

HIGH SCHOOL MATHEMATICS TEACHERS' KNOWLEDGE AND VIEWS OF CONDITIONAL PROBABILITY

by

ROBERT ADAM MOLNAR

(Under the Direction of Jennifer J. Kaplan)

ABSTRACT

In this study, United States high school mathematics teachers were interviewed about topics related to conditional probability. U.S. mathematics teachers need to understand conditional probability to help students learn to make reasoned decisions under uncertainty and because many school curricula now include the topic. Researchers have identified probabilistic misconceptions held by learners and have developed some useful instructional approaches, but existing research on teachers' knowledge of probability is sparse. Therefore, I investigated (a) how teachers solve conditional probability tasks, (b) how teachers respond to student misconceptions in conditional probability tasks, and (c) teachers' perceived needs to teach conditional probability.

Between May and July 2014, I interviewed 25 teachers from Georgia, Pennsylvania, and South Carolina. The sample contained a few experienced probability instructors, but three-quarters of the sample had not taught a course on probability and statistics. The interview protocol included 9 task-based questions. Participants solved problems, identified potential student misconceptions, and offered responses to misconceptions. After the tasks, participants answered open-ended questions about curriculum, teaching concerns, and requests for assistance.

Problem solutions, identified misconceptions, and participants' responses to misconceptions were catalogued. Open-ended remarks were analyzed using thematic analysis.

Participants avoided most misconceptions in their solutions. Independence was an exception; most participants erroneously defined independent events. Participants recognized known student misconceptions on less complicated tasks, but had fewer ideas about more complex problems. They generally would respond to misconceptions with legitimate arguments. Their responses incorporated explanations of vocabulary, confrontations about computational errors, and some innovative approaches such as physical representations, classroom demonstrations, and analogies.

Participants considered probability practical and relevant, but courses often covered little probability because it was the last chapter in the book and a minor part of standardized tests. Participants divided teaching concerns and requests for assistance among subject matter, pedagogical, and curricular needs. The most commonly expressed needs were for classroom-ready tasks, task-based pedagogical training, and subject matter instruction.

The study results provide evidence to frame many research activities. The most pressing areas are unraveling misconceptions about independence, creating practical classroom tasks, and developing teacher training about multiple ways of responding to student misconceptions.

INDEX WORDS: Mathematics education, conditional probability, teachers, content knowledge, pedagogy, professional development

HIGH SCHOOL MATHEMATICS TEACHERS' KNOWLEDGE AND VIEWS OF
CONDITIONAL PROBABILITY

by

ROBERT ADAM MOLNAR

A.B., Harvard University, 1996

M.S., The University of Chicago, 2010

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

2015

© 2015

Robert Adam Molnar

All Rights Reserved

HIGH SCHOOL MATHEMATICS TEACHERS' KNOWLEDGE AND VIEWS OF
CONDITIONAL PROBABILITY

by

ROBERT ADAM MOLNAR

Major Professor: Jennifer J. Kaplan

Committee: Jeremy Kilpatrick
Patricia S. Wilson

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2015

ACKNOWLEDGEMENTS

The most important people in this dissertation are the 25 teacher participants. Although they sometimes doubted their own ability, they offered many helpful solutions, responses, and ideas. Without them, this dissertation would literally not be possible. Thank you.

The most important people in my life have been my parents, Robert and Gloria Molnar. I'm very glad you met in literature class. For this project, besides introducing me to mathematics teaching through your classrooms, you played the role of the first two pilot subjects.

My committee members have contributed to my development as a statistics educator. My major professor Jennifer Kaplan has provided guidance, edited efficiently, and managed to keep me mostly on task. Jeremy Kilpatrick showed me how to evaluate research literature in a gentlemanly manner. Pat Wilson has offered writing advice. On this project, many people helped recruit participants: Chris Franklin, Hamilton Hardison, Jennifer Kaplan, Clay Kitchings, Gloria Molnar, Robert Molnar, Teo Paoletti, and Ryan Smith. I am also grateful to Jessica Bishop for lending camera equipment; the three anonymous PhD student pilot subjects; and friendly editors Simone Brosig, Suehyeon Paek, and Anna Yu.

Among contributors to my professional life, two stand out. Working for Bill Fenton at Bellarmine, I learned about how to act as a professor in a competently run department. Working for Chris Franklin, I learned about how to present statistics instruction to teachers. Briefly, I will also mention my brother Ray, Wayne Seese, Herman Chernoff, Rob McCulloch, and Ruth Wagoner. To you and hundreds of others, thank you for being a part of my story.

I also acknowledge the permission of The College Board to include the text of the Survey and ELISA tasks from Advanced Placement (AP) Statistics exams. AP is a registered trademark of the College Board, which was not involved in the production of, and does not endorse, this product. Because I do not have permission to quote the full text of the questions in this publicly available version, in two places, I have included links to the AP website where you can find the full text of the problems.

Chapter 1 of this version also contains an additional update about two people, with information that was not available when the dissertation was submitted in July. I thank the relevant people for the updates.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	ix
LIST OF FIGURES	xi
CHAPTER	
1 INTRODUCTION	1
Probabilistic Decision-Making	2
Probability, Literacy, and Standards	4
Overview	6
2 LITERATURE REVIEW	8
History of Probability	9
Connections to Mathematics and Statistics	11
Educational Standards Documents	14
Probability in State Standards	18
Research on Learning Probability	22
Teacher Knowledge Frameworks	29
Research on Teacher Knowledge	32
Research on Teacher Views	35
Summary	37
3 METHODS	38

Selecting Tasks	39
Rash Task.....	43
Lucky Dip Task.....	45
Survey Task	48
Taxicab Task.....	50
ELISA Task	53
Open-Ended Questions	55
Participant Recruitment	57
Interview Protocol.....	59
Participant Demographics.....	60
Initial Analysis	65
Detailed Analysis.....	66
4 TASK RESULTS.....	69
Overall Subject Matter Results	69
Rash Results.....	74
Lucky Dip Results.....	80
Survey Part A Results	88
Survey Part B Results	92
Survey Part C Results	95
Taxicab Results.....	99
ELISA Part A Results	104
ELISA Part B Results	107
ELISA Part C Results	111

5	COMBINED AND OPEN-ENDED RESULTS	117
	Teacher Subject Matter Misconceptions	117
	Repeated Student Misconceptions	119
	Repeated Teacher Response Strategies	122
	Curricular Issues	126
	Expressed Concerns	129
	Sources of Support	133
	Requests for Assistance	134
6	DISCUSSION	139
	Discussion About Teacher Solutions	139
	Discussion About Responses to Student Misconceptions	141
	Discussion About Perceptions and Needs	143
	Delimitations and Limitations	144
	Future Directions	146
	REFERENCES	152

LIST OF TABLES

	Page
Table 2.1: Conditional Probability in State Curricula and Assessments	22
Table 3.1: Summary of Interview Questions	42
Table 3.2: Academic Degrees Listed by Participants	62
Table 4.1: Primary Solution Paths in Rash Question.....	75
Table 4.2: All Solution Paths in Rash Question	76
Table 4.3: Misconceptions Mentioned in Rash Question	78
Table 4.4: Primary Solution Paths in Lucky Dip Question	82
Table 4.5: All Solution Paths in Lucky Dip Question	83
Table 4.6: Misconceptions Mentioned in Lucky Dip Question	86
Table 4.7: Primary Solution Paths in Survey Part A Question	89
Table 4.8: All Solution Paths in Survey Part A Question	89
Table 4.9: Misconceptions Mentioned in Survey Part A Question	91
Table 4.10: Primary Solution Paths in Survey Part B Question	93
Table 4.11: All Solution Paths in Survey Part B Question	93
Table 4.12: Misconceptions Mentioned in Survey Part B Question.....	94
Table 4.13: Primary Solution Paths in Survey Part C Question	96
Table 4.14: All Solution Paths in Survey Part C Question	96
Table 4.15: Misconceptions Mentioned in Survey Part C Question.....	98
Table 4.16: Primary Solution Paths in Taxicab Question.....	100

Table 4.17: All Solution Paths in Taxicab Question.....	101
Table 4.18: Misconceptions Mentioned in Taxicab Question	103
Table 4.19: Primary Solution Paths in ELISA Part A Question	105
Table 4.20: All Solution Paths in ELISA Part A Question.....	105
Table 4.21: Misconceptions Mentioned in ELISA Part A Question	106
Table 4.22: Primary Solution Paths in ELISA Part B Question	108
Table 4.23: All Solution Paths in ELISA Part B Question	108
Table 4.24: Misconceptions Mentioned in ELISA Part B Question.....	110
Table 4.25: Primary Solution Paths in ELISA Part C Question	112
Table 4.26: All Solution Paths in ELISA Part C Question	113
Table 4.27: Misconceptions Mentioned in ELISA Part C Question.....	115
Table 5.1: Concerns Expressed by Teachers	130
Table 5.2: Categories of Requests for Assistance.....	135

LIST OF FIGURES

	Page
Figure 1.1: xkcd comic “Probability”	3
Figure 3.1: Rash task wording	43
Figure 3.2: Lucky Dip task wording	45
Figure 3.3: Survey task wording	48
Figure 3.4: Taxicab task wording	50
Figure 3.5: ELISA task wording	53
Figure 3.6: Participant background questionnaire	61
Figure 3.7: Participant probability and statistics course experience	63
Figure 4.1: Correct responses by teacher	70
Figure 4.2: Correct responses by teacher, not including ELISA task	71
Figure 4.3: Correct responses by question	73
Figure 4.4: Tree diagram for Lucky Dip question	82

CHAPTER 1

INTRODUCTION

To us, probability is the very guide of life.

—Joseph Butler, *Analogy of Religion*

In the 18th century, Bishop Joseph Butler wrote to defend the Christian concept of God against those who pointed out inconsistencies in the idea. Butler argued that the concept was not fully provable because a human argument could never be perfect. Unlike infinite intelligences, who never have uncertainty, finite humans can access only imperfect information. Each person should evaluate the data associated with hypotheses and then select the option that appears best. People must be guided by degrees of likely truth, a “presumption, opinion, or full conviction,” not moral certainty (Butler, 1736/1860, p. 83). After this preface, Butler then presented his argument for Christianity, which he considered “of weight on the side of religion” (p. 90).

As a theologian, Butler evaluated hypotheses using probability. He was not a scientist and did not seek to advance “the nature, the foundation, and measure of probability ... This belongs to the subject of logic, and is a part of that subject which has not yet been thoroughly considered” (Butler, 1736/1860, p. 85). Today, 279 years later, other men and women have considered the logic of probability much more thoroughly. Nevertheless, human knowledge about probability remains incomplete. Incomplete understanding comes not just from humans’ finite capacity; some topics have not yet been fully considered. In this dissertation, I examine one under-developed topic in probability: high school teachers’ knowledge and views about conditional probability.

I selected conditional probability because of its relevance in decision-making. I decided on a population of high school teachers because recent changes in curricular standards have demanded high school teachers know more about teaching probability. In this chapter, I introduce probabilistic decision-making through the vital example of medical treatment. I then briefly describe the teaching context before presenting my three research questions. I conclude by outlining the structure of this dissertation.

Probabilistic Decision-Making

Theologian Butler (1736/1860) used the language of probability to conclude the existence of a supreme being. He recommended that serious persons “set down every thing which they think may be of any real weight at all in proof” (p. 300) which might include claims for the contrary side. Not everyone had the same data, because “God has afforded to some no more than doubtful evidence of religion” (p. 318). To deal with the data problem, he had provided positive evidence and answered negative objections so that people would know that “the moral system of nature, or natural religion, which Christianity lays before us, approves itself, almost intuitively, to a reasonable mind, upon seeing it proposed” (pp. 318–319).

Many scientists have also written about natural theology, the origin and purpose of the world. Discussion about the role of chance persists in science, particularly in the fields of quantum mechanics and evolutionary biology. On the other hand, theologians now rarely invoke probability (Bartholomew, 1988). People now apply probability to worldly questions, not otherworldly ones. Real-world applications of probability include gambling, purchasing insurance, legal testing, evaluating machine defects, and medical testing (Gal, 2005; Rossman & Short, 1995).

As an example, every person must make medical treatment choices based on symptoms, rather than with perfect knowledge of the body. Gal (2005) offered an insurance example in which an overweight 30 year old with a slight heart murmur had to decide whether to purchase life insurance. Would it be better for the decision maker to purchase insurance immediately, or to wait until after a visit with a heart specialist, whose news could affect the risk level? Or should the decision maker try to go on a diet before attempting to purchase insurance at all?

Health decisions with uncertain outcomes are often matters of life and death. Comic artist Randall Munroe became immersed in probability logic after his fiancée was diagnosed with Stage 3 breast cancer. As he pondered the consequences of her disease, he published a comic about probability that appears as Figure 1.1.

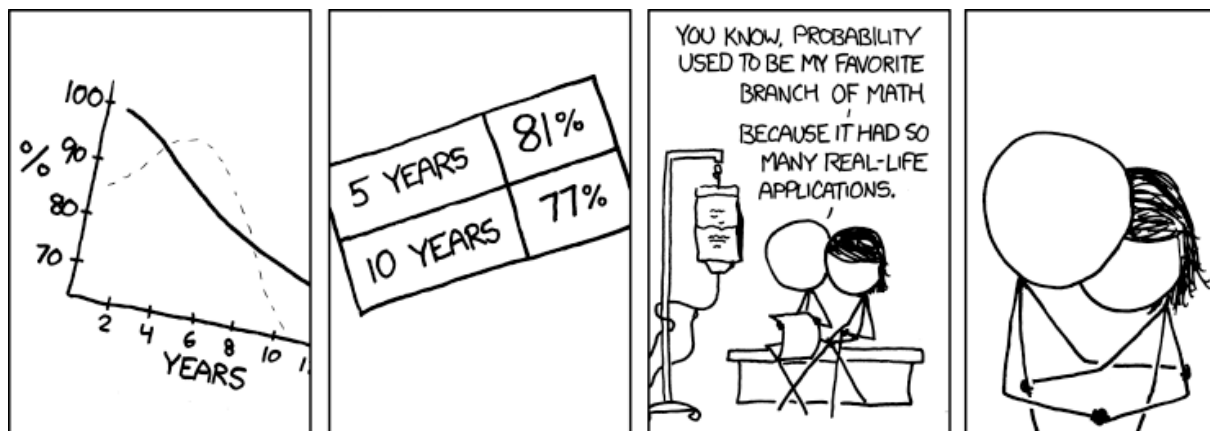


Figure 1.1. xkcd comic “Probability.” From “Probability” (<http://xkcd.com/881/>). Copyright 2011 by Randall Munroe. Reprinted with permission under a Creative Commons Attribution-NonCommercial 2.5 license (<https://creativecommons.org/licenses/by-nc/2.5/>)

Munroe’s story had a cheerful ending; his wife is currently in remission. It’s easy to find more medical decision stories. As I write in May 2015, on Facebook I see pictures posted of my

35-year-old friend in a blue wig, ringing the bell at the end of her chemotherapy for Stage 1 breast cancer. She and her oncologist decided on chemotherapy after considering her condition, age, and family history. I also read treatment updates from my college roommate, who in February was diagnosed with bone cancer in his leg. After his initial diagnosis, we had a conversation about probability and treatment options. He told me that, conditional on the usage of chemotherapy and radiation, the cure rate for similar cases is between 80% and 95%. In pleasant news, it appears that his course of chemotherapy has been successful.

[As an update, the friend with breast cancer completed radiation in July 2015 and now takes an anti-cancer drug. In August 2015, my college roommate was confirmed as free of cancer; he remains healthy as of June 2016.]

Probability, Literacy, and Standards

The term *literacy* is used to describe a minimal set of skills expected of all citizens. For instance, I believe that understanding enough probability to make rational medical decisions is an essential component of probability literacy. Probability literacy skills are commonly combined with data collection, inference, and other statistical topics under the broader heading of statistical literacy. Statistical literacy has been advocated for generations. In 1951, S. S. Wilks, then president of the American Statistical Association, asserted that “statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write” (p. 5). To incorporate statistical literacy into schools, Wilks proposed “eliminating some of the fossilized subject matter from high school algebra, trigonometry, and particularly solid geometry and replacing it by subject matter from elementary probability, statistics, and logic” (p. 12).

United States school mathematics standards have changed since Wilks’s presidency. National committees have consistently recommended more probability and statistics content, but

until 2010 curricular changes occurred slowly, on a state-by-state or district-by-district basis. During 2010, there was a disruptive shift. Over 40 state governments adopted the Common Core State Standards developed by two nonprofit associations: the National Governors Association (NGA) and the Council of Chief State School Officers (CCSSO). These standards, generally called the Common Core, were designed to prepare all students for college or a career (NGA Center for Best Practices & CCSSO, 2010), establishing them as standards of literacy in English language arts and mathematics. The Common Core high school mathematics standards include a section on conditional probability and the rules of probability. I chose to investigate topics in the conditional probability section because the standards include probability literacy; specifically, students should “recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations” (p. 82). If implemented well, knowledge of probability will help people better guide their lives.

For my population of interest, I chose practicing high school mathematics teachers. Working teachers bear the responsibility of implementing curriculum changes like the Common Core standards. They must know how to solve problems. They must also have the pedagogical ability to assist students. Teachers with less training in probability might not have this expertise. Teachers with prior probability training, but no recent practice, can also use support. As one teacher said after our interview, “[Probability is] not one of those things where you throw an equation at me and I can solve it. Seeing some of those problems I knew what I needed to do, but I’m like, ‘oh’, just because you forget.”

My long-term goal is to develop probability problems, guides, and related materials. Instructors can use these tools to better aid students. Unfortunately, I know of no research studies that have asked high school teachers exclusively about probability. We do not know what

teachers do and what needs they have. Therefore, in the language of the American Statistical Association report on using statistics effectively in mathematics education research (Scheaffer & Smith, 2007), this dissertation frames a research program about teacher knowledge of conditional probability. Ideally, hypotheses developed from these results will be examined in more extensive samples.

In order to find information to frame the research program, I decided on a qualitative exploration. I interviewed a total of 25 high school mathematics teachers across Georgia, Pennsylvania, and South Carolina. Interviews contained five tasks that covered Common Core standards on conditional probability and the rules of probability, plus a few open-ended questions. Interview data was analyzed to provide information about three research questions. Two questions were about knowledge—specifically, how teachers solved problems themselves and responded to student misconceptions. The third research question asked about teachers' views about their perceived needs. The research questions are formally stated as follows:

1. How do high school mathematics teachers solve conditional probability tasks?
2. How do teachers respond to student misconceptions in conditional probability tasks?
3. What do teachers perceive as their needs to be prepared to teach conditional probability?

Overview

In this chapter, I introduced probability as a guide for better decision-making and illustrated its importance using examples from the field of medicine. After the adoption of Common Core standards, many high school mathematics teachers need to develop expertise in conditional probability. Since little prior research has been published on teacher knowledge of probability, I chose to qualitatively investigate this subject using face-to-face interviews.

In Chapter 2, I include more background information: more precise definitions for terms related to conditional probability, school curriculum standards, and a review of prior research on probability and teachers. In Chapter 3, I present study methods, including the interview tasks. I also describe the participants and summarize background information from a questionnaire.

Chapters 4 and 5 contain results from analysis. Tasks are individually examined in Chapter 4. For each task, I discuss how the teachers solved the problem and how they responded to student misconceptions. In Chapter 5, I summarize content knowledge results and pedagogical patterns across tasks. Also in Chapter 5, I present the teachers' expressed concerns and requests for assistance to teach conditional probability. I also summarize what the teachers said about current curriculum standards. Finally, in Chapter 6 I reflect on the implications of the analysis, discuss the limits of what this analysis can show, and present future research directions.

CHAPTER 2

LITERATURE REVIEW

As I mentioned in Chapter 1, the research questions in this dissertation were designed to investigate high school teachers' knowledge and views about conditional probability. This chapter contains eight sections of relevant background information, with a summary at the end of the chapter. First, historical references about probability are listed, introducing topics included in this dissertation such as independence, conditional probability, Bayes' rule, and probability axioms. Probability has connections to mathematics and statistics; the connections between the three subjects are described in the second section. These connections matter because in school curricula, probability frequently appears alongside statistics as part of mathematics. In the third and fourth sections, references to probability in curricular standards documents are listed, with an emphasis on the Common Core and official documents from the three states from which participants were drawn – Georgia, Pennsylvania, and South Carolina. Standards related to conditional probability appear in all three state mathematics syllabi, indicating the need for mathematics teachers to know about the subject.

Although a study like this one has not been previously conducted, prior research exists on misconceptions when learning probability, models of teacher knowledge, teacher knowledge about probability, and teacher views about probability. The fifth through eighth sections contain summaries of prior work on each topic.

History of Probability

Although the discipline of probability has a relatively short history compared to parts of mathematics like Euclidean geometry, games of chance have a very long record. For instance, excavated Egyptian tombs from 3500 B.C. contain board games with thrown astragali bones (David, 1998), and Book 1 of the Hindu epic *Mahabharata* refers to a kingdom lost in an unfair game of dice. Nonetheless, few records exist of quantitative study on probability before about 1600. Before this time, many people in Christian Europe perceived future events as outside their control. Future events were the result of luck or the actions of a mysterious power. The philosophical shifts of the Protestant Reformation increased the role of human agency and led scientists to consider chance in a systematic way (Bernstein, 1996). Early writings include Cardano's manuscript on games of chance from about 1560, Galileo's computations about the sum of three dice from about 1620, correspondence between Fermat and Pascal from 1654, and Huygens' book on games of chance from 1657 (David, 1998). Jacob Bernoulli's 1713 *Ars Conjectandi*, which included permutations, combinations, and problems outside games, is now regarded as the beginning of mathematical theory on probability (Stigler, 1986). David (1998) mentioned that *Ars Conjectandi* included a few instances in which Bernoulli pointed out potential errors in reasoning, which is an early example of considering student misconceptions about probability.

Probability theory was extended by De Moivre in *The Doctrine of Chances* (1756). Importantly for this dissertation, De Moivre included definitions for independent events and dependent events because "the terms independent and dependent might occasion some obscurity" (p. 6). Two events are independent when "the happening of one neither forwards nor obstructs the happening of the other" (p. 6). Then, "two Events are dependent, when they are so

connected together as that the Probability of either's happening is altered by the happening of the other" (p. 6). De Moivre also included definitions for *conditional probability* and the *general multiplication rule* (pp. 7–8). Over 250 years later, these definitions remain applicable.

Another work still relevant after 250 years was first read at London's Royal Society in December 1763. Bayes considered a problem about the binomial distribution drawn from De Moivre's book. His solution involved balls thrown with uniform probability on a table. First, a single marker ball was tossed. After that, multiple throws of a second ball were made, with successes recorded if the second ball came to rest to the right of the first ball. His two major conclusions did not explicitly state what modern textbooks call Bayes' theorem. Rather, Bayes showed how to update the subjective probability of success given an initial prior guess and the available data from the second ball (p. 392). He also proposed an inferential rule to find the chance that the unknown true probability of success lies between two values, in modern symbols $P(a < \theta < b)$ (p. 399).

Bayes's work received little attention after its publication. The complex geometric reasoning meant that few scholars applied his inferential argument. Instead, Laplace's later algebraic argument became the foundation of inverse probability and statistical population inference (Stigler, 1986). Bayes's other conclusion—how to update subjective probability—was and still is controversial because it explicitly included a guess. Statisticians often categorize themselves as either subjective Bayesians who accept guessing or nonsubjective frequentists who do not. In the 19th century, the subjective approach dominated practice; in the 20th century, nonsubjective research was prominent (Efron, 2005). The school standards documents described later align with 20th century nonsubjective practice, as did international curricular documents examined by Jones, Langrall, and Mooney (2007). Since subjective probability does not appear

in school mathematics curricula, problem solutions in this dissertation apply Bayes' theorem only with nonsubjective information.

During the 19th and early 20th century, probability theory continued to develop. In 1880, the probabilist Venn expanded the "Eulerian circle" (p. 1) into what is now known as the Venn diagram. Formal axioms did not emerge until the 20th century, when Komolgorov (1933) codified the axioms generally held today: Probability must be nonnegative; the full sample space has probability of 1; and the probability of the union of countably infinitely many disjoint sets equals the sum of the individual set probabilities.

Connections to Mathematics and Statistics

Although modern authors refer to the field of probability, early authors such as Huygens, De Moivre, and Bayes often used the word *chance*. The word *probability* originally did not have a mathematical definition. According to the *Oxford English Dictionary*, beginning in the 1400s the word *probability* referred to the appearance of truth: something being more likely than not. The current definition was adapted from the statement about belief ("Probability," 2015).

The term *statistics* also originally did not have a fully mathematical definition. Statistics in the 1700s and 1800s was considered part of political science. Statistics referred to the collection and classification of facts bearing on the condition of a community or state ("Statistics," 2014). These facts could be numeric or alphabetic. Over time, the word became more strongly associated with numeric information, and areas of application expanded beyond government.

Mathematics, probability, and statistics have an intertwined relationship. Both statistics and probability are mathematical sciences that require mathematical tools such as number sense, arithmetic, and algebra. In probability, mathematical formalism enables axioms free of any

subjective or nonsubjective philosophical interpretation. Probability instruction must utilize sample spaces, power series, integrals, and other things. “A pure experimental approach is not sufficient in the teaching of probability” (Batanero, Henry, & Parzysz, 2005, p. 33).

In addition to tools from mathematics, statistics also requires tools from probability. The *Guidelines for Assessment and Instruction in Statistics Education (GAISE) K-12* report listed probability topics utilized in the recommended school statistics curriculum, such as chance variation, relative frequency, independence, expected values, the normal curve, and other probability distributions (Franklin et al., 2007, pp. 85–87). Probability also draws from statistics, particularly when data collection provides the basis for assignment of subjective prior probabilities.

Many authors have produced results in more than one discipline. Bernoulli’s *Ars Conjectandi*, a foundational text in probability theory, also attempted statistical inference about binomial proportions and introduced Bernoulli numbers into mathematical number theory (David, 1998). Although Bayes’ theorem is probabilistic, Bayes’s (1763) paper also proposed an inferential statistics rule on the true chance of success.

Given crossover among tools and authors, how can the three mathematical sciences be distinguished? One way is through a political quote from former U.S. Secretary of Defense Donald Rumsfeld. Though Rumsfeld intended to describe levels of certainty in military reports, his phrases can distinguish the mathematics, probability, and inferential statistics in school curricula. Each sentence corresponds to a research discipline, noted inside brackets.

There are known knowns; there are things we know we know. [Mathematics]
 We also know there are known unknowns; that is to say we know there are some things we do not know. [Probability]
 But there are also unknown unknowns -- the ones we don't know we don't know. [Statistics] (Federal News Service, 2002, words in brackets added)

The two distinguishing assumptions are knowledge about the population and knowledge about randomness. In mathematics and probability, the population structure is known, but in statistics, only a sample is available and the population is unknown. Mathematics does not have random variation; given a mathematical model, the result of applying the model is known. On the other hand, probability and statistics have random variation; given a probability model or statistical sampling model, the result of applying the model is unknown. Thus, mathematics has known populations and fixed known results; probability has known populations but varying unknown results; statistics has unknown populations and varying unknown results.

Since probability shares one distinguishing assumption with mathematics and one with statistics, intellectual classifications can place it with either discipline. For example, at the University of Georgia, probability instruction is offered by both the mathematics department in Math 6600 and the statistics department in Stat 6810 (University of Georgia, 2015). Most designations group probability with mathematics, because of the type of logical reasoning associated with probability. Mathematics and probability rely on deductive reasoning; they involve starting from a known general process or population and proceeding to make claims about a specific trial or sample. Certain proof can result. Doing statistics involves inductive reasoning, starting from a specific trial or sample and proceeding to make claims about a general process or population. While opening many real-world scenarios to analysis, inductive conclusions cannot assert pure certainty, since part of the population model remains unknown.

Statistics thought leaders have tried to sever basic statistics from mathematics and probability because of the shift in logical reasoning combined with introduction of contextual data. The GAISE K-12 recommendations claimed that at the primary and secondary school levels, learners of statistics need only limited formal mathematics and no formal probability. For

them, precollege statistical training requires only intuitive knowledge of probability. “Probability plays an important role in statistics, but formal mathematical probability should have its own place in the curriculum” (Franklin et al., 2007, p. 9). College professors Cobb and Moore (1997) argued that “first courses in statistics should contain essentially no formal probability theory” (p. 820), since only informal probability is required for an understanding of inference. Cobb and Moore do admit, however, that “probability is important in its own right” (p. 822).

Educational Standards Documents

Despite the distinguishing points and attempts to separate the disciplines, United States school curricula and standards documents have tended to place probability and statistics as fields within the mathematics curriculum. Over the past 50 years, many national committee reports on mathematics curriculum have included sections on probability. For example, the 1963 Cambridge Conference on School Mathematics proposed an ambitious schedule, introducing trigonometric functions in Grade 6 and multiple integration in high school. In this accelerated curriculum, the authors made sure to include probability relatively early in secondary school, during Grade 8, since “an elementary feeling for probability and statistics” is “of particular importance” for all students (Educational Services Incorporated, 1963, p. 9). Probability contributes to liberal education: “It can raise the level of sophistication at which a person interprets what he sees in ordinary life, in which theorems are scarce and uncertainty is everywhere” (p. 70). In a 1977 position paper, the National Council of Supervisors of Mathematics included elementary notions of probability as part of the basic skill of prediction needed by all students. The National Council of Teachers of Mathematics (NCTM) included probability and statistics content in all grades of the 1989 *Curriculum and Evaluation Standards*, separating probability from statistics beginning in Grade 5. The 1989 NCTM writers noted that

probability theory underpinned the modern world, which made an understanding of it an essential aspect of citizenship. One decade later, the *NCTM Principles and Standards* included “Data Analysis and Probability” as one of the five content strands in all grades, suggesting that conditional probability topics such as mutually exclusive, joint, and conditional events be studied in high school (National Council of Teachers of Mathematics, 2000).

The most recent attempt at nationwide guidelines was the Common Core State Standards Initiative, generally known as the Common Core. Two nonprofit associations, the National Governors Association (NGA) Center for Best Practices and the Council of Chief State School Officers (CCSSO), sponsored the initiative. Association members are listed on the NGA website (www.nga.org) and the CCSSO website (www.ccsso.org). Although the CCSSO includes a representative from the Department of Defense, both the NGA and the CCSSO are nonfederal, nonpartisan associations. Neither has a mandate to make policy or law.

The Common Core website includes an official project history (Common Core State Standards Initiative, 2015). Preliminary discussions in 2007 and 2008 led to a report called *Benchmarking for Success: Ensuring U.S Students Receive a World-Class Education* (NGA, CCSSO, & Achieve, Inc., 2008). After providing evidence suggesting the need to improve American education, including international test score comparisons, the report proposed five action steps: (a) Upgrade state standards, (b) Ensure materials are aligned to the new standards, (c) Revise teacher support policies, (d) Hold schools accountable, and (e) Measure student performance in an international context. The Common Core standards are developed from the first action step, “Upgrade state standards by adopting a common core of internationally benchmarked standards in math and language arts for grades K-12 to ensure that students are

equipped with the necessary knowledge and skills to be globally competitive” (NGA, CCSSO, & Achieve, Inc., 2008, p. 24).

Actions developing standards for mathematics and English/language arts took place during 2009 and 2010. Although many people served on development and validation committees, just three men worked as principal writers in mathematics: Phil Daro, William McCallum, and Jason Zimba (Garland, 2014). The authors attempted to write standards with three characteristics common to the standards of high-performing nations. First, the Common Core would have greater focus on fewer topics each year. Second, topics across grades would have an orderly, coherent progression. Third, the new guidelines would have authentic rigor in combining conceptual understanding, procedural skills, and application (NGA, CCSSO, & Achieve, Inc., 2008, p. 24). After comments and revisions, final standards were released in June 2010 (Common Core State Standards Initiative, 2015).

Initial adoption proceeded very rapidly, in large part driven by the federal government. The Race to the Top Initiative offered states a total of 4 billion dollars for education reform. The scoring system awarded extra points to states who adopted college and career readiness standards before August 2010. Given the short time frame, 41 states chose to adopt the Common Core mathematics standards. Further adoptions followed; in late 2012, only five states had not adopted the Common Core mathematics standards (ASCD, 2012). The three states represented in this dissertation—Georgia, Pennsylvania, and South Carolina—had all signed on.

Since 2012, protest has arisen around the Common Core guidelines, and political considerations have led several states to withdraw usage of these standards. In this dissertation, only changes in states where interviews occurred are described. During the 2013–2014 school year, the year immediately preceding these interviews, all three states operated under standards

aligned with the Common Core. Nevertheless, between August 2013 and April 2015, all three states made changes, either in guidelines or in assessment methods. In the remainder of this section, probability in the Common Core is described. The following section details standards and recent changes in Georgia, Pennsylvania, and South Carolina.

The Common Core writers introduced a strand called “Statistics and Probability” in Grade 6. In earlier grades, a small number of topics under “Measurement and Data” related to statistics, but none referenced probability. Topics in Grades 6 and 8 are primarily statistical. Grade 7 contains the first reference to probability. Four standards introduced probability models, tables, tree diagrams, and simulation (NGA Center for Best Practices & CCSSO, 2010, pp. 50–51).

In high school, the Statistics and Probability strand includes four areas. Two areas, “Interpreting Categorical and Quantitative Data” and “Making Inferences and Justifying Conclusions,” do not reference probability. The other two sections, “Conditional Probability and the Rules of Probability” and “Using Probability to Make Decisions,” do reference probability (NGA Center for Best Practices & CCSSO, 2010, pp. 82–83). In the sections on probability, 9 of 16 standards begin with a (+) symbol, which indicates importance for advanced mathematics preparation but not necessarily for all students. The first seven standards in the conditional probability section are intended for all college and career ready students. As mentioned in the introduction, the goal of the present study was to investigate a topic needed by adults for effective citizenship. Because the Common Core authors considered conditional probability necessary for all students, selected problems are related to the section on rules of probability. The relevant standards are listed below.

1. Describe events as subsets of a sample space (the set of outcomes) using characteristics (or categories) of the outcomes, or as unions, intersections, or complements of other events (“or,” “and,” “not”).
2. Understand that two events A and B are independent if the probability of A and B occurring together is the product of their probabilities, and use this characterization to determine if they are independent.
3. Understand the conditional probability of A given B as $P(A \text{ and } B)/P(B)$, and interpret independence of A and B as saying that the conditional probability of A given B is the same as the probability of A , and the conditional probability of B given A is the same as the probability of B .
4. Construct and interpret two-way frequency tables of data when two categories are associated with each object being classified. Use the two-way table as a sample space to decide if events are independent and to approximate conditional probabilities. *For example, collect data from a random sample of students in your school on their favorite subject among math, science, and English. Estimate the probability that a randomly selected student from your school will favor science given that the student is in tenth grade. Do the same for other subjects and compare the results.*
5. Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations. *For example, compare the chance of having lung cancer if you are a smoker with the chance of being a smoker if you have lung cancer.*
6. Find the conditional probability of A given B as the fraction of B 's outcomes that also belong to A , and interpret the answer in terms of the model.
7. Apply the Addition Rule, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$, and interpret the answer in terms of the model.
8. (+) Apply the general Multiplication Rule in a uniform probability model, $P(A \text{ and } B) = P(A)P(B|A) = P(B)P(A|B)$, and interpret the answer in terms of the model.
9. (+) Use permutations and combinations to compute probabilities of compound events and solve problems.

(NGA Center for Best Practices & CCSSO, 2010, p. 82)

Probability in State Standards

The three states where data were collected for this study included some rules of probability in their 2013–2014 mathematics standards, though not necessarily all of the Common Core probability topics as worded in the document. In Georgia, the Common Core Georgia

Performance Standards include all nine standards worded almost exactly as they appear in the Common Core State Standards, with only minor notation changes (Georgia Department of Education, 2011). Standards 1 through 7 appear in the class titled Analytic Geometry (Georgia Department of Education, 2012b). For accelerated students, Standards 1 through 7 appear in Accelerated Geometry B with Advanced Algebra (Georgia Department of Education, 2012a, 2012c). Advanced Mathematics Standards 8 and 9 are postponed until Pre-calculus (Georgia Department of Education, 2012b, 2012d).

To graduate from high school in Georgia, students must attain a passing grade in Analytic Geometry. During 2013–2014, Georgia had a mandatory End-of-Course Test in Coordinate Algebra and Analytic Geometry. The test counted for 20% of a student’s final mark, and included conditional probability questions on the syllabus (Georgia Department of Education, 2013). In 2014–2015, the state switched to Georgia Milestones assessments, which still counted for 20% of the final course grade. Statistics and Probability, almost completely conditional probability, made up approximately 11% of the content on the Analytic Geometry exam (Georgia Department of Education, 2014).

Pennsylvania began the 2013–2014 school year with the Pennsylvania Common Core Standards (Pennsylvania Department of Education, 2013). During the school year, Pennsylvania dropped the word *common* and published the Pennsylvania Core Standards in multiple subjects. In both documents, probability appears as part of Section 2.4, “Measurement, Data, and Probability.” Bivariate data using frequencies appears in Grade 8, similar to Common Core Standard 4. Other topics related to the rules of probability are placed under high school standards. For this dissertation, the two relevant items are CC.2.4.HS.B.6, “Use the concepts of independence and conditional probability to interpret data,” and CC.2.4.HS.B.7, “Apply the rules

of probability to compute probabilities of compound events in a uniform probability model” (Pennsylvania Department of Education, 2013, p. 16; 2014a, p. 16). The wording is not the same as in the Common Core standards.

Each relevant Pennsylvania standard has assessment anchors for standardized tests, called Keystone Exams. Lawmakers mandated that exams be developed in multiple subjects, including Algebra I, Algebra II, and Geometry. Starting in the 2016–2017 school year, in order to graduate from high school a student must ordinarily demonstrate proficiency through an objectively validated assessment in English Literature, Biology, and Algebra I. (Academic Standards and Assessments, 2014). Before 2017, in order to graduate a student must demonstrate proficiency on some state or local mathematics exam, which could include the Algebra I Keystone Exam.

Almost all assessment anchors related to conditional probability in the Pennsylvania standards are on the non-mandatory Algebra II exam. Only one anchor appears in Algebra I, A1.2.3.3.1, “Find probabilities of compound events (e.g., find probability of red and blue, find probability of red or blue) and represent as a fraction, decimal, or percent” (Pennsylvania Department of Education, 2014b, p. 13). This Algebra I anchor incorporates content from Common Core Standards 1, 4, and 7, but not independence or conditioning. Independence and conditioning appear in Algebra II.

In August 2013, South Carolina was in the final year of transition to Common Core standards, but political debate was serious. In May 2014, the legislature passed and Governor Nikki Haley signed a bill which mandated new standards for the 2015–2016 school year. After the new standards were approved in March 2015, school districts had to make another program change (South Carolina Department of Education, 2015).

All South Carolina participants taught at schools within one school district, which had transitioned to course progressions based on the Common Core standards. In South Carolina, students must attain four high school mathematics credits to graduate. State regulations list courses that schools must offer. South Carolina had a state high school exit exam in 2013–2014 (which was eliminated by another legislative act), but not end-of-course tests in mathematics (Defined Program, Grades 9-12 and Graduation Requirements, 2013).

The South Carolina Department of Education began with the Common Core standards as written. It created a spreadsheet, assigning each standard a place in two pathways – integrated mathematics for the technologies, and Algebra 1/Geometry/Algebra 2. No high school probability standards were placed in the first three courses of either high school pathway. South Carolina provided four options for the fourth course at the high school level: Fourth Course, Probability and Statistics, Discrete Math, and Pre-Calculus. All nine conditional probability standards appeared in the first three options, but not in Pre-Calculus (South Carolina Department of Education, 2013). Therefore, unlike Georgia and Pennsylvania, in South Carolina a high school student could graduate without ever having studied conditional probability.

Table 2.1 summarizes state curricula and assessments. It indicates whether the state included Common Core guidelines primarily as written or with modifications, the primary course location, and how much conditional probability was included on mandatory student assessments.

Table 2.1

Conditional Probability in State Curricula and Assessments

State	Common Core as written?	Primary location	Mandatory student assessment
Georgia	Yes	Analytic Geometry	11% of Analytic Geometry exam
Pennsylvania	No, modified	Algebra II	Minor part of Algebra I exam
South Carolina	Yes	Fourth courses, but not Pre-Calculus	None

Research on Learning Probability

As mentioned earlier, the field of probability has connections to mathematics and statistics. Therefore, it is not surprising that research about teaching and learning probability sometimes appears in mathematics education journals, sometimes in statistics education journals, and sometimes elsewhere. No journals exclusively cover probability learning. Sessions on teaching and learning probability occur at statistics and mathematics conferences, such as the International Conference on Teaching Statistics (ICOTS) and the International Congress on Mathematical Education (ICME). Although multiple perspectives provide better background, the lack of a defined research home might lead to a lack of attention. Egan Chernoff, co-chair of the next ICME topic study group on probability, wrote recently, along with Gale Russell, that there is a “documented dearth of research on teachers’ probabilistic knowledge” (Chernoff & Russell, 2013).

Before becoming instructors, teachers learned conditional probability as students. Following the same order, I discuss research on learning probability before research on teaching. This review is not comprehensive, focusing on seven misconceptions and biases related to conditional probability topics within state curricula: equiprobability, randomness, independence,

conjunction, time-axis causality, confusion of the inverse, and missing base rate. More general reviews include Kahneman, Slovic, and Tversky (1982), Fischbein and Schnarch (1997), and Jones, Langrall, and Mooney (2007). Also, research conducted on individual tasks is discussed in the next chapter, accompanying the introduction of each task.

Equiprobability. As I mentioned in the section on history, early authors on probability wrote frequently on games of chance. For devices in games of chance like cards and dice, generally every outcome is assumed to have the same probability. For example, if there are six balls inside a bag, published solutions almost always presume the chance of selecting each ball is exactly $1/6$. Such devices are called *fair*. The assumption that all situations have fair outcomes is called *equiprobability*. In some cases, such as fair gambling devices, equiprobability holds. In other cases, authors including Laplace have applied equiprobability to simplify calculations. Nevertheless, many random processes are not fair; each outcome does not occur with equal chance. In these situations, assuming equiprobability will lead to incorrect answers. Many young children incorrectly believe random experiments should always be fair. A significant minority of high school students, even those with substantial probability background, persist in this belief (Lecoutre, 1992). On the positive side, Lecoutre demonstrated that a planned sequence of problems could lessen this bias.

Randomness. Many participants in Lecoutre's (1992) studies argued for equiprobability because "random events should be equiprobable by nature" (p. 561). These participants did not have a proper probabilistic view of the term *random*. Students often struggle with the word *random*, because in everyday language it has another meaning. In probability, a process becomes random when the sample space of possible outcomes is known and a probability distribution is defined on possibilities in the sample space. Individual outcomes are unknown, but come from a

known regular distribution. Building a random variable is not haphazard or unplanned. For example, possessing a physical six-sided number cube is not sufficient to define a random process. The sample space consists of the numbers $\{1, 2, 3, 4, 5, 6\}$, but a probability distribution is still needed. A number cube typically has equiprobable outcomes, but an incorrect assumption could lead to great losses, as in the Mahabharata.

In everyday colloquial English, the word *random* stands for haphazard, weird, or unusual. Randomness does not come from a known process; it arises without known reason. When asked, college students most frequently offered the everyday definition (Kaplan, Fisher, & Rogness, 2010). Although student understanding improved when an instructor repeatedly contrasted definitions of the term within an introductory statistics course (Kaplan, Rogness, & Fisher, 2014), doing so required careful attention.

Independence. Another term requiring careful attention to definitions is *independence*. For the probability of events, the words *dependence* and *independence* have the same meanings as they did in the 1700s. Two events are dependent if the “probability of either’s happening is altered by the happening of the other” (De Moivre, 1756, p. 6). Two events are independent if the occurrence of one event does not affect the probability of the other. When two events A and B are independent, the joint probability equals the product of the individual probabilities: $P(A \text{ and } B) = P(A) * P(B)$ (NGA Center for Best Practices & CCSSO, 2010, p. 82). When events have nonzero probabilities, independence can also be expressed conditionally; the conditional probability of Event A, given the occurrence of Event B, must equal the unconditional probability of Event A: $P(A | B) = P(A)$.

In colloquial English, although *independence* can refer to freedom from conditioning, the first dictionary definition refers to control. The adjective *independent* refers to something “not

depending on the authority of another, not in a position of subordination or subjection; not subject to external control or rule; self-governing, autonomous, free” (“Independent,” 2015) Authority has a stronger implication than effect. Rumsey (2008) hypothesized that the everyday concept of independence as autonomy leads people to a common error, incorrectly defining independent events as mutually exclusive events that cannot occur simultaneously.

To make things more challenging, *independence* also has several other definitions in mathematics and statistics. When introducing algebraic functions, mathematicians refer to independent and dependent variables. In the Common Core standards, one sixth-grade standard asks students the following:

Write an equation to express one quantity, thought of as the dependent variable, in terms of the other quantity, thought of as the independent variable. Analyze the relationship between the dependent and independent variables using graphs and tables, and relate these to the equation. (NGA Center for Best Practices & CCSSO, 2010, p. 44)

The functional expression and the value of the independent variable completely specify the value of the dependent variable. This level of control is closer to the colloquial definition than to the probabilistic sense.

Probability texts also apply the word *independence* to continuous random variables, creating a phrase where both *random* and *independence* have noncolloquial meanings. The test for independent random variables includes multiplication, as does the test for independent events, but now the multiplication test applies to cumulative distribution functions, not events (Wackerly, Mendenhall, & Scheaffer, 2008). This usage is very similar to the event sense, but has a continuous, not discrete, context.

In statistics, the word *independence* is applied in two more situations. First, some hypothesis tests require independent samples. Independence for samples has the same verbal

description as independence for events, lack of effect—where values from one sample provide no information about values in the other sample. No formulaic test for samples exists, though; claiming independent samples requires knowledge about data collection and is usually decided subjectively (Truran & Truran, 1997). Second, variables in predictive models such as regression are classified as independent or dependent. One or more independent variables enter the model as predictors; an experimenter can manipulate their values. The dependent variable is the response, to be explained by patterns in the other variables (Wackerly, Mendenhall, & Scheaffer, 2008). This definition is similar to the algebraic definition, but it has two differences. A statistical model can have more than one independent variable, unlike the one-to-one functional relationship described in the Common Core standards. Also, statistical models include unknown error, unlike mathematical models.

In total, students must juggle up to six related definitions for independence: colloquial, algebraic variables, probabilistic events, probabilistic random variables, statistical samples, and statistical model variables. Students must pay close attention in order to avoid confusion. Although statisticians know about multiple meanings of *independence*, I found no published study that examined student understanding of the word's lexical ambiguity.

Conjunction. Researchers have examined the *conjunction* fallacy (Kahneman, Slovic, & Tversky, 1982; Gigerenzer, 1991). Humans are drawn to coherent detailed stories. People frequently increase the estimated likelihood of an event after hearing information that aligns with a mental picture, forming a more coherent mental image. In actuality, any additional condition that makes an event more complex will never increase its probability. Symbolically, $P(A) \geq P(A \text{ and } B)$, no matter how cogent A and B sound together. To challenge this fallacy, students should work problems involving compound events. Requiring students to record numeric count

frequencies for each option, such as Y out of 100, has been shown to reduce the conjunction fallacy (Gigerenzer, 1991). Conceptually, posing analogous problems, mathematically similar yet psychologically different situations, can also help students make judgments more probabilistically (Fast, 1997).

Time-axis causality. Stories also matter when discussing *time-axis causality*, sometimes called the Falk phenomenon after the researcher who first documented the fallacy (Falk, 1986). Time-axis causality refers to the inability of learners to view problems outside the flow of time. In causal logic, time only flows forward. If Event A occurs before Event B, Event A might influence Event B, but B can never affect A. Unlike causal logic, in conditional probability, information from a second event can help find the probability of the earlier event. For instance, if two balls are taken sequentially from an urn, knowing the color of the ball drawn second modifies probability calculations for the color of the earlier draw. When asked to solve for the probability of the first ball, Falk (1986) reported that many children did not use information about the color of the second ball. A few children called the problem logically not possible. Falk suggested that teachers remind students of cases where recently obtained information changes previous belief about past uncertain events, such as when archaeological excavations modify historical understanding. Physical demonstrations can also help students. For instance, a student could first draw a ball and place it a bag without looking, and then draw a second ball. The student could then compute possibilities about the ball in the bag.

Confusion of the inverse. Falk's 1986 paper also described *confusion of the inverse*, a fallacy that other authors had previously demonstrated. People often confuse a desired conditional probability $P(A|B)$ with $P(B|A)$, the conditional probability with the other event given. This fallacy occurs in child and adult reasoning. Watson (2005) reported on the struggles

of middle school students, while Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, and Woloshin (2008) demonstrated that when doctors are faced with mammogram test results, they frequently overestimate the probability of a patient having breast cancer. Confusion of the inverse frequently occurs with medical tests. Just because the probability of a positive test given occurrence of a disease is high, perhaps even 95%, the probability of having the disease given a positive test result does not automatically also equal 95%.

Multiple researchers have investigated ways to reduce this confusion. Falk (1986) suggested not relying on shortcut terms, instead stating situations symbolically and writing data in frequency tables. Gigerenzer et al. (2008) supported the use of frequencies; they demonstrated that dealing with fractional proportions causes much of this confusion. They asked people to think about frequency counts out of 100 or 1000. Utts (2003) proposed a larger theoretical population for finding frequencies—a “hypothetical hundred thousand” that makes fractional computation unlikely (p. 77).

Missing base rate. Another fallacy common in medicine, also appearing in other situations, is the *missing base rate*. When presented with specific information (such as a conditional probability of a positive test, given a disease) and general information (such as the prevalence of the disease), a person may make this fallacy if he or she neglects the general information. Sometimes this fallacy looks like confusion of the inverse, since in both cases a person incorrectly applies conditional probability, but the omitted information differs. For example, imagine a scenario in which an adult named Tom lives in New York City. Tom rode horses as a child, visits art museums on international vacations, and enjoys listening to opera music. Given this information, is it more likely that Tom is a professional violinist or a banker? Although Tom’s characteristics may sound more like those of a violinist, New York City has

many more bankers than it does violinists. Therefore, the base rate makes banking a more likely profession for Tom.

In medicine and other scientific tests, authors neglect the base rate when they report relative change, but not absolute value; this can cause societal consequences (Gigerenzer et al., 2008). Reading that a medical pill doubles the risk of a life-threatening complications may sound alarming. Doubling an absolute risk from 1 in 5 to 2 in 5 might lead many people to stop treatment, but if absolute risk doubled from 1 in 7,000 to 2 in 7,000, people might be less likely to change their behavior. Converting relative rates to absolute frequencies often clarifies the true effect. In practice, people need to rely on probability calculations, not conditional judgments. Bar-Hillel (1980), a researcher on the base-rate fallacy, noted that “an entire methodology of experimental control has been conceived to guard against this prevalent side effect of the base-rate fallacy” (pp. 213–214).

Teacher Knowledge Frameworks

High school mathematics courses cover many topics. It is unrealistic to expect teachers to have a level of understanding equivalent to probability researchers like Bar-Hillel or Gigerenzer. Researchers in education have attempted to define the types of information that teachers should know about a subject. In a frequently cited article, Shulman (1986) proposed three categories of content knowledge: subject matter, pedagogical, and curricular. Subject matter knowledge helps people solve problems. Pedagogical knowledge relates specifically to teaching, not solving problems. It includes knowing learners’ backgrounds and misconceptions about a subject, and then representations, analogies, demonstrations, examples, and other ways of making topics comprehensible to others beyond one’s self. Curricular knowledge includes the various

instructional materials available for a given topic, lateral knowledge from other concurrent classes, and vertical knowledge of earlier and later topics.

Other researchers have proposed more complex frameworks. Ball, Thames, and Phelps (2008) presented a model with six types of mathematical knowledge for teaching; this model has been adopted by some researchers studying statistics pedagogy. In this model, there are three domains related to content knowledge: common content knowledge used outside teaching, specialized content knowledge of relationships and procedures for teaching, and horizon content knowledge about earlier and later topics. The other three domains are considered parts of pedagogical content: students' backgrounds and misconceptions; teaching representations, analogies, and examples; and curriculum materials available for a topic. Although Ball, Thames, and Phelps expanded the amount of available information, their model also created a more complicated structure. Shulman's model offered more straightforward delineation, and has had high levels of uptake.

Turning to probability, a few authors have attempted to conceptualize needed teacher knowledge. In 2012, the Conference Board of the Mathematical Sciences (CBMS) published an updated edition of a report titled *The Mathematical Education of Teachers*, originally published in 2001. In the more recent edition, recommended statistics and probability knowledge changed because of the new Common Core standards. The CBMS recommended course content that primarily followed the Common Core topics. The CBMS (2012) noted that teachers would need chances to study new content not previously taught, "particularly in the areas of statistics and probability" (p. 68). Although these reports have helped promote teacher training in probability, they are not comprehensive.

Kvatinsky and Even (2002) offered a more comprehensive framework for probability subject matter knowledge. In their framework, teachers should know about the contrast between nonsubjective and subjective approaches to uncertainty, multiple representation forms to frame a basic repertoire of techniques, and examples that demonstrated the strength of probability. Recommended subject matter knowledge also included the difference between classic theoretical computation and experimentally found frequencies.

Although useful to categorize subject matter knowledge, the Kvatinsky and Even (2002) model did not include pedagogy or curriculum. A few years after Kvatinsky and Even, Papaieronymou presented two conference reports. The first report (Papaieronymou, 2008) described content knowledge needed by secondary school teachers, synthesizing information from NCTM's 2000 *Principles and Standards*, ten published state standards, recommendations from professional organizations, and mathematics textbooks. The second report (Papaieronymou, 2009) identified aspects from all three of Shulman's types—subject matter, pedagogical, and curricular—making it the more valuable framework. Unfortunately, in the second report Papaieronymou reviewed only recommendations from professional organizations. Because state standards direct what teachers help students learn, neglecting state standards provided a limited view of activities needed in the classroom. Nonetheless, the ideas Papaieronymou identified help frame discussions. Teachers should be able to accomplish tasks such as the following: distinguish theoretical and experimental probability, conduct simulations with and without technology, represent probabilities through multiple models, define fair games, confront common misconceptions, and make decisions and predictions.

Most recommendations from professional organizations have dealt with subject matter; Papaieronymou (2009) noted the lack of suggestions about pedagogy and curriculum. State

standards also provide primarily subject matter guidance. I know of no existing comprehensive theoretical model for teacher knowledge of probability. Instead of proposing a model, this dissertation's research questions are set within Shulman's framework. Research Question 1, how teachers solve problems, asks about subject matter knowledge. Research Question 2, how teachers respond to student misconceptions, asks about part of pedagogical content knowledge. Research Question 3 allows teachers to express perceived needs in any part of the framework.

Research on Teacher Knowledge

Instead of far-reaching theoretical proposals, researchers on teacher knowledge about probability have written on specific topics. In the 2005 edited volume *Exploring Probability in Schools*, Stohl wrote that “compared to the many chapters in this volume dedicated to students’ understanding of probability, there has been significantly less research on teachers’ knowledge *of* probability and their knowledge *for* teaching probability” (p. 351). Nevertheless, some publications exist. Studies on practicing high school teachers are most relevant, but research about non-high-school teachers and preservice teachers can also provide information. Additional studies, particularly on technology, are described in Stohl’s chapter.

Batanero and Díaz (2012) wrote about their experiences training preservice teachers in Spain. They mentioned difficulties including informal definitions, equiprobability, and time-axis causality. They made more training suggestions, such as project work and technology. Chernoff and Russell (2013) wrote about probability rule misconceptions among 54 preservice elementary teachers in Canada. Carnell (1997) investigated 13 prospective North Carolina middle school teachers’ subject matter understanding of conditional probability; the results indicated that the prospective teachers had trouble defining the conditioning event and avoiding time-axis causality. Carter and Capraro (2005) gave an online probability and statistics examination to over

200 prospective Texas elementary school teachers. A majority of the participants did not explain conjunction correctly, and the most common wrong solution incorrectly applied equiprobability.

In Colorado, Dollard conducted 26 task-based interviews, averaging about 45 minutes each, for a dissertation on preservice elementary teachers' conceptions about probability. In a 2011 article, Dollard stated that before completing coursework on probability, a majority of subjects did not have an adequate understanding of theoretical vs. frequentist probability. Half of the participants incorrectly applied equiprobability in a situation with unequal chances. Begg and Edwards (1999) reported results from a convenience sample of 22 practicing and 12 preservice primary school teachers. Just over two-thirds of the New Zealand respondents showed familiarity with issues behind the equiprobability fallacy, roughly half understood randomness, and less than half understood independence.

Publications about practicing teachers have generally consisted of qualitative studies with fewer than ten participants. Groth (2010) investigated the learning environments that three seventh-grade teachers constructed around conditional probability and independence, which included key roles for language, fractions, and combinatorial ideas. Zapata-Cardona (2008) conducted task-based interviews with one expert and one novice high school teacher on topics in reasoning about chance including sample spaces, independence, and randomness. In the interviews, Zapata-Cardona asked about student approaches to each task; she then presented either an incorrect student solution or research result and asked for comment. The expert teacher was able to identify student difficulties more quickly than was the novice. The two teachers applied response strategies differently; the expert used simulation more effectively and confronted student misconceptions more often. Liu and Thompson (2007) conducted a summer teaching experiment with eight high school statistics teachers. The teachers had trouble viewing

some problems as stochastic, often resorting to deterministic interpretations that did not involve chance or probability. Mojica (2006) examined class lessons prepared by four middle school teachers. The teachers had trouble connecting theoretical probabilities with experiential results. Carlson and Doerr (2002) observed high school mathematics teachers conducting activities in probability. Four of the six participants felt unprepared to teach probability. In the classroom, the teachers concentrated on parts they understood, such as physical mechanics and exponential growth. Similarly to Mojica's teachers, they neglected connections. Haller (1997) observed four middle school teachers. Teachers with less subject matter knowledge made more errors, and demonstrated more misconceptions, when compared with teachers with higher content knowledge. In summary, these researchers have provided information on individual topics through small qualitative studies, but did not establish a comprehensive frame about conditional probability and decision-making.

A few authors have taken more extensive samples of practicing teachers. Lee and Lee (2009) looked at how teachers and learners can utilize technology, particularly through a software tool called *Probability Explorer*. Vermette and Gattuso (2014) presented 12 Quebec high school mathematics teachers with six tasks related to variability. The participants were first tested on subject matter, and then asked to demonstrate pedagogical knowledge by reacting to potential student solutions. Some tasks included probabilistic topics. Vermette and Gattuso identified three types of teaching intervention: explanation of the concept involved, confrontation about erroneous reasoning, and experimentation with alternative numbers or physical systems. Probabilistic situations led more frequently to experimentation recommendations.

The most extensive survey of practicing teachers was published in 2001 by Jane Watson. Her profiling instrument about chance and data was designed to cover all three of Shulman's

(1986) knowledge categories. Forty-three Australian school teachers participated, 15 from primary schools and 28 from secondary schools. About half were interviewed; the other half submitted written answers. When asked to choose a topic in chance and data, the secondary school teachers chose probability more often than they did any other topic. The secondary school teachers generally had stronger mathematics backgrounds and expressed more confidence about their teaching ability than the primary school teachers had, with six of the nine topic comparisons statistically significant. Watson asked the teachers to suggest student responses for two tasks about probability and four tasks about data analysis. In the task about odds, about half of the secondary school teachers provided a correct student response, but none of the primary school teachers did.

Across all the studies, at least a substantial minority of prospective elementary and middle school teachers held probabilistic misconceptions; in many cases, a majority committed errors. Watson's 2001 study combining primary and secondary teachers indicated that although secondary school teachers exhibited greater knowledge and confidence, they still made errors. Very few studies with more than ten in-service teachers have been published, and no known study has included results about a large range of high school probability topics.

Research on Teacher Views

Relatively reliable instruments exist to measure attitudes about statistics, but not probability. The University of Georgia Statistics Department has given students one attitude instrument, the Survey of Attitudes Towards Statistics (SATS) (Schau, Stevens, Dauphinee, & Del Vecchio, 1995). Unfortunately, SATS questions ask about "statistics" and "statistical" concepts. Although words could be modified, results about probability would not be comparable, and the validity of these instruments would be highly questionable. A recent project has

introduced a scale measuring statistics teachers' self-efficacy (Harrell-Williams, Sorto, Pierce, Lesser, & Murphy, 2014). Although promising for statistics teaching, the instrument asks about the GAISE statistics framework (Franklin et al., 2007), not probability instruction.

One instrument measuring attitudes about probability has been published: the Probability Attitude Inventory of Tan, Harji, and Lau (2011). These Malaysian researchers created the instrument by taking a Mathematics and Science Attitude Inventory published online (Rochester Institute of Technology, 1999) and changing *mathematics* to *probability* in the mathematics questions. Although efficiently constructed, their research did not provide much evidence of validity. One other paper (Veloo & Chairhany, 2013) reported on results from the Probability Attitude Inventory in a group of students, but did not measure teachers. Given the lack of verified validity and teacher results, the Probability Attitude Inventory was not appropriate for the present study. Attitudes and other views were investigated through open-ended questions, not through a written questionnaire.

Watson (2001) asked Australian teachers about the types of professional development that would benefit them the most and who should lead these professional development initiatives. Regarding the desired type of professional development, 40% of the interviewed teachers preferred school-based sessions, 19% preferred personal reading or a university course, and the remainder offered no preference or multiple options. Regarding the desired professional development leader, 51% of the teachers requested an outside expert, 21% requested a teacher at the school, and 21% requested a government curriculum officer; in response to these questions, many teachers indicated more than one choice (p. 323). The teachers wanted to ensure that the professional development leader understood classroom realities. Almost half the participants had never participated in professional development on the subjects of chance and data.

Summary

In Chapter 1, three research questions on teacher knowledge and views about conditional probability were introduced. The first part of this chapter contained information on the historical development of the field of probability and topics in conditional probability, placing probability in relationship to statistics and mathematics. In the next part, national and state school standards were examined. Conditional probability appears in the Common Core standards and state standards of the three states covered in this dissertation, albeit in different courses in each state. This makes knowledge about conditional probability relevant for high school mathematics teachers.

Researchers in probability have identified many misconceptions held by learners, with some success in overcoming problems. According to the Shulman model, teachers should have subject matter knowledge, pedagogical content knowledge, and curricular knowledge. Since no better comprehensive model for teacher knowledge of probability has been presented, the research questions were composed under Shulman's framework of subject matter and pedagogy.

Existing literature on teacher knowledge and views about probability is limited. Many researchers have investigated subject matter knowledge of preservice elementary teachers, not inservice high school instructors. Publications about practicing teachers have generally consisted of qualitative studies with fewer than ten participants. A few researchers, including Dollard (2011), Vermette and Gattuso (2014), and Watson (2001), conducted task-based interviews on moderately sized groups, but no known researchers have reported exclusively about the probability knowledge of high school mathematics teachers. Chapter 3 describes the present study, which was designed to fill part of this gap.

CHAPTER 3

METHODS

In the literature review in Chapter 2, I established that little research exists on high school teachers' knowledge about conditional probability as listed in the Common Core and individual state standards. In this chapter, I set the stage for analysis by describing the research design of this study. This chapter includes information on the interview materials, the interview process, the teachers who participated, and the analysis process. The interview materials consisted of tasks and open-ended questions. Tasks appear first, building from the literature review in Chapter 2. I decided to use conditional probability tasks that had prior information about misconceptions. Trial interviews were used to select tasks that met the aims of the interview process. The five selected tasks, containing a total of nine questions, are presented, each with standards addressed, a solution, relevant research, and potential pitfalls. Following the tasks, the next section describes the open-ended question starters selected to investigate the third research question, teachers' perceived needs to teach conditional probability.

After presenting tasks and open-ended questions, I describe the study protocol in sections that chronicle the teacher recruitment process and how interviews were conducted. I then present information about the participants from the background questionnaire completed by all interview subjects. The last two sections of this chapter describe the analysis process. One section details initial analysis conducted after recording each interview and before transcription. The last section describes the detailed analysis process through which I generated the task analysis contained in Chapter 4 and combined analysis contained in Chapter 5.

Selecting Tasks

Prior researchers have conducted task-based interviews with teachers (Chernoff & Russell, 2013; Dollard, 2011; Vermette & Gattuso, 2014; Watson, 2001; Zapata-Cardona, 2008). Interviews allow participants to explain task solutions more fully than written examinations. Interactive discussion would also provide answers for Research Question 2 about responding to student misconceptions. Written examinations on pedagogy exist for some well-researched topics, but conditional probability has not yet been sufficiently researched. Nevertheless, prior studies can provide good tasks. When considering potential tasks that covered at least one Common Core standard, I found that having at least one published result made a problem more appealing because prior research would provide information on misconceptions. All selected tasks had information from at least one prior administration, either in a research paper or as an Advanced Placement® (AP) test question. Some questions had more than one prior finding.

Other factors considered in task selection included time and difficulty. Selected tasks should cover as many standards as possible, given teacher time constraints. School class periods rarely extend for more than about 90 minutes. Therefore, total interview time including administering consent forms needed to remain around 90 minutes to make it possible to conduct an interview during a long block schedule planning period. Reserving 10 to 15 minutes for consent forms and a background questionnaire, and 10 to 15 minutes for open-ended questions about needs, left 60 to 70 minutes for task discussion. This time estimate fell in the range of prior studies. Dollard (2011) averaged 45 minutes in interviews; Zapata-Cardona (2008) averaged about an hour; Watson (2001) explicitly asked teachers to take no more than 90 minutes.

Within the 60 to 70 minutes, participants would complete questions at different speeds. To balance depth and breadth, the goal was to find questions that took roughly three to five

minutes to solve. This length would enable discussion about solutions and student misconceptions to take place within 8 to 10 minutes per question. All participants would complete at least six questions, with more rapid problem solvers finishing up to ten.

The selected questions needed to have heterogeneous difficulty. Some teachers have taught AP® Statistics and might be expected to correctly answer questions from those exams. Other teachers have had almost no probability training and might struggle with table-reading problems. Potential questions were divided into three difficulty levels: Direct, Computation, and Multistage. Direct questions could be answered by forming ratios based on information given in the problem, without requiring formulas. Solving Computation questions required one piece of mathematical knowledge not provided in the question, such as the formula for combinations or the multiplication rule to test independence. Multistage questions were more difficult than Computation problems because participants needed to apply more than one piece of mathematical knowledge.

Proposed tasks were discussed during meetings of the University of Georgia statistics education research group led by Jennifer J. Kaplan. Group members made suggestions about wording, possible solutions, and potential student misconceptions. After research group discussions, during Thanksgiving weekend 2013, I piloted five tasks with two retired secondary mathematics teachers who have known me literally my entire life, my parents. My parents met as undergraduates in mathematics education, although they met in a literature class. My father taught school mathematics for 35 years, shifting between high school and middle school. He also worked as an adjunct college professor in computer science. My mother taught high school mathematics for 27 years. During the 1990s, she served on the test development committee for an earlier version of Pennsylvania's standardized exams.

Initial task trials went well. With the exception of the most difficult task, my parents worked through each problem in less than 5 minutes. Including discussion, full task times ranged from 6 to 15 minutes. My parents thought that teacher participants could complete at least four tasks in 60 minutes. To further assess task timing, I administered four tasks as part of my final project in QUAL 8410, Designing Qualitative Research. Three University of Georgia graduate students volunteered for audio interviews. Although University of Georgia class restrictions prevent discussion of how the graduate students performed, their feedback was also helpful in finalizing wording choices and question order.

After adjustments based on parent and class trials, the final set of five tasks contained nine questions. Each task had a short name: Rash, Lucky Dip, Survey, Taxicab, ELISA. The Survey and ELISA tasks each consisted of three question parts; the other three tasks had only one question part. Trial subjects categorized Lucky Dip, Rash, and Survey as easier than Taxicab and ELISA. For interview order, I decided to place easier tasks first, so as to not agitate participants with a difficult start. As the Rash and Survey tasks both included numeric tables, those two tasks were separated, with the single-part Rash task first and the multiple-part Survey task third. On the more difficult side, both the Taxicab and ELISA tasks had challenging computations, correctly solved by only a minority of participants in prior studies. Because researchers have examined variants of the Taxicab task for over 40 years, but ELISA has less history, I placed Taxicab fourth and ELISA fifth. ELISA would be the task dropped in the event of time constraints. Dropping ELISA due to time constraints occurred in four interviews. In two more interviews, the teacher and I chose not to attempt the most difficult final part of the task.

Table 3.1 contains a summary of the questions, listed in interview order. Each question appears with an abbreviated name, difficulty level (Direct, Computation, or Multistage), and

brief description of the Common Core standard addressed. Taken together, the tasks cover all nine standards.

Table 3.1

Summary of Interview Questions

Question name	Difficulty	Primary standard
Rash	Direct	4. Reading tables; 5. Everyday situations
Lucky Dip	Computation	1. Describing events; one possible solution involves 9. Combinations
Survey A	Computation	7 Addition rule
Survey B	Direct	6. Definition of conditional probability
Survey C	Computation	2. or 3. Independence
Taxicab	Multistage	1. Describing events; 6. Conditional probability
ELISA A	Direct	1. Describing events; 6. Conditional probability
ELISA B	Direct	6. Definition of conditional probability
ELISA C	Multistage	1. Describing events; possible solutions can involve 7. Addition rule and 8. Multiplication rule

The following five sections provide more information on each task. Each section begins with a figure showing the task wording. Next appears at least one solution for each part; most problems have notes about multiple solution paths. After the solution, I include the Common Core standards addressed, published reference or references, and information from the literature. Based on research literature and analysis, each problem has potential student errors and misconceptions. Because categorization between computational errors and conceptual misconceptions is not uniform, in this dissertation the terms *misconception* and *error* are used interchangeably. For the first four tasks, not including ELISA, one common student

misconception in each question was defined as the key pitfall. During the interview, if the teacher had no idea what errors students might commit, or failed to mention the common pitfall, I tended to introduce the key pitfall into the discussion when time permitted. On the first three tasks, I introduced the key pitfall most of the time; on Taxicab and ELISA I did so less than half the time.

Rash Task

Medical researchers have developed a new cream for treating skin rashes. New treatments often work but sometimes make rashes worse. Even when treatments don't work, skin rashes sometimes get better and sometimes get worse on their own. As a result, it is necessary to test any new treatment in an experiment to see whether it makes the skin condition of those who use it better or worse than if they had not used it.

Researchers have conducted an experiment on patients with skin rashes. In this experiment, one group of patients used the new cream for two weeks, and a second group did not use the new cream. The total number of patients in the two groups was not the same, but this does not prevent assessment of the results.

	Rash Got Better	Rash Got Worse
Patients who DID use the new skin cream	223	75
Patients who DID NOT use the new skin cream	107	21

Is using the new cream better, the same, or worse than not using the new skin cream?

Figure 3.1. Rash task wording.

Using probability, the answer is that using the new cream is worse than not using the new skin cream. The most common solution path compares the improvement rate for the $(223 + 75) = 298$ patients who did use the new skin cream, $223/298 = 74.8\%$, against the improvement rate for

the $(107 + 21) = 128$ patients who did not use the new skin cream, $107/128 = 83.6\%$. Because the new skin cream has a lower percentage of people who got better, using the new skin cream is worse than not using the new skin cream. Alternatively, one could compare percentages of people who got worse, about 25% with the new cream against about 16% without the new skin cream. Solutions that compare column percentages are also possible, though in prior results column comparisons occurred less often than row comparisons. An alternative solution path without computing marginal totals for each group compares the likelihood ratio of improvement to worsening. For people who did not use the new cream, the ratio of 107 to 21 is about 5 improvements for every 1 worsening. The ratio of 223 to 75 for people who did use the new cream, about 3 improvements for every 1 worsening, shows that the new skin cream performs less well.

A person familiar with statistical significance testing might conduct a chi-square test. Fisher's exact test yields a p -value of .0574; the Pearson chi-square yields a p -value of .0472. A statistician who conducted a Fisher's exact test might claim no significant difference from the new skin cream. When a participant mentioned the chi-square test, I asked the participant to consider results in the sample using probability, not conduct the hypothesis test about the population. Examining hypothesis testing was not part of this study.

This task addresses Common Core Standard 4, Reading tables. Standard 5 on Everyday situations also applies, as I first read about this problem in a news report. The task comes from a 2013 study about political views published by Kahan, Peters, Dawson, and Slovic. In their study of several hundred adults, less than half answered the rash question correctly. Even among those in the top 10% on a numeracy test, only 75% gave the correct response. A similar question also appeared in the Comprehensive Assessment of Outcomes in a first Statistics course, frequently

used to judge learning outcomes after introductory statistics courses (delMas, Garfield, Ooms, & Chance, 2007). For more information on solution paths, see a blog post I wrote (Molnar, 2014).

Misconceptions in 2-by-2 table problems have been previously studied. As noted by Kahan et al. (2013), common incorrect approaches compare 223 against 75, upper left cell against upper right cell, or 223 against 107, upper left against lower left. Either approach would give an incorrect answer. School students might have trouble comprehending the introductory paragraph and not even reach a meaningful cell comparison. For the interviews, the key pitfall was comparing the counts of those improved, 223 against 107, without considering proportions or percentages.

Lucky Dip Task

Dominic has devised a simple game. Inside a bag he places 3 black and 3 white balls. He then shakes the bag.
He asks Amy to take two balls from the bag without looking.

If the two balls are the same color, then Amy wins.
If they are different colors then Dominic wins.

Is the game fair, meaning Dominic and Amy have equal probability of winning?
If not, then who is most likely to win?

Figure 3.2. Lucky Dip task wording. From Modeling Conditional Probabilities 1: Lucky Dip (beta version) (p. S-1), by Mathematics Assessment Resource Service, 2012, Nottingham, UK: University of Nottingham. Copyright 2012 by the University of Nottingham. Reprinted with permission under a Creative Commons Attribution, Non-commercial, No Derivatives License 3.0 (<https://creativecommons.org/licenses/by-nc-nd/3.0/>).

The game is not fair. It favors Dominic, with Dominic having a $\frac{3}{5}$ probability of winning and Amy having a $\frac{2}{5}$ probability of winning. The probability can be shown in many ways. The most direct approach is sequential conditional logic. No matter what color Amy draws first, two balls of that color will remain in the bag for the second draw out of five possibilities.

Alternatively, students might draw a tree diagram to explain possible outcomes and find the probabilities.

There are several ways to enumerate the possible outcomes. Considering draws sequentially would yield four mutually exclusive possibilities: black then black, black then white, white then black, and white then white. The probability that both balls are black is $(3/6) * (2/5) = 1/5$. Similarly, the probability that both balls are white is also $1/5$, leading to a total victory probability for Amy of $1/5 + 1/5 = 2/5$. Instead of imagining sequentially drawn balls, considering the two drawn balls as a set would yield three mutually exclusive possibilities: two black, one black and one white, and two white. Probabilities of the three possible sets can be found with combinations. There are $\binom{6}{2} = 15$ possible combinations of two balls. Selecting two of the three black balls can be done in $\binom{3}{2} = 3$ ways, with 3 different ways to select two of the three white balls. Amy's winning probability is therefore $(3+3) / 15 = 6/15 = 2/5$. Since there are $\binom{3}{1}\binom{3}{1} = 9$ combinations of one black and one white ball, Dominic wins with a probability of $9/15 = 3/5$. More complex enumeration strategies label black and white balls separately, such as B1, B2, and B3; proper counting of the 30 possibilities will lead to a correct result.

This task addresses Common Core Standard 1, describing the sample space. One solution path uses combinations from Standard 9, though the use of combinatorics is not necessary. As shown, this task was published as a formative assessment lesson developed at the University of Nottingham in England (Mathematics Assessment Resource Service, 2012), although similar versions have appeared earlier (Zapata-Cardona, 2008). Contributors to the Mathematics Assessment Resource Service project trialed the lesson in several U.S. classrooms. The task authors provided two correct approaches: drawing a tree diagram and enumerating the possibilities in the sample space. The teacher guide also includes a sample lesson plan, which

asks students to critique one of three incorrect solutions. Anna enumerated the sample space in terms of colors: BB, BW, WB, and WW. Instead of finding the probability of each option, she erroneously assumed equiprobability and concluded that the game was fair. Ella tried to write out the entire sample space with six balls, but failed to notice that the same ball could not be selected twice. Jordan attempted a tree diagram, but made the denominator for second ball draws 6 instead of 5.

Besides the three misconceptions provided in the teacher guide, another common error occurs when students assume replacement after the first draw as Ella did, but compute probabilities with fractions instead of enumeration. Assuming the second ball has probability $\frac{3}{6}$ of being black or white yields the same incorrect conclusion of fairness as equiprobability. For the interviews, the key pitfall was Anna's approach of equiprobability, a common mistake provided in the teacher guide.

Survey Task

An advertising agency in a large city is conducting a survey of adults to investigate whether there is an association between highest level of educational achievement and primary source for news. The company takes a random sample of 2,500 adults in the city. The results are shown in the table below.

Primary Source for News	Not High School Graduate	High School Graduate but Not College Graduate	College Graduate	Total
Newspapers	49	205	188	442
Local television	90	170	75	335
Cable television	113	496	147	756
Internet	41	401	245	687
None	77	165	38	280
Total	370	1,437	693	2,500

Question wording is not available in this version of the document, due to copyright restrictions.

To see the question wording, visit the 2010 Form B free response questions at http://apcentral.collegeboard.com/apc/public/repository/ap10_frq_statistics_formb.pdf

Figure 3.3. Survey task wording. Source: AP® Statistics Free-Response, 2010, Form B Question 5. Copyright © 2010 The College Board. Reproduced with permission. <http://apcentral.collegeboard.com>.

This task contains three questions. In Part (a), there are 693 college graduates plus 687 people who get news primarily from the Internet, minus 245 people in both groups, for a total of $(693 + 687 - 245) = 1135$ distinct people that meet at least one criterion. Randomly selecting from the 2500 people in the sample, the probability becomes $1135/2500 = .454$. Alternatively, a solver might mark and count cells from the appropriate row and column, making sure not to include the 245 twice, for $41 + 401 + 245 + 188 + 75 + 147 + 38 = 1135$ people. Dividing by the sample size of 2500 again yields .454.

Solving Part (b) involves recognizing the desired probability is conditional. There are 693 adults who are college graduates, of whom 245 obtain news primarily from the Internet. Selecting randomly from the sample yields a probability of $245/693 = .354$.

The Common Core standards include two ways to test for independence; Part (c) can be solved with either approach. Using multiplication from Standard 2, one tests if the product of the probability of being a college graduate and the probability of obtaining news from the Internet is equal to the joint probability of the two events. Since $(693/2500) * (687/2500) \neq (245/2500)$ because $.076 \neq .098$, the events are not independent. Using Standard 3, one checks if the conditional probability of obtaining news from the Internet given college graduate status, computed in part (b), equals the unconditional probability. Since the unconditional probability of obtaining news from the Internet of $687/2500 = .275$ does not equal the conditional probability $.354$, the two events are not independent.

This task was originally on an AP® Statistics exam, Question 5 on Form B in 2010 (The College Board, 2010). The exam included another part about the chi-square test of association, but the chi-square test does not appear in the conditional probability standards. The entire problem requires students to read tables as in Standard 4. The three parts selected for this dissertation study examine additional Common Core standards. In Part (a), I expected most solvers would follow a path using the addition rule from Standard 7. In Part (b), the solution includes computing a conditional fraction as in Standard 6. Part (c) requires a test for independence, either multiplication as in Standard 2 or conditional comparison as in Standard 3.

AP® Statistics scoring guidelines (The College Board, 2010) include information on some of the common student pitfalls listed here. In Part (a), the most frequent error and key pitfall is failing to remove the doubly counted group, those college graduates who obtain news

from the Internet. This yields an answer of $1380/2500 = .552$. Problem solvers might also confuse the word *or* with *and*, resulting in an answer of $245/2500$. In Part (b), people might confuse the inverse and reverse rows and columns to yield $245/687$. The key pitfall is neglecting the conditional part of the statement, resulting in an answer of $245/2500$.

In Part (c), the participants might not remember the definition of probabilistic event independence. They might instead offer a verbal nonnumeric justification. Confusing independence with mutual exclusivity occurs relatively often, when someone says two events are independent if and only if they share no observations. Mutual exclusivity was intended to be the key pitfall, but as described in Chapter 4, the participants' responses quickly changed this discussion into one without an external misconception.

Taxicab Task

A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data.

(a) 85% of the cabs in the city are Green and 15% are Blue.

(b) A witness identified the cab as Blue.

(c) The court tested the reliability of the witness under the same circumstances that existed on the night of the accident, and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time.

What is the probability that the cab was actually Blue? Explain your reasoning.

Figure 3.4. Taxicab task wording.

At least eight solution paths exist, which I described for a similar problem in a blog post on Dan Kahan's Cultural Cognition blog (Molnar, 2014). Probability textbooks often include a formulaic solution based on Bayes' Rule. Let Event B be that the cab was actually blue and

Event A be that the witness identified the cab as blue. Statement (b) in the problem data indicates that Event A occurred.

$$P(B | A) = P(B \text{ and } A) / P(A). P(B \text{ and } A) = P(A | B) P(B) = (.80) (.15) = .12.$$

$$P(A) = P(A | B) P(B) + P(A | B^c) P(B^c) = (.80) (.15) + (.20) (.85) = .29.$$

$$\text{So } P(B | A) = .12 / .29 = .41.$$

There are several alternative solution paths. A student could construct a two-level tree diagram with probabilities. The first level could be the actual taxicab color, with the second level whether or not the witness correctly identified the color. As another visual approach, a student could create a grid of 100 cells and shade in probabilities to find the answer. Creating a table with taxicab counts would work, but the problem does not provide the number of taxicabs in the city. Some authors have advocated assigning a hypothetical population size and computing counts. Choosing a large divisible number, like 1,000 or 10,000, leads to integer cell values, which are easier to understand than probabilities (Gigerenzer et al., 2008).

This task addresses Common Core Standard 1 by asking participants to describe events. Solutions always include finding a conditional fraction, as in Standard 6. Other standards can apply. Some solution paths involve creating a table as in Standard 4; others utilize the general multiplication rule described in Standard 8.

The taxicab witness problem is one of the most cited in research on personal probability, with at least a dozen papers investigating variations. Originally introduced by Kahneman and Tversky in 1972, the wording used in this project is closest to that of Bar-Hillel (1980). Summaries of results have been provided by Tversky and Kahneman (1982) and Krynski and Tenenbaum (2007). Without training, only a small minority of people achieve a correct response. Bar-Hillel (1980) asked Israeli students this question on a college entrance exam; only 6% of

those students answered this question correctly. Only 8% of surveyed doctors got a similar problem right (Gigerenzer, 2002).

The modal response is 80%, generally given by 40% to 50% of people. Not considering the underlying proportion of cab colors and relying only on the witness rate persisted in a wide variety of problem formulations (Bar-Hillel, 1980). A response of 80% illustrates the missing base rate fallacy; when asked to explain, people do not mention the base rate of 15% blue cabs. Other problem solvers attempt to include the base rate, but do not know Bayes' formula or another approach. Often, participants can identify one of the two scenarios that would lead the witness to claim the cab was blue, where the taxicab was blue and the witness identified the color correctly. The blue and correct scenario occurs $(15\%) (80\%) = 12\%$ of the time. A response of 12% neglects the other possible scenario, where the witness incorrectly identified a green cab as blue. This scenario neglect can occur in any representation.

Over the years, researchers have discovered ways to improve results. Changing the wording of the question to introduce a causal implication, by implying the failure of the observer was due to faded paint (not random chance) or that the cab proportions were due to accidents and not registrations, appeared to help people think logically and give the correct answer (Krynski & Tenenbaum, 2007; Tversky & Kahneman, 1982). In one study, grid shading helped a majority of participants obtain a correct answer (Cosmides & Tooby, 1996). Because I anticipated that most interviewees would not answer this question correctly, I expected to have more of an instructional role in this problem. I chose a relatively simple key pitfall: the modal 80% response from the base rate fallacy.

ELISA Task

The ELISA tests whether a patient has contracted HIV. The ELISA is said to be positive if it indicates that HIV is present in a blood sample, and the ELISA is said to be negative if it does not indicate that HIV is present in a blood sample. Instead of directly measuring the presence of HIV, the ELISA measures levels of antibodies in the blood that should be elevated if HIV is present. Because of variability in antibody levels among human patients, the ELISA does not always indicate the correct result.

As part of a training program, staff at a testing lab applied the ELISA to 500 blood samples known to contain HIV. The ELISA was positive for 489 of those blood samples and negative for the other 11 samples. As part of the same training program, the staff also applied the ELISA to 500 other blood samples known to not contain HIV. The ELISA was positive for 37 of those blood samples and negative for the other 463 samples.

Question wording is not available in this version of the document, due to copyright restrictions.

To see the question wording, visit the 2010 Form B free response questions at http://apcentral.collegeboard.com/apc/public/repository/ap09_frq_statistics_formb.pdf

Figure 3.5. ELISA task wording. Source: AP® Statistics Free-Response, 2009, Form B Question 2. Copyright © 2009 The College Board. Reproduced with permission. <http://apcentral.collegeboard.com>.

Like the survey task, this problem contains three parts. Solving Part (a) requires realizing the question asks about results from blood samples without HIV in the training data. In the training data, there were 500 blood samples without HIV, of which 37 incorrectly tested positive, an estimated probability of $37/500 = .074$. Part (b) also requires forming a conditional probability fraction from data provided in the problem. There were positive ELISA results from 489 blood samples with HIV and 37 without, a total of 526. Of these, 489 actually had HIV, a proportion of

$489/526 = .930$. In these first two parts, representing the data as a tree diagram or 2 by 2 table might make it easier to find the correct probabilities, but no additional representation is required.

Part (c) is substantially more challenging. Let Event E indicate a positive ELISA result on a single test. If the blood sample does not contain HIV, the probability of a false positive ELISA result was found in Part (a), $P(E) = .074$. To reach the more expensive test, the first ELISA must be positive, and then at least one of the next two must also be a false positive. One possible way this occurs is with a false positive ELISA test and then another false positive test. This scenario has probability $P(E E) = .074 (.074) = .00548$. Sending the sample for further testing also occurs with three tests run with positive, not positive, and then positive results, in that order. This scenario has probability $P(E E^c E) = .074 (.926) (.074) = .00507$. Combining the two mutually exclusive scenarios yields a probability of about .0105. There are other ways to represent the outcomes besides the two scenarios described above. A three-level tree diagram would work, although it is complex. Formulaically, one could write out $P(E E E) + P(E E E^c) + P(E E^c E)$ and compute three probabilities, or find $P(E) * P(\text{not both negative}) = .074 (1 - .926^2)$.

This task was originally on an AP® Statistics exam, Question 2 on Form B in 2009 (The College Board, 2009). The false positive and false negative ELISA test rates are very similar to borderline case results described in a 1985 article by Weiss et al. HIV tests have become much more precise since 1985, with current accuracy above 99.9% (Malm, von Sydow, & Andersson, 2009). Because false result rates have changed so much, this question lacks currently realistic context and does not meet the everyday situation criteria in Common Core Standard 5. If anything, the unrealistically high error rates might hurt everyday thought by inducing fear about HIV test accuracy. Question (a) asks test takers to describe events as in Standard 1 and find a fractional conditional probability, as in Standard 6. Question (b) also asks for a fractional

probability as in Standard 6. In Question (c), the solution must describe events as in Standard 1. As described above, the probabilistic solution relies on the general multiplication rule in Standard 8 and the addition rule in Standard 7.

This task has the least research about misconceptions. The AP® scoring guidelines include a few errors, including an incorrect tree diagram in part (a) (The College Board, 2009). Reading comprehension plays an important role in this task. With many numbers floating around and no convenient table, participants might confuse values in Parts (a) and (b) and form incorrect fractions. Part (c) is quite difficult. Solvers must determine ELISA test combinations that result in the more expensive screening. When doing so, people often forget the mandated first negative test. The wording “at least one” leads some people to apply a binomial distribution with $n = 3$ trials, which does not work. Other people try to enumerate all possible outcomes, but fail to include at least possible scenario, computing results such as $P(E E E) + P(E E^c E)$. Calculation errors are also frequent because of problem difficulty.

I developed no key pitfalls for the ELISA task. Because this task was placed last in the interview order, I expected many interviewees to not attempt it, and most others to have only a small amount of time remaining. I wanted to allocate the time to teacher responses. Additionally, this task has less prior research than any other, making it more difficult to identify common mistakes.

Open-Ended Questions

The final 10 to 20 minutes of each interview were reserved for discussion about the third research question, teachers’ perceived needs to teach conditional probability. As described in the literature review, little background exists on teacher attitudes and views about probability, even less than research on solving probability problems. Because so little research exists, these

questions have a lightly structured format. Although I wrote some open-ended question starters , the teachers' experiences also influenced discussion in this section. For example, if a teacher commented about standardized testing, I asked about that state's standardized tests. AP® Statistics teachers were asked about probability in their training for the advanced course.

The list of question starters went through several iterations. At first, the list included two questions about statistics because I believed teachers would want to talk about probability and statistics together. Comments from my parents and my dissertation committee helped me realize that high school teachers could separate probability from the larger field of statistics. Therefore, the questions shifted to three areas about probability: standards, concerns, and assistance. The list of question starters contained nine questions:

1. What do you know about probability in the Common Core State Standards in mathematics?
2. What do you think about probability in the standards?
3. What probability topics would you include in the school curriculum?
4. How prepared do you think you are to teach probability?
5. What concerns do you have about teaching probability?
6. What past experiences have you had with conditional probability?
7. How prepared do you think you are to teach conditional probability?
8. What concerns do you have about teaching conditional probability?
9. What assistance would enable you to teach probability topics to your students?

The first three questions asked about curriculum standards. Responses did not necessarily address the research question about teacher needs, although the teachers could express a need for better curricular knowledge, one of the areas of content knowledge in Shulman's framework. For

many interviewees, curricular questions provided a transition from the recently completed tasks to larger issues. Teachers have thought about standards and generally can discuss them. Question 1 specifically mentioned the national Common Core, as all three states involved in this dissertation had adopted the Common Core standards at the time of the interview.

Questions 4 through 9 inquired about the core issues of teacher concerns and preparations. Separate questions about preparation and concerns appear for probability (Questions 4 and 5) and conditional probability (Questions 7 and 8), because state standards and the Common Core contain probability topics not related to conditional probability. In the Common Core, Grade 7 and high school standards cover random variables, probability distributions, and expected value (NGA Center for Best Practices & CCSSO, 2010). Originally, Question 6 about experiences with conditional probability appeared on the written questionnaire administered before the interview, but I deemed it important enough to move into the discussion. Some of the participants had offered incidents earlier; others told of their experiences in the open-ended portion of the interview.

Participant Recruitment

Having selected interview tasks and questions, the next step was to recruit current high school mathematics teachers. A current high school mathematics teacher was defined as someone who taught at least one class of mathematics or statistics in the 2013–2014 school year, to students in Grades 9 through 12. I did not have access to a sampling frame that would enable a probability sample. The participants formed a convenience sample. Nevertheless, I attempted to gather a relatively balanced representation of probability teaching experience. For instance, it might have been possible to request participants through the AP® Statistics examination reader group on Facebook, but AP® Statistics readers likely have greater experience and skills than

typical high school teachers. Oversampling highly skilled teachers would have overestimated teacher content knowledge and might have underestimated teacher needs.

To determine sample size, I considered prior interview-based studies of practicing teachers. Watson (2001) had 21 in-person interviews as part of her 43 total participants. Begg and Edwards (1999) talked to 22 practicing primary school teachers. Vermette and Gattuso (2014) interviewed 12 high school mathematics teachers. After considering these studies and available resources, I set the desired sample size at 25, similar to the larger investigations.

To recruit participants, I contacted mathematics teachers I knew and encouraged them to spread the offer to others at their schools. I also reached out to administrators. Several people mentioned in the Acknowledgements aided in recruitment. In addition to personal requests, messages were also sent to a Georgia statistics teacher list, a Georgia AP® Calculus teacher discussion group, and a local list of mathematics education graduates from the University of Georgia. Professional contacts through my parents proved particularly helpful.

Participants were given a monetary incentive, a \$50 gift card from Amazon. Generous donors contributed \$1,250 to the Statistics Department Discretionary Fund in order to purchase the gift cards; a petty cash account in the Department of Statistics was established to track disbursements. I decided to offer incentives, and find the necessary funds, because the time of working professionals should be respected. I also thought monetary incentive might also encourage participation from more undecided people with less experience, one of my goals in sample selection.

Though not intended, distribution across states was balanced, with 9 teachers from my childhood state of Pennsylvania, 8 from my parents' current residence of South Carolina, and 8 from my current residence of Georgia. Confidentiality restrictions prevent naming schools or

teachers, so the following descriptions provide general location information. In three schools, a majority of the mathematics teachers participated in the study, a fortunate outcome because it increased the representativeness of probability teaching experience.

The Pennsylvania participants all lived in the central part of the state. At one school, all five mathematics teachers volunteered. In a neighboring school district, three of the four mathematics teachers participated, excluding the department chair, who covered classes while the other teachers participated in the study. The final Pennsylvania participant came from a district about two hours away. In the South Carolina Lowcountry, six teachers came from one high school, six of the eight who taught mathematics at that school. Of the two teachers unable to participate, one was the AP® Statistics instructor. Two other participants worked at schools in the same county. In Georgia, all the teachers worked within 90 minutes of the University of Georgia, with no more than two participants from any school or three from any school district. The Georgia group included two private school teachers, the only ones in the sample.

Interview Protocol

I conducted the 25 face-to-face interviews between May and July 2014, with a majority of the interviews in early June. I traveled to a variety of locations. In Georgia, six interviews took place in Aderhold Hall on the University of Georgia campus, and two took place at a school site. In South Carolina, seven interviews took place in school buildings and one in a community room. Eight of the nine Pennsylvania teachers were interviewed in school classrooms. The final Pennsylvania teacher and I spoke in a conference room at a summer conference. In total, five interviews occurred during school hours, warranting my preparation for time constraints.

Each interview consisted of five parts: written confirmation of consent, distribution of gift card, background written questionnaire, tasks, and open-ended questions. The first three

parts were not recorded. Tasks and discussion were videotaped, with a single camera pointed at the participant's paper. The intent was not to create videos for outside display; rather, the video camera allowed me to reference gestures in transcripts. Recorded interview time ranged from 43 to 101 minutes, with a mean of 75 minutes and median of 74. Most of the interviews lasted 60 to 90 minutes, with four more than 90 and two less than 60.

For each task, I asked the participants to write their answers on the provided task paper. The written work does not contain the name of the subject, just an interview number. To increase confidentiality, the participants selected their own natural numbers, as long as no prior interview had the same number. I offered pencils, pens, and a TI-83 calculator, but some participants used their own writing implements and calculator. I took handwritten notes about solution paths, misconceptions, and interesting quotes. As we discussed each task, I recorded my impression about the accuracy of the solution. My impression occasionally changed during subsequent review of the data, as mentioned later, but served as a starting point for the content analysis. I did not automatically indicate if an answer was correct, but if a teacher asked, I would respond. I also helped participants work through problems if they asked for assistance.

Participant Demographics

After receiving a gift card, each teacher filled out a one-page questionnaire while I set up camera equipment. The questionnaire, shown as Figure 3.6, asked for background information on teaching history, coursework, training, and probability experience.

1. How many years have you been teaching mathematics?	
2. Have you taught mathematics at levels besides high school, Grades 9-12? If so, at what levels?	
3. Have you taught any subjects other than mathematics? If so, which subjects?	
4. Have you ever taught a separate course on probability or statistics? If so, for how many years? And was this course an AP course?	
5. In which state or country did you first teach?	
6. Besides the place where you first taught, have you taught in any other states or countries? If so, which ones?	
7. Circle the college degrees that you hold.	
Bachelor's in Education	Bachelor's in Mathematics
Master's in Education	Master's in Mathematics
Other (please list)	
8. How many college courses have you taken on probability or statistics?	
9. In what year was the most recent course, if any?	
10. Specifically relating to probability, circle the coursework you have taken.	
Topics as part of a statistics course	A standalone course on probability
More than one standalone course (please explain)	
11. How much training, such as workshops or professional development, have you had related to probability or statistics?	
12. How much training was specifically related to probability?	

Figure 3.6. Participant background questionnaire.

Of the 25 participants, 8 were male and 17 were female. Table 3.2 contains a list of academic degrees listed by the participants. Nineteen teachers had a masters' degree or higher. A few teachers listed only their highest degree, so the count of bachelors' degrees is less than 25.

Table 3.2

Academic Degrees Listed by Participants

Type	Field	Count
PhD	Mathematics education	1
JD	Law	1
Specialist	Education	3
Masters	Education	17
Masters	Computer science	1
Masters	Mathematics	1
Bachelors	Education	9
Bachelors	Mathematics	9
Bachelors	Accounting	1
Bachelors	Industrial engineering	1

For most of the participants, classroom exposure to probability had occurred as part of a statistics course. The year of most recent course ranged from 1977 to 2013, with a slight majority (15) during this millennium. Only three participants had taken a stand-alone probability course; none had taken more than one. Figure 3.7 visually displays statistics and probability course counts. For participants with a separate probability course, their total includes that course; for instance, the box above 3 in the dot plot stands for 2 statistics courses plus 1 probability course.

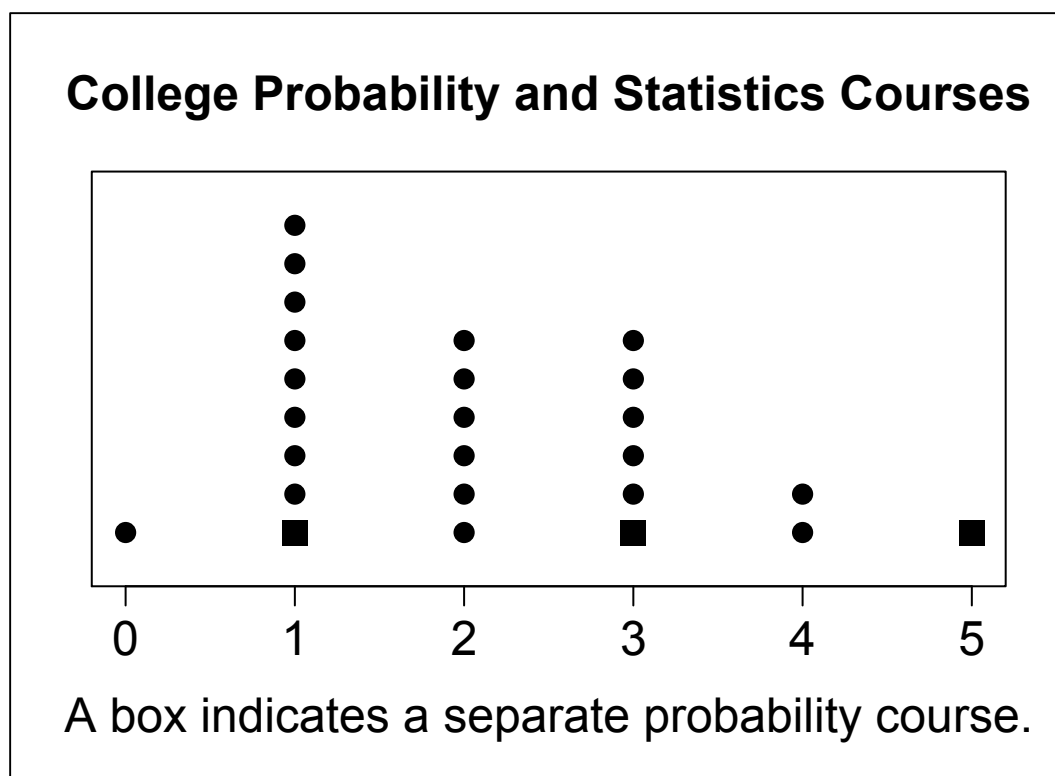


Figure 3.7. Participant probability and statistics course experience.

The participants reported little training in the teaching of probability and statistics, with only 9 of the 25 teachers ever having any workshops or professional development. Three of the teachers taught AP® Statistics; they had all participated in at least one week of training, which includes one to two days on the probability portion of the syllabus. One Pennsylvania teacher had received two days of probability instruction as part of a school district initiative; he described it as helpful. Three other teachers reported multiple days of statistics training, but only one of them had more than a half day specifically on probability. The final two teachers had short workshops, with just a couple hours discussing probability.

Shifting to teaching experience, the number of years taught ranged from 2 to 33, with a mean of 12.48 years. The study participants had a comparable distribution of teaching experience to that described in a national report about secondary public school mathematics teachers (Snyder & Dillow, 2013). In the national survey, 46% had less than 10 years of experience, 35% had 10 to 20 years, and 19% had more than 20 years. In this study, 11 participants (44%) had less than 10 years of teaching experience, 9 (36%) had 10 to 20 years, and 5 (20%) had more than 20 years.

At the time of the interview, all participants taught mathematics to students in grades between 9 and 12, but 12 teachers had teaching experience other than high school mathematics. Three teachers had previously taught college mathematics; 8 teachers had previously led mathematics classes with students in Grades 6 through 8. Eight participants had taught classes other than mathematics. Physics was most common with 3 responses; computer science had 2; study skills, personal finance, English, French, physical education, and biology each had 1 response. Five teachers had shifted states during their career, but no prior location was mentioned more than once.

Seven teachers (28%) reported prior experience teaching high school statistics. The three AP® Statistics teachers resided in different states; four participants had taught a non-AP® statistics course, at least one in each state. Unlike years of experience, government data do not exist on courses taught. In four schools from which teachers volunteered for this project, where I know all the teachers, 4 of 29 people (14%) teach statistics. Based on this limited information, I somewhat oversampled statistics teachers. Nevertheless, 13 of the 25 participants had no statistics teaching experience and no professional development in probability or statistics. Having a majority of inexperienced respondents is important because state instructional

standards do not restrict conditional probability to specialized courses. I compare results from experienced and inexperienced instructors, but in order to propose ideas to help all teachers prepare to teach the subject, examining the general situation is more important.

Initial Analysis

Examination for analysis began after each interview, as I transferred the video recording from the camera memory card to a secure hard drive. During each interview, I had written an impression about the correctness of each response. An answer was considered correct if the teacher's solution had no more than miniscule calculator errors, plus some appropriate form of logic. For instance, I marked ELISA Part C correct when I witnessed a calculator error after the participant properly wrote out all possible cases with the proper probabilities. On the other hand, miscopying a number from earlier in the problem made a solution incorrect. Choosing the right option in the Rash and Lucky Dip tasks, but making an error when stating the reason, was also not considered correct. When my notes indicated uncertainty about response correctness, I watched the appropriate portion of the interview to decide. After deciding on initial content correctness, I recorded results in a spreadsheet.

During the initial analysis, I more fully examined how the participants solved problems by categorizing wrong answers in the three easier tasks: Rash, Lucky Dip, and Survey. The more difficult Taxicab and ELISA tasks had more errors, so I saved them for detailed analysis. When I considered an answer incorrect, I recorded a brief description of the teacher's error. For example, on the Rash task five teachers gave answers initially considered incorrect. Three of the five improperly compared proportions out of the grand total, $223/426$ against $107/426$. One person made the key error, directly comparing the counts of persons who got better under both

treatments, 223 against 107. The last person misread the question and gave a response that did not meaningfully answer the question.

Initial codings were never intended as final responses; they had two intermediate purposes. First, I wanted to check if results from this sample reflected success rates found in prior research. For most questions, the results aligned with the literature. Independence in Survey C was an important exception; the participants struggled more than I had expected with the probabilistic meaning of independence. Second, I wanted preliminary results to publicize. Based on the initial analysis, I presented a poster at ICOTS in July 2014 and a contributed talk at the August 2014 Joint Statistical Meetings.

Detailed Analysis

Later, I searched through each interview transcript to find information on the three research questions. The interviews generated a large data corpus, with 21,000 transcript lines containing about 220,000 words. My goal was to reduce the data corpus into manageable answers about solution paths and misconceptions of the participants, suggested responses to students, and needs of the participants for teaching. In methodological terms, I performed thematic analysis, searching across the data to find repeated patterns of meaning (Braun & Clarke, 2006).

I first recorded solution paths, student misconceptions with replies, and expressed needs. Some solution paths might have been more thoughtful than others, but all attempts were documented. Since I did not have a second coder, I used intra-rater reliability by evaluating the correctness of responses a second time. When my two evaluations differed, I reviewed interview videos and notes to make a final determination. I made seven changes (3.3% of questions) between initial and final analysis.

After recording answers, I searched for patterns in the data that answered the research questions. Thematic analysis made sense, rather than the more open alternative of grounded theory, because some potential codes are known from prior studies. For instance, mutually exclusive events are frequently confused with independent events (Rumsey, 2008), so I marked each instance when mutually exclusive was confused with independent. When looking through coded transcripts, my overall goal was to describe the teachers' patterns of experience and suggest uniting threads, the goal of thematic analysis (Ayres, 2008).

In the next few paragraphs, I set out more specifics for each research question and explain how I arranged results. Throughout the results, I use quotes from the transcripts. Instead of creating 25 pseudonyms, I indicate speakers by their self-selected number.

To answer Research Question 1, how teachers solve conditional probability tasks, I recorded solution paths participants used or mentioned during discussion. Chapter 4 contains two tables of content solutions for each task. In the first table, I list each *primary solution path*, the solution used in the teacher's stated answer. Attempted questions always had a primary solution path for each teacher, even if it was getting stuck. Primary paths are ordered by frequency, with indications if the eventual solution was correct or incorrect. The second table contains all mentioned solution paths; this table does not include correctness because not all proposed solutions were worked out. In addition to the tables, I discuss common solution paths and compare participants' solutions to prior research.

To answer Research Question 2, how teachers respond to student misconceptions, I catalogued mentioned misconceptions for each question, creating one table for each question. After each table, I describe teacher responses to common misconceptions, illustrating responses with teacher quotes. Tasks except ELISA had key misconceptions drawn from the literature,

which I generally—but not always, depending on time and conversation flow—introduced into the discussion if not mentioned by the teacher. For questions with a key misconception, I describe responses teachers said they would make on that question. In Chapter 4, I report on replies at the individual task level. Connections across questions are described in Chapter 5.

To answer Research Question 3, what teachers perceive as their needs to teach conditional probability, I read through the transcripts to classify comments. Most comments about needs occurred during the concluding period of open discussion. The teachers spoke about subject matter, pedagogical, and curricular knowledge, the three categories proposed by Shulman (1986). I grouped comments under four major areas: (a) curricular issues, (b) current sources of assistance, (c) teacher concerns about their ability to teach conditional probability, and (d) requests for assistance. In Chapter 5, I describe themes in each major area, with extra attention to themes mentioned in multiple states.

CHAPTER 4

TASK RESULTS

In previous chapters, I introduced the problem of understanding high school teacher knowledge about conditional probability; chronicled curricular documents and prior research; and then described interview materials, participant information, and the analysis plan. In this chapter, I begin presenting results from the interviews. This chapter contains question-by-question results about the first two research questions, teachers' content knowledge about conditional probability and teachers' knowledge of student misconceptions. Results that combine information across multiple tasks appear in Chapter 5.

I begin this chapter with a summary of overall subject matter results, including correct responses by participant, correct responses by question, and performance comparisons to other groups who attempted these tasks. After the overall results, the chapter contains one section for each of the nine questions on probability that comprise the five tasks. Each section contains (a) an introduction with content results for that question, (b) a table of primary solution paths, (c) a table of all solution paths, (d) a discussion of common solution paths, (e) a table of listed misconceptions, and (f) a discussion of teacher responses to common misconceptions, including the key misconception when applicable.

Overall Subject Matter Results

Overall, the participants answered 108 of 211 (51%) questions correctly. No teacher got all nine questions right; one AP® Statistics instructor had eight correct answers, missing only ELISA Part C. Results by participant are graphed in Figure 4.1. Because four interviews omitted

the entire ELISA task, and two others omitted ELISA Part C, I display the percent correct instead of the number of questions answered correctly. In the figure, black circles represent the 3 AP® Statistics teachers, gray circles represent the 4 non-AP® probability and statistics teachers, and white circles represent the 18 teachers without probability and statistics course teaching experience.

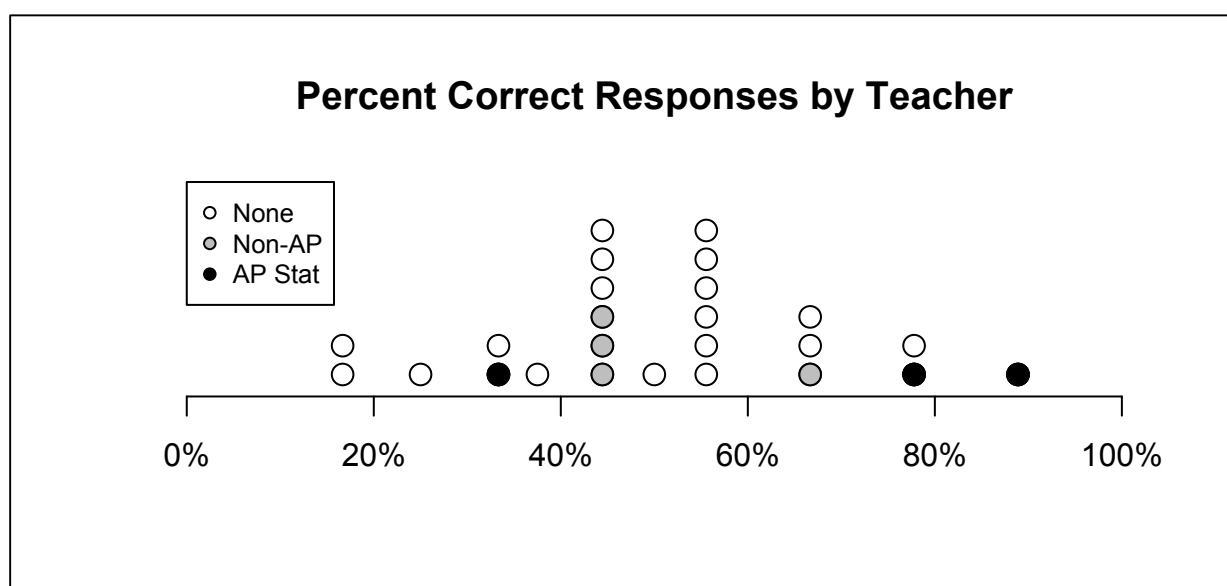


Figure 4.1. Correct responses by teacher.

I also examined results on the first six questions, not including any parts of the ELISA task. In Figure 4.2, I display the number of questions correct (out of six) for each participant. The only two participants with more than four correct answers from the first six questions were both experienced AP® Statistics teachers. The Multistage Taxicab problem and Survey Part C about independence were very challenging for almost all of the participants.

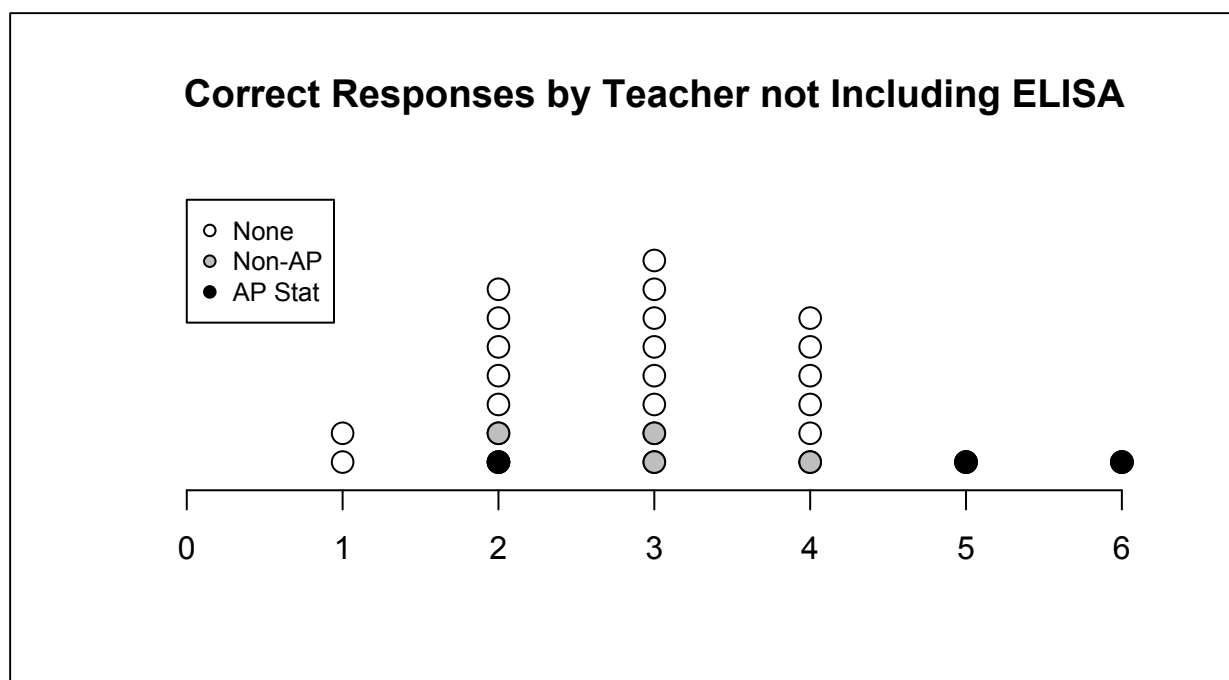


Figure 4.2. Correct responses by teacher, not including ELISA task.

Because of the broad range of interview lengths, from 43 to 101 minutes, I checked whether interview time had a relationship with the number of correct responses (not including ELISA). I found no pattern between interview time and subject matter correctness. The participant with the shortest interview had only one correct answer, but the second quickest interviewee had four correct answers and the other participant with just one correct answer took 87 minutes. The two participants with more than four correct answers took 78 and 82 minutes. The Spearman rank correlation between interview time and number of correct responses (not including ELISA) was close to zero, $r(23) = -.051$. I also noticed no pattern between time taken and quality of responses to student misconceptions. This made sense, because the teachers generally offered prompt responses to potential student misconceptions. They tried to react as they would in the classroom, even though the interview format only approximated a class. The

prime contributor to interview time differences was problem solving speed, which as noted above had no discernable relationship with correctness. Digressions on subjects outside probability also occasionally extended interviews.

I compared results for the participants with and without probability and statistics teaching experience on all of the questions, including ELISA. Those participants with probability and statistics teaching experience correctly answered 36 of 63 (57%) questions. Those participants without probability and statistics teaching experience correctly answered 72 of 148 (49%) questions. I had expected a larger gap between the two groups.

Across the three levels of question difficulty, the participants correctly answered 75 of 92 (82%) Direct noncomputation questions, 31 of 75 (41%) Computation questions, and 2 of 44 (5%) Multistage complex questions. The percent correct by question is shown in Figure 4.3. In the figure, Direct questions have black bars; Computation questions have gray bars; Multistage questions have white bars. The participants did best finding a conditional fraction in the Survey B question, making only one error. They were least often correct on the Taxicab and ELISA Part C questions, giving only one correct answer on each part.

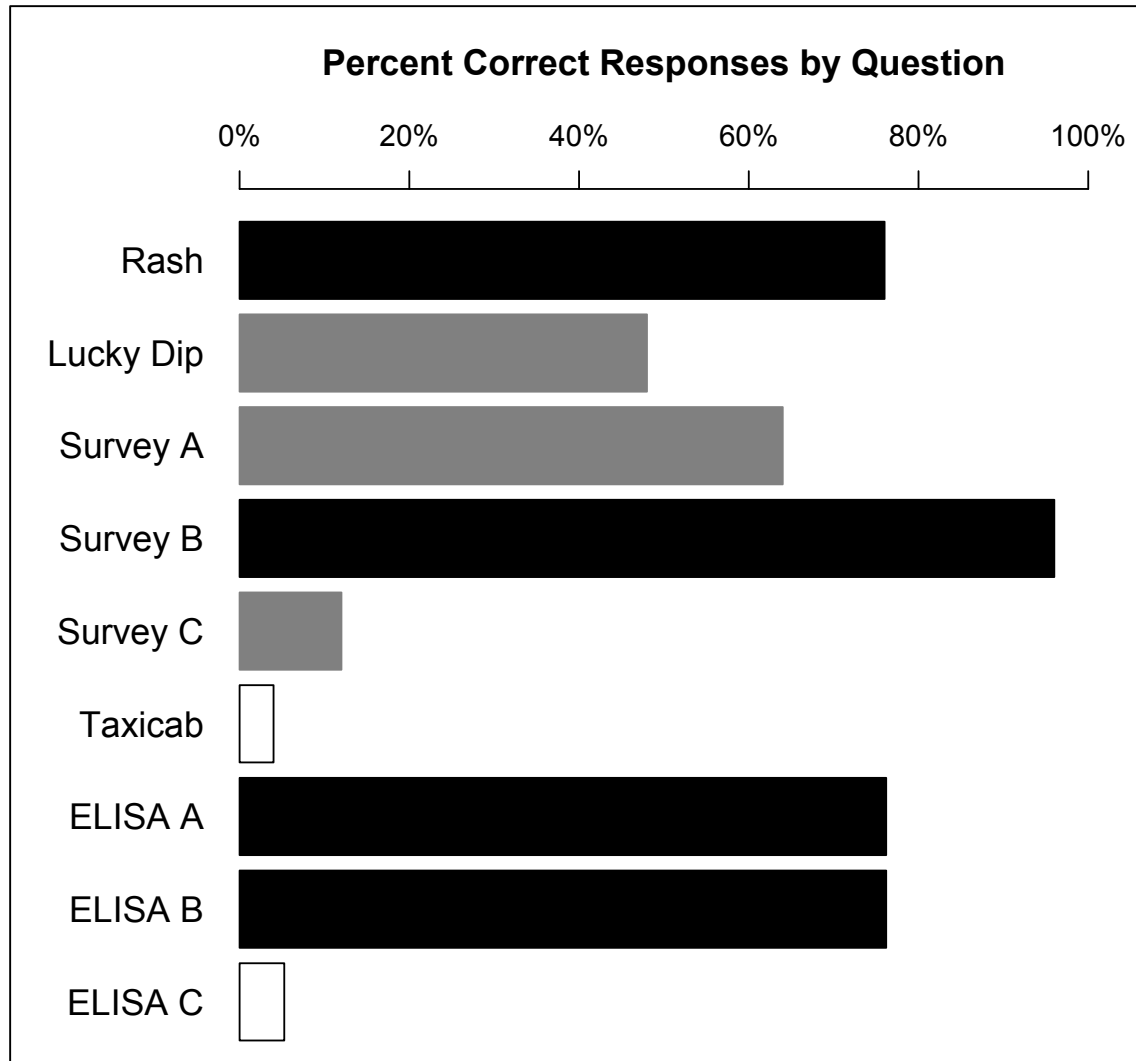


Figure 4.3. Correct responses by question.

The study participants can be compared to previously studied groups on every task except Lucky Dip. On the Rash question, Kahan et al. (2013) reported that 41% of their nationally representative sample of adults answered correctly. The participants' performance in this study, with 76% correct, matches the performance of people in the top decile of numeracy in the Kahan et al. sample. On the Taxicab question, the 4% correct rate in this study is similar to the 6%

correct rate on an Israeli college entrance exam reported by Bar-Hillel (1980) and the 8% correct rate of medical doctors reported by Gigerenzer (2002).

The Survey and ELISA problems were taken from old AP® Statistics exams. For comparative purposes, I graded the teacher participants' responses using the published scoring guidelines. On the ELISA task, the mean teacher score was 1.76 out of 4. On the Survey task, the mean teacher score was 1.86, but the problem administered on the exam had a fourth part that I did not include because the part did not ask about conditional probability. Because both tasks appeared on an alternate form of the exam, score distributions are not available. I can make a rough comparison by considering general AP® Statistics scores. Problems tend to have a mean score around 1.75 and mean scores rarely exceed 2.0. The participants' average ELISA score is comparable to those averages. The participants' average Survey score appears better than average because the participants had a maximum possible score of 3, not 4. A mean score of 1.86 out of 3 would have likely become more than 2.0 out of 4 if the participants had attempted the fourth part of the Survey task.

In summary, the study participants did notably better than a representative sample of U.S. adults on the Rash task, appeared to exceed an average AP® Statistics score on the Survey task, performed comparably to other samples on the Taxicab task, and appeared to have about an average AP® Statistics score on the ELISA task.

Rash Results

This question tested participants' ability to compare rates of improvement between two groups when the results of the experiment are displayed in a table. Of the 25 participants, 19 answered this Direct question correctly by determining that in the experiment, the new cream was less effective than not using the cream. Most of the teachers compared proportions between

the two groups shown in the rows of the table. Half the teachers computed proportions for both patients who got better and patients who got worse, although only one of the two computations would have sufficed. A few of the participants mentioned that they did the second set of computations to check their first results. Others used both comparisons in their explanation. When categorizing solution paths, I distinguished between explanations that mentioned only the patients who got better, explanations that mentioned only the patients who got worse, and explanations that mentioned both sets of patients, those that got better and those that got worse. This distinction appears in the two tables of results, primary solution paths in Table 4.1 and all suggested solution paths in Table 4.2.

Table 4.1

Primary Solution Paths in Rash Question

Primary solution path	Correct	Incorrect
Group comparison (better only)	11	0
Group comparison (better and worse)	8	0
Cell counts out of grand total 426	0	5
Comparing cell counts, 223 against 107	0	1

Table 4.2

All Solution Paths in Rash Question

Solution path	Count
Group comparison (better only)	14
Group comparison (better and worse)	9
Cell counts out of grand total 426	5
Group comparison (worse only)	3
Chi-square hypothesis test	2
Bar graph	1
Column comparison	1
Comparing cell counts, 223 against 107	1
Comparing marginal proportions	1
Decision tree	1
Relative risk	1

Three noteworthy issues arose during the solving of this problem: marginal totals, sample sizes, and hypothesis testing. Since the source problem (Kahan et al., 2013) did not provide marginal totals for the rows and columns, I did not include marginal totals in the problem statement. About half the teachers (12 of 25) added marginal row totals or column totals to the table. Teacher 21 spoke about why she found both row and column totals, calling it a routine: “I realized this [column margins] has nothing to do with it. But I would always tell them because I remember these problems when I taught it, is you always total everything, even if you’re not going to use it.” A few of the teachers touched upon the lack of row and column totals during the discussion; these teachers would have preferred that the table already have totals provided because it would simplify the problem.

Seven teachers made a comment about the vastly different sample sizes in the two groups, 298 with the new skin cream treatment against 128 without. This comment was not surprising, as pilot subjects had been asked about the disparity. I had added a sentence to the problem, “The total number of patients in the two groups was not the same, but this does not prevent assessment of the results,” to try to alleviate this concern. Nevertheless, some of the participants still talked about the discrepancy. Three teachers qualified their conclusions because of the disparity, including Teacher 23, who said,

Because there wasn't the same amount [of] testing in both experiments, when I'm coming up with the percentage then it's not the same people, so I guess that makes a difference too, but I think I could make a better judgment if it was the same amount of people in both cases.

I considered qualified solutions correct if the teacher correctly computed rates and made a statement that the group without the new skin cream recovered more often. I did not ask why people thought equal sample sizes were necessary; although interesting, that question falls under the statistical content areas of experimental design and inference, not the scope of this study.

Two teachers ventured further into inference by initially suggesting a hypothesis test, and two more asked me about potential statistical significance. Suggesting a hypothesis test is not a misconception; it is a different assumption: that patients in this experiment came from a larger population. During the interview, I asked participants to focus on only the 426 experimental patients and base their decision on conditional probability. Thinking about how population assumptions affect the choice of statistical technique is important, but outside the scope of this study.

Shifting from teacher content knowledge to the ability of teachers to anticipate student misconceptions, Table 4.3 contains misconceptions mentioned by the teachers. Four misconceptions were mentioned by at least five teachers: the key misconception of comparing

cell counts without computing proportions, poor reading comprehension, incorrectly forming fractions within the table, and computation issues.

Table 4.3

Misconceptions Mentioned in Rash Question

Misconception	Count
Comparing cell counts, 223 against 107 (key misconception)	11
Reading comprehension	9
Forming fractions from incorrect table values	7
Computation and rounding issues	6
Cell counts out of grand total 426	3
Trouble finding marginal totals in table	3
Column comparisons instead of groups in rows	2
Results not in line with expectations about medicine	2
Comparing column marginal totals	1
Using counts not fractions	1
Attempting to use permutations or combinations	1

To respond to the key misconception, comparing cell counts without forming fractions, most of the teachers suggested some kind of reminder about the remaining information in the question. Some would be direct, asking about or pointing to the table column of people who got worse. Others would verbally mention the two separate groups, like Teacher 10 who said,

Obviously it's not out of this total. We have two separate groups, so we would need to compare the group of the rash that got better out of the total of the groups that used the cream. So you can't compare apples to oranges.

Three teachers would make a different response; they would ask students to fill in marginal row totals on their papers and then consider the distinct sample sizes.

Reading comprehension worries came up on most questions, but were mentioned most often on this task and the ELISA task because of their wordy scientific nature. When I asked Teacher 17 about her concern, she replied that students tend to struggle to find the key information:

Is this extraneous information, is this not extraneous information, do I need to know this, is this trying to lead me somewhere? And my kids have been taught to take tests so much that that's almost their first response is, is this the extra information, do I need this?

Most of the teachers stressed careful reading in their responses. Two teachers suggested making the problem more relatable to counteract misunderstandings and the wordiness of the problem.

Teacher 17 wanted to shift the context from skin cream to going to the movies. Teacher 16 offered an example with students' eye color.

For the third common misconception, forming fractions, several teachers mentioned that students might haphazardly write down visible numbers as numerators and denominators.

Teacher 12 offered an example, when students "might just put 223 over 107. That would be the students that don't have much understanding of what they were doing and [were] just trying to copy whatever they saw the teacher doing without understanding what was going on." To respond to this misconception, some of the teachers stressed context, understanding the purpose of the numerator and denominator. For example, Teacher 12 would ask students to think about what they are trying to find:

Well, if you're trying to find a percentage of something, what is a percent? Um, a percent is supposed to be the part out of the whole. And I would try and through questioning lead them to figuring out: Oh okay, the [107] isn't the whole that they're trying to figure out.

One misconception I expected the teachers to mention more often was the presumption that new medical treatments make things better. I chose this problem in part because the numbers did not support the idea that medical treatment improves outcomes. I thought that teachers might mention this when solving the problem, but none did. Only two interviewees brought it up as a potential student misconception, including Teacher 104, who said “Just by reading it, [students have] already decided if they’re using the cream they’re going to get better. They almost use the given information to make their decision instead of relying on the math.” Most of the teachers in the study relied on data instead of contextual conceptions about medicine; I wonder if students would think more subjectively and weigh prior beliefs about medicine more heavily.

Lucky Dip Results

This question tested participants’ ability to determine which player, Amy or Dominic, had a better chance of winning a ball-drawing game. Of the 25 participants, 12 answered this Computation question correctly, identifying the game as unfair and determining that Dominic had a 60% probability of winning against Amy’s 40%. Because the game had to have a winner, computed probabilities needed to total 100%. I considered answers incorrect when the probabilities did not total 100%, even though Dominic might have been identified as the favorite.

The teachers used five distinct correct primary solution paths, more than any other question. Primary solution paths appear in Table 4.4. There were many solution paths, as shown in Table 4.5, with six approaches mentioned five or more times. The most common correct primary solution path was direct conditional logic, in which a teacher wrote down probability for the second ball given an outcome for the first ball. The most common incorrect primary path was computing only one color option instead of two, resulting in victory probabilities of 20% for Amy and 30% for Dominic.

Seven teachers mentioned a decision tree. In probability, a decision tree, or tree diagram, is a visual model that displays all possible outcomes of a sequential experiment. Tree diagrams are included in the Common Core standards in Grade 7 (NGA Center for Best Practices & CCSSO, 2010). A decision tree begins at a root node. From the root node, one branch is drawn for each possible outcome of the first event. The probability of travelling to each node is written along the path to that node. I present a tree diagram for this task in Figure 4.4. The first ball selected could be white with probability $3/6$, or black with probability $3/6$, so the decision tree has two first level nodes.

To continue the tree diagram, branches at the second level are drawn conditional on results at the first level node. In this task, each second level decision has two outcomes, black and white, but the event probabilities differ based on the first drawn ball. The probability that the second draw is black given the first draw was black is $2/5$, but the probability that the second draw is black given the first draw was white is $3/5$. If there are more than two levels of branches, the process repeats, with probabilities along each branch computed conditional on prior branches in the tree. In the completed diagram, each path from the root node to an end node represents a mutually exclusive outcome. The probability of an outcome is found by multiplying the probabilities shown on each branch of that path. For example, in Figure 4.4 the probability of two black balls is $(3/6 * 2/5) = 6/30$.

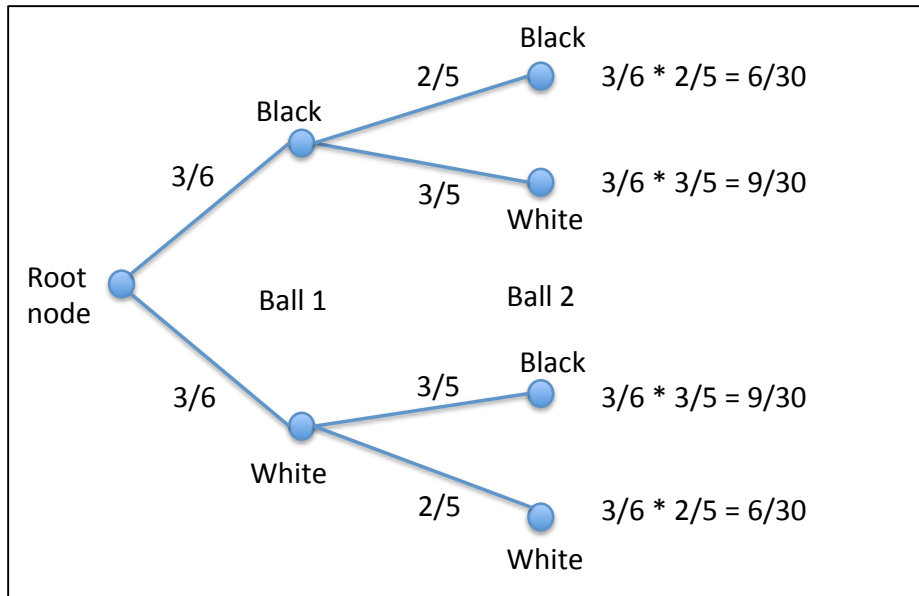


Figure 4.4. Tree diagram for Lucky Dip question.

Table 4.4

Primary Solution Paths in Lucky Dip Question

Primary solution path	Correct	Incorrect
Direct conditional logic	6	2
One-color option, 20% and 30%	0	6
Enumeration of outcomes	2	3
Outcome space of size 3, BB BW WW	0	2
Outcome space of size 4, BB BW WB WW	2	0
Decision tree	1	0
Permutations and combinations	1	0

Table 4.5

All Solution Paths in Lucky Dip Question

Solution path	Count
Direct conditional logic	10
Enumeration of outcomes	8
Decision tree	7
One-color option, 20% and 30%	7
Physical experimentation	6
Permutations and combinations	5
Outcome space of size 4, BB BW WB WW	5
Paper and pencil drawing	3
Outcome space of size 3, BB BW WW	2
Computation with complements	1
Conditional probability formula	1
2 by 2 table	1
Venn diagram	1

Both correct conditional logic and the incorrect one-color option involve sequential statements. One important realization differentiates them: whether the color of the first ball matters. Some participants correctly realized that the color of the first ball did not matter, like Teacher 12:

The first ball that is drawn, it doesn't really matter what color is drawn, so we could call the ball that she draws color one. And so then there's five left in the bag, there's two left of the color one and there's three left of the opposite color. So if Amy pulls another one that is the same as color one then she wins, but there's only two left out of the five. The other three would make Dominic win, which means that he would have a three out of five probability of winning.

In contrast, those teachers who specified a first color wound up in the incorrect one-color option.

Teacher 64 computed probabilities of .2 for Amy and .3 for Dominic. In her explanation, she took the same logical steps as Teacher 12, but chose a color for the first ball. This led her to an incomplete sample space:

On the first, if Amy picks, the probability that she would get, let's say, a black ball would be one out of two. And then the probability, given that she got a black ball, that she gets another black ball would be two out of five. So then I would multiply those probabilities together, and I would get two out of ten, or one fifth. That's if the balls are the same color.

If they are different colors, then the probability that ... Amy gets a black ball is still one out of two, and then the probability that she would get a white ball given that she got a black ball would be three out of five. So the probability would be three out of ten. So Dominic has a better chance of winning then because his probability is three out of ten and hers is two out of ten.

Teachers who used enumeration also had trouble completing the sample space; three of the five teachers who attempted to list all possible outcomes failed to complete the space correctly. Some of the teachers were not happy taking this solution path. Teacher 21 used enumeration as her primary solution path even though she thought it was “not the teacher way to do it. Like you wouldn't count, you would use the formula.” Teacher 12 said that a pictorial solution was not something “that I would go to normally, but I know that's something that would make more sense to my students.” Teacher 42 successfully found all possible outcomes via enumeration, but believed students would have trouble because the balls were not distinct. He said enumeration “only truly works well when you have different numbers or different objects.”

I made a few other notes about solution paths. The sample solutions provided in the task (Mathematics Assessment Resource Service, 2012) did not rely on formulas, so I tallied nonformulaic primary solution paths. Six teachers used a nonformulaic primary solution path: five enumerations and one decision tree. As teachers worked, I tracked whether they drew any

kind of diagram on their paper. Seven teachers did so. Three teachers mentioned drawings as a possible solution path, although only one of those three actually had made a drawing. Next, both teachers who used an outcome space of size 3—two black balls, one of each color, and two white balls—made the equiprobability error. They claimed that the three outcomes had equal probability; under this claim, Amy had greater likelihood of winning because two outcomes favored her. Finally, slightly over half the teachers (13 of 25) asked me if this question involved replacement of the first drawn ball. I replied that Amy put two balls on the table. The prevalence of the replacement question surprised me, as I had thought the wording of the question was not ambiguous and thus did not modify the original source.

Perhaps the teachers' questions about replacement anticipated their thoughts about misconceptions, as sampling with replacement was mentioned by half the participants. The second most common misconception noted was confusing addition ("or") with multiplication ("and"). Only three teachers mentioned equiprobability of outcomes, my selected key misconception. I list all stated misconceptions in Table 4.6.

Table 4.6

Misconceptions Mentioned in Lucky Dip Question

Misconception	Count
Sampling with replacement	13
Adding (OR) instead of multiplying (AND)	6
Computation errors with fractions	5
One color option, 20% and 30%	4
Equiprobability (key misconception)	3
Incomplete enumeration	3
Confusing ball with color type	2
Picking 2/6 as 2 balls out of 6 in bag	2
Finding counts not probabilities	1
Guessing	1
Improper labeling	1
Outcome space of size 3	1
Symbol mistakes in formulas	1
Time-axis causality	1

When responding to a student who sampled with replacement, the teachers most often said they would ask the students about what remained in the bag after seeing one ball. As an alternative, several teachers would want students to physically select balls from a bag; Teacher 17 would expose the mechanism by beginning with the balls on the table. Nevertheless, most teachers offered a response similar to the following imaginary scenario provided by Teacher 77:

[I'd] kind of give them a scenario ... What color did she pick the first time? So, I don't know, she picked a black one. And I'd say, "Okay, both three blacks and three whites, if she picked the black the first time how many are left?" Two. Okay, so that means that the three they put is wrong.

And then I'd say, "Does it say anything about putting the ball back?" [They'd say,] "No, it's basically picking two at the same time." So you're down to one less ball than you started with, so it can't be three out of six.

In this question, which does not rely on vocabulary, the teachers preferred to ask questions, not give direct guidance. Teacher 66 described his strategy by saying, "You just can't ask right away, if they had 20 and 30, say well, why is it not 40 and 60? ... The point of asking questions is to get them to answer the questions along the way." The teachers would also use enumeration, but as mentioned earlier, they generally did not prefer this option. Teacher 13 used enumeration as a backup in the following response to the one color misconception:

If we've done a lot up to this point with the ands and the ors and everything else up to this point and they're still not seeing it, I would actually have them explain it by: "Okay, let's list out all the scenarios. What could you possibly pick, or what are all the possibilities that can be picked out of this?" And once they actually see all the possibilities, then they would have a better sense of why they missed this second [color] part.

Reluctance about enumeration did not extend to the key misconception, however. In most interviews, I introduced the misconception by writing the four possible outcomes on paper while describing the situation. I thought the teachers would respond by computing probabilities and showing the probabilities not equal. Only a minority of the teachers took this approach. A slight majority of teachers treated the misconception as an enumeration error and responded by clarifying the enumerative process. Teacher 64 gave a representative reply:

You would technically have to say you could get Ball 1, Ball 2, so you would put subscripts under [each ball]. Well, this [pointing to a ball] could be Ball 2 but it wouldn't be. You're not getting all the possibilities, because what if the balls had the number 1, 2, and 3 on them, then they would be different balls [with more possibilities].

Misconceptions and response strategies mentioned by the participants aligned with those in prior research. For instance, Zapata-Cardona (2008) asked two teachers about a similar problem with bags and marbles. Those teachers also suggested enumeration, physical

experimentation, and trees as possible solution paths. In the teacher's guide for this task, the original authors provided a list of common issues (Mathematics Assessment Resource Service, 2012). Listed issues included equiprobability, replacement, and computation trouble, all issues mentioned by multiple participants.

Survey Part A Results

This question tested the participants' ability to find the probability of occurrence of the union of two events that are not disjoint when the data are given in tabular form. Of the 25 participants, 16 correctly answered this Computation question, computing a probability of $1135/2500$ that a randomly selected adult is a college graduate or obtains news primarily from the Internet. The most common approach was writing a formula that included marginal totals for both college graduates and Internet users, $693 + 687$. I call this solution *marginal fractions* in Table 4.7 of primary solution paths and Table 4.8 of all solution paths. When a participant wrote down one or zero marginal totals and added cell counts, like $693 + 41 + 401$, I label the solution *counting cells*.

In the tables, I distinguish two ways of writing fractions. The term *separate* indicates writing two or three fractions with separate denominators, like $\frac{693}{2500} + \frac{687}{2500} - \frac{245}{2500}$. The term *together* indicates writing two or three cell counts over a single denominator of 2500, like $\frac{693+687-245}{2500}$. The separate and together approaches appeared with roughly equal frequency.

Interestingly, those participants who wrote separate denominators made errors more often. Because I did not notice this distinction until performing the detailed analysis, I did not follow up during interviews.

Table 4.7

Primary Solution Paths in Survey Part A Question

Primary solution path	Correct	Incorrect
Marginal fractions (separate)	6	8
Marginal fractions (together)	7	0
Counting cells (together)	3	1

Table 4.8

All Solution Paths in Survey Part A Question

Solution path	Count
Marginal fractions (separate)	16
Marginal fractions (together)	10
Counting cells (together)	7
Subtracting cell counts from 2500	4
Venn diagram	3
Decimal probabilities	1

Textbooks such as Wackerly, Mendenhall, and Scheaffer (2008) provide a formula for the probability of one event A or another event B, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$. Most of the teachers attempted to apply this formula using marginal fractions. Seven of the nine errors made were the key misconception, failing to remember the last term of the formula and subtract the 245 people who are both college graduates and obtain news from the Internet. I frequently have seen this error when teaching probability, so this result was consistent with my expectations.

One unexpected issue was fraction simplification. About half the teachers (11 of 25) either reduced fractions like $1135/2500 = 227/500$ or asked if it was necessary. Several teachers mentioned that they automatically simplify fractions in their classes; Teacher 16 said, “That [simplification] is because ... we preach it to them.” When I teach probability, I want the denominator to represent the total larger group, and thus I do not simplify fractions. When needed, I use decimal answers. Teacher 3 expressed a similar idea when I asked why he simplified his fraction; he said

I almost don’t prefer my students to write the simplified form, I kind of do it almost out of habit. But I actually prefer them to do the 1380 out of 2500 because that seems more understanding than the 69 out of 125.

These discussions illustrated a curricular issue. Probability pedagogy prefers not reducing fractions, but simplifying serves the needs of topics like factoring and proportions. Mathematics teachers must balance these needs.

Turning to misconceptions in this problem, the most commonly mentioned mistake was the most common participant error: failing to subtract out those counted twice. The second most common error was using only those counted twice, confusing *or* with *and*. I catalog all mentioned misconceptions in Table 4.9.

Table 4.9

Misconceptions Mentioned in Survey Part A Question

Misconception	Count
Failing to subtract out 245 people counted twice (key)	16
Confusing OR with AND, yielding 245/2500	7
Reading comprehension	4
Wrong denominator	4
Computation errors	3
Including marginal totals in cell counting summation	2
Multiplying fractions instead of adding	2
Thinking problem is conditional, 245/693 or 245/687	2
Believing OR is not inclusive	1
Dividing marginal counts, 693/687	1
Guessing	1
Taking just Internet people, 687/2500	1

To respond to the key misconception, failing to subtract out the 245 people counted twice, three strategies were mentioned by at least four teachers each: (a) directly ask a question, if anyone was counted twice; (b) have students highlight the appropriate cells in the table and solve the problem with cell counting; (c) do an in-class example with physical counts, with two groups such as females and people wearing green shirts.

Because confusing an *or* (union) with an *and* (intersection) is a vocabulary issue, responses to this misconception tended to involve direct reinforcement about the words. For example, Teacher 23 said, “I would point out the fact that it stated or, separating those two categories, versus if it said a college graduate who obtained news from the Internet. They’re just

talking about one group of adults there versus two separate.” Two teachers suggested drawing a Venn diagram to illustrate the difference.

I combined the less common replies of 245/693 and 245/687 because both replies incorrectly assume the problem is conditional, more like Survey Part B than Survey Part A. After talking about this misconception, one teacher chuckled when she saw actual conditional phrasing in Question B. Responding to this error was tougher, because these student answers indicated a lack of conceptual knowledge. As Teacher 50 said, “To me that’s an indication that they all they’re doing is getting accustomed to plugging numbers in, where I worked one like this before, so this number went here; it must happen again.” Teacher responses to this misconception were more conceptual, such as using a situation with students to demonstrate the concept of union, or drawing a Venn diagram.

Survey Part B Results

This question tested participants’ ability to find a conditional probability from values in a table. It had the most correct answers and the fewest solution paths. Of the 25 participants, 24 answered this Direct question correctly, finding that the fraction of college graduates who obtained news from the Internet was 245/693. No formula was necessary, since both numerator and denominator were numbers in the table. Almost all of the interviewees directly wrote the correct conditional fraction. As an alternative, two teachers mentioned the definition of conditional probability found in Common Core Standard 3, the joint probability divided by the marginal probability. The only other suggested solution path was the key misconception, failing to use conditioning to reduce the denominator and thus writing 245/2500. Table 4.10 contains primary solution path counts; Table 4.11 contains total counts for each solution path.

Table 4.10

Primary Solution Paths in Survey Part B Question

Primary solution path	Correct	Incorrect
Writing conditional fraction	23	0
Definition of conditional probability	1	0
Writing fraction without conditioning, 245/2500	0	1

Table 4.11

All Solution Paths in Survey Part B Question

Solution Path	Count
Writing conditional fraction	24
Writing fraction without conditioning, 245/2500	3
Definition of conditional probability	2

Almost all of the teachers talked about how to find the correct denominator. Teacher 9201 described the dilemma, “Two hundred forty-five selected adult obtains news primarily from the Internet. So is it out of 693 or out of 2500 is my question.” Ten teachers physically marked or covered part of the table to help show the conditioning. Instead of a physical mark, Teacher 6 offered a hallway analogy to help students understand:

Imagine this [table] as a building and all of these as different halls or rooms, or however way you want to look at it, and then you know when we say given that they’re a college grad, that means we’re going to go down this hallway, out of everybody in this hallway how many are in this room?

Neither teacher who mentioned the conditional probability definition preferred that approach. The teacher who used the definition as primary path said that she preferred that students use the table, because the table is more obvious. Teacher 13, who offered the definition as a secondary solution path, was more negative. The textbook in her school had the conditional definition as the primary solution path for tables, but she was not pleased:

I do understand why they're doing it, but this is the reason why people who struggle with mathematics hate it. ... Because when you stick too closely to a formula and you throw all logic out the window, you make it complicated for no reason. And it just confuses the heck out of the kids, and they just get so frustrated. And then it's what turns kids off to mathematics, and it's like, well, stop doing that.

She illustrated her point by putting a large X over the formula on her paper.

The expected key misconception of failing to condition and answering 245/2500 was mentioned by a majority of the participants. Several people mentioned other incorrect fractions, with five talking about the case with a denominator of 687 Internet users, not 693 college graduates. I list all stated misconceptions in Table 4.12.

Table 4.12

Misconceptions Mentioned in Survey Part B Question

Misconception	Count
Failing to condition, yielding 245/2500 (key misconception)	18
Wrong marginal total for denominator, yielding 245/687	5
Reading comprehension	5
Incorrect fraction numerator	2
Taking union instead of intersection	1
Wrong buttons on calculator	1

To respond to the key misconception, failing to condition and selecting 2500 as the denominator, two suggestions were made by at least one-third of the teachers. The first was indirect, suggesting that the student read the question again. Question rereading was also suggested when the student used the Internet count of 687 instead of the college graduate count of 693. Two teachers said that the first word of the question, *if*, might pose a problem, because some students do not interpret *if* as introducing a conditional statement. These teachers wanted to change the first word to *given that* to clarify the conditioning.

The second common suggestion for dealing with the key misconception was more direct, asking the student what the denominator should be. For example, Teacher 42 offered a more direct suggestion which incorporated Part A:

I would preface it by looking at this lead-in to the question and making sure that they can only grab a college graduate, but you also have to yell at them that the denominator's not the same every time. Just because that's [2500] your denominator in the first part doesn't mean that it works for all these.

Survey Part C Results

This question tested the participants' ability to determine whether two events were independent. Of the 25 participants, only 3 answered this Computation question correctly; all 3 used a formula to determine that the college graduate and news from Internet events were not independent. Either the multiplication formula from Common Core Standard 2 or the conditional formula from Common Core Standard 3 could have been used, but the majority of the teachers did not use a formula. Instead of a formula, they offered a verbal explanation. I describe solution paths using their explanation in Table 4.13 of primary solution paths and Table 4.14 of all suggested solution paths. The most frequent wrong answer was what I had defined as the key misconception, claiming the events were not independent because they were not mutually exclusive. There were two incorrect responses in which participants concluded that the events

were independent. In an answer I labeled *subset*, teachers claimed independence because neither group was a subset of the other group, such as Teacher 7, who said yes, “Because there are people who received news from the Internet who are not college graduates.” In the other common incorrect answer, the teachers talked about the lack of *effect* of one variable on the other, but did not refer to subsets of full groups.

Table 4.13

Primary Solution Paths in Survey Part C Question

Primary solution path	Correct	Incorrect
No, not mutually exclusive	0	10
Yes, there is no subset	0	6
Multiplication formula	2	2
Yes, lack of effect	0	3
Conditional formula	1	1

Table 4.14

All Solution Paths in Survey Part C Question

Solution path	Count
No, not mutually exclusive	8
Yes, there is no subset	6
Multiplication formula	5
Yes, lack of effect	3
Conditional formula	2
Hypothesis testing	1
Checking to see if more likely	1

Many of the interviewees seemed to have confidence in their verbal explanations; relatively few of them asked me about their answer. Many explanations were succinct, like that of Teacher 7 above. Teacher 66 said no, “My thoughts would be, are all their possibilities separate from each other?” Teacher 21 explained a no answer with just three words: “Because they cross.” When I asked for further detail, she mentioned the song “Miss Independent” by Kelly Clarkson. She said, “I would always sing that song, and I’d say what does she sing about, what does that mean, you know being on her own. And so if they cross, I would always tell them then they’re not independent.”

Some of the teachers mixed correct and incorrect language. For example, Teacher 77 had a correct definition for mutually exclusive events, but not independence. “Often times I’ll teach independence with mutually exclusive information, and I just kind of talk about that and kind of . . . tie them together,” he said. Teacher 13 used the word *effect*, a synonym for independence I often use, stating, “If they [the events] were truly independent, one would have no effect on the other.” She gave one correct example, rolling a die and picking a card. Nonetheless, she also incorrectly claimed that when drawing a card, the events “draw a king” and “draw a diamond” were not independent because with the king of diamonds both events occur together. Because graduating from college and obtaining news from the Internet were overlapping, like the king of diamonds, the two events were not independent. Teacher 44 also used the idea of effect, but she said the events were independent because “one does not affect the other. So if you just go in and, ‘Yeah, they don’t have any,’ that would be what I say. They don’t affect each other.” Both Teachers 13 and 44 used synonyms I would use, but both unfortunately drew an incorrect conclusion—opposite incorrect conclusions.

The high error rate and low level of doubt worried me. I had expected this to be a relatively easy question, because independence appears in standards and textbooks. In one interview, the teacher and I looked up the correct definition in her algebra book. My prior expectation was not correct; most of the participants had a misconception about the term's meaning. Further research will be needed to determine how to deal with the error.

Given the subject matter results, it is likely not a coincidence that the student misconception most frequently mentioned by the teachers was vocabulary, as shown in Table 4.15. The teachers who attempted a formula-based solution generally mentioned potential trouble with the formula, but as described earlier, only a minority of the participants attempted any computation in this problem. The number of misconceptions is lower than in prior problems, because about one-third of the teachers made no suggestions.

Table 4.15

Misconceptions Mentioned in Survey Part C Question

Misconception	Count
Vocabulary	8
Trouble with formula	3
Mutually exclusive	2
Algebraic definition of independence	1
Confusing independence and intersection	1
Finding a number, not yes/no answer	1
No, because of causation	1
Wrong conceptual meaning	1
Yes, because numbers are large	1
Yes, there is no subset	1

When responding to vocabulary problems, the teachers tended to offer students direct information, a suitable response when asked about basic terminology. Teacher 17 identified one issue described in the literature review: multiple meanings of mathematical terms. In the following excerpt, she talked about *independence* from probability and *range* from statistics.

Teacher 17: Just getting this stuff mixed up. That's not my input. What do you mean independent? I thought independent and dependent were input and output. Domain and range. And so that, that's also very problematic for my kids when—.

Adam: The wording with algebra is what you're talking about.

Teacher 17: Yeah, algebra wording, yeah. So the algebra definition of independent as opposed to the statistical definition of independent, and kids expect, I mean, I've experienced that with range this year. When we talk about range with stats and creating box-and-whisker plots, they're doing range. And so when I'm doing mixed review, and it could be anything, to go back to a range problem after just doing a stats problem [means] all right, What is range? And we have to go through the whole process again. So yeah, confusing vocabulary.

I had planned to ask about mutually exclusive events as the key misconception, but given the large number of participants who made the mutually exclusive error, I abandoned that plan. Therefore, this problem had no key misconception. The question brought trouble enough on its own.

Taxicab Results

This question tested the participants' ability to use three given statements about a courtroom scenario to eliminate impossible outcomes and then compute a conditional probability based on the possible outcomes. Of the 25 participants, only 1 answered this Multistage question correctly, finding that the probability the taxicab was actually blue equaled $12/29$, or about 41%. Four participants were stumped and unable to offer any answer. As shown in Table 4.17 of all suggested solution paths, at some point over half the participants found the probability the cab was blue and the witness identified cab color correctly, $15\% * 80\% = 12\%$. This joint probability

was also the most common primary solution, as listed in Table 4.16. The second most common answer was 80%, the proportion of the time the witness identified the cab color correctly. The one participant who was correct used a tree diagram.

Table 4.16

Primary Solution Paths in Taxicab Question

Primary solution path	Correct	Incorrect
Blue cab and identified correctly, $15\% * 80\% = 12\%$	0	8
Witness correct identification rate 80%	0	4
No answer	0	4
Base rate of blue taxicabs 15%	0	3
Decision tree	1	2
2 by 2 table	0	1
Union of witness rate and blue cab base rate, $80\%+15\% = 95\%$	0	1
Witness rate 80% and equiprobability, so 40%	0	1

Table 4.17

All Solution Paths in Taxicab Question

Solution path	Count
Blue cab and identified correctly, $15\% * 80\% = 12\%$	16
Witness correct identification rate 80%	8
Decision tree	4
Hypothetical population of cabs	4
Base rate of blue taxicabs 15%	3
2 by 2 table	2
Bayes' rule	1
Conditional probability formula	1
Outcome space of size 4	1
System of equations	1
Union of witness rate and blue cab base rate, $80\% + 15\% = 95\%$	1
Witness rate 80% and equiprobability, so 40%	1

Previous researchers found that less than 10% of untrained participants would answer this question correctly (Bar-Hillel, 1980; Gigerenzer, 2002). Therefore, I warned the interviewees that this question was more difficult and told them a majority of teachers did not get it right. This made some of the participants nervous; on the other hand, a few looked forward to the challenge. As the teachers worked, they agreed that this problem was much harder than earlier ones. Teacher 50 claimed, “I haven’t had a student in 25 years that’s going to get [the right answer].” Teacher 13 thought the problem required more than just mathematics, saying “in some sense it’s a math problem, and in some sense it’s a riddle.”

Four of the teachers gave no answer to this riddle. Many others had little confidence in their answers, unlike the independence problem, where the teachers felt more confident. For example, Teacher 42 began his explanation about the witness validity rate of 80% with a qualification:

If I'm thinking about this right, it wouldn't matter how many of the cabs are green or how many of the cabs are blue. My thought was if it was blue, then there's an 80% chance that she's right. If it was green, there's a 20% chance that she's wrong. So if she said it's blue, there's an 80% chance that it really was blue.

As another example, in the following excerpt, Teacher 89 described her solution path, the blue cab and identified correctly path leading to 12%, and then expressed doubt:

Teacher 89: Well, my initial reaction would just be to take 15% of them are all blue, and he correctly identified them 80% of the time, so 80% of the time he's correct that it was blue. I would just take 15% times 80% to get my answer. And come up with what is the probability the cab was actually blue. I'll say 12 percent. Okay, so is that right or wrong?

Adam: Okay, no that's not it.

Teacher 89: I figured it wasn't.

On the positive side, several of the teachers believed their 12% answer was too low, even though they could not think of a better solution path. Their hesitation showed some intuition about the problem. The four teachers who suggested hypothetical populations similar to those of Gigerenzer et al. (2008) also had good instincts, even though two of the four teachers wondered if hypothetical populations were allowed.

The only correct solution was provided by Teacher 1142, an experienced AP® instructor, via a decision tree. He also suggested placing a hypothetical set of counts in a 2-by-2 table, mentioning that in this type of problem, "counts tend to be easier." Although he considered Bayes' rule, he said he does not ordinarily cover the formula in his classes, because he believes trees and tables lead to fewer student errors.

Only nine teachers suggested potential student misconceptions in this problem. Those teachers with no answer had no basis to suggest misconceptions; many others had so much trouble that they offered no ideas. I list mentioned misconceptions in Table 4.18. The most frequently identified misconception was reading comprehension. The key misconception, using just the witness's 80% identification rate, was given by three teachers. Two other common mistakes made by interviewees, the blue cab correctly identified response of 12% and the base rate of 15%, were also mentioned more than once.

Table 4.18

Misconceptions Mentioned in Taxicab Question

Misconception	Count
Reading comprehension	4
Witness correct identification rate 80% (key misconception)	3
Blue cab and identified correctly, $15\% * 80\% = 12\%$	3
Base rate of blue taxicabs 15%	2
Blue correct + Blue incorrect, total 15%	2
Equiprobability	1
Finding a number, not fractional answer	1
Forming final conditional fraction incorrectly	1

The teachers thought that understanding the problem would be difficult for students, given the introductory text plus three factual statements. The composite picture was hard to create, as Teacher 50 noted, “The other [problems], even with the big numbers, I can break it

down into smaller numbers and kind of make up something. [It's] very difficult to get a clear picture of this is what's going on." To help students create the composite picture, several teachers suggested drawing a visual diagram such as a decision tree or hypothetical hundred taxicabs. Teacher 4 stressed the importance of labels; she would tell students to "label each branch as to what color the car was." Teacher 1142 has had better success helping students once they drew a tree diagram. To finish the solution, he would ask them to "find me the outcomes where the cab was identified as blue, find me those places. ... So that'll sometimes get past that [misconception]."

I introduced the key misconception of 80% witness correctness just over half the time, when a teacher gave an answer that was not 80%. Because an answer of 80% does not use all the information in the problem, most of the respondents would respond by telling students to make sure they used all the given information. Two teachers indicated they would specifically mention the base rate to students struggling to complete the problem. Three teachers would go further than a reminder and would suggest that students use a visual solution path.

ELISA Part A Results

This question tested participants' ability to form a fraction from appropriate numbers found in problem text. Because of time considerations, four interviews omitted this question. Of the 21 participants who attempted this problem, 16 answered this Direct question with the correct fraction of 37 false positives out of 500 samples known to not contain HIV. The majority of those teachers directly wrote down a fraction; a minority created a table or other chart before attempting to find the fraction. Table 4.19 lists primary solution paths in this problem. Only five teachers offered any alternate solution paths; those suggested alternatives are incorporated with primary paths in Table 4.20.

Table 4.19

Primary Solution Paths in ELISA Part A Question

Primary solution path	Correct	Incorrect
Directly writing fraction	12	1
Creating 2-by-2 table	2	2
Writing false positive and false negative fractions	2	1
Decision tree	0	1

Table 4.20

All Solution Paths in ELISA Part A Question

Solution path	Count
Directly writing fraction	14
Creating 2-by-2 table	5
Writing false negative and false positive fractions	3
Conditional probability formula	1
Decision tree	1
False positives out of all positives	1
Hypothetical population	1

Because I told the participants this problem came from an AP® Statistics exam, many were surprised when the answer did not require many steps. Teacher 5 gave a representative correct response: “The blood sample not containing HIV is 37 out of 500. Is that it? That’s all I’m to do on that one?” Teacher 37 did a little more work; she first constructed a 2-by-2 table containing positive and negative counts. Nevertheless, she could also describe her solution path

with one statement: “Using the data, conditional that we were not HIV positive, so I said positive intersecting with HIV, which is 37 over 1000, not HIV which is 500 out of 1000, which is 37 out of 500, which is .074.”

The five incorrect answers all had an incorrect denominator: 37/463 (twice), 37/526, 37/1000, and 48/1000. In the discussion of misconceptions, only one of these wrong answers was mentioned, 37/1000. The teachers suggested wrong numerators just as often as wrong denominators, as shown in Table 4.21. The table has fewer entries than most of the problems discussed previously because only 10 teachers suggested misconceptions. As noted above, four interviews omitted the ELISA task, and in five other interviews, lack of time prevented a full discussion of student issues. Other teachers without probability teaching experience made no suggestions. Because I had expected these constraints, I did not select a key misconception.

Table 4.21

Misconceptions Mentioned in ELISA Part A Question

Misconception	Count
Reading comprehension	7
All positive test results, 526/1000	2
Correct positives out of known HIV, 489/500	2
False positives out of full sample, 37/500	2
Taking count as percentage, 37%	1
Unable to create table	1

The teachers worried about the length of the text in the ELISA task. Teacher 33 made a comment: “My kids would have been like, ‘Ah, I’m not doing this, here you go.’” Teacher 50 spoke about the trouble he had, like many students, with long unfamiliar problems:

To me, that’s the difficulty I experience with statistics, because the words of the problem, the way the problem is written, can have a significant variant on the outcome that you get in the end. And maybe other people see that in calculus as well, but I see it in statistics because my brain, I don’t think in those terms often, I sometimes evaluate in a different way.

The responses to reading troubles were usually instructions to reread the question; the teachers did not have specific reading tips. For example, teacher 77 said, “Just tell them to reread it, and if I knew they were wrong, just tell them to make sure and use the correct sample.” The only teacher who suggested an alternative approach was Teacher 89, who wanted to show a table. With a table, she said, “I don’t think it would have been as confusing. ... That would have made it way easier to pull that apart and label it.”

ELISA Part B Results

This question tested the participants’ ability to form a fraction from appropriate numbers found in a problem text, similar to Part A. This question differed from Part A because of the process to find the denominator. In the previous question, the denominator was a number found in the problem text. This solution required adding 489 correct positive test results and 37 false positive test results to get a total of 526 positive tests. The correct fraction was $489/526$, the number of positive samples with HIV over the total number of positive tests.

Due to time considerations, four interviews omitted this question. Of the 21 participants who attempted this problem, 16 answered this Direct question correctly. The majority of those teachers directly wrote down a conditional fraction; a minority created a table or other chart before attempting to find the fraction. Table 4.22 lists primary solution paths in this problem. As

in Part A, alternate solution paths were rare. Only two teachers offered any alternative solution paths; those two are incorporated with the primary paths in Table 4.23.

Table 4.22

Primary Solution Paths in ELISA Part B Question

Primary solution path	Correct	Incorrect
Directly writing conditional fraction	14	0
Positive tests out of all 1000 samples	0	3
Creating 2-by-2 table	2	0
Multiply positive test proportions, $489/500 * 37/500$	0	1
Positive tests out of 500, $526/500$	0	1

Table 4.23

All Solution Paths in ELISA Part B Question

Solution path	Count
Directly writing conditional fraction	15
Positive tests out of all 1000 samples	4
Creating 2-by-2 table	2
Multiply positive test proportions, $489/500 * 37/500$	1
Positive tests out of 500, $526/500$	1

Those teachers who wrote out the conditional fraction tended to give a straightforward explanation. For example, Teacher 33 described her solution path as follows:

Among the blood samples examined, basically there was 1000 altogether. It says “that provided positive results,” so positive results was 37 and 489, then what proportion actually contained the HIV? Well, of the first 500, those are the ones that contained it, so 489 actually contained it. To where 37 did not, so I had 489 over the 526 for the total and got about 93%.

Teacher 6 revisited his hallway metaphor from Part B of the survey task. He said the key part was what was given, “You’re not providing that they have HIV, you’re providing that they were positive, which means that again that whole hallway thing comes in, and says now we’re going down this hallway. Down this hallway, how many actually have HIV?”

In contrast, Teacher 77 had trouble finding the right values; her solution path led to an incorrect answer of 526/1000. She wasn’t comfortable with her solution:

I took the 489 positive from the ones that were known to contain, and then I took the 37 that were positive from the ones that didn’t contain the HIV virus, and I just added them up, and I got 526 out of the 1000 total. Obviously that’s not correct.

After we went through the correct solution, she commented about the trouble she had, saying, “I guess just going back and forth between the different paragraphs, obviously this middle one is the most important so, just kind of lots of numbers being thrown [together] and overthinking it.” Because she had difficulty reading the problem and finding the correct information, she made a error in her solution.

Similar to Part A, only about half the teachers suggested any misconceptions, owing to time constraints or lack of experience. As in the previous part, I anticipated time and experience constraints and therefore did not prepare a key misconception. The complete list of suggested misconceptions appears as Table 4.24. Three teachers mentioned the error that Teacher 77 made, taking positive test results out of 1000. Reading comprehension and using only the known positive samples were most frequently mentioned. I broke out comments about one specific

reading issue, the word choice of *proportion*, because the comments surprised me. I describe comments about proportion as part of the reading comprehension discussion.

Table 4.24

Misconceptions Mentioned in ELISA Part B Question

Misconception	Count
Reading comprehension	4
Using only known positive samples, 489/500	4
Positive tests out of all 1000 samples	3
Proportion word choice	2
Multiply positive test proportions, $489/500 * 37/500$	1
Positive tests out of 500, 526/500	1

Responses to reading comprehension problems were similar to those for Part A; the teachers would encourage careful rereading and make a suggestion if needed. Some of the teachers believed that many wrong answers came from not understanding the question. For instance, Teacher 42 believed that using only the known positive samples was a comprehension issue. He would tell students, “You just have to look at what the wording is, and what you actually need to look for, positive results not positive sample.”

Four teachers brought up a wording issue, the word *proportion* in “what proportion actually contained HIV?” They did not consider a single fraction to be a proportion. For instance, Teacher 11 asked for clarification: “When they say proportion, I’m thinking of two fractions set

equal to one another. Is that what they mean, or are they talking about just a fraction here?”

Teacher 10 thought the question wording was flawed because “proportion suggests an equivalency between two ratios.” She believed the desired answer was actually a ratio between two values, not a proportion. Two of the four teachers thought that students would struggle because of the choice of words. I clarified that statisticians tend to intermix the words *ratio* and *proportion*; a single fraction was acceptable. My reply resolved the problem in the interviews. Nevertheless, I do not know the prevalence of this belief about proportions in high school mathematics classrooms; a larger sample would be necessary.

In this sample, a few teachers responded to conceptual misconceptions. Teacher 1142 described possible thoughts of a student who wrote down 489/1000, the correct numerator but an incorrect denominator. The logic might be as follows: “Okay, I’ve identified that these are the people who got positive results and really have HIV,” but that hypothetical student did not “really understand [what] I was conditioning on; I was only looking on a particular set of people and not all participants in the study.” Teacher 66 also emphasized the denominator in a possible response to a student who wrote down 526/1000:

I would just go through and say, You know the base, your denominator is always out of the possible candidates. And I would say that you have your 526 in the wrong place because the 526 is actually the total candidates. It didn’t ask what’s the proportion of tests that came positive with HIV, it was out of the HIV tests. And it doesn’t say [tests] but it says among the blood samples.

ELISA Part C Results

This question tested the participants’ ability to compute the probability of a complicated event that would send a blood sample without HIV for further testing. There were several possible ways to deconstruct the description into computable parts; all ways required independence and the probability of a false positive from Part A. Because of time considerations,

six interviews omitted this question. Of the 19 participants who attempted this Multistage question, only 1 had an answer considered essentially correct. Actually, that person made a minor calculator error, so no participants wrote down the correct probability of .0105.

As in the Taxicab problem, I classified some solution paths by the teacher's argument. For instance, three teachers found the probability of two consecutive false positive tests. The most successful approach, used by the teacher who gave the correct answer and two other people who made only small errors, was to find possible cases that involved three HIV tests. In Table 4.25 of primary solution paths and Table 4.26 of all solution paths, I call this the *cases with three HIV tests* approach. As with Parts A and B of this task, alternate solution paths were rare. Only two teachers offered any alternatives beyond their primary solution path.

Table 4.25

Primary Solution Paths in ELISA Part C Question

Primary solution path	Correct	Incorrect
Cases with three HIV tests	1	5
Decision tree	0	3
No answer	0	3
One false positive test, 37/500	0	3
Two consecutive false positive tests	0	3
Binominal distribution	0	1

Table 4.26

All Solution Paths in ELISA Part C Question

Solution path	Count
Cases with three HIV tests	7
Decision tree	3
No answer	3
One false positive test, 37/500	3
Two consecutive false positive tests	3
Binomial distribution	2

Four teachers had answers close to .0105, with only one change needed to arrive at the exact answer. As described earlier, one teacher made a minor computation error. Two teachers found the correct cases, but did calculations using the probability of a positive result in the entire sample (526/1000) instead of the probability of a false positive (37/500). The fourth teacher, Teacher 11, also found the two correct cases, but multiplied the two case probabilities instead of adding them. Nonetheless, his thought process is a good example of the logic involved in a solution path. He started by writing 37/500 for the first test, which had to be a false positive.

Then he continued:

They're going to do it again because of that [first false positive], so I'm going to keep the same [probability of 37/500], and then if at least one of the two, so let's assume that one was. So this one is positive, this one, the one out of the two, is positive, so they would get a percent. Let me see here [calculating]. About .5 percent on that, but that would be the case if the second one was. Then what if it came up, first one did, then your second one didn't, but then your third one did, so they'd have to do it again.

In his description, he described the two possible results that would lead to further testing on a sample without HIV: two consecutive false positives, or a false positive, a correct negative, and then a false positive.

Many of the teachers had trouble understanding all the possibilities in the scenario, so they developed incomplete logical arguments. For example, Teacher 64 stared at the problem for a minute or two, then ventured an answer: “So you have 37 that tested positive that were actually negative, so since it was independent, I would say 37 out of 500. Is that right?” As she thought, her argument was not completely correct. In the process of developing logical arguments, three teachers helped themselves by constructing a decision tree. None of the three wound up with a completely correct answer, but they all showed relative promise—they needed only one or two more steps to complete the argument successfully.

I suspected that this problem would challenge the interviewees; I did not prepare a key misconception because I expected most teachers to not have strong ideas about solving the problem. I was not surprised when only six teachers offered a suggestion; these six suggestions are listed in Table 4.27. Reading and comprehending the scenario was the only repeated misconception.

Table 4.27

Misconceptions Mentioned in ELISA Part C Question

Misconception	Count
Reading comprehension	2
Calculating with .7 instead of .07	1
Confusing OR with AND	1
Forgetting one of the cases	1
Misapplying binomial formula	1

When the teachers spoke about responding to these misconceptions, they talked in terms of their own problems. They did not identify more general concerns. For instance, Teacher 66 talked about his solution: “I probably, I was going to miss one of these two [cases] just because it’s been a while, but once I saw one, the other kind of [was there].” Teacher 4 said that for her, the phrases “at least one and then independent” were critical to read and recognize. “It’s those key words that [students] aren’t comfortable with and recognizing, but at least one and independent, that was key for me to think binomial,” she said after she had applied the binomial distribution in her answer.

The results for this question were in line with my expectations. Because this part required multiple steps, I expected the teachers to struggle with the content and not have many ideas about misconceptions. I told the teachers that I believed less than a quarter of AP® test takers would get this part correct. Teacher 1142, the most experienced AP® instructor, also thought the number of correct answers would be “very, very small, even among AP students.” In addition to the problem being difficult, the teachers had been solving problems and discussing students for

40 to 70 minutes; fatigue was a factor. For many of the participants, moving to open-ended discussion was a relief.

CHAPTER 5

COMBINED AND OPEN-ENDED RESULTS

In this chapter, I continue presenting results from the interviews. Chapter 4 contained content knowledge and pedagogy results for each individual interview question. In this chapter, I present results from combining information across multiple questions. To begin, I examine misconceptions that the teachers made when solving problems. After the section on teacher misconceptions, there are two sections about pedagogy. In the first pedagogy section, I describe two misconceptions mentioned across multiple questions: arithmetic computation and reading comprehension. I then detail frequently repeated response strategies, including analogies and experimental representations.

After the sections on pedagogy, the next four sections contain results from open-ended questions. Each section summarizes responses on one of four major topics: curricular issues, teacher concerns, current sources of support, and requests for assistance.

Teacher Subject Matter Misconceptions

In this section, I examine how the participants fared with the seven misconceptions and biases related to conditional probability topics within state curricula that I described in the literature review on pages 23–29.

- **Equiprobability.** Two teachers incorrectly used equiprobability of outcomes in the Lucky Dip question, and one teacher made the error in the Taxicab question. On the positive side, many of the teachers offered helpful responses when asked about equiprobability of outcomes, the key misconception in the Lucky Dip question.

- **Randomness.** I read through every occurrence of the word *random* in the transcripts.

In four cases, a participant used *random* in a nonprobabilistic sense—three times as haphazard and one time as weird. On the positive side, six teachers explained the word in the probabilistic sense; interestingly, two correct explanations came from teachers who also used *random* in a nonprobabilistic sense.

- **Independence.** In the task analysis of Survey Part C, a majority of the participants did not give the probabilistic definition of *independence*. Incorrect definitions included mutual exclusivity, subset, and lack of effect. The high prevalence of this misconception is problematic because two Common Core standards rely on the definition.

- **Conjunction.** None of the teachers succumbed to this misconception. Additionally, several teachers brought up counts instead of probabilities as a potential solution in the Taxicab task. Counts are one way to reduce the prevalence of the conjunction fallacy, so the teachers' introduction of counts was a positive sign.

- **Time-axis causality.** This was a very minor concern. It was not a problem for any interviewee on the Lucky Dip task. One teacher briefly mentioned this misconception during his discussion of potential Lucky Dip pitfalls.

- **Confusion of the inverse.** The participants almost never reversed the direction of conditioning. One teacher did in Survey Part C; none did in Lucky Dip or Survey Part B. Additionally, five teachers recognized this potential student misconception in Survey Part B and were prepared to offer assistance to students making this mistake.

- **Missing base rate.** Four teachers used only the witness identification rate in the Taxicab task and neglected the base rate of blue cabs. Though not optimal, the 16% prevalence for this misconception was lower than the 40% to 50% reported by Bar-Hillel (1980).

Additionally, three teachers identified this key misconception on the Taxicab task. Most of the teachers I asked were able to offer useful responses when I introduced the misconception into the discussion.

In summary, the participants avoided most of the misconceptions in their solutions. They had almost no trouble with conjunction, time-axis causality, and confusion of the inverse. They had a little trouble with equiprobability, randomness, and the missing base rate. The one exception was independence, where three-quarters of the teachers offered an incorrect definition.

Repeated Student Misconceptions

When the teachers thought about how students might make errors, they repeatedly mentioned deficiencies in arithmetic computation and reading comprehension across multiple problems. The teachers explicitly identified computational issues on the Rash, Lucky Dip, Survey Part A, and Survey Part B problems. They brought up concerns about reading comprehension on all problems except Lucky Dip and Survey Part C. On the more difficult tasks—Taxicab and ELISA—reading comprehension was the most frequent response. In this section, I first detail what the teachers said about computation, then turn to reading comprehension.

Because probabilities range between zero and one, students who have trouble computing with fractions and decimals are likely to make errors on conditional probability problems. Many of the teachers spoke about errors with fractions. Teacher 13 expressed frustration: “One sixth times one sixth becomes two twelfths, and I want to scream.” Teacher 37 had watched her students enter numbers incorrectly into the calculator, flipping the numerator and denominator “so they end up with a probability that’s greater than one.” This error persisted: “It doesn’t

matter how many times I say ‘Probability can’t be bigger than 100%,’ they don’t see the correlation between one and 100%.”

Sometimes student misconceptions have been shaped by earlier instruction. Students do not always fully understand prerequisites, as Teacher 3 said: “Just because you got a 70 in a course doesn’t necessarily mean you’ve quote unquote mastered the standards.” Teacher 7 gave a specific example: “As a ninth grade teacher, we have a big issue with children having fear of an improper fraction.” Although probability does not use improper fractions, this stumbling block indicates that her students did not fully comprehend fractions and thus might have had more stumbling blocks. South Carolina Teacher 9201 gave an example: Students might talk about the correct column in a table, but use the wrong numbers. This error might occur less often if students had had more problem-solving practice in earlier grades. She wished she had more time to reinforce computation problem solving, but with the Common Core standards, “we go so far in depth with so many other topics to get ready for pre-calc and calculus that we don’t get to the problem-solving strategies that really would help our students of probability and stats.”

The teachers also wanted to help their students practice reading and make sense of written mathematical language. One language skill was vocabulary, since problem solvers must apply definitions in a probabilistic context. As described in the task results, the teachers raised this issue with the term *proportion* in ELISA Part B; the problem called for a proportion as a single fraction, but some of the teachers defined *proportion* as an algebraic equation between two fractions. On Part C of the Survey task, vocabulary was the most frequently mentioned concern. The teachers pointed out that the term *independent* has many meanings, both mathematical and nonmathematical. Teacher 17 noted that she had to reinforce vocabulary

definitions in context because her students would think of algebraic independence when doing probability problems, and vice versa.

At other times, the teachers worried about problem phrasing. Teacher 17 pointed out the wording of the Rash problem: “I think the whole ‘total number of patients in the two groups was not the same doesn’t prevent assessment of the results’ would confuse the heck out of them.”

When I asked why, she replied that her students usually believe problem statements are designed to give them new information, not clarify the ability to assess results. The students rarely saw clarifying statements and thus might get confused.

Teacher 44 commented on the long introduction to the ELISA task. “There’s a lot of, when you’re like positive and not positive, contains and doesn’t contain it,” she said. On her first pass, she didn’t understand the problem: “When you get to that point, you realize there’s no possible way to know, then you realize I must have read something incorrectly or missed something.” She had to reread the question to find the necessary information. Teacher 44 and several other teachers believed some students would not always take the time to read carefully. Sometimes “[they] read it wrong and [do] not interpret it. They just see numbers and start working,” Teacher 44 noted. Occasionally, the teachers made the same mistake. For example, on ELISA Part B Teacher 42 almost made a reading comprehension error: “[I] was ready to do the same thing that the kids would do and just go 489 out of 500. It took me a while to reread provided positive results for HIV, and not came from known positive samples.”

Several of the teachers gave response strategies to help students read more carefully.

Teacher 64 provided a general approach in her discussion of the Survey task:

Just reading the question carefully, which I think is a problem with these because sometimes, and even with me ... I have to see it, and take piece by piece the problem. And students just want to just rush through it, I think. And slow down. ... [I would be] taking them through each step of the sentences in here to help them.

A few teachers offered specific reading strategies. Teacher 37 encouraged her students to use highlighters. Two different highlighters could be used for clarification, such as marking the college graduates (the column) separately from the news from Internet people (the row) in Survey Part A. Teacher 16 stressed looking for the total number when she reviewed with students for the state-wide standardized test. “You know, find your total and most times that’s where most mistakes are, in the wrong total of people,” she said.

Repeated Teacher Response Strategies

Just as the teachers repeated student misconceptions, they returned to certain response strategies across multiple problems. In this section, I begin by categorizing interventions within a general framework. I then elaborate on two common specific responses—experimental representations and analogies—and one less-common yet promising technique: extreme cases.

Vermette and Gattuso (2014) identified three types of teaching intervention in their analysis of responses from Quebec high school mathematics teachers: *explanation* of some aspect of the problem, *confrontation* of part of the student’s response, and *experimentation* by changing the problem’s numbers or representation. In my analysis, I found many occurrences of all three types of intervention.

Explanation occurred most often with vocabulary. A student cannot determine whether two events are independent without knowing the definition of independence, for example. When their students confused union (OR) with intersection (AND), most of the teachers would reinforce the association with addition and multiplication, respectively. Teacher 21 had a saying, “I always told them if it’s AND you multiply; if it’s OR you add. And I said to remember it, these two [and, add] start with A but they don’t go together.” Another misconception to which the teachers tended to respond with an explanation was drawing marbles without replacement in

the Lucky Dip problem. For example, Teacher 16 said, “Remember that you’ve already taken out that first black marble; you’ve pulled it. ... You know how many are left in that bag.”

When a student had attempted a solution, the teachers preferred to confront the mistake instead of explaining the correct way. As Teacher 50 remarked, “I try to let them come to the conclusion instead of me giving them the answer when at all possible.” On the Rash task, when a student would compare counts and not fractions, Teacher 12 would “try and through questioning lead them to figuring out, ‘Oh okay, the [107] isn’t the whole that they’re trying to figure out.’” In the task results, the teachers often preferred confrontation to explanation. On the Survey A question, when students failed to subtract people in both groups, eight of the teachers would ask the students if anyone was counted twice; none of them said they would explain the formula. On the Taxicab question, when students would incorrectly use only the 80% witness identification rate, six of the teachers would remind the students about other information, whereas just two would offer an explanation about the base rate.

Despite the teachers’ preference for confrontation over explanation, they sometimes proposed an explanation when confrontation might have been possible. For example, in the ELISA Part B discussion, Teacher 66 would respond to an answer of 526/1000 instead of the correct 489/526 with an explanation. He would inform the student that “your denominator is always out of the possible candidates. And I would say that you have your 526 in the wrong place because the 526 is actually the total candidates.” I did not detect any general reason why teachers chose between explanation or confrontation, although the more difficult questions seemed to have more explanation.

Experimentation techniques, the third type of teaching intervention, include classroom demonstrations, computer simulations, and physical representations. During the open-ended

discussion, the teachers frequently mentioned conducting classroom experiments using physical objects such as cards, coins, and dice. On the tasks, about half the teachers suggested the use of an experiment, with at least one such response given on five questions: Rash, Lucky Dip, Survey Part A, Survey Part C, and Taxicab. In total, there were ten suggestions of physical representation, six of classroom demonstration, and two of computer simulation.

Some of the teachers outlined classroom demonstrations for the Rash and Survey Part A questions. On the Rash question, some would divide the class into two visible groups. Teacher 1142 created a hypothetical class with 16 girls and 12 boys. He would then assign a trait to some students in both groups, such as 4 girls and 4 boys receiving an A at the end of the semester. The groups would have the same number of As. But he explained, “Really girls have more chances to get As” because there are fewer boys. He continued, “Students seem to relate to that when you start to actually look at students in your room, divide them into two groups, and start looking at counts and frequencies and relative frequencies.” After the classroom demonstration, the students would become more likely to use relative frequencies, the correct solution path.

On Part A of the Survey task, some of the teachers would have their students stand and be counted in order to illustrate the mistake of double counting. Teacher 7 liked to use groups consisting of girls and people wearing sweatshirts for similar problems in her classroom, though she first ensured that the groups overlapped. By counting people, students “can see the combination, so it doesn’t have to be just the girls, and it doesn’t have to be just the boys. And it includes all the girls, and we don’t count the girls who are wearing sweatshirts twice.” When large numbers appear on paper, as in the Survey task, students sometimes forget about double counting. When that misconception occurs, “We always refer back to that scenario. They’re like, ‘Oh, we don’t count those people twice.’”

Some of the teachers also suggested physical manipulation for the Lucky Dip and Survey tasks. Two teachers thought that playing cards could help demonstrate independence in Part C of the Survey task. On the Lucky Dip task, in which marbles are taken from a bag, six teachers wanted their students to physically select marbles; two others suggested computer simulation. Some of the teachers listed advantages of working with physical objects. Teacher 33 thought drawing from a bag would help visual learners. Teacher 4 believed experimentation would help students struggling to understand marble selection order. Teacher 17 would place the six marbles on the table, not hidden inside a bag, because “that would help get to the two fifths and three fifths if you had actual things you were picking up and putting down.”

In all three types of response—explanation, confrontation, and experimentation—the teachers often tried to help students understand the question context by creating a link to a more familiar situation. I called these parallels *analogies*. Teacher 37 tried to give students analogies for every class topic. When she had a class with many athletes, she liked to talk about sports. She has also used analogies not related to sports, such as illustrating independence through shopping for headphones. When deciding what to buy, students “take into account certain conditions. And some of them are independent, and some of them are dependent. And we can wrap that around, so [independence is] not just a completely separated thing.”

Most classroom demonstrations, like Teacher 7’s activity with girls and sweatshirts, were analogies because they shifted a probabilistic concept into a more familiar context. There were also many verbal analogies. The task results included Teacher 6’s hallway analogy about conditional probability. Among other examples, Teacher 66 would convert rash skin cream counts to test questions, where “you got 223 questions right but you got 75 questions wrong.” Teacher 1142 had demonstrated the importance of defining which event is conditional to his

students with an example about women in America and United States Supreme Court justices. In one direction, given that a person is a Supreme Court justice, the probability of being a woman is 3 out of 9. In the other direction, given that the person is a woman, the probability of being a Supreme Court justice is 3 out of roughly 160,000,000 (U.S. Census Bureau, 2015).

In the female Supreme Court analogy, the fraction denominators are radically different, with a fraction having a single-digit denominator compared to one with a nine-digit denominator. This example demonstrates an *extreme case* wherein a teacher used vastly varied numbers to emphasize a difference. I have taken the term *extreme case* from Vermette and Gattuso (2014). In their study, three teachers presented an extreme case about sampling variability in spinners. I use extreme cases from time to time and expected to find that many of the teachers did as well, but the teachers offered only three extreme case examples, including the Supreme Court analogy. Teacher 64 proposed an extreme case analogy in the Rash task:

Let's say you had 500 kids try out for a particular team and so many made it. ... [Also] you had a smaller school, and you have smaller percentages and number of students you were working with. What's the probability that you would make the team at either of these schools?

Extreme cases do not have to be analogies. Teacher 7 would change the number of black and white balls to respond to the equiprobability misconception on the Lucky Dip task: "Is that still equal? There could be 2 black and 100 white, and would it still be fair?" This seemed like a promising confrontation. During my analysis, I wondered if extreme cases should be responses more often, which might be worth investigation in future research.

Curricular Issues

In the section on repeated misconceptions, Teacher 9201's comment about the lack of time for problem-solving practice and Teacher 16's comment about finding denominators on standardized tests referenced curricular demands. The two teachers tended to make decisions

based on a proscribed course of study rather than professional preference. Their choice was the standard choice. Although participants in all three states spoke positively about probability, probability was often relegated to minor status because other mathematical topics were considered more important in standards and on standardized tests.

Two of the open-ended scripted questions (p. 56) asked about probability in the curriculum. All of the teachers said that probability should be included in mathematics standards. When I asked them why, they commonly replied using terms like *relevant*, *practical*, and *real-world*. Teacher 44 thought probability should be a major topic because of its relevance: “I mean, just the tasks that we did right there show you how important it is.” Teacher 9201 brought up decision-making: “[Making] informed decisions from data is a great life skill and is increasingly in demand.” Teacher 17 contrasted algebra and probability in the minds of her ninth-grade students: “While I totally get that exponents have great practical applications for kids in interest and money, they don’t understand those things yet. If I can draw something out of a sack and hand it to somebody, that’s very different.” Teacher 12 wanted to integrate more probability into the curriculum because she thought “probability is a very useful type of math, and it’s also something that I feel like our students are more interested in than some of the other topics that we teach.” Most teachers agreed that students tended to like probability topics, with a few considering them fun. One exception was Teacher 77, who said that students dislike “straight up probability” topics when the relevance was unclear.

There was one state-specific response about probability in the standards. Those teachers from South Carolina expressed more confusion about standards than the teachers from Georgia or Pennsylvania. This response reflected the political situation; the Georgia and Pennsylvania standards were stable, but South Carolina politicians were publicly debating the Common Core

standards. I conducted most of the interviews in South Carolina in June 2014. At that time, a law mandating new standards had been passed, but new standards had not yet been drafted. Because it was unclear what changes would need to occur, confusion was a logical response.

In about two-thirds of the interviews, the teachers brought up the position of probability in the curriculum. For many of the teachers, the amount of probability content was more than they had experienced previously. Teacher 50, a department chair, commented that he saw many “teachers that are not comfortable with [probability and statistics], possibly because it wasn’t covered in their math book 5 and 10 and 15 and 20 years ago.”

Nevertheless, despite the increasing amount of content, probability did not have prominence in the curriculum. Across all three states, probability was a topic located at the end of the book, to be covered if time permitted. In South Carolina, Teacher 23 spoke about her algebra class: “[Probability] is in the last chapter, so usually we don’t get to it. There’s a lot of algebra skills that we don’t get to, so prob and stat is usually not a priority when we’re in the algebra class.” In Pennsylvania, Teacher 89 was a little sheepish: “This sounds horrible, but with the level of students that I had this year, with one group I didn’t even go there ... The other group we did touch on [probability] because they were a little bit higher level.” In Georgia, Teacher 10 told the story of her geometry class, where the final unit contains conditional probability:

We spent the first three months on the first unit, so we skimmed this unit. We didn’t even actually teach conditional ... We showed them two-way tables, we showed them independence and how to check with multiplication, that kind of stuff ... We had like two days to do it, and so we didn’t even really get into it.

I heard several more versions of this story and physically confirmed that probability was in the last chapter in Teacher 7’s Algebra 1 textbook. Chapter 10 had several nice sections on probability, but her class had only managed to complete Chapter 8.

Standardized tests had substantial influence in Georgia and Pennsylvania, the two states with mandatory state-wide mathematics exams. The teachers stressed topics that appeared frequently on the mandatory tests. As detailed in Chapter 2, the probability topics examined in this dissertation comprise about 10% of Georgia's Analytic Geometry end-of-course test. Probability makes up less than 10% of Pennsylvania's Keystone Exam in Algebra 1, which students must pass in order to graduate from high school. Probability is not a major topic in either state; therefore, the teachers chose to spend relatively little time on the topic.

The interviewees were candid about their decisions. Georgia Teacher 64 thought teachers might say, "Well, I wouldn't spend so much time on it" because "it would be more important to spend your time on the other part, which makes up a larger portion of the test." Georgia Teacher 44 had just received her school's standardized test results. Out of about 40 questions, "five or six problems in each course were statistics and probability. So as an administrator, where are you going to invest your money?" Pennsylvania Teacher 5 offered a farm analogy: "When we're teaching to the Common Core standards for the state test, there might be one question on the whole bloomin' test. This isn't where you put your eggs, because you've got more eggs in other baskets." Her choice of baskets made sense to me.

Expressed Concerns

When asked, 22 of the 25 participants expressed at least one concern about teaching probability or conditional probability. Six teachers gave two concerns, so the compilation in Table 5.1 summarizes 28 responses. Only two teachers specifically mentioned a concern about conditional probability; almost all teachers subsumed conditioning into the larger field of probability. In Table 5.1, I have classified responses by Shulman's (1986) categories of teacher

knowledge—subject matter, pedagogical, and curricular; lines separate categories in the table.

Approximately one-third of the responses fell into each category.

Table 5.1

Concerns Expressed by Teachers

Concern	Count
Subject matter knowledge of material (no specific subtopic)	4
Forgetting common sense about probability	1
Visualizing problems	1
Differences between conditional, theoretical, and experimental probability	1
Solving problems in conditional probability	1
Choosing appropriate relevant class activities	3
Knowing ways students understand probability	2
Finding flaws in student reasoning	1
How to train students in logic about finding answers	1
How to help students visualize problems with large numbers	1
How to help students with tricky problem wording	1
Generating on-the-spot extension questions	1
Keeping credibility after teacher mistakes	1
Knowing what should be covered in curriculum	3
Determining what prerequisites students know	2
Students lack prerequisite knowledge	2
Not enough probability in current curriculum	1
How to keep topic difficulty level appropriate for students	1

Most of the concerns about probability subject matter knowledge were general, but Teacher 42 specifically referred to conditional probability. Earlier, he had said conditional probability was a complex topic. When I asked why, he talked about relationships between given and conditional information: “It’s hard for me to see the link between [parts].” In contrast, “when it’s independent probability, and you can just do some straight multiplication, or even if there’s some addition that’s in there, it just seems to make more sense.” Teacher 104 also worried about complexity. Earlier, she had said probability was her least favorite topic in mathematics. When learning probability, she got frustrated like a lot of her students. She said, “I thought too much about it; I thought too much about replacement and combinations and permutations” and thus got “overwhelmed with all the different ideas of probability, and sometimes I would forget common sense.” She wanted to develop her common sense about probability before teaching the topic.

The teachers’ pedagogical concerns included activities, student thinking, and teacher-student interactions. Three of the teachers mentioned the need to have enough relevant activities (several more teachers requested activities when I asked for requests, which I describe later). Teacher 4 explained that there was “so much you can do with [probability],” but she was “kind of limited by what’s available.” She wished she could develop activities and Excel computer simulations, but said she never had the time.

Two of the teachers had general anxiety about not understanding student thought, without mentioning any specific area. Teacher 44 described a more specific concern: how she might find the flaw in a student’s argument. She might be confident in her ability to solve a problem, but said that she “can so easily listen to what a student is saying and it sounds rational to me. So to be able to go back and find the flaw in their thinking. ... I’m not that good.”

Five of the teachers expressed concern about some part of their interaction with students. Three wanted to know better ways to help students accomplish tasks; two wanted to improve their own actions. Teacher 66 pointed out a difference between probability and other mathematical topics when instructor errors occur:

In geometry, if I make a mistake the students usually catch me on it, ... and it's easily fixed, and it's clear that I just made a mistake. Whereas in probability and stats, you make a mistake, it looks like you don't know what you're doing. ... If I wasn't sure how to explain it, that would be setting myself up for failure.

Shifting to curricular concerns, four teachers worried about their students being set up to fail because the students did not have the necessary prerequisite knowledge. This concern was more common than I expected, though the reasons for it were solid. Teacher 3 reminded me that a previous passing grade of 70 does not necessarily indicate mastery of the standards; he would need to know what students actually could do. Teacher 37 did not believe her students would have prior exposure to Venn diagrams. Teacher 77 had students who had not experimented with the basic tools of cards and dice. Teacher 64 from Georgia explained that students who had the wrong mix of earlier standards and the current Common Core Georgia Performance Standards would have no prior experience with probability and statistics. She had “to start from the bare very beginning.”

The other subcategory under curriculum knowledge was topic selection. Several teachers were unsure of standards related to probability. They wanted to scrutinize the documents before teaching students. One teacher wanted to add more probability, but most just wanted to understand topics and topic sequencing. Teacher 13, who told the story about the complicated conditional probability formula in Survey Part B, wanted to prevent students from becoming frustrated. She said teachers had to “keep it simple to the extent that, yes, the mathematics is there, but you have to bridge a gap for those folks who are not going to look at mathematical

equations and get it.” Because the goal of probability literacy is to help all students, “if you want people to be won over, you have to keep it simple with the explanations.”

Sources of Support

The interviewees had mixed opinions about their level of preparation to teach probability and conditional probability. Unsurprisingly, those teachers with probability teaching experience generally felt more prepared. During the discussion, a few teachers mentioned sources that they used in preparation. Because I wanted to hear more about sources, if the teacher had not mentioned any materials I would often propose a hypothetical situation like the following: “If the principal came in and said that in the fall we’re going to offer a probability and stats course for one semester, and you’re assigned to it, what would you do?” The teachers gave ideas based on sources they had used to prepare for other mathematics classes. In this section, I report on those ideas. Because only three teachers mentioned sources without prompting, I present a single combined list.

Three types of reference were mentioned by at least five participants: teachers, textbooks, and technology. The teachers generally turned first to their colleagues for reference. Relatively new Teacher 12 said, “Pretty much any time that I have any kind of struggles with any of my classes, I usually go down the hall to one of the mentor teachers and talk to them.” More experienced teachers concurred, such as Teacher 104: “A good resource is another teacher who’s taught it before.” To plan a new course, Teacher 16 would reach out to other schools in the area.

The teachers also wanted printed resources. Multiple teachers stated that they would not teach a course without first completing the problems in the course textbook themselves. They also would look for examples in other texts. At one school where I interviewed, the teachers kept

a closet of old books for reference. Teacher 13 went so far as to order a statistics textbook from Amazon because she felt she needed “to get a book that didn’t look like stereo instructions.”

A majority of the respondents mentioned the Internet; some teachers provided specific sites. YouTube was the most frequently mentioned website, with four references. For instance, when Teacher 16 could not figure something out on her own, she searched YouTube to “find some videos of professors explaining it to their [students], and a lot of people do post those videos.” If she found a worthwhile video, she would tell students who had been absent for an extended period of time to watch the explanation online. Two of the teachers relied on Google; the College Board, Khan Academy, and Georgia Standards websites were each mentioned by one teacher.

Requests for Assistance

When a teacher voiced a concern about teaching conditional probability, the effect was to identify a need. Concerns are not needs like food or sleep, of course. Research Question 3, about teachers’ perceived needs, uses the word *need* in the sense of want, not requirement. The teachers wanted ways to allay their concerns about teaching probability.

Another way to discover the teachers’ needs is to ask about assistance that they might desire. Teachers have sources of support, as detailed in the last section, but they might not be able to find everything they need. Therefore, the final scripted question (p. 56) asked each teacher: “What assistance would enable you to teach probability topics to your students?” Every teacher received a version of this question. In some of the early interviews, the teachers did not understand my intent; they wondered what I meant by *assistance*. Therefore, in some later interviews, I personalized the question and asked what sort of assistance the teacher would want

from a university. This change in the question better expressed my intent. The teachers responded with a multitude of requests. Only one person offered no needs.

I grouped requests into categories, including an Other category for requests made by one person. In Table 5.2, I list the categories and number of requests in that category. The two largest categories were pedagogical knowledge requests; the next two largest were about subject matter knowledge.

Table 5.2

Categories of Requests for Assistance

Request category	Count
Activities for classroom use	8
Professional development on pedagogical topics	8
Professional development on probability subject matter	6
College coursework on probability subject matter	5
Better training for preservice teachers	2
Guest appearances by college faculty	2
Other	3

In the requests for classroom activities, the teachers spoke about the challenges they had in finding quality tasks. For Teacher 6, “the hardest thing is getting tasks that are meaningful but actually doable.” Teacher 64 requested good datasets; she “can look online, but sometimes the datasets are so big” that she has trouble pulling things out of them. Classroom activities should get students involved. Teacher 77 wanted games and activities that students can do in “20

minutes, half an hour, that gets them excited and gets them interested and sees it hands on.” As an example, Teacher 33 brought up the Grand Prize Game from a children’s television program, *The Bozo Show* (Hall, 1980). In the game, the player tried to toss a ball into each of six buckets, starting with the closest Bucket Number 1 and progressing towards Bucket Number 6 for the grand prize. The game could be fun for students when played and modeled in a high school class.

Teacher 6 pointed out that “as a country or as a state we want to move towards task-based learning” with activities like the Grand Prize Game. Multiple teachers requested professional development on teaching with active learning. Teacher 12 provided a rationale: “A lot of us teach the way that we learned, which is not necessarily the best way to do things.” Teacher 104 wanted to “go through the material that you’re going to be presenting to the kids” with a structure like our interview. The training would “go over common mistakes, go over the best teaching practices, and focus on what information we’re going to be teaching versus much higher concepts where the kids aren’t going to be.” Some of the participants wanted instructors from colleges; Teacher 10 offered an alternative with veteran teachers, “who could mind and think about best ways to teach [probability] to students in a high school classroom.”

Other participants wanted to develop subject matter knowledge, not pedagogical knowledge. Teacher 17 preferred more general subject matter instruction, not discussion of problem pedagogy, because “it’s impossible to work all the types of problems you could work.” Teacher 16 felt she was relatively weak with probability, since “it’s not one of those things where you throw an equation at me and I can solve it. I’ve got to go through every question.” She wanted to improve her own problem-solving skills. Some of the teachers wanted an inservice format for developing subject matter knowledge, but others would rather take a college class. Teacher 5, a department chair, said that she and the other mathematics teachers at her

school knew they were weak in probability. She said that “we would love to have a local professor ... come here and actually give us a refresher course ... a whole day in-service.” On the other hand, Teacher 9201 requested a college course. She explained why: “I like being a student. I like being in the classroom. I like the connection with peers and instructor. So for me, I would prefer a classroom setting.”

A few of the participants talked about college courses. I heard three stories about terrible college statistics classes, although only two of the three explicitly appealed for better college instruction. One of the two requests for guest appearances was for college faculty to lead a few class activities. The other, by Teacher 3, was more comprehensive. He mentioned the community of policy makers, textbook writers, mathematics coaches, professors, and others who help develop productive teachers. He would “like to see them more in the current classroom,” not the university or corporate office. The community should be “looking at how the students are responding in the courses ... because that would help explain the misconceptions that we may miss as a teacher.”

During the discussions, I was impressed by the teachers’ candor. They acknowledged their weaknesses. They made requests to try to improve the learning environment for their students, including the three requests I grouped as others: textbook recommendations, examples of probability’s usefulness outside the classroom, and trips to visit college classes.

I compared the requests in Table 5.2 to the concerns in Table 5.1. The teachers made statements about teacher subject matter, pedagogical, and curricular knowledge in both situations. Roughly one-third of concerns and one-third of requests were about teacher subject matter knowledge; most expressed needs were general, with a few specific concerns about conditional probability. Pedagogical requests were more prevalent than pedagogical concerns.

The teachers expressed desires for pedagogical training in both their requests and their concerns. The extra requests were for classroom activities, something college staff could provide. Additionally, six teachers asked me for the interview tasks, although I did not count those requests in Table 5.2.

There were proportionally fewer requests than concerns on curricular subjects. Perhaps the teachers felt comfortable using their own sources on student prerequisites and standards. Alternatively, the teachers might have been responding to my position as a relative expert on probability tasks and pedagogical techniques. Regardless, the teachers provided extensive detail on their needs to teach conditional probability, more than enough for me to draw conclusions and frame questions for further research. Those conclusions and questions appear in the next chapter.

CHAPTER 6

DISCUSSION

In the last two chapters, I presented results from the interviews. In this chapter, I summarize findings about the three research questions:

1. How do high school mathematics teachers solve conditional probability tasks?
2. How do teachers respond to student misconceptions in conditional probability tasks?
3. What do teachers perceive as their needs to be prepared to teach conditional probability?

I discuss each research question in a separate section. After those three sections, I briefly describe constraints that limit the validity of this study. In the final section, I introduce future directions for research that might be explored based on the frame established by these results.

Discussion About Teacher Solutions

The participants solved the conditional probability tasks pretty well. They did not prepare for the interviews, and most did not have experience teaching probability. Nonetheless, their results compared favorably to those of AP® Statistics test takers. It would be unwarranted to use the argument that about half of their solutions were incorrect to claim that the teachers lack problem-solving ability. When I discussed the initial results with Dan Kahan (personal communication, July 23, 2014), lead author of the paper containing the Rash task (Kahan et al., 2013), he concluded that “people with a critical reasoning disposition are selecting in to be high school teachers. It’s heartening!” I agree. The included Multistage questions were formidable. Furthermore, the 8% gap in correct answers between teachers with and without probability

teaching experience was smaller than I anticipated; I had thought experience would lead to greater improvement in results.

The participants often elected to take additional steps to confirm their solutions. Over half the teachers wrote down marginal totals in the Rash task, computed both better and worse percentages in the Rash task, and clarified the lack of ball replacement in the Lucky Dip task. Most of the time, the teachers used solution paths similar to those found in prior research. Occasionally they touched upon fields outside probability, such as the statistical chi-square test in the Rash task and arithmetic fraction reduction in Part A of the Survey task.

Many of the teachers preferred to solve questions with formulas and avoid nonformulaic approaches such as enumeration and decision trees. A few teachers explicitly stated this preference. I found further evidence in the teachers' primary solution paths. On the Lucky Dip task, a mere six teachers used nonformulaic approaches, although the suggested solutions provided by the task authors were nonformulaic. On Part A of the Survey task, only four teachers solved the problem by counting cells instead of a formula, but half the responses to student misconceptions suggested nonformulaic solution paths.

Turning to misconceptions described in the literature review, the participants had a great deal of trouble with the concept of independence; 19 of the 25 participants gave an incorrect definition. They had a little trouble with equiprobability, randomness, and the missing base rate. They had almost no trouble with conjunction, time-axis causality, and confusion of the inverse. Overall, the participants avoided misconceptions pretty well. Given preparation time before leading a class in probability, they would likely improve.

Discussion About Responses to Student Misconceptions

Classroom teachers must do more than solve conditional probability problems; they must also assist students who hold misconceptions and make mistakes. To offer assistance, teachers need “knowledge of the strategies most likely to be fruitful in reorganizing the understanding of learners” (Shulman, 1986, pp. 9–10). I asked the participants about student misconceptions and potential responses to investigate their knowledge of reorganizational strategies.

In general, the participants recognized common student misconceptions on the easier Direct and Computation problems, but had fewer ideas about the more difficult Computation and Multistage problems. On the first four questions, all of which were Direct or Computation, the participants averaged more than one stated misconception per question. The teachers most frequently gave a problem-specific misconception cited in prior research. On the last five questions, which included two Multistage questions and the hardest Computation question, the stated misconception rate dropped below one per participant. The teachers tended to either suggest less focused misconceptions like vocabulary and reading comprehension, or talk about potential errors in their own work. This contrast between easier and harder questions makes sense because the teachers also demonstrated less subject matter knowledge on later questions. A teacher who has trouble solving a problem might be able to describe the area of confusion in his or her solution, but is less likely to be able to identify a new problem-specific misconception.

Responding to an incorrect solution is also more involved than finding a correct answer, because as Teacher 23 noted, a teacher has “to come up with a legitimate argument for why that may be incorrect, instead of just a process.” The participants generally demonstrated the ability to come up with legitimate arguments about misconceptions they recognized. In many instances, they were also able to offer potential responses on the spot when I introduced a misconception.

For example, on the Taxicab task about half the teachers gave logical responses to the missing base rate misconception, even though only one had computed the correct answer. Immediate responses were not always optimal, as on the Lucky Dip task, where half the teachers responded to the equiprobability of outcomes misconception with a clarification about enumeration. Nevertheless, the ability to respond on the spot impressed me. The teachers could reply to one part of a problem without understanding the complete situation.

The teachers demonstrated three different types of teaching intervention: explanation, confrontation, and experimentation. When a student had a vocabulary issue, such as replacement in the Lucky Dip task, the teachers tended to suggest an explanation of the term in question. When a student had attempted a solution, the teachers expressed a preference for confronting the mistake instead of explaining a correct solution path. Most, but not all, of their responses would attempt to lead the student to recognize the mistake without a direct decree. About half the teachers suggested experimentation techniques as a task response, such as physical manipulation and classroom demonstration.

About half the teachers tried to help students understand question context by proposing an analogy to a more familiar situation. A few participants suggested creative approaches, such as hypothetical populations, computer simulations, and extreme cases. Most of the teachers did not appear to know about some of the innovative interventions. Teachers might benefit from professional development opportunities to learn about the innovative approaches. Not coincidentally, professional development about pedagogy was one of the teachers' most frequent requests.

Discussion About Perceptions and Needs

High school mathematics courses are primarily organized around textbooks and published state standards. The teachers generally liked probability as a topic because they found it practical and relevant for students. Nevertheless, probability did not have a prominent location in their courses. Repeatedly, the teachers told me that probability was the last chapter in the book, covered in the rare event the class had time. In Georgia and Pennsylvania, the two states with mandatory standardized tests, probability comprised no more than about 10% of any mandatory exam. A few of the teachers wanted to increase the amount of probability content, but all proceeded with the standards as written. The teachers said that they wanted to be familiar with probability topics in current standards, even if their teaching assignments did not include probability. A few of the teachers were concerned about their lack of standards knowledge.

During the discussion, 22 of the 25 teachers expressed at least one concern about teaching probability or conditional probability. Every teacher except one had a request for assistance. About half the requests were on pedagogy, about one-third on subject matter, and the remainder on curriculum. In the next three paragraphs, I outline expressed needs by category.

Most of the teachers' subject matter concerns and requests did not specify a specific subtopic, although two teachers had specific concerns about conditional probability content. In their requests, the teachers were divided about the best method of subject matter instruction. Some teachers wanted inservice training; others favored a classroom environment.

Their pedagogical concerns were divided into three subcategories: understanding student thinking, improving teacher-student interactions, and finding high quality activities. Eight teachers requested professional development on teaching probability with task-based strategies. Several other teachers did not express a need for training, but requested meaningful, interesting

class activities designed for high school students. Three teachers were concerned with their ability to find activities without assistance.

The participants' curricular concerns were divided into two subcategories. The first was knowledge of standards, as described earlier. The second was information about what students understood from prerequisite classes. Without assurances that their students had experience with Venn diagrams, cards, dice, and other tools, the teachers felt obligated to start from the very beginning. Curricular requests were rare, perhaps because the teachers had sources for information on standards. Three teachers told me about their terrible college statistics course experiences; two of them considered improved college instruction a need. With better college courses, future mathematics teachers would be more comfortable teaching probability and statistics.

Delimitations and Limitations

As stated in the introduction, my long-term goal is to develop probability tasks, guides, and related materials that would make mathematics educators more comfortable when teaching probability. These results provide a considerable amount of information towards that goal. Nonetheless, every study, including this one, has constraints that inhibit its usefulness. Some constraints restrict external validity, the ability to generalize results. External validity constraints, called delimitations, are the result of my choices about study scope. Other constraints restrict internal validity, the ability to make causal conclusions. Internal validity constraints, called limitations, are weaknesses I am able to identify but not control (Ellis & Levy, 2005). In this section, I list delimitations and limitations of this study.

I chose practicing high school mathematics teachers as my population of interest. This population choice was a delimitation. I might have selected middle school mathematics teachers,

pre-service mathematics teachers, college instructors, or students. Each possibility would likely have had different research questions and led to different results. The results of this study will generalize poorly to any group except high school mathematics teachers.

Another delimitation was my choice of interview locations. I selected three locations within the United States where I had connections who would help me recruit participants. I do not believe this delimitation is as restrictive as my population choice, because the main interview areas were separated by at least 200 miles. My choice of location should not restrict the extension of results to United States teachers, but the findings might not apply in other countries.

Because prior research on teacher knowledge of probability has been limited, as described in Chapter 2, there were few established hypotheses to test. I therefore decided to conduct exploratory interviews. Not having hypotheses is a limitation because I could not use statistical inference to provide evidence—thus there are no hypothesis tests in this dissertation.

I recruited a convenience sample of 25 volunteers. Voluntary participation is a limitation restricting general conclusions. Although I offered \$50 gift cards to encourage participation from the broad population of teachers, and most of the participants did not have probability course teaching experience, teachers who felt unprepared to discuss conditional probability were unlikely to sign up to be interviewed. My voluntary sample was not representative of the entire population of U.S. high school mathematics teachers. Therefore, I could not write about confidence intervals for the population; I could only describe results for these participants.

I decided to interview the participants one time, for approximately 90 minutes including time to gain consent and fill out the background questionnaire. The time limit was useful because I interviewed five teachers within class periods and seven more during inservice days. By not requesting too much time, I increased my response rate. Despite the utility of this choice, it was a

delimitation. The time limit restricted the number of questions I could ask; the single interview prevented follow-up questions after I reviewed transcripts. Also, I had to rely on what the teachers said about how they would respond to students in their classes; I did not visit their classrooms to observe responses given during the practice of teaching.

I introduced another delimitation by selecting five tasks that examined topics in the Common Core standards on conditional probability and the rules of probability. There were other possible standards and task choices. I decided on the conditional probability standards because the Common Core standards included probability literacy, one of my areas of interest. I chose tasks that covered topics in those standards. My selections restricted the topics on which I gathered information. This dissertation still contains information on multiple tasks and misconceptions, but different tasks would have generated different data.

Qualitative analysis introduces a limitation: the lack of internal consistency. Almost all quantitative techniques will generate the same result if the same data are analyzed multiple times. Qualitative analysis, on the other hand, might produce different results. Sometimes this limitation is overcome by the use of two or more analysts who establish inter-rater reliability, but multiple analysts did not exist on this project. I transcribed all 220,000 words and completed all analyses, except for two consultations about potentially correct answers. I did use intra-rater reliability by comparing my judgments about correctness between initial and final analysis.

Future Directions

After 140 pages of exposition, the section on delimitations and limitations might cause disillusionment. The constraints on this study are not incidental. The limitations prevent me from making any generalized statistical conclusions, because the sample does not represent a known population. The delimitations restrict the scope of the project. From this modest sample of

teachers and tasks, I have evidence to make statements about selected standards and misconceptions related to conditional probability and the rules of probability. The statements might eventually generalize to high school mathematics teachers in the United States.

Despite the constraints, there are three reasons these results are of consequence. First, the population of interest is very large. In the United States, more than 250,000 people teach high school mathematics (Snyder & Dillow, 2013). Estimating from state populations, over 20,000 high school mathematics teachers work in Georgia, Pennsylvania, and South Carolina. Even a minor improvement will have a major impact when implemented by thousands of teachers.

Second, the teachers and I covered a substantial amount of material. Most of the interviewees attempted five tasks with a total of nine questions. I found a future research idea in every question. Based on the task results, I was able to supply evidence about the prevalence of seven misconceptions related to probability. I also compared teacher subject matter knowledge to that of AP® Statistics test takers and the general population. In the combined analysis, patterns emerged about teachers' knowledge of misconceptions and responses to student errors. Additionally, the lists of teacher concerns and requests provide direction to people involved in teacher training and professional development.

Third, I did not design this study to generate generalized conclusions. The ASA report on using statistics effectively in mathematics education research defined five components of a research program. From start to finish, ideas are (a) generated, (b) framed into feasible constructs, (c) examined systematically, (d) generalized, and (e) extended (Scheaffer & Smith, 2007). In the literature review, I established that prior researchers had generated ideas about teacher knowledge of probability, but had not framed or examined most of them. Therefore, I designed this study primarily to frame hypotheses about teacher knowledge of probability and

conditional probability. I also examined the prevalence of misconceptions about probability and surveyed teacher needs, but framing ideas remained the primary purpose.

I now propose research questions about teacher knowledge of probability deduced from the results of this study. Other researchers and I might examine these research questions systematically and then move to generalize some of the results. Most proposed research ideas relate to the tasks. Some task-based questions explore statements made by the teachers, such as comments about sample size on the Rash task. Other task-based questions consider potential interventions, such as how to help problem solvers who make the one-color option error on the Lucky Dip task. After the task-based questions, I offer questions that I deduced from the combined analysis.

- On the Rash problem, why do teachers attend to the different sample sizes and how does this affect their solution paths? Also, what are the effects of considering this task a hypothesis testing problem, not a conditional probability problem?
- On the Lucky Dip problem, how might a teacher effectively respond to a student who makes the one-color option error?
- On Part A of the Survey problem, does the choice of writing fractions as separate or together indicate a different perception of the problem? Is it easier to solve the problem with one representation or the other?
- On Part B of the Survey problem, how might a teacher effectively respond to a problem solver who confuses the inverse and writes down $245/687$?
- On Part C of the Survey problem, most participants did not give the correct meaning of independent events. What definitions, analogies, and other tools might help teachers and students better understand the concept of independence?

- On the Taxicab problem, researchers have shown that nonformulaic approaches improve success rates (Cosmides & Tooby, 1996; Gigerenzer et al., 2008). How might teachers acquire knowledge about these innovative approaches?
- On Part A of the ELISA problem and other complex problem statements, what techniques might help students parse through the text and comprehend the question?
- On Part B of the ELISA problem, some teachers had previously defined proportion as an equation, not a single fraction. How prevalent is this definition?
- On Part C of the ELISA problem and other Multistage questions, what techniques might help teachers better define the outcome space?
- When asked about student misconceptions, do teachers use computation and reading comprehension as default replies when they cannot think of a problem-specific issue?
- Why might teachers decide on an explanation when confrontation is possible? Is that decision affected by the difficulty of the problem?
- What factors make analogies about probability more or less effective with students?
- What are the effects of responding to student misconceptions with extreme cases?
- Other academics and I know how to develop activities for classroom use. How should these activities be distributed so teachers can easily access them?
- What kinds of training about task-based learning do teachers need?
- How might the community of people who help develop teachers implement the expressed teacher requests?

To simultaneously investigate all of these questions would require more resources than I—or any academic research group—have available. I must select priorities. The participants offered suggestions about their concerns and needs, but most of their requests were general. For

example, the participants did not request pedagogical training on analogies and extreme cases because they did not know about these innovative approaches. They did not ask about specific solution paths for the Taxicab and ELISA Part C problems because they were stumped and did not have ideas about the alternatives. My job as a researcher is to identify specific topic needs within the general requests. I can then develop solutions and present the new methods.

My primary research concern is the study of misconceptions about independence. Independence is a vital topic in probability and statistical inference, including two Common Core standards. Independence is also a complex topic because the term has multiple meanings across the mathematical sciences. The high error rate on Survey Part C—three-quarters of the participants gave an incorrect definition of independent events—and low level of participant doubt disturbed me. If misconceptions are this prevalent in the larger teacher population, accurate teaching is not occurring in many classrooms. I want to ask more teachers and extend the study to students. To reach large samples, I will need a written or online survey; the information on solution paths and misconceptions from this study should help me design such an instrument. I also want to explore definitions, analogies, and examples that might make the multiple meanings of independence more understandable. Because I know of no published study that compares approaches to teaching independence, the pedagogical investigation will need to begin with interviews or case studies.

I am also interested in the connection between conditional probability and hypothesis testing raised by several teachers in the Rash task. High school teachers, who usually cover topics from probability and statistics, would benefit from increased curricular knowledge of connections between the fields. In order to determine what connections teachers perceive, I would need to interview more instructors who teach conditional probability and hypothesis

testing. I could also write an explanatory article about the connections that would support classroom instructors.

My final priority is to increase the uptake of task-based activities. The identified misconceptions will help me write problems that challenge student errors; the responses provide advice I can incorporate into teacher guides. After tasks have been developed, suggestions from the teachers in the present study might help the teacher support community design professional development sessions. Ideally, thorough training on carefully designed activities leads to high-quality instruction and learning about conditional probability.

Although probability might not be the “very guide of life” that Butler claimed in 1736, the concepts of probability defined in the 18th century have been expanded over the generations. Probabilistic logic is now an important part of medical treatment choices and other decisions made by all citizens. Near the end of our interview, Teacher 50 expressed a sentiment about probability and statistics literacy similar to that of Wilks in 1951: “If we don’t understand statistics, we’re going to be led by those people that do.” All citizens should understand enough probability to not need to be led. Hopefully, information from this dissertation will lead to better classroom instruction and advance society towards that goal.

REFERENCES

- Academic Standards and Assessments. 22 Pa. Code § 4 (2014).
- ASCD. (2012). *Filling the promise of the Common Core State Standards: Moving from adoption to implementation to sustainability*. Alexandria, VA: Author.
- Ayres, L. (2008). Thematic coding and analysis. In L. M. Given (Ed.), *The SAGE Encyclopedia of Qualitative Research Methods*. Thousand Oaks, CA: Sage.
- <http://dx.doi.org/10.4135/9781412963909.n451>
- Ball, D. L., Thames, M. & Phelps. G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59, 389–407.
- Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, 44, 211–233.
- Bartholomew, D. J. (1988). Probability, statistics and theology. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 151, 137–178.
- Batanero, C., & Díaz, C. (2012). Training school teachers to teach probability: Reflections and challenges. *Chilean Journal of Statistics*, 3(1), 3–13.
- Batanero, C., Henry, M., & Parzysz, B. (2005). The nature of chance and probability. In G. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 15–37). New York, NY: Springer.
- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions*, 53, 370–418. doi:10.1098/rstl.1763.0052

- 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. Retrieved from <http://www.aare.edu.au/data/publications/1999/beg99082.pdf>
- Bernstein, P. L. (1996). *Against the gods: The remarkable story of risk*. New York, NY: Wiley.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77–101. doi:10.1191/1478088706qp063oa
- Butler, J. (1860). *Analogy of religion, natural and revealed, to the constitution and nature*. New York, NY: Harper & Brothers. (Original work published 1736)
- Carlson, L., & Doerr, H. (2002). Secondary teachers' attitudes toward probability and their teaching strategies. In B. Barton, K. C. Irwin, M. Pfannkuch, & M. O. J. Thomas (Eds.), *Mathematics education in the South Pacific* (Proceedings of the 25th annual conference of the Mathematics Education Research Group of Australasia [MERGA], Auckland, pp. 171–178). Sydney, Australia: MERGA.
- Carnell, L. J. (1997). *Characteristics of reasoning about conditional probability (preservice teachers)*. Unpublished doctoral dissertation, University of North Carolina-Greensboro.
- Carter, T. A., & Capraro, R. M. (2005). Stochastic misconceptions of pre-service teachers. *Academic Exchange Quarterly*, 9, 105–111.
- Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, 104, 801–823.
- Chernoff, E. J., & Russell, G. L. (2013). Comparing the relative likelihood of events: The fallacy of composition. In Martinez, M. & Castro Superfine, A (Eds.). (2013). *Proceedings of the 35th annual meeting of the North American Chapter of the International Group for the*

- Psychology of Mathematics Education* (pp. 333–340). Chicago, IL: University of Illinois at Chicago.
- The College Board. (2009). *2009 AP Statistics Free-Response questions (Form B)*. New York, NY: Author.
- The College Board. (2010). *2010 AP Statistics Free-Response questions (Form B)*. New York, NY: Author.
- Common Core State Standards Initiative. (2015). *Development process*. Retrieved from <http://www.corestandards.org/about-the-standards/development-process/>
- Conference Board of the Mathematical Sciences. (2012). *The mathematical education of teachers II*. Providence, RI: American Mathematical Society.
- Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all?: Rethinking some conclusions of the literature on judgment under uncertainty. *Cognition*, 58, 1–73.
- David, F. N. (1998). *Games, gods, & gambling: A history of probability and statistical ideas*. Mineola, NY: Dover.
- delMas, R. C., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, 6(2), 28–58.
- De Moivre, A. (1756). *The doctrine of chances* (3rd ed.). London, England: Millar.
- Defined Program, Grades 9-12 and Graduation Requirements. South Carolina Regulation R43-234 (2013).
- Dollard, C. (2011). Preservice elementary teachers and the fundamentals of probability. *Statistics Education Research Journal*, 10(2), 27–47. Retrieved from <http://iase-web.org/Publications.php?p=SERJ>

- Educational Services Incorporated. (1963). *Goals for school mathematics: The report of the Cambridge conference on school mathematics*. Boston, MA: Houghton Mifflin.
- Efron, B. (2005). Bayesians, frequentists, and scientists. *Journal of the American Statistical Association*, 100, 1–5.
- Ellis, T. J., & Levy, Y. (2009). Towards a guide for novice researchers on research methodology: Review and proposed methods. In E. B. Cohen (Ed.), *Growing information, Part 1: Issues in informing science and information technology* (pp. 323–337). Santa Rosa, CA: Informing Science Press.
- Falk, R. (1986). Conditional probabilities: insights and difficulties. In R. Davidson & J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics* (pp. 292–297). Victoria, Canada: International Statistical Institute.
- Fast, G. R. (1997). Using analogies to overcome student teachers' probability misconceptions. *Journal of Mathematical Behavior*, 16, 325–344.
- Federal News Service. (2002, February 12). *DoD News Briefing – Secretary Rumsfeld and Gen. Myers* [News transcript]. Retrieved from <http://www.defense.gov/transcripts/transcript.aspx?transcriptid=2636>
- Franklin, C., Kader, G., Mewborn, D. S., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A preK-12 curriculum framework*. Alexandria, VA: American Statistical Association.
- Gal, I. (2005). Towards “probability literacy” for all citizens: Building blocks and instructional dilemmas. In G. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 39–63). New York, NY: Springer.

- Garland, S. (2014, December 29). The man behind Common Core math [blog post]. Retrieved from <http://www.npr.org/blogs/ed/2014/12/29/371918272/the-man-behind-common-core-math>
- Georgia Department of Education. (2011). *Common Core Georgia Performance Standards CCGPS: Mathematics standards, high school*. Atlanta, GA: Author. Retrieved from <https://www.georgiastandards.org/Common-Core/Pages/Math-9-12.aspx>
- Georgia Department of Education. (2012a). *Common Core Georgia Performance Standards CCGPS: Mathematics standards, accelerated CCGPS Analytic Geometry B / Advanced Algebra*. Atlanta, GA: Author. Retrieved from <https://www.georgiastandards.org/Common-Core/Pages/Math-9-12.aspx>
- Georgia Department of Education. (2012b). *Common Core Georgia Performance Standards CCGPS: Mathematics standards, accelerated CCGPS Pre-Calculus*. Atlanta, GA: Author. Retrieved from <https://www.georgiastandards.org/Common-Core/Pages/Math-9-12.aspx>
- Georgia Department of Education. (2012c). *Common Core Georgia Performance Standards CCGPS: Mathematics standards, CCGPS Analytic Geometry*. Atlanta, GA: Author. Retrieved from <https://www.georgiastandards.org/Common-Core/Pages/Math-9-12.aspx>
- Georgia Department of Education. (2012d). *Common Core Georgia Performance Standards CCGPS: Mathematics standards, CCGPS Pre-Calculus*. Atlanta, GA: Author. Retrieved from <https://www.georgiastandards.org/Common-Core/Pages/Math-9-12.aspx>
- Georgia Department of Education. (2013). *Content descriptions, Analytic Geometry*. Atlanta, GA: Author. Retrieved from <http://www.gadoe.org/Curriculum-Instruction-and-Assessment/Assessment/Pages/EOCT-Content.aspx>

Georgia Department of Education. (2014). *Georgia Milestones content weights: 2014-15*.

Atlanta, GA: Author. Retrieved from <http://www.gadoe.org/Curriculum-Instruction-and-Assessment/Assessment/Pages/Georgia-Milestones-Assessment-System.aspx>

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond “heuristics and biases.” In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology* (Vol. 2, pp. 83–115). Chichester, UK: Wiley.

Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. New York, NY: Simon & Schuster.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2008). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, 8(2), 53–96.

Groth, R. E. (2010). Teachers’ construction of learning environments for conditional probability and independence. *International Electronic Journal of Mathematics Education*, 5(1).

Hall, A. (Producer). (1980). *The Bozo show* [Television series]. Chicago, IL: WGN.

Haller, S. K. (1997). *Adopting probability curricula: The content and pedagogical content knowledge of middle grades teachers*. (Unpublished doctoral dissertation). University of Minnesota, Minneapolis.

Harrell-Williams, L. M., Sorto, M. A., Pierce, R. L., Lesser, L. M., & Murphy, T. J. (2014). Validation of scores from a new measure of preservice teachers’ self-efficacy to teach statistics in the middle grades. *Journal of Psychoeducational Assessment*, 32, 40–50.

Independent. (2015). In *Oxford English dictionary* (2nd ed.). Retrieved from <http://www.oed.com/view/Entry/94325>

- 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.
- Kahan, D. M., Peters, E., Dawson, E., Slovic, P. (2013). *Motivated numeracy and enlightened self-government* (Cultural Cognition Project, Working Paper No. 116). New Haven, CT: Yale Law School. doi:10.2139/ssrn.2319992
- Kaplan, J. J., Fisher, D. & Rogness, N. (2010). Lexical ambiguity in statistics: How students use and define the words: association, average, confidence, random and spread. *Journal of Statistics Education*, 18(2). Retrieved from <http://www.amstat.org/publications/jse>
- Kaplan, J. J., Rogness, N., & Fisher, D. (2014). Exploiting lexical ambiguity to help students understand the meaning of *random*. *Statistics Education Research Journal*, 13(1), 9–24. Retrieved from <http://iase-web.org/Publications.php?p=SERJ>
- Komolgorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung* [Foundations of the theory of probability]. Berlin, Germany: Springer.
- Krynski, T. R., & Tenenbaum, J. B. (2007). The role of causality in judgment under uncertainty. *Journal of Experimental Psychology: General*, 136, 430–450.
- Kvatinsky, T., & Even, R. (2002). Framework for teacher knowledge and understanding about probability. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on the Teaching of Statistics* (pp. 1–6). Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://iase-web.org/Conference_Proceedings.php?p=ICOTS_6_2002
- Lecoutre, M-P. (1992). Cognitive models and problem spaces in “purely random” situations. *Educational Studies in Mathematics*, 23, 557–568.

- Lee, H. S., & Lee, J. T. (2009). Reasoning about probabilistic phenomena: Lessons learned and applied in software design. *Technology Innovations in Statistics Education*, 3(2).
Retrieved from http://escholarship.org/uc/uclastat_cts_tise
- Liu, Y., & Thompson, P. W. (2007). Teachers' understandings of probability. *Cognition and Instruction*, 25, 113–160.
- Malm, K., von Sydow, M., & Andersson, S. (2009). Performance of three automated fourth-generation combined HIV antigen/antibody assays in large-scale screening of blood donors and clinical samples. *Transfusion Medicine*, 19, 78–88.
- Mathematics Assessment Resource Service. (2012). *Modeling conditional probabilities 1: Lucky Dip* (beta version). Nottingham, UK: University of Nottingham.
- Mojica, G. F. (2006). *Middle school teachers' conceptions of empirical and theoretical probability: A study of pedagogy* (Master's thesis, North Carolina State University).
Retrieved from <http://www.lib.ncsu.edu/resolver/1840.16/1658>
- Molnar, A. (2014, September 5). Teaching how to teach Bayes's Theorem (& covariance recognition) – in less than 2 blog posts! [Web log post]. Retrieved from <http://www.culturalcognition.net/blog/2014/9/5/teaching-how-to-teach-bayess-theorem-covariance-recognition.html>
- Munroe, R. (2011, April 4). Probability [Cartoon]. Retrieved from <http://xkcd.com/881>
- National Council of Supervisors of Mathematics. (1977). *NCSM position paper on basic mathematical skills*. Denver, CO: Author.
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards: Mathematics*. Washington, DC: Authors.
- National Governors Association, Council of Chief State School Officers, and Achieve, Inc. (2008). *Benchmarking for success: Ensuring U.S. students receive a world-class education*. Washington, DC: National Governors Association.
- Papaieronymou, I. (2008). Essential knowledge of probability for prospective secondary mathematics teachers. In *Proceedings of the Annual Conference of the Special Interest Group of the Mathematical Association of America on Research in Undergraduate Mathematics Education* (SIGMAA-RUME). San Diego, CA: San Diego State University. Retrieved from <http://sigmaa.maa.org/rume/crume2008/Proceedings/Proceedings.html>
- Papaieronymou, I. (2009). Recommended knowledge of probability for secondary mathematics teachers. In V. Durand-Guerrier, S. Soury-Lavergne, & F. Arzarello (Eds.), *Proceedings of the Sixth Congress of the European Society for Research in Mathematics Education* (CERME, pp. 358–367). Lyon, France: Institut Francais de L'Éducation. Retrieved from <http://ife.ens-lyon.fr/editions/editions-electroniques/cerme6>
- Pennsylvania Department of Education. (2013). *Academic standards for mathematics: Grades PreK—High School, January 2013*. Harrisburg, PA: Author.
- Pennsylvania Department of Education. (2014a). *Academic standards for mathematics*. Harrisburg, PA: Author. Retrieved from <http://www.pdesas.org/Standard/StandardsDownloads>

Pennsylvania Department of Education. (2014b). *Keystone Exams: Algebra I*. Harrisburg, PA:

Author. Retrieved from

http://www.portal.state.pa.us/portal/server.pt/community/state_assessment_system/20965/keystone_exams/1190529

Probability. (2015). In *Oxford English Dictionary* (3rd ed.). Retrieved from

<http://www.oed.com/view/Entry/151690>

Rochester Institute of Technology. (1999). *Mathematics and science attitude inventory, Project*

EDGE. Retrieved from <http://oerl.sri.com/instruments/up/studsurv/instr129.html>.

Rossman, A., & Short, T. (1995). Conditional probability and education reform: Are they

compatible? *Journal of Statistics Education*, 3(2). Retrieved from

<http://www.amstat.org/publications/jse>

Rumsey, D. J. (2008). Random thoughts: Letting go to grow: Independent vs. mutually

exclusive. *Journal of Statistics Education*, 16(3). Retrieved from

<http://www.amstat.org/publications/jse>

Schau, C., Stevens, J., Dauphinee, T. L., & Del Vecchio, A. (1995). The development and

validation of the Survey of Attitudes Toward Statistics. *Educational and Psychological Measurement*, 55(5), 868–875.

Scheaffer, R., & Smith, W. B. (Eds.). (2007). *Using statistics effectively in mathematics education research*. Alexandria, VA: American Statistical Association.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.

Snyder, T. D., & Dillow, S. A. (2013). *Digest of Education Statistics 2012* (NCES 2014-015).

Washington, DC: National Center for Education Statistics.

- South Carolina Department of Education. (2013). *Compilation of course progressions*. Retrieved from <http://ed.sc.gov/agency/programs-services/190/ProgressionDocuments.cfm>
- South Carolina Department of Education. (2015). *South Carolina College- and Career-Ready Standards for Mathematics*. Retrieved from <http://ed.sc.gov/scde-grant-opportunities/NewSCStandards.cfm>
- Statistics. (2014). In *Oxford English Dictionary* (3rd ed.). Retrieved from <http://www.oed.com/view/Entry/189322>
- Stigler, S. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge, MA: Harvard University Press.
- 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, NY: Springer Science+Business Media.
- Tan, C.-K., Harji, M. B., & Lau, S.-H. (2011). Fostering positive attitude in probability learning using graphing calculator. *Computers & Education*, 57, 2011–2024.
- Truran, J. M. & Truran, K. M. (1997). Statistical independence: One concept or two? In B. Phillips (Ed.), *Papers from Statistical Education Presented at ICME 8* (pp. 87–100). Hawthorne, Australia: Swinburne Press. Retrieved from <http://iase-web.org/documents/papers/icme8/Truran.pdf>
- Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 153–160). Cambridge, England: Cambridge University Press.
- U.S. Census Bureau. (2015). State & County QuickFacts. Retrieved from <http://quickfacts.census.gov/qfd/states/00000.html>

- University of Georgia. (2015). *Spring 2015 UGA Bulletin*. Retrieved from <http://www.bulletin.uga.edu/index.aspx>
- Utts, J. (2003). What educated citizens should know about statistics and probability. *The American Statistician*, 57(2), 74–79.
- Venn, J. (1880). On the diagrammatic and mechanical representation of propositions and reasonings. *Philosophical Magazine and Journal of Science (Series 5)*, 10, 1–18.
- Veloo, A., & Chairhany, S. (2013). Fostering students' attitudes and achievement in probability using teams-games-tournaments. *Procedia – Social and Behavioral Sciences* 93, 59–64. doi:10.1016/j.sbspro.2013.09.152
- Vermette, S., & Gattuso, L. (2014). High school teachers' pedagogical content knowledge of variability. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA* (pp. 1–6). Voorburg, The Netherlands: International Statistical Institute.
- Wackerly, D., Mendenhall, W., & Scheaffer, R. L. (2008). *Mathematical statistics with applications* (7th ed.). Boston, MA: Cengage Learning.
- Watson, J. M. (2001). Profiling teachers' competence and confidence to teach particular mathematics topics: The case of chance and data. *Journal of Mathematics Teacher Education*, 4, 305–337.
- Watson, J. (2005). The probabilistic reasoning of middle school students. In G. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 145–169). New York, NY: Springer Science+Business Media.

Weiss, S. H., Goedert, J. J., Samgadharan, M. G., Bodner, A. J., Gallo, R. C., & Blattner, W. A.

(1985). Screening test for HTLV-III (AIDS agent) antibodies: Specificity, sensitivity, and applications. *JAMA: Journal of the American Medical Association*, 253, 221–225.

doi:10.1001/jama.1985.03350260073029

Wilks, S. S. (1951). Undergraduate statistical education. *Journal of the American Statistical Association*, 46, 1–18.

Zapata-Cardona, L. (2008). *Teachers' understanding of students' conceptions of chance* (Unpublished doctoral dissertation). University of Georgia, Athens.