated with the theory. This result coincides with results from previous research that shows that novice teachers do not have well-developed pedagogical content knowledge that allows them to construct detailed explanations (Borko & Livingston, 1989) and that expert teachers' strategies are rich in examples and representations (Leinhardt, 1989).

The third research question was addressed in chapter 4 and revisited in chapter 5. In those chapters I described the strategies that the two AP Statistics teachers suggested using to help students to transform their intuitive knowledge into more formal thinking about chance. Both teachers concurred in saying that students learn better what they practice doing. Using diagrams and graphical representations, doing simulations, constructing the theoretical distribution, and contrasting the experimental with the theoretical probability were considered by the teachers as some essential strategies that they used in their practice. Although both the expert and the novice remarked on the value of simulations as essential tools in their teaching, the novice teacher's conception of simulations corresponded to an inductive perspective of experimentation associated with the popular wisdom that "one learns by manipulating and observing what happens as a result." The novice teacher's connections between the experimentation with the manipulatives and the construction of knowledge were not clearly stated. The expert teacher, in contrast, was more versatile in his considerations about simulations. Besides physical counters, he considered that the introduction of technology such as graphic calculators and statistical software should be an alternative for the introduction of simulation and a necessary aspect for exploring and visualizing data. The expert teacher conceived the simulation not only as an experience but also as a tool to facilitate the understanding of the concepts, promote critical thinking, and develop students' research skills.

One of the biggest differences between the novice and expert teachers' strategies was associated with the consideration of students' prior knowledge in the design of the interventions. The expert teacher several times showed great interest in knowing students' previous ideas about chance as a starting point for his interventions. The instances in which the novice teacher considered the students' prior knowledge were almost nonexistent. The data also revealed that the expert teacher was skillful in suggesting strategies in which the students could confront their misconceptions, work in cooperative groups, and contrast misuses of statistics with formal instruction. The novice teacher, on the other hand, did not make any reference to these aspects in her suggested pedagogical strategies.

My primary goal in this study was to explore four core ideas that are present in the AP Statistics curriculum. These foundational ideas are essential to understanding data exploration, the design of experiments, probability, inference, sampling, and chance variation, among other topics. Initially, I was not separating mathematical probability from statistics. I was looking at these two domains as parts of the same unit; however, what I found was that these four constructs and the tasks I used to explore them were more closely related to probability than to statistics. I am referring not to mathematical probability but to the informal probability that is sufficient for a conceptual grasp of inference.

Benefits from Participation

At the beginning, I thought that the participants were not going to benefit from this study. I thought that their participation would be more of a burden than a benefit for them. I found out, however, that the discussions we had in the interviews made the teachers more aware of students' difficulties. After some interviews had taken place, the novice teacher, for example, reported that she had used episodes from the interview protocol in her classroom to explore students' reasoning about chance. In a later interview, she commented on the different ways students had analyzed and worked out the problems and on her perceptions of students' difficulties. She said she was becoming more aware of students' answers, reasoning, and difficulties. By engaging in the interviews, she reflected on her practice and on better ways to help students overcome their difficulties.

Limitations

One limitation of this study is the characteristics of the participants. The AP Statistics learning community where this research took place was a voluntary gathering, which makes this group of teachers a team with special characteristics. Members of the team were concerned with the improvement of teaching and learning AP Statistics. This concern is not a universal characteristic of AP Statistics teachers. Consequently, the conclusions from this study should not be extended to all novice and expert teachers. These conclusions, however, might be helpful in describing characteristics of experts and novices to understand the development of expertise.

Another limitation of this study was the length. One semester is limited time to see changes or growth in the teachers' practices and reflections that suggest that expertise is being developed. This limitation helps to support my claim that the research literature needs to be enriched with longitudinal studies that explore the development of expertise in ways that help in identifying the evolutionary trajectory of the process of becoming expert.

The tradition in expert-novice research has been to consider student teachers in their last years of college as the novices, and their respective cooperating teachers as the experts. The selection of the participants following such criteria shows a big gap in the continuum of expertise because the participants are at the extremes. This study was not able to close the gap, but it was moved further along in the continuum of expertise. The novice teacher in this study was not a student teacher but a practicing teacher, which places her more toward the middle of the continuum of the expertise. In addition, the expert teacher was a very experienced teacher who was far beyond a typical AP Statistics expert teacher. He was highly qualified; besides his credentials as an AP Statistics teacher, he was engaged in other professional activities that helped him to build up his expertise. For instance, he had conducted workshops in AP Statistics for several years. The preparation of the workshops and the interaction with other AP Statistics teachers had given him unique opportunities to develop his expertise. He had unusual expertise that is not easily found in an AP Statistics expert teacher with the same number of years of experience. This study, in spite of the limitations associated with the unusual characteristics of the participants and with the gap between the expert and the novice, contributes to the research literature on the development of expertise by describing the characteristics of a beginning teacher near the middle of the continuum of expertise in the process of becoming an expert.

Only a small number of teachers participated in this study. Although interviews and observations provided a rich basis for exploring teachers' perceptions of students' thinking about chance, one should not lose sight of the limitation of this study related to the number of participants. Interviewing and qualitative analysis of responses is a time-consuming methodology and necessarily restricts the number of participants to be included; but the researcher is usually rewarded with greater richness in the data.

Significance of the Study

Describing how expert and novice teachers perceive students' difficulties in reasoning about chance and how they deal with such difficulties is important for several reasons. First, without an understanding of the differences between expert and novices we cannot understand the development of the expertise that could orient teacher education and teacher professional development programs. Second, the contrast can help us identify exemplary practices of expert teachers and also the kind of support novice teachers need in the process of becoming experts.

Most studies that have explored reasoning about chance have focused mainly on students (Batanero & Serrano, 1999; Jones et al., 2007; Jones et al., 1999; Kahneman & Tversky, 1972, 1996) and only few have focused on teachers (Begg & Edwards, 1999). Those who have studied teachers have looked at the teachers' content knowledge and almost never at teachers' pedagogical content knowledge. Consequently, this study contributes to two main branches of the research literature: The pedagogical content knowledge of AP statistics teachers was explored, and the novice teacher was an AP Statistics practicing teacher and not a preservice teacher as has been the tradition in expert-novice research.

Implications for Teacher Education Programs

The results from this study suggest several directions for teacher education and professional development programs. In chapters 4 and 5, I said that the novice teacher was more reflective and more open to considering students' difficulties after being shown specific students' struggles with the tasks. Providing beginning teachers with opportunities to examine real students' difficulties can assist the teachers in thinking reflectively about students' struggles and in the design of efficient pedagogical strategies oriented to help students overcome such difficulties.

I showed in the analysis that the novice teacher was highly interested in analyzing those students' difficulties that resembled her students' difficulties. The exercise of having teachers analyze their own students' difficulties seems to promote deeper reflection about students' thinking.

I observed that the beginning teacher was eager to collect activities to do in her classroom; this is not only a characteristic of this particular teacher. Literature (Moir, 2004) has shown that beginning teachers are mainly concerned about building the curriculum by means of collecting activities to do in their classroom. Apparently the collection of activities gives them security. Teacher preparation programs and teacher development programs can promote the collection of resources only if that is combined with systematic reflection about the critical aspects of the activities, potential students' difficulties, contribution to the development of statistical concepts, and other teachers' experiences with the activities. If novice teachers have the opportunity to share their experiences about specific activities with other teachers and discuss the problematic features of the activity as well as the aspects that need to be redesigned, the teachers' collecting of activities would not be done just to have a repertoire of activities to pick from but also might help them to start constructing a critical pedagogical content knowledge. This study convinced me that deep, early, and continued reflection about students' difficulties may be a key element in teacher preparation. When teachers have reflected about specific matters in teaching, they are more likely to be aware the next time the situation is present.

This study was conducted in a learning community setting. Members of the learning community explored their own issues related to assessment, curriculum, pacing, and students' motivation, and they actively tried to resolve those issues. The teachers' positive responses to the learning community suggest that a professional development program following a similar format could be promising. Teachers would to meetings specific needs that might be discussed by a team of professionals with different levels of experience.

This analysis suggests that novice and expert teachers deal differently with students' difficulties which may reveal that they see the complexities of teaching differently. An approach to improving teacher education programs would be to explore novice teachers' ways of operating and then confront them with the work of expert teachers. This action could have different purposes; it could be to demonstrate exemplary practice, providing guidance for their teaching while the novice teachers work on gaining expertise.

Ideally, teacher preparation programs should be designed to guide novice teachers according to the stages of the expertise they are at. However, the evolutionary trajectory in the development of expertise is not sufficiently well defined to confidently design teacher's preparation programs.

The work in the development of expertise could start at the beginning of the teachers' preparation programs, when the teachers are doing student teaching so that they can have experience teaching a topic in several contexts and reflecting on the strategies that work well for their students and those that need to be redesigned. Teaching a topic for the first time is time consuming even for expert teachers, but the possibility of being systematic in collecting experiences and evidences of what work and what does not is crucial for development of expertise.

Results of the present study suggest that teacher preparation does not end when one obtains a degree. Novice teachers face new challenges in their classroom each day, but the scenario is especially hard for AP Statistics teachers who in most cases are alone in their schools. They do not have other AP Statistics teachers with whom to share their teaching experiences. Teacher professional development programs for AP Statistics teachers should be structured so that teachers have time for reflecting on their practice and time for feedback from other teachers. These programs should also consider offering the novice teachers support in different formats: (1) matching novice teachers with expert teachers during at least the first year of teaching so that the novice teacher has full responsibility for the classroom but relies on the expert teacher as a source of advice when needed, (2) creating support groups like the learning community described in this study in which the teachers benefited from the other teachers'' participation, (3) deciding to have the class taught by a pair of a novice and an expert AP Statistics teachers, which could be highly beneficial for the development of expertise.

Recommendations for Future Research

Many of the variables in this study might be investigated further in future research. The length of the study, for example, the topics explored, or the number of expert and novice teachers could be modified to get a better, more detailed picture of the nature of expertise in teaching AP Statistics.

Many studies exploring the contrast between expert and novice teachers have focused on the description of the characteristics of well-defined novices, usually student teachers, and well-defined experts at a single point in time or over a short period of time (including this study). There is a lack of longitudinal studies that explore the process by which novices become experts. We need to know the path of the process of becoming an expert to be able to orient teacher preparation programs and teacher professional development programs.

We know very little about the role of teacher preparation and professional development programs in the acquisition of expertise. It is necessary to explore how those programs contribute to the development of expertise. Expertise is associated not with the number of years of teaching but rather with the number of opportunities teachers have when they face the same situation over and over again and how they reflect on those situations to improve their understanding in their interaction with students.

The role of experience in the development of expertise is recognized but not understood. Longitudinal studies of teachers at various levels of expertise and experience might help us to understand the effect of experience in the development of expertise. These types of studies might help in identifying whether *the principles of teaching and learning ideas of chance* that came up in this study are consistent over time. Longitudinal studies might also help in recognizing the evolutionary trajectory of the process of becoming expert and when the *characteristics of expertise* generated in this study start becoming stable across different levels of expertise.

Future research in teachers' perceptions of students' ideas of chance could be enriched by classroom observations of teachers dealing with actual students' difficulties. There is a great deal of difference between what teachers say they do and what they actually do. Contrasting teachers' reflections about students' difficulties with the way they deal with real students' difficulties that appear in the classroom would be a way to validate findings from this study. It is also necessary that the distinctions I stated here among expert and novice in anticipating and explaining students' difficulties and in suggesting pedagogical strategies are investigated further with teachers at different levels of expertise and with other subjects to determine whether the characteristics identified in this study are representative of expertise.

Finally, there is a widespread belief that students have a better performance with expert teachers than with teachers who are not experts. It would be interesting to investigate this belief further and explore the effect of teachers' expertise on students' performance and whether students whose teachers have different levels of expertise perform differently.

REFERENCES

Agresti, A., & Franklin, C. A. (2006). *Statistics: The art and science of learning from data*. Upper Saddle River, NJ: Pearson Prentice Hall.

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

Begg, A., & Edwards, R. (1999, December). *Teachers' ideas about teaching statistics.* Paper presented at the combined annual meeting of the Australian Association for Research in Education and the New Zealand Association for Research in Education, Melbourne, Australia.

Benner, P. E. (1984). *From novice to expert: Excellence and power in clinical nursing practice*. Menlo Park, CA: Addison-Wesley.

Borko, H., Eisenhart, M., Brown, C. A., Underhill, R. G., Jones, D., & Agard, P. C. (1992). Learning to teach hard mathematics: Do novice teachers and their instructors give up too easily? *Journal for Research in Mathematics Education, 23*, 194–222.

Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal, 26*, 473–498.

Borovcnik, M., & Bentz, H.-J. (1991). Empirical research in understanding probability. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters: Probability in education* (Vol. 12, pp. 73–105). Dordrecht, The Netherlands: Kluwer.

Carter, K., Sabers, D., Cushing, K., Pinnegar, S., & Berliner, D. C. (1987). Processing and using information about students: A study of expert, novice, and postulant teachers. *Teaching and Teacher Education, 3*, 147–157.

Castro, C. S. (1998). Teaching probability for conceptual change. *Educational Studies in Mathematics, 35*, 233–254.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55–81.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152.

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *American Mathematical Monthly, 104*, 801–823.

College Board AP. (2007). *AP Statistics course description*. Retrieved March 3, 2008. from http://apcentral.collegeboard.com/apc/public/repository/52269_apstatistics_lo_4328.pdf

Crotty, M. (2004). *The foundations of social research: Meaning and perspective in the research process*. Thousand Oaks, CA: Sage.

DeSessa, A. A., & Wagner, J. F. (2005). What coordination has to say about transfer. In J. Mestre (Ed.), *Transfer of learning from a multidisciplinary perspective* (pp. 121–154). Greenwich, CT: Information Age Publising.

Drier, H. (2000). Children's probabilistic reasoning with a computer microworld. *Dissertation Abstracts International, 61*(06), 403A. (UMI No. 9975511)

Falk, R., & Bar-Hillel, M. (1983). Probabilistic dependence between events. *Two-Year College Mathematics Journal, 14*, 240–247.

Franklin, C. A., & Garfield, J. B. (2006). The GAISE project: Developing statistics education for grades pre-K–12 and college courses. In G. Burrill & P. C. Elliott (Eds.), *Thinking and reasoning with data and chance (Sixty-eighth Yearbook of the National Council of Teachers of Mathematics,* pp. 345–375). Reston, VA: NCTM.

Gal, I. (2005). Towards "probability literacy" for all citizens: Building blocks and instructional dilemmas. In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 39–64). New York: Springer.

Garfield, J. (1995). How students learn statistics. *International Statistical Review, 63*, 25–34.

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.

Garfield, J., & Ben-Zvi, D. (in press). The research on teaching and learning statistics. In J. Garfield & D. Ben-Zvi (Eds.), *Developing students' statistical reasoning: Connecting research and teaching practice*. Dordrecht, The Netherlands: Springer.

Garfield, J., & delMas, R. (1991). Students' conceptions of probability. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (Vol. 1, pp. 340–349). Voorburg, The Netherlands: International Statistical Institute.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Hidi, S., & Klaiman, R. (1983). Notetaking by experts and novices: An attempt to identify teachable strategies. *Curriculum Inquiry, 13*, 377–395.

Jones, G. A., Langrall, C. W., & Mooney, E. S. (2007). Research in probability: Responding to classroom realities. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 909–955). Charlotte, NC: Information Age.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1999). Students' probabilistic thinking in instruction. *Journal for Research in Mathematics Education, 30*, 487–519.

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

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review, 130*, 582–591.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction, 6*, 59–98.

Konold, C. (1995). Confessions of a coin flipper and would-be instructor. *American Statistician, 49*, 203–209.

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.

Lecoutre, M. P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics, 23*, 557–625.

Leinhardt, G. (1986). Expertise in mathematics teaching. *Educational Leadership, 43*(7), 28–33.

Leinhardt, G. (1989). Math lessons: A contrast of novice and expert competence. *Journal for Research in Mathematics Education, 20*, 52–75.

Leinhardt, G., & Greeno, J. G. (1986). The cognitive skill of teaching. *Journal of Educational Psychology, 78*, 75–95.

Leinhardt, G., & Smith, D. A. (1985). Expertise in mathematics instruction: Subject matter knowledge. *Journal of Educational Psychology, 77*, 247–271.

Leinhardt, G., Weidman, C., & Hammond, K. M. (1987). Introduction and integration of classroom routines by expert teachers. *Curriculum Inquiry, 17*, 135–176.

Livingston, C., & Borko, H. (1990). High school mathematics review lessons: Expert-novice distinctions. *Journal for Research in Mathematics Education, 21*, 372–387.

Magel, R. C. (1998). Using cooperative learning in a large introductory statistics class [Electronic version]. *Journal of Statistics Education*, *6*. Retrieved November 23, 2006 from http://www.amstat.org/publications/jse/v6n3/magel.html

Maxwell, J. A. (2005). *Qualitative research design: An interactive approach* (2nd ed.). Thousand Oaks, CA: Sage.

Moir, E. (2004). Phases of first-year teaching. *Teachers' Support Program.* Retrieved June 4, 2008, from http://www.teachersupportprograms.org/phases_first_year_teaching/

Moutier, S., & Houdé, O. (2003). Judgment under uncertainty and conjunction fallacy inhibition training. *Thinking and Reasoning 9*, 185–201.

Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children*. New York: Norton.

Polaki, M. V. (2002). Using instruction to identify key features of Basotho elementary students' growth in probabilistic thinking. *Mathematical Thinking and Learning, 4*, 285–313.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*, 211–227.

Raymond, A. M. (1997). Inconsistency between a beginning elementary school teacher's mathematics beliefs and teaching practice. *Journal for Research in Mathematics Education, 28*, 550–576.

Reading, C., & Shaughnessy, M. J. (2004). Reasoning about variation. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 201–226). Dordrecht, The Netherlands: Kluwer.

Ross, B. M., & DeGroot, J. F. (1982). How adolescents combine probabilities. *Journal of Psychology, 110*, 75–90.

Rossman, A., & Chance, B. (2003). Random babies. *Activity and web-based post-calculus probability and statistics project*, from http://www.mathspace.com/NSF_ProbStat/Teaching_Materials/rowell/final/prob%20random%20babies%20allan%20sp02%20v2.doc

Schwab, J. J. (1973). The practical 3: Translation into curriculum. *School Review, 81*, 501–522.

Schwandt, T. A. (1994). Constructivist, interpretativist approaches to human inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 118–137). Thousand Oaks, CA: Sage.

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*, 295–316.

Shaughnessy, J. M. (1992). Handbook of research on mathematics teaching and learning. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465–494). New York: Macmillan.

Shaughnessy, J. M. (2003). Research on students' understanding of probability. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 216–226). Reston, VA: National Council of Teachers of Mathematics.

Sternberg, R. J., & Horvath, J. A. (1995). A prototype view of expert teaching. *Educational Researcher, 24*(6), 9–17.

Stohl, H. (2005). Probability in teacher education and development. In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 345–366). New York: Springer.

Swanson, H. L., O'Connor, J. E., & Cooney, J. B. (1990). An information processing analysis of expert and novice teachers' problem solving. *American Educational Research Journal, 27*, 533–556.

Tentori, K., Bonini, N., & Osherson, D. (2004). The conjunction fallacy: A misunderstanding about conjunction? *Cognitive Science, 28*, 467–477.

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.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 30*, 293–315.

Von Glasersfeld, E. (1984). An introduction to radical constructivism. In P. Watzlawick (Ed.), *The invented reality* (pp. 17–40). New York: Norton.

Von Glasersfeld, E. (1987). Learning as a constructive activity. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics* (pp. 3–17). Hillsdale, NJ: Erlbaum.

Wagner, C. H. (1981). Paradoxes in sampling. In A. P. Shulte & J. R. Smart (Eds.), *Teaching statistics and probability* (1981 Yearbook of the National Council of Teachers of Mathematics, pp. 162–167). Reston, VA: NCTM.

Wagner, J. F. (2006). Transfer in pieces. *Cognition and Instruction, 24*, 1–71.

Watson, J. M., Kelly, B. A., Callingham, R. A., & Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology, 34*, 1–29.

Well, A. D., Pollatsek, A., & Boyce, S. (1990). Understanding the effects of sample size on the mean. *Organizational Behavior and Human Decision Processes, 47*, 289–312.

Yates, D. S., Moore, D. S., & Starnes, D. S. (2003). *The practice of statistics: TI-83/89 graphing calculator enhanced* (2nd ed.). New York: W.H. Freeman.

Zawojewski, J. S., & Shaughnessy, J. M. (2000). Data and chance. In E. A. Silver & P. A. Kenney (Eds.), *Results from the seventh mathematics assessment of the National*

Assessment of Educational Progress (pp. 235–268). Reston, VA: National Council of Teachers of Mathematics.