UNDERSTANDING THE PROBABILITY LITERACY OF HIGH SCHOOL STUDENTS

by

Fred A. Coon, IV

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Curriculum and Instruction

Charlotte

2019

Approved by:

____
Dr. Anthony Fernandes

____
Dr. Vic Cifarelli

____
Dr. Allison McCulloch

____
Dr. David Pugalee

ABSTRACT

FRED A. COON, IV. Understanding the probability literacy of high school students. (Under the direction of Dr. ANTHONY FERNANDES)

This multi-case study examined high school students' probability literacy, with focus on randomness, independence, and sample space. Task-based interviews were used with ten high school students. A Levels of Understanding Matrix (LUM), that was created based on pilot data using known literature, was deployed to gauge the students' understanding of randomness, independence and sample space. All data was processed through the LUM to create a consistent analysis while also providing a means to group the students into groups: Beginner, Intermediate, and Advanced. Each group held common characteristics such as the Advanced group's ability to build and use sample space while the Beginner group did not do either. The Intermediate group could use the sample space with some prompting. More general results show that students do not access positive or negative recency when dealing with random events but are still developing their understanding of randomness as an unordered list. Students used representativeness when dealing with independence while mostly being able to identify separate events. Further, when students could create and use sample spaces, their apparent understanding of randomness, and to a lesser extent independence, was more developed. Lastly, it was found that students did not see outcomes as independent when there was the perception that skill was a factor in the activity, such as basketball. Students viewed the trials as being related since the basketball player was building skill as he or she played the game. The study indicates that developing and using the sample space plays a key role in students' understanding and should be a focus of teaching.

Keywords: probability literacy, recency, representativeness, high school students

DEDICATION

This dissertation is dedicated to my wife Kristie and my children Isabella and Zachariah. The time and sacrifices that the three of you made for me to complete this lifelong dream of mine is not unnoticed nor will it be forgotten. I thank you for your patience and understanding as I know I am not easy to live with under the best of circumstances. It is my hope that this achievement will enable me to take better care of my wife, daughter, and son as well as make the world a better place for my children.

I would also like to make special note of my grandmothers, who both acted with grace while being firm and direct in their actions that often involved ignoring “silly” rules. And, I would also like to make special note of my grandfathers for their examples of how true men believe in strong women.

ACKNOWLEDGEMENTS

I would like to thank Dr. Anthony Fernandes for guidance through this process. I appreciate the way you maintained constant pressure on me when I needed it and provided me space to figure things out when necessary. Your input on methods and topic development will serve me in my professional life for years to come. I would like to thank Dr. Allison McCulloch as her suggestions helped in the development of what I believe to be a robust Levels of Understanding Matrix (LUM). Thank you to Dr. Vic Cifarelli for your presences as a constant advocate on my behalf. Thank you, Dr. David Pugalee. Your calm and encouraging demeanor have been both inspiring and honest.

A special thanks goes to my friends and classmates of the Random Dragons – John, Mandy, and Prem. The words of encouragement and laughter helped the tunnel seem a little bit shorter, if not an illusion completely.

Lastly, my praise goes to Yahweh for His continued blessings in all aspects of my life as I continue to learn to walk by faith.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

Probability is essential to modern-day existence and has been ingrained in societies for ages. Probability is viewed as the study of chance or the likelihood of an event occurring. People have not always properly connected chance with probability, viewing them as separate processes. Additionally, chance and probability have been used to describe different levels of likelihood. The understanding of chance and probability has evolved over the ages with gradual progress to a deeper understanding (Hald, 1990).

Probability arises in many applications of modern life and is important for all citizens to understand. Examples include weather systems and insurance policies that are dynamic and chaotic. Probability models provide a reasonable level of predictability of such chaotic and dynamic phenomena (Horvath & Lehrer, 1998). The probabilistic thinking required to understand weather systems is true for both meteorologist and consumers alike (Gal, 2005; Horvath & Lehrer, 1998). Data applications are ingrained into the fabric of United States' (U.S.) culture. The democratic process is saturated with data and without proper training one cannot properly participate in the process or discussion around the process (English & Watson, 2016). The need for understanding data and its interpretations, such as exit polls and how they use probability models, points to need for more research on how students understand probability.

Statistics is another area where an understanding of probability is key. Probability is integral to inferential statistics using p-values, which assist in decision making about data sets (Lodge, Alhadad, Lewis, & Gašević, 2017). The p-value gives the probability that allows the research to reject the null hypothesis, or the condition that is associated with no change in the events (Ross, 2000). The p-value allows the research to state that the no change event is false to a reasonable level of accuracy. This empowers the use of big data sets to make predictions about groups, populations, or medications (Lodge et al., 2017). The ability to make accurate predictions based on inferences gives people the ability to make decisions about real-life scenarios. Scenarios may have different levels of uncertainty and students need the ability to correctly navigate this terrain (Batanero & Serrano, 1999).

Probability in the School Curriculum

Given the importance of probability in real life, it is not surprising that it is featured in school curriculum. Multiple research studies have stated that probability belongs in school curriculum as early as primary school (Jones, 2007; Watson, 1997). Gal (2005) and Watson (1997) suggest that students should learn about randomness in such a way as to allow for students to be able to transfer this understanding of randomness to ideas outside of the problems presented in class problem sets. Additionally, there should be a strong emphasis on chance and independence. Students need to be able to correctly "compute and interpret the expected values of random variables" (Batanero, Chernoff, Engel, Lee, & Sanchez, 2016, p. 14).

The National Council of Teachers of Mathematics (NCTM) stated in their 2000 standards that students should be able to: 1) collect, organize, and display data; 2) determine and analyze data using statistical methods; 3) use data to make inferences and prediction; and 4) use and understand basic probability (NCTM, 2000). The last two of these standards are directly related to probability education in the U.S. The standards require students to be able to use probability to solve problems. The idea of making predictions based on data may start with statistical analysis but making predictions ends with students using probability to expand the findings of those statistical findings. Thus, demonstrating the need for students to have a robust understanding in probability. The Standards encourage students to develop skills related to making predictions and drawing inferences based on data (Wasserman, 2015). High school students, in particular, are expected to be able to understand and apply sample space and probability distributions, build distributions based on empirically based simulations, use expected value of random variables, be able to distinguish between conditional probability and independent events, and be able to calculate compound events.

The National Governors Association Center for Best Practices, Council of Chief State School Officers, and Common Core Standards, built on the NCTM Standards. The Common Core Standards make direct references to students' abilities to use computational skills to solve problems (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010). The CCSSM focuses on conditional events compared to independent events and two-way frequency tables. There is also a focus on individual skills that students are expected to know such as the Additional Rule or the Multiplication Rule for probabilities (NGA, CCSSO, 2010).

Currently, students are being exposed to wide variety of topics of probability. Many of these topics are connected such as independence and compound events or distribution and sample spaces. Jones, Langrall, and Mooney (2007) stated that learning probability, namely sample space and compound events, is a growing aspect of high school curriculum. Independence would be considered an aspect of compound events thereby independence ia a topic of discussion by default. At the high school level, the focus is mostly on theoretical and experimental approaches of probability. Additionally, probability distributions are a focus of all high school documents involving probability (Jones, Langrall, & Mooney, 2007). The absence of randomness from the above list of topics being directly taught do not mean it was not being taught or even that it was not studied as the list was not exhaustive. Furthermore, it should be noted that randomness fits in nicely with more than one of the above as compliment topic. The focus at this point in the writing is to note that students are being exposed to probability and to point out that the discussion is not to determine if students have been exposed to enough probability or statistical training only that exposure is occurring.

A Brief Historical Look at Probability

Most of the early knowledge about probability arose from games of chance (David, 1962). Chance games, such as backgammon, were the earliest occurrence of probability being used as a form of entertainment (Batanero, & Serrano, 1999; David, 1962). The practice dates back to Egypt, although some think the games started further east and came west some 5,500 years ago where the games were played using objects that resembled a pair of dice (David, 1962). The first 'die' were symmetrical, four sided objects with that were hollowed out (see Figure 1). The Roman version did not contain the numbers two and five (David, 1962). The dice results were associated with favorable or unfavorable outcomes, thus leading to rules that dictated how to play, win, and lose.

(Figure 1: Massimo, P. (2013). Roman dice in bone and ivory. www.ancient.eu/image/1204/)

Chance games developed rules of play based on the players' intuitive beliefs about of how often different events occurred. Konold (1993, p. 393) noted that people tend to develop a "set of heuristics", or a set of rules, when faced with decisions involving probability, even if they are unaware of the probabilities at play. Konold (1993) found that adults and students alike make use of heuristics as they seek to discover regularities in patterns. There is evidence that heuristics have always been present and can be more powerful than accepted knowledge about a probabilistic situation. In other words, people create rules for observed patterns sometimes out of necessity and other times out of curiosity. Individuals playing dice games developed rules to help them win, which is an example of necessity. However, people develop incomplete knowledge about probability just based on their observations and through their mental filters (Nacarato & Grando, 2014). Thus, the dice players' rules may or may not have been based on established or legitimate mathematical rules but rather their own "feelings" about the game.

During Aristotle's time, probabilities were commonly labelled one of three categories: certain, probable, or unpredictable (Hald, 1990). People would engage in games and other activities that involved these outcomes of certain, probable, and unpredictable. The labels were assigned to outcomes based on computed probabilities as much as general observations; people were often making intuitive decisions about probabilities.

Later, during the Renaissance, people commonly held non-numerical ideas (Hald, 1990). This is much like the people of Aristotle’s time in that intuition was the primary source of decision making. During both time periods, people played dice games. However, the probability of the dice games would not be known. The people only knew that some outcomes were “certain” while others were “probable” (David, 1962).

Antoine Arnauld and Pierre Nicole are given credit for first determining numerical values for probabilities in the late 1600’s (Hald, 1990). Antoine Gombaud, Chevalier de Mere, Blaise Pascal, and Pierre de Fermat formalized the ideas that went on to become combinatorial chance (Hald, 1990). Combinatorial chance was accepted as the elementary aspects of probability (David & Barton, 1963). The premise of combinatorial chance is counting spaces and the primary method to dealing with or understanding probability of an event begins with building and understanding sample spaces (David & Barton, 1963).

The Probability Mindset

Probabilistic thinking differs from the deterministic thinking in most other areas of mathematics (Jones, Langrall, & Mooney, 2007). Solving a linear equation for a given variable in algebra is an example of deterministic thinking. After following a set of steps, there is a fixed answer. On the other hand, probability problems do not necessarily have a fixed answer and, in some cases, do not have fixed processes as algebra often does. While probabilistic thinking is based on a set of rules (Horvath & Lehrer, 1998), the order in which the procedures are applied can change mildly based on the situation, forcing the individual to think differently about how to solve the problem.

Probability of an event is a calculable value used to provide insight into predictions based on a known pattern or a set of characteristics. A symmetrical die, for instance, has six equally likely outcomes. This allows the throwers to know that there are only six possible outcomes. However, the randomness of the outcomes means that the person will not know which number will come up on a given trial. Probabilistic thinking can be centered around determining all possible outcomes rather than a single solution. The addition of multiple outcomes increases the complexity of the thinking and uncertainty of solution(s) (Horvath & Lehrer, 1998). Thus, probabilistic thinking is reliant on understanding of concepts rather than procedures. For instance, a student may decide that the probability of rolling a one on a six-sided dice is ⅙ because the student was asked for the probability of the one and the sides are labeled one through six. This method would provide the probability of the side labeled two as having a probability of two out of six and demonstrates no substantial probability understanding even though the student may have some understanding. The student may have determined the probability of the event and based the prediction for the next trial on what they have observed rather than understanding the next trial maintains probabilities for outcomes but is not predictable in a particular outcome. Compared to algebra, where a student finds the x-intercept to be four and thereby knows exactly where the curve crosses the x-axis, the die roll maintains a level of uncertainty that algebra problems do not. Therefore, one aspect of thinking a student would display, if he or she can think probabilistic, is the ability to communicate this uncertainty of the event.

The Gambler's Fallacy

Probabilistic thinking can be challenging to students since it seems to go against their intuitions. A common error associated with predicting random outcomes is known as the "gambler's fallacy" (Konold, Pollatsek, Well, Lohmeir, & Lipson, 1993). The idea is that if one has not won in a long time, then the losing streak will end and turn to "good luck" for the gambler. Gambler's fallacy is the result of an individual's lack of understanding of randomness and variation which is a complex thought process (Albert, 2003). The gambler's fallacy is well documented in sports. For instance, if a baseball player has been in a slump and not had a hit in a while, people might comment that he is due a hit. The further implication is that a "good luck" or a "bad luck" streak is strictly a chance event with no component of skill, injury, or illness. Furthermore, people might use a streak to make predictions about what will definitely happen next. Probabilities are governed by chance and chance; random events do not allow for predictions of the next outcome only a likelihood of the next possible outcomes. The probabilities in such situations are not a predictor of future events in that if someone's probability of hitting a baseball is 30% and the person has missed hitting the ball the last seven times, they should get a hit at the next three at bats. But, people ruled by the gambler's fallacy fall into the trap of believing that the next three at bats will be hits because the last seven were misses. However, each at bat, or attempt to hit the ball, is separate from the previous at bat which includes components of independence and randomness. Baseball is skill-based activity and the “probability” of hitting the ball is based on the players statistics of past events (Chernoff & Sriraman, 2014).

Modern professional baseball teams access statistics of players to make personnel decisions (Chernoff & Sriraman, 2014). Probability can help make long term predictions like the Oakland A’s did in creating a championship team. The A’s turned a losing team into a championship team in a single season strictly by using probability models (Chernoff & Sriraman, 2014). Oakland made decisions based on large datasets for prediction of what would happen in another large dataset; the next season. Players were chosen for their ability to score runs. The players were chosen based on their combined scores as predicted with statistical models. This approach differed from other teams where there was a focus on acquiring homerun hitters. Homeruns are exciting to watch, but do not necessarily translate into winning games. The problem fans fall into is not looking at each at bat as strictly independent. The fan’s intuition is the same as the gambler’s in that he or she was not viewing the next roll as independent of the last roll and making predictions by focusing on a small dataset, such as a homerun, rather than the full dataset, such as two earned runs per game but fouling out constantly over the season.

Components of Probabilistic Understanding

There are many components to a student’s understanding of probability. Gal (2005) noted a collection of abilities involving probability understanding as probability literacy. Probability literacy is discussed in greater detail in chapter two, but the premise that Gal (2005) used was to separate probability literacy into main topics of knowledge elements and disposition elements. Randomness, independence, and sample space belong to the knowledge elements and were major points of concerns for this study. The primary purpose of focusing on randomness, independence, and sample space, which will often be referred to as big ideas, was to ensure that each component of probability literacy was carefully investigated. Gal (2005) outlined eight components of probability literacy with each component being complex in of itself. Proper investigation required each component be research at an exhaustive level to provide a complete understanding of a high school student's probability literacy.

As stated before, the big ideas items of the knowledge elements are randomness, independence, and sample space. A deeper discussion about the big ideas is discussed in chapter 2, but an overview is presented here. Randomness is characterized by the next outcome of an event being unknown in spite of known probabilities (Jones, 2005). For instances, if referring back to the previously stated example of rolling a die, the probability for each roll is a known quantity in that it will be one out of six, but the actual outcome for the next roll is uncertain. The frequency of a large set of trials can be determined theoretically or through the repetition of many trials but does not necessary provide the insight for predicting outcomes of events as the idea of the random events implies that the outcome is not predictable. Furthermore, the development of a set of trials progresses to the creation of sample space which will aid in determining a student's understanding of probability due to samples space's ability to be used as a tool in probability. Additionally, the ability of a student to determine the independence of the trial has further implications of a student's understanding of probability. All or some of these ideas were pushed to the front of student thought, as based on the task used, regardless of whether it was directly mentioned (Jones, 2005). Students can hold an implicit understanding of randomness, independence, and sample space without realizing their use or the connection that exist between independence, randomness, and sample space (Jones, 2005).

Purpose of this study

The study of probability in its current form is more than 30 years old and not much has changed (Jones, Langrall, & Mooney, 2007). There is little research on probability literacy at the high school level.

The purpose of this study is to understand the probability literacy of high-school students. The term probability literacy will be defined and discussed in great detail in chapter two, but for now it is enough for it to be understood as a collection of abilities that are associated with a student's ability to properly complete tasks involving probability, in both the student's understanding and application (Jones, Langrall, Thornton, & Mogill, 1997).

Research Questions

What is the probability literacy of high-school students? In particular, what do high school students understand about the probability concepts of chance, randomness, independence and sample space?

Conclusion

The next chapter will discuss the literature related to probability literacy, and in particular the prior research related to the concepts of randomness, independence, and sample space. The chapter will also discuss the frameworks of probabilistic understanding as presented by Gal (2005) and Jones, Langrall, and Mooney (2007). The overlap between the two frameworks will be discussed and how they relate back to the big ideas of randomness, independence, and sample space.

CHAPTER TWO: LITERATURE REVIEW

The purpose of this study is to understand the probability literacy of high-school students. Chapter 2 will outline the literature in this area and will include a discussion on probability literacy and its primary components; knowledge elements and dispositional elements. The literature review will focus on knowledge elements including sample space, randomness, and independence. The language used in probability, a major aspect of the knowledge element, will be discussed. Next, I will discuss the two major frameworks of probabilistic understanding; Jones' and Kelly and Watson's frameworks for understanding. These frameworks are important for setting a baseline for comparing students' understanding within my study. The frameworks also provide a means to compare probability understanding of students between studies.

Purpose

The purpose of this study is to understand the probability literacy of high-school students. This ultimately was the reason why this topic was chosen; students need a complete and thorough understanding of probability to be able to function in the highly technological and data driven world.

Research Questions

What is the probability literacy of high-school students? In particular, what do high school students understand about the probability concepts of chance, randomness, independence and sample space?

Experimental, Theoretical, and Subjective Probability

Experimental, theoretical, and subjective probability are important aspects of the development of probability understanding. Ideally, children would have a "multifaceted conception” of probability (Jones, 2005, p. 3) with lessons and experiences dealing with probability that are experimental, theoretical, and subjective and lessons that explore the relationships between the three different concepts (Jones, 2005). Experimental probabilities are usually based on frequency of observed events (Jones, 2005). Once the sample space becomes large enough, the experimental probabilities start to resemble the theoretical probabilities. The theoretical probabilities are calculated using ideal scenarios with given numerical values that are not found via observed data (Jones, 2005). Subjective probability is more abstract than either experimental or theoretical probabilities (Jones, 2005). Subjective probability maintains a separation from experimental probability, which is grounded in observations, and theoretical probability, which is grounded in calculations. Subjective probability is an adjustment or rather an interpretation and thereby subject to judgement (Horvath & Lehrer, 1998). Neither experimental nor theoretical probabilities are open to such judgements as they are known quantities. Subjective probabilities might have a known quantity but that value was found through means that are questionable. For instance, the probability that a particular illegal behavior is committed but not reported leaves, or even creates, error in the numerical value associated with the event. The number of non-reported events is unknown and thus any probability models based on that data are estimates at best and are subjective probabilities (Jones, 2005).

Theoretical probabilities allow students to explore probabilities in a static situation away from uncontrollable variables (Jones, 2005). This allows students to focus on computations. However, if a student’s understanding of probability stops at figuring computations, the student will be missing the larger picture of probability. For instance, Aspinwall and Tarr (2001) demonstrated that experimenting with random events helped sixth graders develop deeper understandings of sample size. The premise was that small sample sizes result in "unusual" sets compared to large samples where the sample sets more closely resemble the "parent distribution" (Aspinwall & Tarr, 2001, p. 235). The findings from Aspinwall and Tarr's study found that most individuals or groups, historically, use frequencies with large sample sizes to determine the primary method of figuring probabilities of events (Jones, 2005. p. 6).

Theoretical probability and experimental probability are both important and useful. But, without a conversation about how theoretical probabilities mimic and describe experimental probabilities, students miss an opportunity for a deeper understanding of probability. Experimental probabilities are also important in developing students' understanding of randomness and sample space (Aspinwall & Tarr, 2001; Jones, 2005). Furthermore, it is important for students to understand the randomness of experiments (Horvath & Lehrer, 1998). The experiment itself introduces possible variations to probabilities (Horvath & Lehrer, 1998). For instance, how hard one was to spin a spinner or roll a dice might change the result. Therefore, if the dice were rolled differently for trials the resulting probabilities might be affected (Horvath & Lehrer, 1998). One study found some second graders believed that variations of throwing die had no effect on outcomes while other second graders believed that an experiment was invalid as a result of rolling the die differently than it was previously rolled (Horvath & Lehrer, 1998). Horvath & Lehrer (1998) compared to Aspinwall & Tarr (2001) point out how experimental probabilities play an important role in the development of students' understanding of sample space and randomness and how sample space can be a tool to develop a deeper understanding of randomness (Aspinwall & Tarr, 2001).

Probability Literacy

The idea of literacy is a well-documented aspect of U.S. schooling. The word literacy generally conjures up images of spelling test, reading books, essays on books, and class discussions about books. Literacy is not one skill, but several skills used in conjunction with each other that help a person read and explain a book, for example. Probability literacy is similar. This study builds on Gal's framework (2005).

Probability literacy, is a collection of abilities that involve a student's ability to successfully complete problems that involve probability combined with a broad understanding of probability and its components (Gal, 2005). It consists of two parts of knowledge and beliefs related to probability. According to Gal (2005), probability literacy involves "the knowledge and dispositions that students may need to develop to be considered literate regarding real-world probabilistic matters" (p. 40). The knowledge components include big ideas, figuring probabilities, language, context and critical questions. The dispositional elements include critical stance, beliefs and attitudes, and personal sentiments regarding uncertainty and risk. This includes the formal mathematical thinking that would be involved to solve problems mathematically such as formal mathematical processes and algorithms. Probability Literacy also includes informal thinking, like trial and error processes which can be asymptomatic and irregular. These processes can allow the student to find the correct answer to a particular problem, but the process may not be transferable to other problems.

Probability literacy involves understanding the underlying concepts that go beyond getting the correct solution to a problem. For example, Konold, Pollatsek, Well, Lohmeier, and Lipson (1993) found that students correctly answered probability task but the explanation provided by the students provided little useful insight into the theories that were used to arrive at the answer. Furthermore, their studies found that students confused equal probability and simple chance.

There are other components to consider when assessing a student's probability literacy such as beliefs, attitudes, and habits of mind (Gal, 2005). The beliefs and attitudes of a student could possibly have an impact on how long the student works on a problem before deciding they cannot solve the problem. The beliefs and attitudes of a student could affect his or her approach to the problem. For instance, students that believe they are good at rolling dice might believe that rolling a six was more likely because the student knows "how" to roll a particular desired outcome. This is an aspect of gambler's fallacy that refers to the belief that someone holds about their ability to alter chance events using acquired skill sets (Konold, 1993).

This study focuses on Gal's (2005) work on probability literacy where he split probability literacy into main categories: knowledge elements and dispositional elements. Knowledge elements are subdivided into 5 groups: big ideas, figuring probabilities, language, context, and critical questions. Dispositional elements are separated into 3 groups: critical stance, beliefs and attitudes, and personal sentiments regarding uncertainty and risk. However, for the purpose of this study the focus was centered around knowledge elements, especially the big ideas. In addition, I made observation about the language component of probability, context, and beliefs of students although all of this was done passively through the lens of randomness, independence and sample space.

Knowledge Elements

According to Gal (2005), the big ideas were a major component of the knowledge element of probability literacy and vital for all students. Randomness, independence, and variation were identified as three big ideas of probability (Gal, 2005). I added two more 'ideas' to this definition such as conditional probability and sample space as these are identified as big ideas in both Common Core standards (NGACBP, CCSSO, 2010) and in NCTM Standards (NCTM, 2000). The addition of conditional probability was based on a great deal of readings I have done on conditional probability including the before mentioned Common Core standards and NCTM standards. The readings collectively imply the need for students to identify conditional events and independent events. Thus, conditional probability seemed like a natural fit to include here as I was already looking at the opposite side, that being independence. But, to maintain focus of the study, independence was investigated apart from conditional probability.

In addition to Gal's definition of probability literacy, there were frameworks that outline levels of understanding and how to gauge these levels of understanding such as Jones, Langrall, Thornton, and Morgill's (1997). According to these researchers, students who demonstrate probability literacy should be able to use major concepts in probability to: (a) solve problems, both theoretical and experimental, (b) use simulations to approximate probabilities, (c) use random variables and interpreting probability distributions, including the normal distribution, and (d) apply random variables to generate and interpret distributions such as the binomial, uniform, normal, and chi-square. The four topics listed here were not identical to Gal's definition of probability literacy, but there was enough overlap to verify Gal's point that probability literacy was a collection of skills and attitudes rather than one piece of collectible knowledge. For instance, Jones, et al.'s (1997) first part of probability framework stated that students should use probability to solve problems, both theoretical and experimental. Gal's (2005) knowledge elements and dispositional elements can be correlated to this single aspect of Jones, et al.'s (1997), see Table 1 in Appendix A. As can be seen in the table, Gal divided the traits of probability into not only different categories but also more categories than Jones, et al.'s (1997). The effect is that Gal's framework allows for a deeper investigation into probability whereas Jones framework provides a structure for gauging understanding. Solving problems within Jones, et al.'s (1997) framework fits with Gal's as they both reference the basic skills of students and the students' ability to use the skills to solve problems. Gal's items that fit into Jones, et al.'s (1997) problem solving are in the blue column. The same format is continued across the table with each column being a category of Jones, et al.'s (1997) framework and the rows show which of Gal's aspects are overlapping. The primary concerns for my study are the Gal's "Big Ideas" that are aligned with Jones, et al.'s (1997) problem solving category. In the sections below, I outline the big ideas of the knowledge elements of randomness, independence and sample space.

Randomness. While randomness is not trivial to explain or to fully understand, it is essential to a deeper understanding of probability. Randomness is both a "property of an outcome" and a "process" by which these outcomes cannot be predicted (Gal, 2005, p. 918). However, a complete understanding of randomness can lead to people having powerful tools that will lead to the ability to correctly comprehend uncertain events (Horvath & Lehrer, 1998). Piaget and Inhelder (1975) found that students' understanding of randomness can be without proper processes in place as a study found in children ages eight and up (Piaget & Inhelder, 1975). He found that students use prelogical thinking and make decisions about random events based on intuitive regulations. Additionally, these regulations are not hierarchical and thus are not soundly seated in actual probabilistic process or thought. Piaget and Inhelder (1975) went on to find that there were two more stages of development with the second stage ending at age twelve and the third beginning at age twelve. This implies that high-school students, in the third stage of development, should have a robust view of randomness. Thus, I will be looking for the students to have set processes in place that are hierarchical and grounded in some mathematical reasoning.

The term random event is used loosely in society at large. Sometimes, in everyday talk, randomness or a random event is used in reference to events that are not random possibility adding to people's misunderstandings regarding randomness (Kaplan, Rogness, & Fisher, 2014). For instance, randomly meeting an acquaintance at the store might not be random at all. Both you and your acquaintance have reasons and means to be at the store. Additionally, there is a likelihood that you keep similar hours at work and thus both of you would be attending the store in regular intervals anyway. Thus, never seeing a coworker might be more unlikely rather than seeing each other. Randomness enters into the situation when trying to predict when the next encounter occurs, but seeing the acquaintance is not random in that it is a known possible outcome. In this situation, people use random in place of likelihood, although there is randomness in this situation what people were referring to was the likelihood of seeing their acquaintance.

The concept of randomness in mathematics differs from the store example and in this study random events will refer to those events that cannot be described by a process or set of rules that dictate the next outcome but rather provide a method for possible next outcomes with given likelihoods (Beltrami, 1999; Gal, 2005). Furthermore, random events are characterized by uncertainty of individual outcomes (Kaplan et al, 2014). Piaget and Inhelder (1975) tried to answer the question of whether the intuition of chance was a trait that individuals were born with or was it developed over time? He used the flipping of a coin and the drawing of marbles from a jar to study randomness in young children. The nature of coin flipping and drawing of marbles are events that are not governed by a law or rule that determine the next flip or marble selected, making the trials truly random compared to the supermarket example where daily schedules caused the creation of rules of outcomes. Piaget and Inhelder (1975) found that young children struggled with randomness possibly because they were trying to create rules to account for the outcomes. The children, in their efforts to understand the random events, tried to force the events into a particular domain regardless of the event and regardless of whether the domains were proper fits or not. The coins had a circle on one side and a cross on the other. One child said there were more crosses after twenty flips because the crosses were bigger and the next flip would be a cross for the same reason. The child was trying to create a rule to explain the randomness. The fact that children struggle with the concept of randomness is possibly due to a lack of rules that govern outcomes. The development of randomness is a key to probability literacy due to its fundamental role in developing a probabilistic understanding.

Randomness also plays a key role in the development of students' probabilistic thinking. English and Watson (2016) conducted a study involving 91 fourth grade students ages nine and ten from middle class homes in Australia where English was the second language for 43% of the students. English and Watson (2016) focused their study on variation and expectation, using it as the foundation to gauge the probability literacy of the students. The study was multi-year and longitudinal and made use of multiple activities to develop and assess student knowledge. One activity that was conducted made use of TinkerPlots and had the students conduct trials of coin tosses, both a single coin toss and multiple coins simultaneously tossed. The study concluded that students need to have experiences with random trials to develop the relationship that exists between expectation and variation. The experience with random trials allows for understanding random events on an intuitive level, thereby building a more thorough understanding (English & Watson, 2016). This intuition allows one to understand how meeting someone at the grocery store and rolling the dice are random events, just different types. English and Watson's (2016) young students are capable of developing a proper understanding of randomness if properly framed and large frequencies are used to demonstrate the variation that occurs. Until students had the experience with large quantities of coin flips, many did not fully grasp the randomness of coin flipping.

Gilovich, Vallone, and Tversky (1985) studied a perceived phenomenon called "hot hand" found that at all levels of play there is a belief that players can have a "hot hand". Hot hand was studied in three stages: 1) statistics of misses and hits, 2) experiments involving players shooting basketballs and having to make predictions for their shots from a location that yielded roughly fifty percent accuracy, and 3) fans' beliefs (Gilovich, et. al., 1985). The fan data simply showed that people believed players get "hot" or are more likely to make their next shot if they had made their last shot and the belief strengthened if multiple shots were made in a roll. The observations were done by watching games and tracking the hits and misses for each player. Then, the researchers checked for the number of streaks in the data sets per player (Gilovich et al., 1985). The greater the number of streaks, the less likely the "hot hand" hypothesis. Finally, the researchers had players take shots and make predictions for each shot. Players' hits and misses were analyzed for streaks again (Gilovich et al., 1985). It was found that players are no more or less likely to make the next shot than their predicted percentage. All but one player had a high number of streaks and that player's streaks showed he was less likely to make multiple shots in a row than his average predicted. The implication was that every shot was random, whether the basketball player made or missed the shot was based on that player's shooting percentage. The truly interesting aspect of this was that in spite of the skill that professional athletes have, every shot maintained an element of randomness and was not predictable. The discussion here was not about the skill level that sets the percentage of made shots but rather that making a shot was random with a certain level of predictability, not unlike dice.

In addition to mistaking non-random events as random, people also tended to separate their intuitions from their theoretical knowledge. A study found that people, even mathematicians, would choose heads over tails because heads "comes up more" for them (Watson & Moritz, 2017, p. 272). This did not point to an issue of not understanding randomness but rather the belief that one can exert some control over the event or that the event has slightly different rules for one person over another. This idea of beliefs is outside the scope of this research, but the beliefs are noteworthy here because they impact intuitions about randomness that determine the probability literacy of a person.

Students tend to believe that rolling a die was fair and random. Fairness here refers to the fact that each face of the dice was equally likely and the event was random. The contradiction lies in the fact that while between 67% and 86% of high school age students believe that dice were fair and random, there was still a belief that the six was the least likely side to occur (Watson & Moritz, 2017). Kerslake (1974) suggested that this may be due to the desired outcome associated with rolling a 6 with certain board games. The suggestion to overcome the contradiction was to have students complete large quantities of trials of dice rolling to demonstrate the randomness of the event (Kerslake, 1974).

People's mistaken beliefs about randomness were caused by one of two effects: positive recency effect and negative recency effect (Bryant & Nunes, 2012). Children often hold positive recency beliefs about events (Bryant & Nunes, 2012). Positive recency effect is characterized by the belief that the next outcome will be the same as the previous string of outcomes. Negative recency effect was mostly associated with adults and is characterized by the belief that the next outcome will be different from the previous outcome because the previous outcome has occurred more often. For example, if a fair coin was flipped several times resulting in seven heads in a row, a person experiencing negative recency would have believed the next outcome would most likely be tails.

Fischbein and Gazit (1984) found that when students were asked if consecutive numbers had a higher chance of winning the lottery, only about one third of them correctly stated that there was no relationship between the numbers. The study was conducted by using two questionnaires. Questionnaire A was administered to the experimental group only with the purpose of determining how much information had been assimilated by the students through the lessons. Questionnaire B was given to both the experimental and control groups and required no special knowledge of probability but rather was assessing the students' intuitive misconceptions. The experimental group was composed of 285 students in grades five through seven with 160 in grade six. The control group was composed of 305 students in grades five through seven with 200 in grade six (Fischbein & Gazit, 1984). The experimental group received lessons in randomness and chance, relative frequency, and compound events. The experimental group increased their understanding of probability at a higher level than the control group, but it should also be noted that students' understanding increased with the grade level. One answer choice that is noteworthy was "random numbers have a higher probability of winning" (Fischbein & Gazit, 1984, p. 16). Approximately forty percent of the students chose this answer implying that there was a sizable student population with a developing level of randomness understanding.

The coin example was similar to an example of a lottery in relating to recency, the lottery numbers are equally likely to occur and their occurrence is random. Fischbein and Gazit (1984) found that students viewed random numbers as more likely than consecutive numbers. The belief was that consecutive numbers imply a pattern of sorts or recency might have been present here, meaning that students might have believed that since the last number was five the next number being six was predictable and thus not random. Or, due to recency, the next number needed to follow a different pattern to be random rather than be similar to the previous number; the last number was picked randomly so the next number must be random as well discounting the idea of consecutive numbers, at least in the students' minds. The idea that consecutive numbers can be randomly generated was rejected by students since they could be grouped into an ordered system. Students' understanding of randomness was highlighted here in relationship to the generation of numbers and the implied linkage to an ordered set of numbers even when one did not exist. Rubel (2007) reinforced this idea when students were asked to classify something as the "most likely" event. She found that students failed to see randomness when events were ordered. Additionally, Rubel (2007) noted that students would state that events were equally likely when events were random and therefore unpredictable. Students were connecting unpredictability with randomness. The problem was that students believed that random meant equally likely. Konold (1993) presented a task where students were asked to discuss the likelihood of certain events with a dice with five black sides and one white side. Students were asked about the likelihood of getting one color in a row several times and then getting that same color once again. This is a random event, but it is far from equally likelihood events. There is a component of independence, but it still points out that random events do not have to be equally likely. Students did not view consecutive numbers as equally likely as all other possibilities, the consecutive numbers were not random they were ordered from the viewpoint of the students. The issue is the randomness of each event and students' inability to see each event as random.

Conclusions from the Literature on Randomness. Randomness is poorly understood at all levels of schooling and beyond. But ultimately, a primary barrier to proper understanding of randomness is to apply or create rules for random events when there are not any (Fischbein & Gazit, 1984; Piaget & Inhelder, 1975). In general, students struggled to accept the unpredictable nature of randomness. Piaget and Inhelder (1975) noted that students might believe that one side of a coin weighed more and thus landed up more. Arbitrary ideas like coin weight is an example of students trying to create rules to explain why an event occurs more often than expected when said event is random. Fischbein and Gazit (1984) noted that students might decide that an event is random because of the unpredictability. Additionally, students believed that there was a level of control that can be exerted over the random event making it more predictable. A goal for this study, and an area that is not well studied, was to build an understanding of how high school students understand randomness and the topics discussed above in relation to randomness.

Independence. Independence states that events do not affect each other or rather one event's outcome has no bearing on the outcome of another event (Gal, 2005). The gambler's fallacy is a popular example of misconceptions related to independence. Here the gambler believes that the last roll of the die will provide insight into the outcome of the next throw. Note that there is a connection between the concept of randomness and independence. Students' belief in the gambler's fallacy is corroborated by a study by Fischbein, Nello, and Marino (1991) that found that students in grades 4 through 8 believed that one could learn to control the flipping of a coin. Independence cannot coexist with the belief that one can control the flip of a coin. Additionally, one cannot completely separate independence from conditional probability nor randomness. In fact, the study suggested that viewing independence as a special case of conditional probability could be more intuitive for students (Jones, 2005). Therefore, for a student to understand conditional probability, the student needs to understand independence as well, with the possibility that independence understanding needs to come before conditional understanding. However, it is possible that the concepts could be built parallel to each other. Independence might allow the student to understand conditional probability better and conditional probability would help one's understanding of independence. The literature that follows centers on independence. The relationships that exist between conditional probability and independence is a worthwhile topic but is outside the scope of this research. Although teaching conditional probability and independence in parallel with each other might prove to be a method to help improve students' understanding of probability, the parallel teaching would be its own study and one I will likely do as an extension to the current research. The research being conducted here is to uncover the levels of probability literacy of students where independence is a component. The task that students will complete will focused on drawing out their understanding of independence.

Jones (2005) summarized several studies about independence and other probability topics. One such study was completed by Fischbein, Nello, and Marino (1991) conducted with 618 students in grades 4 through 8. Students were asked to state the likelihood of obtaining a certain order of heads or tails from multiple coin flips

(Jones, 2005). Fischbein, Nello, and Marino (1991) found in their study that students might incorrectly justify outcomes while correctly determining the likelihood of the flipping of a coin. The reasoning demonstrated that students lacked understanding of independent events. For instance, some students believed that the fact that any outcome was possible was the reasoning for the sequences of two tails being equally likely to the sequence heads followed by tails. Both events are equally likely, but the belief that both are possible therefor the probabilities are equal is deeply flawed. Students also did not recognize that two tails and heads followed by tails are equally likely events because the second tails was independent of the first flip and thus has no bearing on the result. Other studies are cited as reinforcing this concept in the United States National Assessment of Educational Progress in Mathematics in 1988 (Jones, 2005). Again, the researcher asked students to determine the likelihood of a particular sequence of coin flips. Here the majority of students, 53%, failed to pick the correct response of equally likely (Jones, 2005). In contrast, a study conducted of 2930 students in the United Kingdom revealed that 75% of the students correctly picked equally likely (Jones, 2005). Konold (1993) conducted a similar study with college students finding that while 61% of undergraduate, remedial mathematics students correctly answered the equally likely case, only 35% of them answered the least likely case. A collective implication to all of this is multipronged. First, most of the research is outside of high school age students. Second, both middle school and college age students struggle with understanding independence. Lastly, the possibility might exist for students to correctly answer or pick appropriate probability outcomes without a formal or complete understanding of probability, in particular independence.

Representativeness is one way that students display their lack of understanding of independence. Representativeness is “the belief that a sample or even a single outcome should reflect the parent population” (Jones, 2005, pp. 220). The implication of representativeness is that students struggle with separating an event characteristic from the population characteristics (English & Watson, 2016). Bryant and Nudes (2012) found that adults and children tend to ignore independence of random events, looking at them more as a collection meaning the independent events were viewed as compound events or worse as a single event instead of a series of events. Thus, one can easily see how students have difficulty processing that an outcome or a set of outcomes may look very different from the parent population. For instance, if an urn contains three red marbles, four blue marbles, and five green marbles, selecting four marbles one at a time with replacement might yield four red marbles. Representativeness states that students would then believe that the urn was comprised of only or mostly red marbles.
Representativeness also has a link to sample space that will be discussed later. But, here I am referencing only to students’ lack of understanding about independence and how the lack of understanding prevents students from being able to see how independence dictates each event is unique and separate from other events meaning that the idea of the second red marble having the same probability as the first is lost on students. Granted the probability of four red marbles may be low, the fact that four red marbles occur leads students with limited understanding to think that they know the makeup of the urn. In actuality, students only know one possible color for the marbles in the urn. The concept of independence states that the next event is not related to the previous event and therefore the previous event is irrelevant to future events. This would be true regardless

of how many times in a row a red marble was pulled from the urn. The next draw has a probability that is based on the composition of the urn not the history of previous draw.

Independence is useful in probability as well as the bigger domain of statistics. Independence is useful in interpreting data sets, but the use of data sets is also useful in helping students learn about independence (Jones, 2005; Watson & Callingham, 2014). Students must work to link data and chance to develop independence (Jones, 2005). A study found that having students complete simulations will develop their understanding of independence (Jones, 2005). However, care must be taken to not reinforce incorrect thinking. Instead, teachers should focus on "predictions over the long term" to allow students to develop solid independence thinking (Jones, 2005, p. 234).

Watson and Callingham (2014) found two-way tables useful for this purpose as well. Watson and Callingham (2014) conducted a study with 110 students from three areas of Australia from grades six through eleven. Students were given two-way tables about lung disease and allergies. The students were then asked a series of questions, a slightly different version of the tables and questions had been used before by the same researchers. When Watson and Callingham (2014) compared the results between grade levels, the study found that the six and seventh graders scored a mean of 3.52, eighth and ninth graders earned a 3.98 and the tenth and eleventh graders earned a 3.21 for the lung disease problem. However, the authors did not find this statistically significant. The content is taught in grades seven and eight, which might explain the spike. The previous mentioned scores translate to students lying between having inconsistent conclusions or justification and having limited justifications. In general students failed to separate the various components of the tables or they struggled with independence of the items in the table seeing the table as in its entirety but not its parts. Thus, while high school students learn about independence as part of their curriculum, they are challenged to relate these ideas to statistics. The focus here on a few key concepts, even in a statistical setting, can give good insight into the level of a student's understanding of independence (Watson & Callingham, 2014). The level of understanding of independence in statistics and probability should be transferable between the two sub-worlds which are intimately related thereby giving insight to a student's probability literacy. Additionally, independence is important to understanding when conducting multiple independent trials of an experiment (Batanero et al., 2016). The context here is different but the idea is not. Therefore, if a student understands independence as a part of probability literacy, then there are far reaching effects into other aspects of the mathematics.

Sample Space. Sample space is the collection of all possible outcomes from a random event (Ross, 2000). Sample space is an important aspect of probability literacy due to its impact on the figure probability of an event occurring. Studies show that some students only focus on a small subset of possible outcomes rather than all outcomes (Batanero et al., 2016). A lack of understanding sample space does not automatically mean that a student cannot arrive at the correct answer. But, English and Watson (2016) point out a study were the students arrived at the correct solution to a coin flip problem without addressing the sample space as a part of their justification. Although students might have arrived at the correct answer, the justification was incorrect. The implication of the disconnect between answer and justification can be examined by studying the sample space understanding, or lack thereof in students.

A student's understanding of probability can be deeper expressed by their use of sample space thus making sample space an indicator of the student's knowledge of probability (Bryant & Nunes, 2012). The chief concern was that the student's justification was not always based on procedural process, but rather the justification was based on the student's assumption that the events are equally likely (Bryant & Nunes, 2012). English and Watson (2016, pp. 31) state that sample space yields a collection of "equally likely outcomes." Sample space understanding provides a method to access probability understanding that does not exist otherwise and thereby preventing mistakes or by providing students with a means to self-correct (Bryant & Nunes, 2012). Some students lack a complete understanding of what elements belong in the sample space. This brings up a cognitive question as well as the question of whether the sample space issue is due to a lack of proper instruction or exposure (Bryant & Nunes, 2012). The other possibility is that some students lack the ability to understand sample space. However, the topic of cognitive issues has not been studied extensively (Bryant & Nunes, 2012) and goes beyond the scope of this research as I am primarily concerned with what is a student's current understanding of sample space rather than why they do or do not understand sample space. Chiefly, I am concerned with the student's ability to create a complete sample space and can students use the sample space as an aid to answer questions about the task at hand. Sample space also provides insight into students' understanding of independence which furthers enlightenment to understanding students thinkings about probability at large (Watson & Callingham, 2014). Furthermore, there is a great deal of impact on understanding other aspects of probability in having a complete understanding of sample space (Kazak, Wegerif & Fujita, 2015); if one is literate in probability then one will have a thorough understanding of the development of sample space as a basic skill. Understanding independence and randomness is partially dependent on being able to develop a thorough sample space for an event. Thus, sample space might provide insight into a students' understanding of independence and randomness (Kazak, Wegerif & Fujita, 2015).

The need to understand sample space is centered in the theoretical model of probability (Kazak, Wegerif, & Fujita, 2015). Some studies have found that students have difficulties creating the entire sample space for even simple events such as flipping a coin (Kazak, Wegerif, & Fujita, 2015). One place where students struggled was viewing flipping two coins as a compound event (Fischbein Nello, & Marino, 1991). Furthermore, some students mistakenly apply the commutative property to the rolling of two dice and fail to see a five and a six as being a different event from a six and a five (Kazak, Wegerif, & Fujita, 2015). Both the coin example and the dice example serve to reinforce the idea that sample space is an important part of probability literacy. The ultimate issue here is if a student cannot systematically create a given sample space, the student lacks the ability to properly visualize the concrete ideas that occur (Horvath & Lehrer, 1998). Thus, the inability to create a proper sample space is an obstacle in solving probability problems because the distribution of the elements in the sample space provide students insight into understanding the probability model (Horvath & Lehrer, 1998). Additionally, students must often classify the elements of the sample space in order to correctly solve probability problems (Bryant & Nunes, 2012). Classifying issues become evident when students begin to look at the sum of two dice where the outcomes are not equally likely. The disequilibrium is only apparent to students once the sample space for the roll of two dice is properly created and classified (Bryant & Nunes, 2012). Representativeness is "the belief that a sample of even a single outcome should reflect the parent population" (Jones, 2005, p. 220). The implication of representativeness is that students struggle with separating an event characteristic from the population characteristics (English & Watson, 2016).

Using the dice example again, the separation between randomness and non-random events might be difficult to understand if the complete sample space for two dice does not include all combinations of numbers. If students are strictly looking at the matching pairs of numbers, they will miss most of the set, thereby thinking that all events are still equally likely. The incomplete sample space might make the difference between one and two dices hard for students to see.

Other Considerations. Bryant and Nunes (2012) pointed out that adults and children are more likely to correctly work probability problems when given what they called the "absolute numbers" (pp. 7) rather than the decimals. They suggested that the issue here is that people work better, at least in the beginning stages, with ratios rather than fractions. Thus, instead of presenting someone with 0.4, the problem will state four out of ten. The distinction is small but apparently important as the person's understanding of fractions is no longer a determining factor in understanding probability. This study will use task that keep this idea in mind to ensure that the student's understanding of probability is not hidden behind their misunderstanding of fractions.

Language of Probabilities

The language of probability refers to the how probability is communicated to others and the terminology itself. Chiefly, the aspects that are of concern here are the terms used when talking about probability and manners in which probability likelihoods are communicated to others (Gal, 2005). The later deals with how one conveys that something is likely or unlikely and to what degree of likelihood. Likelihood is the idea behind the doctor saying the cough is “likely nothing to worry about” or “this cough might kill you.” The complexity of interpreting these scenarios is tied up in the quantitative scale of where these two statements are on the continuum. Students need to have significant understanding of the terms used to convey these ideas. Students also need to have experiences where these terms are paired with the mathematical experience of what likely is and is not. By mathematical experience, I am referring to the student’s exposure and knowledge of probability combined with the application of said skills. Probability could be communicated strictly under mathematical descriptions such as 50% likely or 90% likely. But, even this creates issues for individuals not familiar with the numerical values’ meanings. Common language or non-mathematical language uses terms like “certain” or “highly likely” to communicate the chance of an event occurring. The problem exists in the fact some people lack the numerical understanding while other lack the understanding of vocabulary (Gal, 2005). Therefore, all students must have the same understanding of both the numeric and terminology of probabilities for communication purposes.

The vocabulary themselves were an item of interest as well. Here, the discussion is not about the vocabulary used in communicating probabilities but the terminology of probability itself (Gal, 2005). For instance, students need to understand what is meant by random, independent, sample space, or normal distribution, with the concern centered around the concept knowledge the student has in reference to the terms compared to the contextual knowledge that the student is able to both draw from and add to the communication aspect of the language of probability. Students should be aware that the terms, when used in class, have a stricter and narrower meaning than the same terms might have outside the classroom (Gal, 2005). Therefore, terminology is important for students to have complete knowledge of and understanding of probability.

Conclusion

There have been numerous studies on what can be considered probability literacy at the elementary and middle school levels (e.g. Aspinwall & Tarr, 2001; Falk & Wilkening, 1998; Fischbein & Gazit, 1984; Bryant & Nunes 2012; Piaget & Inhelder, 1975) and at the college level (e.g. Konold, 1993). However, there are few studies that specifically address high-school students' understanding of randomness, independence, and sample space. Reubel (2007) is one of the few exceptions, though there is no focus on probability literacy directly, rather independence and conditional probability. In this study, I will be focusing on how high school students understand the knowledge elements of probability literacy: independence, sample space, and randomness.

Levels of Understanding

Gathering a usable and reasonably complete understanding of students' understanding will be a necessity for assessing students' level of probability literacy. There are multiple frameworks in the literature such as Gal's for probability literacy and Jones for particular components of probability at the middle school level. Jones' frameworks divided components into four levels with level four being the highest level of achievement implying the highest-level of understanding. The Jones framework is separated by topic as well as level, providing topic specific reasoning for each level

(Jones, 2005). An advantage that Jones (2005) provided was that the framework simplifies the process of assessing the student's understanding. The drawback is that Jones did not create a framework for every topic and for the purposes of this study independence was the only topic provided. Hence, randomness and sample space required justification to be created and added to the framework.

The creation of the probability literacy framework for high school students provides researchers and practitioners a viable tool to gauge a student's levels of understanding of probability literacy. The framework will provide practitioners with valuable insight to inform their instruction. Researchers will be able to use the tool to investigate methodology, tasks development and usage, and possibility other unforeseen purposes that will ultimately benefit student learning.

Jones' Framework

The Jones (2005) and Jones, Lagrall, Thornton, and Mogill (1997) frameworks used 4 levels to distinguish student understanding of probability for understanding of independence. The first level was termed the subjective level, level two was the transitional level, level three was the informal quantitative level, and level four was numerical. Jones (2005) created a listing to classify student among the different levels for conditional and independent events. At the subjective level, or level 1 (L1), students are prone to ignore the data presented in the problem and make predictions based on no actual probability calculations. The predictions at this level are intuitive at best. The transitional level, or level 2 (L2), is marked by students making some connections between events but the methods are inappropriate for the context while there is some evidence of representing the mathematics symbolically. The informal quantitative level, or level 3 (L3), is marked by students not practicing precision in their calculations. However, students have created complete sample spaces and can distinguish between independence and conditional probabilities. The numerical level, or level 4 (L4), is marked by the students' ability to distinguish between independent and dependent events in relation to with and without replacement scenarios. Similar frameworks of understanding were created by Jones et. al (1997) for sample space, probability of an event, probability comparisons, and condition probability. Not all of these frameworks are useful to this study and one framework will need to be created for randomness.

Connecting Gal and Jones' Frameworks

Jones' framework provides a fairly clear structure to follow to assess students' understanding of probability. Gal's framework provides a means to examine the various facets of probability but not necessarily the quality of each facet. Jones, Langrall, and Mooney (2007) talk about the ability to use probability to solve problems, both theoretical and experimental which connects to multiple aspects of Gal's (2005) framework. Particularly, Gal's knowledge elements of big ideas, figuring probabilities, context, and language aspects as well the dispositional elements such as critical stance, beliefs and attitudes, and personal sentiments are all connected to solving problems as described by Jones, Langrall, and Mooney (2007) as they will have an influence on the way a student will interpret, attempt, and work the probability problems. Jones, Langrall, and Mooney (2007) also talked about the use of simulations to approximate probabilities which are connected to Gal's (2005) knowledge elements of figuring probabilities and critical questions and dispositional elements of personal sentiments and critical stance. Jones, Langrall, and Mooney (2007) stated that the use of random variables and interpreting probability distributions are an important aspect of probability and run parallel to Gal's (2005) knowledge elements of context and critical questions and dispositional elements of personal sentiments and critical stance. Lastly, Jones, Langrall, and Mooney (2007) talked about applying random variables to generate and interpret distributions which are connected to Gal's (2005) knowledge elements of figuring probabilities, critical questions, and language and dispositional elements of personal sentiments and critical stance. See table 1 in the appendix A for a visual of how the components fit together.

The connection between the two frameworks is drawn here to emphasize that neither framework is sufficient for purposes of this study. The purpose of this study was to gather an understanding of the current levels of probability literacy among high school students. Two aspects emerge: 1) the different facets of probability literacy and 2) the levels of understanding. Gal (2007) provides a structure to separate the facets of probability literacy while Jones, Langrall, and Mooney (2007) provide a basis for assessing the levels of those facets. Collectively the two frameworks provide this study with the structure required to properly complete this assessment.

The Watson and Kelly (2007) framework was based on increasing levels of appropriate and quantified response accompanied by the correct solution and was used for reference purposes in the creation of the missing aspects of assessing levels of understanding. The Watson and Kelly (2007) framework also helped provide the needed adaptations for the Jones, Langrall, and Mogill (1997) levels of understanding for high school students. The Watson and Kelly (2007) framework was not particular to high school students but rather to particular questions or tasks. The study used tasks that were related to task that Watson and Kelly (2007) used although none were directly used. Watson and Kelly (2007) was used as a guide to make needed adaptation.

Summary

Probability literacy is a collection of abilities that students can access to solve problems, interpret, and make predictions about the world around them. The knowledge elements of independence, sample space, and randomness are fundamental aspects of understanding and making sense of probability. Even slight confusion on one of these aspects has carry over effects to the other components. For instance, independence is pivotal in realizing that two random events like the rolling of dice are unrelated and are thereby unpredictable beyond knowing the likelihood of certain events occurring. A student’s understanding of randomness and understanding of independence are connected. Furthermore, just as a lack of understanding one aspect effects complete understanding the other aspect, the understanding of one aspect might overshadow a student’s lack of understanding of the other. The interplay between topics requires the need to look at both topics both collectively and as isolated concepts.

CHAPTER THREE: METHODOLOGY

The purpose of this study was to understand the probability literacy of high school students. The primary question guiding the study was:

> What is the probability literacy of high-school students? In particular, what do high school students understand about the concepts of chance, randomness, independence and sample space?

Research Design

This multi-case study was conducted with ten students using task-based interviews (Goldin, 2000; Rubel, 2007, Patton, 2002). A pilot study was initially conducted to select and refine the tasks that were used in the interviews. Further, a rubric to measure the students' understanding of randomness, independence and sample space was designed and tested in the pilot study. This rubric - Levels of Understanding Matrix (LUM), was refined during the study with the ongoing collection of data. Based on their scores in the LUM, the students were classified into three categories - Beginner, Intermediate, and Advanced. A visual representation of the process is provided in Figure 2: Flowchart of analysis. Each aspect of the study is discussed in greater detail later in the chapter.

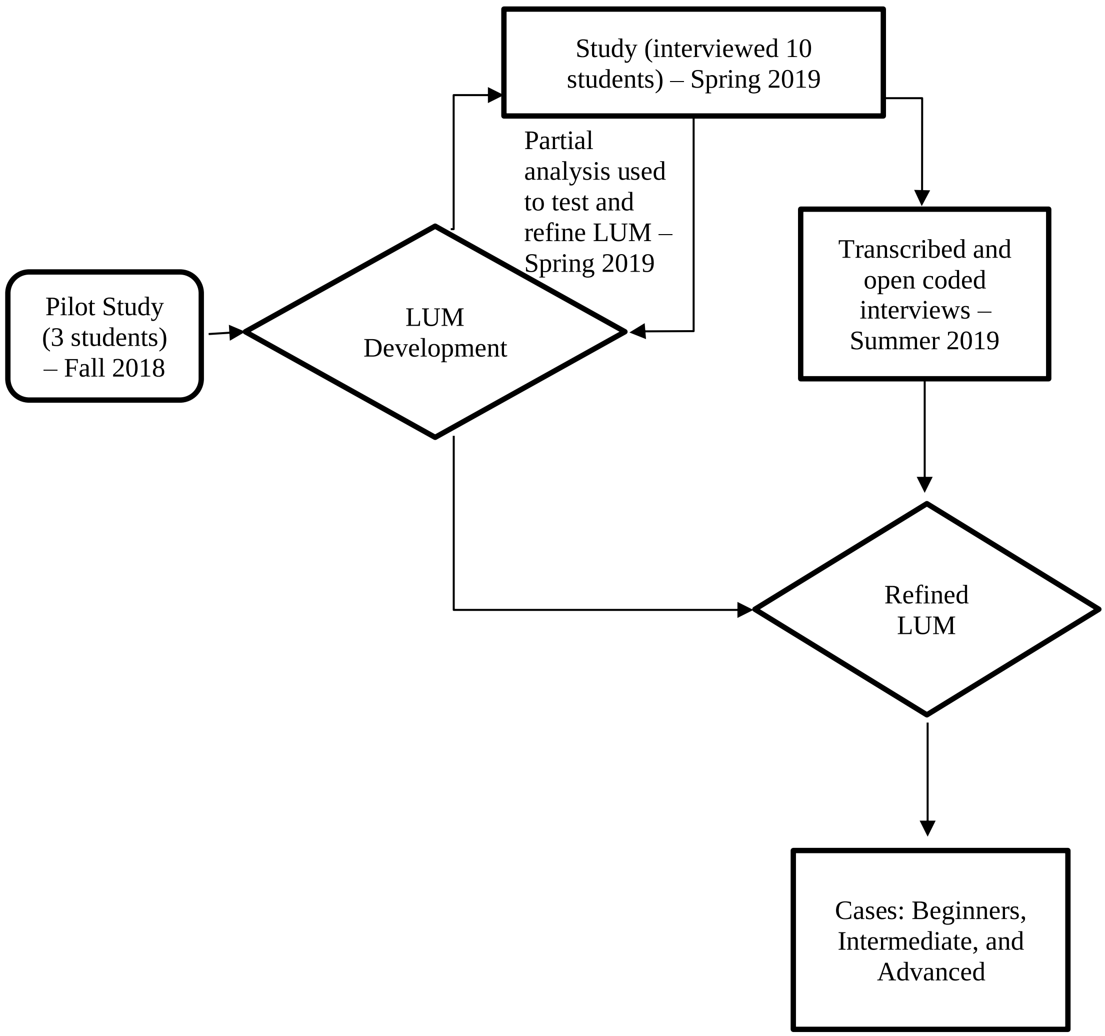

Figure 2: Flowchart of the analysis

Task-based interviews

Probability understanding is tied to language and meaning. Given that students can interpret written questions in various ways, using task-based interviews to get an in depth understanding of their probability literacy was the appropriate approach. According to Goldin (2000), task-based interviews allow for interaction between a researcher and subject, or the participant, creating discussion regarding how a problem was solved. The tasks used were chosen based on topic and ability to generate conversation and they were presented by the researcher in such a way to as to encourage thought and discourse between the researcher and participant (Goldin, 2000; Rubel, 2007). The complexity of the students' understanding was uncovered using the chosen task, allowing the researcher to develop an understanding of the students' understanding of probability in the context of randomness, independence, and sample space (Patton, 2002). I developed a list of probing questions that were designed to reveal the students' underlying thinking about probabilistic concepts. I developed the questions based on a review of the literature and the students' performance on the same tasks. In addition, I allowed wait time for students to develop a unique approach to solve the problems that could possibly differ from prior research reports. The method described above allowed for a "systematic observation" of student knowledge about probability (Goldin, 2000, p. 520).

Ruebel (2007) used task-based interviews to determine the probability strategies of middle and high school students in order to determine probabilities using sample spaces of simple and compound events. A pool of 10 tasks were used across the interviews, though each interview consisted of four tasks (Ruebel, 2007). My study built on Ruebel's study, who is one of a few researchers in probability to examine the probabilistic thinking of high school students.

According to Goldin (2000), selecting the right tasks is key to eliciting the students' thinking. Rich tasks allow multiple entry points for the students, build on various modes, and allow opportunities for the researcher to make interpretations about the students' thinking. In this study, up to three tasks were used to assess students' understanding of each of the three aspects of Big Ideas from Gal's (2005) probability literacy. Overall, the students completed five tasks in an interview. All the selected tasks are adapted from previous studies. Adaptations were made for clearer context and language for students as many of the tasks were designed for students of a different setting and time period. The goal for each task was to engage the participants in rich discussion so that a deep understanding of the participants thinking could be inferred about randomness, independence, and sample space. Familiar contexts, like coins and die, provided students with entry points into the tasks and discussions (Goldin, 2000). The way the students handled the concrete objects also allowed for interpretations about the students' thinking (Goldin, 2000). The manipulation of the concrete objects also allowed students to bridge possible challenges they encountered with using the language in probability (Goldin, 2000). For example, when participants were working on the task involving the black and white dice, I provided the participants with a die with 5 black sides and 1 white side. The students were able to roll the die, study the die, and even build a sample space from physical rolls. While the participant was doing this, building their explanation for the task, I observed the gestures and discussion of the participant. This context allowed for the presence and use of real dice that could prompt the student to experiment (Goldin, 2000).

The tasks selected for the study were rich in content, allowed the students to have multiple approaches, which allowed for participants to be able to talk extensively about them (Goldin, 2000). The students were engaged in probability tasks that relate to randomness, independence and sample space and their interactions with the interviewer were audio and video recorded. The students' written work related to the tasks was also collected. The combined formats required students to explain their reasoning in greater detail, yielding greater insight into student thinking. The duo form of recording for student thought is based on the idea that students can express different ideas based on the format they are asked to explain (Ruebel, 2007). Students use different reasoning as based on the problem at hand and the method of delivering the answer (Pollatsek, Well, Konold, Hardiman, & Cobb, 1987). The duo form of recording information about students provided two perspectives to understand what participants think about probability. Two perspectives of thought allowed development of a richer, deeper, model of student thought even in instances where the written and oral explanations matched.

The Pilot

A pilot study was conducted in the fall of 2018 with the goal of refining the interview tasks and the probing questions. Further, the pilot was also used to develop and refine the rubric for understanding the students' thinking about the concepts of randomness, independence and sample space. Three students, two White females, ages 17 and 18, and one Asian male, age 18, volunteered from an after-school math club run by the researcher. The students were part of a dual enrollment high school taking both high school and college credit classes. None of the students were current students of either the main researcher or his advisor. All students were members of the National Honor Society for their high school and had completed advanced mathematics courses. All three students indicated an interest in pursuing science, engineering, mathematics, or technology (STEM) fields as a part of their college studies. While I recognized that these students were not typical of students I intended to select in the main study, personal constraints in my schedule required the selection coming from a different pool. The selected pilot study students had not studied probability recently and were open to interact with me about probability tasks. The students provided me with detailed insight into the probability thinking of high school students. For instances, the pilot study allowed me to narrow down and eliminate several tasks from consideration for use. Many of the tasks that were eliminated used language that was not well understood by the students. Ideally, there would not have been any explanation of the tasks used with students as this used time that could be used to gather the students' understanding about one of the three topics; randomness, independence, and sample space. It might be more important to point out that if the students provided evidence that they did not have much understanding of the three topics, the issue might have been with the task itself. The student's understanding of probability may have been significant but, if the task was poorly constructed or inappropriately used, the student's understanding of the task might have clouded my perception of their probability understanding, an issue that I was trying to avoid. The students themselves provided me with feedback that helped me refine the interviews and tasks to draw out more information about my desired topics.

Students completed questions adapted from Fischbein, Nello, and Marino (1991), Lecourte (1992), Jones (2005), Fischbein and Gazit (1984), and Konold, Pollatsek, Well, & Lohmeier (1993). Given that a goal of the pilot study was to test the tasks, a wide range of tasks were used (see Appendix B for the full list of questions from the initial consideration). Three of the tasks originally up for consideration were two tasks that referenced lottery numbers and one question that referenced marbles in an urn. The focus on these three tasks was to try to uncover student understanding of randomness. Additionally, there was a coin flipping task primarily focused on independence. There were two dice tasks that were focused on sample space. Students were interviewed both orally and in written format (Rubel, 2007). Rubel found that there were inconsistencies between the oral and written responses from the same students and my pilot study found similar results. The rationale here is that more insight will be acquired into students' probability literacy by employing the same methodology.

The interviews were video and audio taped for transcription and analysis. The interviews lasted approximately 45 minutes each. Students were given a page per task to explain their work although none of them used that much space for written explanation. Students were asked a series of follow up questions designed to probe thinking deeper (See Appendix C for the current listing of interview questions).

The analysis uncovered common themes of thinking. The common themes included errors in thinking about randomness, use of recency, communicating randomness as an unordered list, use of frequency, independence, use of representativeness, and creation and use of sample space. Jones, Langrall, Thornton, and Morgill's (1997) framework was used as a basis to calibrate the level of understanding. There is an overlap between Gal (2007) and Jones et al. (2007) so that levels of probability literacy can be found. The levels were not accessed for the pilot.

The pilot study was used to develop and refine a rubric for classifying the students understanding in randomness, independence and sample space. The first iteration of the LUM is listed in Appendix D. The primary discovery for the LUM was that students were scoring artificially high in the levels. For example, Pilot Student 3 (PS3) could be assigned a level 3 (L3) in randomness despite a clear understanding of randomness. PS3 noted that randomness was a "force." Ultimately, I understood that PS3 was referring to inertia of the dice as the "force." PS3's definition was not incorrect, but the primary issue that I had with this definition was that he was not able to describe randomness as an unknown event. The LUM was adjusted to reflect the inclusion of this idea of randomness, namely can the students identify random events as having an unknown outcome. More details of the refinement of the LUM are in the next section on the findings.

Pilot Findings

The pilot study allowed me to determine a few key aspects about how to move forward with the full study. First, originally there were ten tasks that I was going to have the students complete. The ten tasks were reduced to four and then two new tasks were added to the task booklet. Then, one of those new tasks was removed, leaving the final count of tasks used in the study at five. Some of the tasks were removed because they were found to elicit very little discussion. The reason for this seemed to center around the simplicity of the task in that there was not enough rigor for high school students. Others were removed because of the context of the task, such as one of the lottery tasks, where students struggled to properly understand the task and much time was spent helping the student understand the task before any probability understanding data could be collected. Many of the tasks were adapted from elementary and middle school studies so the difficulty level issue was not a surprise. The data gathered help guide the selection of the two new tasks that were more appropriate for high school students.

During the pilot interviews, there were several instances where students made comments or worked problems in a way that aligned with expectations from the literature. The alignment itself is not surprising, but rather an indicator that much of the data collected at the lower levels has an application at the high school level. The tasks that elicited the highest rate of these types of response were kept for the study.

An oversight in the pilot was not bringing in concrete materials for students to manipulate. This was especially true on the dice and coin problems. The responses were somewhat flat and short on those tasks. In retrospect, it was obvious that the students were struggling to express their thoughts on the events. Being able to hand the students dice would have given the students a media to explore and express their thoughts on the tasks.

I used the pilot study to refine the levels on the LUM and developed a second iteration of the LUM (Appendix E). The details of the refinement for each concept is described below.

Levels for randomness. The first iteration of the LUM included language about a student's use of distributions, but there were few refences to distributions, so this language was of little use in the study. The first iteration made note of randomness being an unordered list but did not mention the unpredictability of a random event which was included for the second and final iteration of the LUM. The first iteration of the LUM yielded one student from the pilot study that was marked as a Level 4 (L4) for randomness. However, this seemed to be artificially high for the student based on his responses. PS3 stated that, if events have an equal chance, then those events are random and PS3 went on to define randomness as "pure probability." PS3 said, "nothing has an effect on it" when discussing random events. Although PS3's responses were not conveying incorrect probability, the responses did indicate a tension with justifying why an event is random, most notability PS3 did not note that random events are unordered but PS3 later pointed out in another task that one cannot predict random events. The insight gathered allowed for an accurate reflection of student understanding. The literature highlights randomness as an unordered list of outcomes and students must be able to ignore positive and negative recency. PS3 was able to ignore recency, I believe this was more about his view of the nature of the task being independent. Meaning, his understanding of independence hid his lack of understanding of randomness. PS3 used the word “unbiased” in describing randomness, implying some understanding of randomness. However, when pressed about “unbiased” he stated there was no outside influence on the event. The basic definition, while not wrong, is also not formally defining randomness and hence the reason for reformatting the L4 for randomness to include the idea of an unordered list at least described. Under this new criteria PS3 was a Level 3 (L3) for randomness, which was the highest any of the pilot students achieved. The other two students were L3 and Level 2 (L2), which was expected and appropriate based on the holistic view of the data collected for each student.

Levels for sample space. Initial versions of the LUM did not include the ability for students to use sample space to solve problems unprompted. However, this yielded all of the students scoring artificially high in sample space. I referred to the Common Core standards for probability finding that the ability to make decisions using probability as a required skill as well as correctly figuring probabilities. Sample space is a tool to figure probabilities, as the literature suggests, and therefore the ability to use sample space as a tool was included as L4 understanding as a result (NGACBP, CCSO, 2010). I stated that students scored artificially high because I noticed from the interviews that none of the students knew how to use sample space to solve problems. Additionally, PS3 was the only student that could build a complete sample space systematically, which necessitated the need to include usage of sample space as a component for L4 sample space in the LUM.

Levels for independence. The pilot interviews highlighted that students might have had a basic conceptual understanding of independence but lacked the ability to justifications and communicate that understanding to others. Additionally, the Common Core standards point out the ability to make decisions as an important component of any mathematical ability. Thus, decision making was added to the LUM (NGACBP, CCSO, 2010). PS2 scored L4 for independence where PS1 and PS3 scored a L3 under the second iteration of the LUM. The scores seemed appropriate for each student as PS1 and PS2 were both able to justify their reasoning after properly identifying the events as independent. PS3 was able to identify an event as independent but was not able to explain his reasoning. All three pilot students were able to ignore representativeness. PS2, for instance, stated that each roll of a dice was unrelated to the next roll in reference to the Black and White task. PS2 added an example of the game Yahtzee, a dice-based game, to explain her understanding of independence. PS2 noted that every roll was unrelated to the previous roll. Alternately, PS1 noted that on the Straight Lottery task that “eventually you have to get picked,” by this PS1 was noting that every set of possible outcomes will eventually occur rather than every player will win. Originally, PS1 scored a L4 on independence but this interaction with her made me decide that the LUM must be further modified as she was allowing representativeness to sway her decision making. Therefore, I added the criteria to the LUM that all three behaviors of independence must be satisfied for a L4. The adjustment moves PS1 to a L3 as she was meeting the other two requirements for independence.

The second iteration of the LUM was used in the main study and was refined further based on the ongoing analysis of the interviews. A description will be provided in later sections.

Further Development of the Levels of Understanding

The students' levels of understanding of the big ideas were assessed in terms of qualifying their perceived understanding at a level. Whether the student arrived at the correct answer was secondary to the justification and conceptual understanding that the student can communicate. While I was not trying to measure the level of knowledge the students have about probability, I am concerned with student levels of understanding for randomness, independence, and sample space. The quality of understanding of these three topics determined at what level of literacy the students were for randomness, independence, and sample space. Each topic had multiple aspects that will be of concern in gauging the quantity and quality of understanding. For instances, I look to understand whether students had the ability to ignore recency when dealing with random events or whether students could overcome representativeness when dealing with independence. These are only two of the items that I was looking for among many to inform me of the levels of students' understanding.

The Levels of Understanding Matrix (LUM) in Table 2 was a result of the refinement in the pilot study as was used as the starting point for the main study. The refinement of the LUM also took place during the study. After each interview I tested the characteristics for the various levels and the concepts and made minor adjustments as needed. Table 2 includes these adjustments as required and is the latest iteration of the LUM and reflects the criteria that was used for all ten students interviewed. Once all interviews were conducted, the latest version of the LUM was used in the data analysis as a framework to assign all students levels for the three topics and categorize student data.

Table 2: Levels of Understanding Matrix (LUM)

Theme	Level 1 (L1)	Level 2 (L2)	Level 3 (L3)	Level 4 (L4)
Sample Space	Students makes incomplete list and does not use list to make decisions regardless of prompting	Adopts a strategy to create list of outcomes but must be prompted to do so but cannot use it to solve problems.	Adopts a strategy to create complete list and uses the complete list to solve problems **or** to make decision, but must be prompted to use the list **or** creates an incomplete list but does not need prompting to create it or use it to solve problems.	Adopts a strategy to create complete list and uses list to solve problems or to make decision and student does so unprompted.
Independence	Makes prediction for subsequent outcomes on previous (or future) outcomes rather than current situation: representativeness affects decision making	Student can identify an independent event and can assign numerical probability with or without replacement, **or** can provide correct justifications for answers, **or** does not allow representativeness to sway decision or calculations about independent events (one must be met)	Student can identify an independent event and can assign numerical probability with or without replacement, **or** can provide correct justifications for answers, **or** does not allow representativeness to sway decision or calculations about independent events (two must be met)	Student can identify an independent event and can assign numerical probability with or without replacement, **or** can provide correct justifications for answers, **and** does not allow representativeness to sway decision or calculations about independent events. (all 3 must be met)
Randomness	Makes predictions based on: pattern or haphazard models. Regular disruptions or frequencies are not mentioned. Evidence that student makes decisions about randomness based on either positive or negative recency. Or student allows personal, non-mathematical, beliefs or experiences to sway decisions on a major level.	Partially make predictions based on: patterns, the most recent occurrences, or haphazard models, or can create a mostly irregular distribution and little use of "Qualitative intuitions of proportionality" (Piaget, 1977, pp. 54), some evidence that positive or negative recency influences decisions.	Make predictions mostly based on frequencies or can create distribution that mostly irregular or partially uniform distributions and sometime uses "Qualitative intuitions of proportionality" (Piaget, 1977, pp. 54), little evidence that positive or negative recency influences decisions. Student still does not include "unordered" although they may point out that the event is "unpredictable" in their discussion about randomness.	1) Make predictions based on frequencies or can create regular or uniform distribution for decisions, 2) Can communicate that randomness is an unordered list of outcomes or "individual outcomes are uncertain" (Kaplan, 2014), 3) Makes extensive use of "Qualitative intuitions of proportionality" (Piaget, 1977, pp. 54), 4) No evidence that positive or negative recency influences decisions, and 5) Can explain reasoning for answer(s) or changes to answers.
*stage sets refers to how students deal with compound events such as two dice being rolled at the same time				

Data Collection

As stated before, students in this study were interviewed both orally and in written format (Rubel, 2007). Rubel found there were inconsistencies between the response from the same students. The rationale was that I acquired more insight into students' probability literacy by employing the same methodology. Both aspects of the interview occurred at the same session with the participants given time to write down their thoughts both before and after the oral part of the interview.

Participants

The participants for this study were high school students ages 15 to 18. There were ten students interviewed for the main study – five females and five males. Three students identified as African American, three identified as Hispanic, two identified as White, one identified as Asian and one identified as multiple races. Six students were selected from a rural high school that was approximately an hour drive from the researcher's home and roughly 45 minutes from the researcher's university for the convenience of the researcher. An additional four students were selected from a Saturday enrichment mathematics program sponsored at the university that the researcher attends. This Saturday program is for local urban students in the researcher's city of residence. The Saturday program serves middle and high school students from a large urban school district. The rural high school serves a small community of less than 20,000 people. The rural high school has approximately 1000 students with approximately 61% of the school population being White, 27% Hispanic, 15% Black, and all other groups being less than

7%. The population of the rural high school is 49% female and 51% male. Sixty-one percent of the rural high school population is considered economically disadvantaged.

A math teacher at the high school provided access to six students based on ability levels from Advanced Placement Statistics, Math 1, 2, and 3. The teacher was instructed to provide students of low, middle, and high mathematical ability as based on test scores and teacher recommendations. The school and the teacher did not share any testing information with the researcher. The pool of possible participants was limited by those that returned the proper authorization forms. Interviews were conducted in a conference room of the high school in the main office of the school or on the university campus in a separate room from other activities during the Saturday program to make the process as convenient as possible for the participants.

Sampling was conducted in this manner as a convenience and to reduce the time to build the sample. More importantly, this reduced the need for collaborating teachers to schedule additional meeting times with students for the purpose of administering a questionnaire to students. Originally a questionnaire was to be given to students, but this idea was rejected after talking with collaborating teachers. Therefore, it was on the shoulders of the teacher to supply a varied level of student ability. The study looked for three students in the low level of ability, four in the middle level of understanding, and three in the upper level understanding in general mathematics. A best faith effort was made to achieve the different levels of abilities and diversity in the sample set but, since test score data was not shared, the ability levels were self-reported by students, all of whom self-reported A’s and B’s.

The Interview

Task-based interviews were conducted with ten students. These students were selected by the cooperating teacher or professor for convenience of the study. Eight students were interviewed one-on-one and two students were interviewed as a pair. Note that the paired interview was conducted due to constraints in the students' schedules. The students read, engaged, interacted with, and wrote about the task, then discussed their thinking about the task (Goldin, 2000). Goldin (2000) pointed out that objects related to the task, when provided, will the help students with thinking and explanations. After the students had a chance to record their solution, I asked them to explain their solution. In the process, I asked probing questions to ensure I understood the nuances in the students' thinking. When appropriate, I asked the students to use the concrete objects to demonstrate an explanation. I avoided asking any leading questions that could indicate an approach to the solution and instead I allowed the students' responses to guide the evolution of the interview.

Students were interviewed in three stages with two students for stage one, two students for stage two, and six students for stage three with roughly a month between stages. The purpose of the stages was to allow for refining the interview questions based on the student responses to ensure a rich data set. The LUM was based on the pilot study data as it provided the threshold of understandings that belong in each component of the study: randomness, independence, and sample space. The LUM was then refined and adjusted after each stage of the study to ensure each component was being accurately gauged and all relevant criterion were included in the different levels of the LUM.

The pilot study also served in narrowing the selection of tasks for the interviews as ten tasks were originally collected and used in the pilot study but only five tasks were used in the study. Based on the reflection of the recorded pilot interviews, I adjusted the interview questions to ensure detailed explanations of the students' reasoning was captured. I consulted with my advisor between the stages to adjust the probing process. The conferences between stages allowed for the tasks to be altered slightly to elicit the desired type of responses.

I asked students questions that have additional follow up task questions built in as a part of the oral questions asked in the interview. The follow up task questions were written out as much as possible. However, a bank of questions was maintained to probe deeper into student thought as the need arose.

Responses and Their Implied Understanding

Fischbein, Nello, and Marion (1991) reported students' responses from a study of probability where students correctly answered probability questions with incorrect reasoning. One such response was, "the probability is the same because the result cannot be predicted" (Fischbein, Nello, & Marion, 1991, p. 538). The student's response was to a task involving the flipping of two coins. The student correctly answered the question, but the reasoning points to the fact the student has little understanding of probability. This indicates the need for more thorough understanding of students' understandings on probability in order to better inform instruction to prevent such misconceptions from propagating.

Types of Responses

Listening to a student explain a concept and knowing what that student understands about that topic are not necessarily the same event. Furthermore, the possibility exist that students can achieve the correct solution to a problem while having a completely wrong understanding of the concept and how to solve the problem. Rubel (2007) found that there was a need to go beyond just asking students for answers but to also search for an understanding of their thinking. The premise here is that students may have a procedural understanding but not a relational understanding of the concept.

The Reason for the Comparison of Oral and Written Responses

Fischbein and Gazit (1984) said that there are two intuitions, primary and secondary. These two intuitions can come into conflict with each other as the primary is a more natural development and the secondary being more formally developed such as one might expect in a classroom setting. They can lead to expressing reasoning differently based on the format of the explanation, namely written response versus oral response (Rubel, 2007). Thus, if the information collected during the interview is oral, then there is a chance that the information is mostly primary or mostly secondary thereby eliminating access to another perspective of the student's understanding of the concept. The irregularities in the responses will give insight to any conflicts the student is having about the concept.

Stage One of the Main Study

There were two interviews as a part of stage one. The interviews provided me with the baseline to evaluate the effectiveness of the LUM. The primary findings for the LUM was that independence and randomness themes were not robust enough to cover all scenarios. Thus, the LUM was adjusted to include more details about what each category would entail as stated above. For instance, independence was rebuilt to include three items that both appeared in the literature and appeared in the pilot study as well as the stage one interviews.

The Hot Hand task displayed students' urge to maintain the basketball shooter's 50% shooting percentage. The issue relates to representativeness. An argument could be made for recency, but I suspected that students were trying to maintain the percentage which is more in line with representativeness and thus implying a connection to independence. Therefore, independence in the LUM was adjusted to include more emphasis on representativeness and any display of representativeness would prevent someone from achieving higher than a L2. Much of the literature pointed to students' reliance on recency in a task such as Hot Hand. However, in practice it was found that students are using representativeness to make decisions about Hot Hand task rather than recency.

Stage Two of the Main Study

There were two interviews for stage two. The interviews were used to further develop the LUM to ensure that it was possible to properly evaluate student understanding of probability. The aforementioned artificially high results seemed to be managed with the mentioned updates. Students displayed recency in their discussions. This implies that recency should remain an aspect of the discussion.

Stage Three of the Main Study

Stage three saw the skill of activities in the tasks emerge in discussion when discussing randomness of the outcomes of events. The skill defined here refers to the skill that students perceived an activity would require. It should be noted that the study was not concerned whether the skill was real or not, only that students talked about skill as being a factor in the outcome. Sample space and independence seemed to continue to correspond to the LUM, but randomness needed a richer description to account for the additional details that emerged. In particular, skill appeared to play a role in students' beliefs about randomness. I saw students move away from talking about the frequency of an event to discuss the randomness in favor of the skill level involved in the event itself. For instance, in the basketball task Hot Hand, Student 6 (S6) talked about the skill and “muscle memory” of the player and how that would alter the outcomes. S6 mostly ignored the fifty percent shooting of the player, referring to that frequency as being in the player’s past. The skill level of the player can increase and thus change, either positive or negative, from the fifty percent mark previously stated.

Stage three also saw the emergence of future events to affect a student’s decisions about independence. Student 5, S5, stated when working the “Black and White” task that the white side needed to come first because of the possibilities it allowed for in subsequence rolls of the die. S5 was referring to the possibility that white appears once in six rolls of the die. Typical thought on independence has viewed the idea that past events affect future events. S5 pointed out the need to account for future events as well. Therefore, I amended the LUM for independence to account for future events under L1. The adjustment accounted for what was observed as parallel events, past and future events, while placing the same level of importance on both.

The Interview Tasks

Tasks are central to task-based interviews (Goldin, 2000). Ideal tasks are accessible and provoke thought and discussion. Different types of task provide various affordances to student thinking. Students have been shown to make fewer errors if the task can be translated into a setting that is familiar to the student (Pollatsek et al., 1987). The tasks for the interviews are adapted from previous research and have the potential to build on previous studies about probability allowing for connections to be made between previous and current research.

Tasks

The tasks have gone through several iterations. The initial list was comprised of more than ten tasks. During the pilot study, some tasks were quickly removed from the collection due to their inability to promote discussion between the students and researcher.

Table 3: Tasks used for study		
Task	Rational and usage	Source and previous findings
Task 1 – Hot Hand A particular pro basketball player is a 50% shooter from a particular spot on the court. This means that he hits half of the shots he takes from that spot. He has made his last 3 shots from that location. Is he more likely or less likely to make the next shot? Explain how you arrived at this decision.	This task is purposed with determining participants' understanding of randomness and independence. The rationale behind this question is an article about "hot hand" referencing pro basketball players. The higher number of streaks a player goes on, the fewer "hot hand" events the player had. Possible conflicts that may arise for participants are their beliefs about "hot hand" streaks which might overrule their understandings or the rules they have created about independence and randomness. This will imply that the participants have an incomplete understanding of independence and randomness.	Gilovich, Valllone, & Tversky, 1985 It was found "hot hand" streaks are mostly non-existent. The belief found among fans was that players go on "hot hand" streaks. But, after watching film of hundreds of games, this idea was rejected as untrue.

Task 2 – Straight Lottery - moved Task 5. Ruth plays the lottery using consecutive numbers like 1, 2, 3, 4, 5, 6, although the numbers do not need to be selection in order to win. She claims it increases her chance of winning. On the other hand Jenny claims that the chance of getting six consecutive numbers like 1, 2, 3, 4, 5, 6 is smaller than that of getting a random list of numbers. Jenny says that a lottery is something chancy and therefore there is no chance of getting a list of consecutive numbers. What is your opinion with regard to the two attitudes, that of Ruth and that of Jenny?	This task focuses on randomness and independence. The consecutive numbers suggestion will hopefully lead students to think about the underlying truth of randomness. However, care will be used for this task as to ensure that the interview does not lead the student to believe that the numbers need to appear in order, only that the numbers that appear are consecutive. The recency effect will also be of concern here during this task as it will provide insight as to whether all or some of participants respond with positive or negative recency.	Adapted from Fischbein et. al. 1984. The study found that students struggled with allowing the idea that 6 could come after 5 when the numbers were supposedly random as they believed that consecutive numbers were not random. This task was not used after the second interview of the study as it did not appear to provide useful data about student's understanding of probability.
Task 3 – Black and White Roll A die is painted white on one side and black on the other five sides. If the painted die is rolled six times, what is the most likely result? a) WBBBBB b) BBBBBW c) WWWWWW d) BBBBBB e) BWBWBW	This task is focused on independence. The major point of concern will be whether students use their understanding of independence or their own personal set of rules to pick the answer. Positive/negative recency will be looked for as it is believed that children should be switching between positive and negative recency at the age of my participants. The task will be used to discuss sample space by extending the question to including asking the participant to create the sample space for six rolls. Randomness will be discussed using follow up questions.	Adapted from Konold, 1993, pp. 393 The study was conducted on college students. 5 of 12 students correctly answered Konold's version of the question. Those 5 students did not go on to demonstrate correct thinking but rather were using their "representativeness heuristic".
Task 4 – Double 6 Two dice will be rolled together, this roll is R1. R1 results in a 5 and a 6 being obtained. Both dice are rolled again, this roll is R2. R2 results in two 6's. Explain your views on the following statement, "The chance of obtaining R1 is the same, less, or the equally likely as obtaining R2."	The primary purpose of this task is to investigate independence This task addresses sample space directly by making the participants discuss the possible events that could occur.	Adapted from Lecoutre, 1992 The dice task was included because Lecoutre designed it as a "standard problem" due to its ability to focus on independence only and bring out a participant's understanding of independence.
Task 5 – Fake Flip You are going to pretend to flip a coin 10 times. Create a listing of outcomes in such a way someone will be fooled into believing the list is real.	This task pushes at the idea of the creation of sample space while also asking the participant to explain independence and randomness. The sample space created will provide insight into how thoroughly participants understand independence. Understanding of randomness will come out as a part of the follow up questions. There will be belief concerns that will come out as well. A noteworthy aspect from this question is what percentage	Adapted from Jones, 2005 Study that used the original task that this task is based on found that 47% of elementary students and 61% of college students correctly identify a given combination of coin flips as equally likely. The major findings of the study discovered that students failed to identify the individual coins flips as independent.

	get this question correct and if their reasoning correctly explains the situation.	
Task 6 – Smart Chance - removed from the task booklet, rational stated above. Gill is 10 years old. In his box, there are 40 white marbles and 20 black ones. Lucy is 8 years old. In her box there are 30 white marbles and 15 black ones. Each of them draws one marble from their own box, without looking. Lucy claims that Gill has a greater chance of extracting a white marble because he is the older one and therefore he is the smartest of both of them. What is your opinion about this?	This task was picked to discover more about students' understanding of randomness and what type of constraints effect randomness. Compared to the Basketball Hot-hand task where skills are involved in the process of shooting the ball but not the event itself is a chance event with a particular probability. Here there is no skill involved in the process, but the task introduces an attribute that could "affect" the outcomes. It will be interesting to see how students deal with this tasked compared to the hot-hand task.	Adapted from Fischbein & Gazit 1984. The study was conducted with a control and experimental group. The two groups of 7th graders answered 87.3% and 82.6% correct, respectively. The task's intention was to determine if age and intelligence have an influence on chance events. The study found that 7th graders most do not believe that age nor intelligence affect chance events. It was found that the students that had more exposure to formal probability education performed worse on this task which is the primary reason for its inclusion in my study.

Each task was used to gather a particular aspect of understanding. The Hot Hand task was used to gather understanding about independence primarily with it also aiding in gathering data about students' understandings of randomness. Students' use of representativeness was on display with the Hot Hand task as was their use or lack of use of recency. The Fake Flip task was primarily picked to gather data on students' understandings of sample space, both a student's ability to create a complete sample space and how a student then uses said sample space. The Fake Flip task also played a role in gathering understanding of both randomness and independence. The Fake Flip task provided good contrast to the Hot Hand task with both tasks having 50% probabilities for their respective outcomes thereby allowing for deeper insight to how students use recency and representativeness when dealing with randomness and independence. The Double 6 task provides insight into students' understanding about independence with minor understandings of randomness. The Black and

White task was chosen mainly to gather student understanding of independence. This particular task allowed me to gain insight into students' understanding and, at a minor level, students' access to recency, either positive or negative. The Straight Lottery task was chosen for its ability to uncover student understanding about randomness and independence. However, only one student ended up completing the task and the data collected was minimal. The task was not used after the second interview due to time constraints and the tasks inability to provide the volume of feedback that the other five tasks did. The contrast between responses from students on different tasks will aid in creating the most complete picture of student understanding of randomness, independence, sample space, and the interaction between them.

Data Analysis

My study made use of Grounded theory in that the data collected was used to produce the LUM (Mertens, 2015). The task-based interviews were videotaped with ongoing analysis completed while refining the LUM although there were no major adjustments made after the pilot. After the completion of the interviews, the LUM was used to analyze the data collected as a means to assign students levels of understanding and to assign students to groups based on their assigned levels. Then, the interviews were open coded for salient phrases with special attention given to any possible parallels with the literature. Common themes emerged through the coding process and the responses were processed through the LUM (Mertens, 2015). It should be noted that LUM was established at the point that coding began and was not altered except for wording to increase clarity of the LUM and its intent. The codes were categorized into items such as recency, representativeness, separate events, and other items that appear in the LUM. The categories were then connected to randomness, independences, or samples space and assigned a level from the LUM. Once the levels were assigned to students, analysis began to determine how to group the students into the Beginner, Intermediate, and Advanced groups. Groupings were initially based strictly on LUM levels; however, this left a few students without a clear group assignment. Thus, a closer inspection was done by dissecting the LUM's components to define the characteristics of each group.

The written responses, although few in quality, underwent a similar process that shed some additional insight into the emerging themes, especially when conflicts appeared. The conflicts here refer to when it was difficult to assign a student to a level on the LUM based on student responses from different tasks that did not align with each other. The overlap between the video transcriptions and written explanations provided validation to findings.

While completing the analysis, certain findings from the literature of previous studies were looked at for incomplete understandings of randomness, independence, and sample space. Additionally, it had been found that in reference to randomness some students used intuitions to determine outcomes. The concerns here were that intuitions can be governed by heuristics and thereby led to incomplete or erroneous understandings of randomness (Fischbein et. al., 1984; Jones et. al., 1999; Aspinwall, 2001; Gal, 2002; Chernoff, 2009). As a result, students might have failed to correctly understand an outcome. For example, participants might have confused the ability to acquire skills with meaning that one could have removed all randomness from an event. Student use of patterns was also looked for while investigating randomness as a student's use of patterns provided great insight into their current understanding of randomness (Fischbein & Gazit, 1984).

As stated before, Fischbein and Gazit (1984) found three common misconceptions that reinforced the recency effect: 1) confusion between personal desire and objective probability, 2) confusion between objective and relative certainty, and 3) insufficient data (pp. 8). The connection to my research was how did this affect high schoolers' understanding of independence. I was looking for the existence of these themes as students explained their thinking of various tasks. Jones (2005) noted that simulations helped students develop their understanding of independence as well. Although, the Hot Hand and Fake Flip tasks in Table 3 were not true simulations, the usage of both was partially based on the idea of using them as simulations (Jones, 2005). If students held a complete understanding of independence it should have been evident in their responses to those two tasks. Namely, I was looking for their ability to explain or create a sequence of events that are independent.

I looked for whether students displayed representativeness in their discussion centered around tasks related to independence (Jones, 2005). Students should have held the ability talk about individual and a collection of events if they understood independence. Students that viewed collections and individual events as the same still provided insight into their understanding in that they were demonstrating independence as they understood it.

When assessing sample space, I looked for students' ability to create and use sample space to create solutions to a given task. For instance, a noteworthy aspect was a student's ability to differentiate between rolling a 5 and 6 versus a 6 and 5 (Kazak, Wegerif, & Fujita, 2015). I also looked for a student's ability to classify or group elements of the sample space by type (Bryant & Nunes, 2012) as well as could the student look at the sample space in the proper context of the population of the events (Jones, 2005). Lastly, I was trying to determine if the student correctly interpret the sample space as indicative of the population based on the size of the sample, or the representativeness of the sample space (Jones, 2005).

Tables 4, 5, and 6 are the individual breakdowns for how each student was rated based on the LUM for each task. There are blanks in some spots since not all components of randomness, independence, and sample space were intended as considerations for each topic. The levels that appeared in Tables 4, 5, and 6 were based on student responses but those levels were not averaged together. Rather, they provided a means to verify levels. This will be discussed in greater detail in later sections, but an example is S5 for independence. S5 scored a L1, L1, and L3 for Hot Hand, Black and White, and Fake Flip respectively, but S5's overall score for independence was L1. This event demonstrates why each student completed multiple tasks and why it was important to acquire the best possible perspective on each student's understanding. The L3 that the student achieved was not necessarily a mistake but an indicator that some understanding is there in the context of that problem. The two L1 did not negate the L3, rather the L1 created entry points to examine the students holistic understanding of independence.

Table 4: Randomness Breakdown

	Hot Hand	Black and White	Fake Flip	Double 6	Straight Lottery
S1	2	4	3		DNU
S2	1	1	1		DNU

S3	2	3	4		DNU
S4	4	4	2		DNU
S5	1	3	2		DNU
S6	2	2	1		DNU
S7	3				DNU
S8			2		DNU
S9	1	3	2		DNU
S10	1	2			DNU

Table 5: Independence Breakdown

	Hot Hand	Black and White	Fake Flip	Double 6	Straight Lottery
S1	1		2		DNU
S2			1		DNU
S3	3	3	4		DNU
S4	4	3	2		DNU
S5	1	1	3		DNU
S6		1	1		DNU
S7			2		DNU
S8			1		DNU
S9	1	3	2		DNU
S10	1	2			DNU

Table 6: Sample Space Breakdown

	Hot Hand	Black and White	Fake Flip	Double 6	Straight Lottery
S1			3		DNU
S2			2		DNU
S3		2	4	4	DNU
S4			2	4	DNU
S5				3	DNU
S6			2	2	DNU
S7				2	DNU
S8				3	DNU
S9		1		1	DNU

S10				1	DNU

Table 7: Task to Topic Usage

	Hot Hand	Black and White	Fake Flip	Double 6
Randomness	X	X	X	
Independence	X	X	X	
Sample Space			X	X

As stated before, most of the tasks had an intended focus of examining student understanding of particular components in isolation rather than collectively. Table 7 displayed the components that each task was focused on exploring noting that the Fake Flip task was the only task that examined randomness, independence, and sample space. Double 6 was only focused on sample space however, it should be noted that information about randomness and independence was gleamed from the Double 6 task as a result of the discussion that took place. Table 7 displays the focus of how the task was intended rather than being an all-inclusive explanation of how it was used. Two students received rankings for sample space for the Black and White task, as can be seen in Table 6, even though it was not a focus of the task as can be seen in Table 7. Similar situations occurred with all of the tasks as discussion progressed and students made mention of various perspectives.

Overall Classification of the Students

The data analysis included assigning each student to a level of understanding in each of the three concepts of randomness, independence and sample space. After the classification of the students within each concept, the level was assigned a score (e.g. L4 was 4 points). The points across each row were added to get a probability literacy score for each student.

A further comparison of students with similar scores revealed commonalities that suggested the ten students could be classified into three groups. The classifications were based on each student's perceived, collective understandings on each of the tasks (see Tables 4, 5, and 6). Initially, to get a baseline of aligning students in groups, the scores among the three topic levels was summed up to aid in the revealing of commonalities. It should be noted that the sums were not used to break the students in to groups but were used as jumping off point to look for the distinctions between groups. The sums allowed for a fairly clear separation between the Advanced Group and the Intermediate Group. However, the distinction between the Intermediate Group and Beginner Group was less pronounced and required deeper investigation to determine where the groups should transition. The three groups were defined by the students' levels of understanding when holistically looking at all three topics in context. The comparisons between the levels that follows was used as a platform to begin the comparisons within and between groups. The following description was not the primary reason for the separations but rather it was the beginning point for those separations.

Table 8 shows the sum of the three topics based on the understandings as based on the interviews. The overall score was the sum topic scores. The topic scores were initially based on the average score from each task completed. Figuring the topic score in such a way created problems in placing students into the correct groups, for example S5 has an overall score of 6.667 and an independence score of level 1.667. S5 was not easily identified as member of the Intermediate group or the Beginning group as well as a level 1 or level 2 for independence. Thus, a closer inspection for all students was completed where individual components were compared for grouping purposes. Table 8 was updated to create a more authentic result topics levels and overall groupings; the results are displayed in Table 8.

Table 8: Overall Rankings

	Randomness	Independence	Sample Space	Overall	Group
S1	3.000	1.000	3.000	7	Intermediate
S2	1.000	1.000	2.000	4	Beginning
S3	3.000	3.000	4.000	10	Advanced
S4	4.000	3.000	4.000	11	Advanced
S5	2.000	1.000	3.000	6	Intermediate
S6	1.000	1.000	2.000	4	Beginning
S7	3.000	2.000	2.000	7	Intermediate
S8	2.000	1.000	3.000	6	Intermediate
S9	2.000	2.000	1.000	5	Beginning
S10	1.000	2.000	1.000	4	Beginning

As can be seen in Table 8, the scores for topics were adjusted to account for students like S5 who had 1.667 for independence which was adjusted to a 1. The adjustments were not a rounding of the previous score but rather the result of an in-depth comparison between S5's abilities and the students surrounding her score wise.

Table 4, 5, and6 displays the scores for randomness, independence, and sample space for the five tasks, respectively. Blank cells are represented places where no data was reportable for the topic, but the task was completed by the student. Shaded out cells represented tasks that were not done with a particular student. The Straight Lottery task was completed by S1, S2, and S5. During the analysis, large quantiles of data was collected for randomness and independence, which were the topics that the Straight

Lottery task was intended to provided data for. However, the Straight Lottery task provided little new data that was not already being provided by another task and thus was found not to be as useful as other tasks. An argument could be made that the Straight Lottery tasks would be necessary for S5 since he was the only student that completed the task and was on the bubble between groups. However, I felt enough data was collected from the other task that S5 had completed to make the Straight Lottery task analysis unnecessary.

Recall the LUM in Table 2 depicts the levels for each of the three topics. The sample space row of the LUM is unique in that it required students to build on previous abilities. For instance, to be a level 3 (L3), the student needed to be able to build the sample space completely and correctly or used an incomplete sample space to solve a problem unprompted. Independence was marked by the fact that to achieve a level four (L4) a student needed have done four things starting with ignoring representativeness following by actively completing three other items. The levels for randomness were constructed similarly. The main ideas were that students needed a collection of abilities for each topic.

The Beginner group held an overall understanding of L1. Three of the students, student 6 (S6), student 2 (S2), and student 10 (S10) all had two L1 understandings, one of which was randomness for all three. The students' understanding of randomness separated them from Intermediate Group. The issue with the Beginner group was whether to include student 9 (S9) with the Beginners. S9 was the only student to have an overall sum of 5, thus leaving some judgement that needed to be made about her placement in the groups.

The placement of S9, S8, and S5 could be made by just looking at the levels of understanding for each of the three topics. As it was described there was not a clear point where one of them distinguished themselves from the rest.

The Intermediate Group had four students overall, but while student 1 (S1) and student 7 (S7) were firmly separated from the Advanced group, the distinction for student 5 (S5) and student 8 (S8) was less pronounced. S5 and S8 both had overall sum scores of sixes, which separated them from the Beginner group and the Advanced group. A consideration for the grouping of S5 and S8 into the Intermediate Group was based on the fact that they were a L3 for at least one category. S1, for instance, was a L3 for randomness and sample space whereas the other three students were a L3 for one category. Additionally, independence was the lowest level for each of them with S7 having a L2, which was a tie for S7 with randomness. The other three students had a L1 understanding for independence.

The Advanced Group consisted of two students with overall sum scores of ten and eleven. The highest sum among the next group, Intermediate Group, was seven. Additionally, the students of Advanced Group, student 3 (S3) and student 4 (S4), achieved level four (L4) understanding in at least one topic. S3 and S4 were the only students to achieve L4 in any category. I would hesitate from saying either student was a L4 overall as both were still developing their understanding of independence and S3 had the additional struggles with randomness. S3 and S4 further separated themselves from the rest of the students by being the only two students to achieve a L3 understanding for independence and L4 for sample space. Thus, S3 and S4 are both categorized as L3 overall.

After the groups were established, commonalities among the students in each group were noted. The students of each group held certain common understandings about the big ideas that came be the distinctive characteristics of the groups. For instance, all the students in the Intermediate group held the ability to build a sample space completely but none of the students of the Beginner group held this ability. Commonalities such of these were noted for each group as well as between groups. This analysis helped to further separate the groups from each other and will play a role in future studies when tiring to place students into groups.

Limitations of Study

The research gathered in this study is limited by the numbers of participants and the items researched. There was an effort to ensure that students were from different courses with different levels of achievement. However, this study will only provide a small picture of a small data set. As a result, the data will not be transferable or generalizable to the high school population at large (Shenton, 2004). Rather the data collected will provide the basis of high school students' understanding for future research. The study itself was only directly studying at one aspect of probability literacy which is a multifaceted realm of understanding and an approach to problem solving even though some aspects of other disposition elements were reported. The dispositions elements such as context were reported as a means to better explain the directly observed knowledge elements: randomness, independence, and sample space.

Trustworthiness

The trustworthiness of the data collection will be addressed via member checks as stated before. Earlier member checks were mentioned as a means to gain additional insight into participants' thought process. But member checks will add value to the depth that the response can be trusted (Brinkmann & Kvale, 2015). The participants were supposed to complete member checks of the transcribed video and the written responses via email (Stenton, 2004). However, no students participated in this aspect of the study due to either students not providing emails addresses or not responding to emails.

Additionally, triangulation was achieved by the use of verbal responses during interviews and written responses after the interview (Stenton, 2004). There was the expectation that there would have been inconsistencies as Rubel (2007) found. However, the differences should have provided insight into student thought while any similarities provided affirmation of understanding on the part of the participants.

Summary

My study collected data in two forms from the participants to provide the greatest level of clarity and insight into high school students' understanding of probability. The five tasks served as media, combined with the probing and follow up questions, to uncover student thinking. The analysis was based on the written and oral responses and deepened the richness of knowledge both from the ways that written explanation or oral responses did and did not agree with each other.

Students' levels of probability literacy in reference to randomness, independence, and sample space was gauged on the Jones et al. (1997) Framework. I interviewed students looking to build a model of what they understand about probability topics. The chief areas of concern were randomness, independences, and sample space. Students' understanding of these topics is foundational to their understanding of all probability. Therefore, if my study finds that students have a high level of probability literacy, then current methodology is providing students with the proper understanding. If my study finds that students have a low level of probability literacy, then hopefully I can gleam models of students' understanding. This knowledge will lead to an improvement in the methods used to help students become more probability literate.

CHAPTER FOUR: RESULTS

The purpose of this study is to understand the probability literacy of high school students. The study seeks to answer the following research question:

> What is the probability literacy of high-school students? In particular, what do high school students understand about the concepts of randomness, independence, and sample space?

This chapter will present the findings from the analysis in response to the research question. The chapter will begin with student demographics, including their prior mathematics coursework. Next, the chapter outlines the overall level of students' understanding in the three key probability concepts – randomness, independence and sample space. The students are classified into three groups – Beginner, Intermediate, and Advanced. Each group is distinguished with common characteristics in their probabilistic understanding. After the overall discussion of the categories, the characteristics of the probability literacy of the students is outlined. In addition to the characteristics, other aspects of probability, like beliefs and habits of mind will be included.

Student Background

There were ten students interviewed for the study – five females and five males (see Table 4). Three students identified as African American, three identified as Hispanic, two identified as White, one identified as Asian and one identified as multiple races. There were four students that were currently taking or had taken Advanced Placement Statistics (AP Stats). The interviews spanned from the end of January 2019 to mid-April 2019. The interviews were conducted in order as they are designated. For instance, the first student interviewed was Student 1 (S1), the second student interviewed was Student

2 (S2), and the remaining eight students following the same process. S1 was a few days into an AP Statistics course and S2 was a few days into a Math 3 course, the third level mathematics course for the local and state school system. The interviews for Student 3 (S3) and Student 4 (S4) occurred at the end of March or roughly halfway through the course. The interviews for Student 5 (S5), Student 6 (S6), Student 7 (S7), and Student 8 (S8) took place at the beginning of April when most of the topics in AP Statistics had been covered. S7 and S8 were interviewed as a pair which provided interesting contrast due to the perceived ability differential, S7 was in Math 2 and S8 was in AP Statistics. All the students, except Student 9 (S9) and S6, had recalled learning probability. Additionally, all students had self-reported a grade of B or higher in the last mathematics class taken. As such, the sample may not be representative of all high school students. All the students for the study were chosen by their teachers and there was a conscious decision to choose students who were open to discussing their mathematical thinking in the interviews.

Table 9: Student Demographics

Student	Grade /Age	Highest Math Class/Grade	Learned Probability	Basic Demographics
S1	12/17	AP Stat/B	Yes	African American Female
S2	10/15	Math 2/A	Yes	White Female
S3	12/18	AP Stat/B	Yes	Hispanic Female
S4	12/15	AP Stat/B	Yes	White Female
S5	12/18	Pre-Calculus/A	Yes	Hispanic Female
S6	11/16	Math 3	NA	African American Male

S7	10/16	Math 2/B	Yes	African American Male
S8	10/16	AP Stat/A	Yes	Asian Male
S9	11/16	Math 2/B	No	Hispanic Male
S10	9/15	Math 1/B	Yes	Multiple Race Male

Groups for Probability Literacy

Recall that the students were classified into three groups based on their score after the analysis of their interviews. The ten students in this study were classified into three groups – Beginner, Intermediate, Advanced. Table 8 provides the classification of the ten students. Overall scores were used to begin the classification of the students and were followed up with further in-depth analysis of the videos. The next sections will discuss the classifications of the students' probability literacy by building on their understanding in each of the three big ideas of randomness, independence, and sample space.

Table 8: Overall Rankings

	Randomness	Independence	Sample Space	Overall	Group
S1	3.000	1.000	3.000	7	Intermediate
S2	1.000	1.000	2.000	4	Beginning
S3	3.000	3.000	4.000	10	Advanced
S4	4.000	3.000	4.000	11	Advanced
S5	2.000	1.000	3.000	6	Intermediate
S6	1.000	1.000	2.000	4	Beginning
S7	3.000	2.000	2.000	7	Intermediate
S8	2.000	1.000	3.000	6	Intermediate
S9	2.000	2.000	1.000	5	Beginning
S10	1.000	2.000	1.000	4	Beginning

Table 2: The Levels of Understanding Matrix (LUM)

Theme	Level 1 (L1)	Level 2 (L2)	Level 3 (L3)	Level 4 (L4)

Sample Space	Students makes incomplete list and does not use list to make decisions regardless of prompting	Adopts a strategy to create list of outcomes but must be prompted to do so but cannot use it to solve problems.	Adopts a strategy to create complete list and uses the complete list to solve problems or to make decision, but must be prompted to use the list or creates an incomplete list but does not need prompting to create it or use it to solve problems.	Adopts a strategy to create complete list and uses list to solve problems or to make decision and student does so unprompted.
Independence	Makes prediction for subsequent outcomes on previous (or future) outcomes rather than current situation: representativeness affects decision making	Student can identify an independent event and can assign numerical probability with or without replacement, or can provide correct justifications for answers, or does not allow representativeness to sway decision or calculations about independent events (one must be met)	Student can identify an independent event and can assign numerical probability with or without replacement, or can provide correct justifications for answers, or does not allow representativeness to sway decision or calculations about independent events (two must be met)	Student can identify an independent event and can assign numerical probability with or without replacement, or can provide correct justifications for answers, and does not allow representativeness to sway decision or calculations about independent events. (all 3 must be met)

Randomness	Makes predictions based on: pattern or haphazard models. Regular disruptions or frequencies are not mentioned. Evidence that student makes decisions about randomness based on either positive or negative recency. Or student allows personal, non-mathematical, beliefs or experiences to sway decisions on a major level.	Partially make predictions based on: patterns, the most recent occurrences, or haphazard models, or can create a mostly irregular distribution and little use of "Qualitative intuitions of proportionality" (Piaget, 1977, pp. 54), some evidence that positive or negative recency influences decisions.	Make predictions mostly based on frequencies or can create distribution that mostly irregular or partially uniform distributions and sometime uses "Qualitative intuitions of proportionality" (Piaget, 1977, pp. 54), little evidence that positive or negative recency influences decisions. Student still does not include "unordered" although they may point out that the event is "unpredictable" in their discussion about randomness.	1) Make predictions based on frequencies or can create regular or uniform distribution for decisions, 2) Can communicate that randomness is an unordered list of outcomes or "individual outcomes are uncertain" (Kaplan, 2014), 3) Makes extensive use of "Qualitative intuitions of proportionality" (Piaget, 1977, pp. 54), 4) No evidence that positive or negative recency influences decisions, and 5) Can explain reasoning for answer(s) or changes to answers.
*stage sets refers to how students deal with compound events such as two dice being rolled at the same time				

Beginner Group

The Beginner group consisted of S2, S6, S9, and Student 10 (S10). The analysis showed that the students' thinking about randomness and independence were intertwined. Note that the analysis was guided by the characteristics isolated from the Levels of Understanding Matrix (LUM). Recall that the LUM was built from the literature presented in Chapter 2. The levels of understanding of each concept or topic was based on how the students demonstrated these characteristics as they explained their thinking in the interviews.

Table 10 outlines the characteristics for each student in the Beginner group. The skills listed in Table Beginners were developed from the LUM. Most of the skills on the LUM came from the supporting literature about randomness, independence, and sample space (Jones, 2007; Kaplan, 2014; NGACBP, CCSCO, 2010; Piaget & Inhelder, 1975). However, some of the items listed in the LUM evolved from an analysis of the data. For instance, students' use of frequencies to answer questions emerged as the interviews moved forward with several students making use of frequencies rather than recency, as was expected from the literature. As stated before, the use of frequencies was classified as a part of randomness, but it also provided insight into the students' understanding of independence.

Table 10 is a listing of the traits from the LUM put in table format to better gauge the understanding of students. One table was completed for each of the three groups as a means to better discuss their perceived understandings of randomness, independence, and sample space. An added benefit of the table was it provided supplement to verify that students were properly placed in their respective groups. The table did not take the place of the LUM, rather the table provided a detailed snapshot of each student from this study.

Table 10: Beginner Group	S2	S6	S9	S10
Randomness				
Positive recency	Yes	No	No	No
Negative recency	Yes	Yes	Yes	Yes
Makes predictions based on patterns	Yes	Yes	Yes	Yes
Provides justification for random events	Yes	No	Yes	No
Communicates randomness as an unordered listing	No	No	No	No
Uses qualitative intuitions to make decisions about randomness	No	No	Yes	Yes
Uses frequencies to make decisions about randomness	Yes	Yes	Yes	Yes
Independence				

Displays representativeness	Yes	Yes	Yes	Yes
Identifies independent events	No	No	No	Yes
Provides justification for independent events	No	No	No	No
Sample Space				
Builds sample space methodically	Yes	No	No	No
Builds complete sample space	No	No	No	No
Makes use of sample space	No	No	No	No
Independence and Randomness				
Allows non-mathematical beliefs to sway decisions about randomness or independence	Yes	Yes	Yes	Yes

In general, the four members of the Beginner group demonstrated basic levels of understanding of randomness and independence with limited understanding of sample space. The group relied on negative recency and the use of patterns in developing their responses about randomness. The students were challenged to communicate random events as unordered outcomes. The Beginner group relied on representativeness when discussing independence and failed to see independent events as separate events. For both randomness and independence, the Beginner group students provided some description of their thinking, but could not expand or defend their thinking without being pressed about their responses. The group also made extensive use of frequencies to make decisions about outcomes when dealing with randomness and independence. The analysis of their responses indicates that the Beginner group is still developing the ability to build complete sample spaces and use sample space to solve problems. Some students in the group demonstrated probabilistic intuitions that were different from prior research, like the use of both positive recency and negative recency. The next sections discuss the details about the students' thinking about randomness, independence and sample space.

Understandings of randomness. The Beginner group was developing their understanding of randomness. In particular, the Beginners demonstrated random outcomes as an unordered list of outcomes, a reliance on patterns, and use of recency when making decisions about random events. Students' inability to communicate random events as unordered was evident in the way students built their flips for the Fake Flip task. Instead of attempting to write an unorganized list, the students used patterns to generate their list. The use of patterns indicates that students were thinking in terms of ordered lists. Thus, detecting patterns in student work is one way a student's ability to think about randomness as unordered can be firmly identified. One of the students in this group, S2, was asked to create three list for the Fake Flip task at different points during the interview. S2 developed three lists with the same pattern each time. The first list and the third list were negative images of each other (see Figure 3).

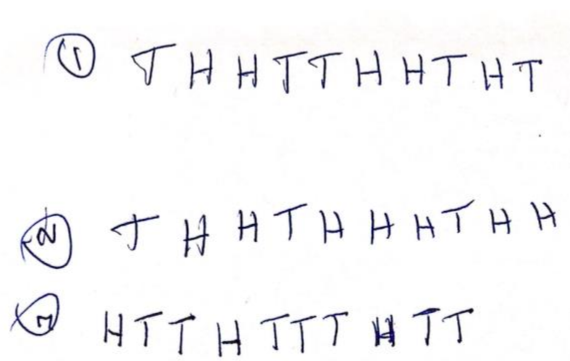

Figure 3: S2 Fake Flip Solution

The lists from Figure 3 was designed by S2 to be somewhat symmetrical in that both list one and list two was made of a repeated pattern of tails, heads, heads, tails. The lists also included an equal number of heads and tails, which was also by design. List 2 in

Figure 3 was created after S2 was pressed about whether it was a requirement to have an equal number of heads and tails. The fact that S2 accommodated the pressing easily implies that her views on randomness are somewhat fluid. S6 appeared to have a similar thought process to S2 in that he created one list initially that was strictly alternating heads and tails. S6 was pressed about the alternating values of heads and tails in the list and why there was no variation in the list. S6 noted that "how it (the list) keeps going back from heads to tails" will fool his friends into believing that the fake flip was real. When asked about changing one tail to a head to create a three head sequence, S6 responded that "they (his friends) would wonder why" the one section was different "because it did not happen before or after that." S6 is eluding to the idea that the list should have some order or even a rule that it follows. S6 was pointing out that the list was following a pattern and deviation from the pattern would cause suspicion. S6 created a second list after the previous discussion (see Figure 4).

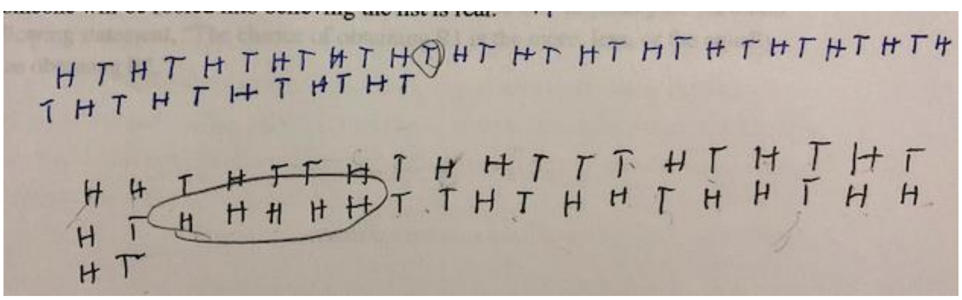

Figure 4: S6 Fake Flip Solution

The second list appeared to be far less pattern generated with a series of five heads appearing. S6 noted his thinking for the second list was that he would be able to "remember" how the coin was flipped every time so the outcome would change. S9 displayed negative recency while completing the Double 6 task as well. S9's description of the task when answering a follow up question revealed that he believed that it was more likely to have an opposite event as the next outcome. When S9 was asked whether 11 or 12 was a more likely sum, S9 stated that the sum of 11 was more likely because it was more likely to get two different numbers rather than the same number twice. Two notes should be made here: 1) S9 had not built the sample space for Double 6, in fact S9 did not ever build the complete sample space for any task presented to him and 2) getting different events on subsequent trials is pure negative recency. Although this is a solid display of S9's understanding of randomness, his understanding of independence is intertwined with randomness, as will be discussed in deeper detail later. S10 created a list that, when carefully studied, is actually a collection of several shorter list each with their own pattern. For instance, S10 completed a series of heads and tails alternating events in five iterations, circled in blue in Figure 5, separating each grouping with either pairs or multiples of both heads and tails, circled in red in Figure 5.

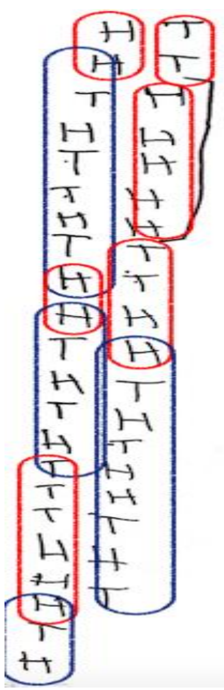

Figure 5: S9 Fake Flip Solution

S9 was the only exception to this on the Fake Flip task as he stated that he was actively trying to avoid anything that looked like a pattern. The list that S9 created reflected this. However, S9 noted that the real list he created by flipping a coin 10 times looked less believable than his fake list because it had two series of tails and one series of heads that made it look "less random", see Figure 6.

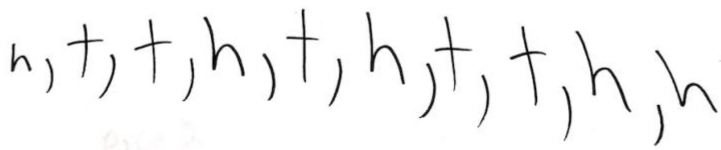

Figure 6: S10 Fake Flip Solution

The implications are that even though S9 tried to actively avoid patterns for the Fake Flip task, he did not completely grasp that random events were unordered. It appeared that S9 was trying to apply a rule to the creation of the list of flips, not that I can speculate on what rule S9 was trying to apply nor do I know for sure that he was using a rule of type. Rather, it appeared that S9 was confining himself to a system as he was observed stopping after each "fake flip" as if S9 was flipping a coin in his head. S9 was asked to explain his thinking of how he created the list. He did not go into the mechanics of the process but his theory. Regardless of the method S9 accessed to create his fake list, S9 appears to believe that randomness has it limits as stated from the comparison to the real list. S9 believed that a perceived pattern that occurs in the list disallows it from being random, he noted as such when mentioned that a series of heads would be "less random" at noted before. A possible alternate perception was that a perceived pattern makes the list look less random and therefore less believable as real.

All the students in the Beginning group displayed negative recency in some form when completing tasks pertaining to randomness. For instances, while completing the

Double 6 tasks S9 mentioned that the 5,6 pairing was more likely than the 6,6 pairing because it was more likely to get a different number on the second roll than the same number on the second roll. S10's responses did not reflect recency while completing the Double 6 tasks but did during the Hot Hand task. S10 noted that as the basketball player makes more shots "he (the shooter) has a lower chance of making the 5th shot...because as he (the shooter) makes more shots, his chances of making the next shot lowers." S9 was explaining that the next outcome should be different from the previous outcome. The fact that S9 believed the percentage of making the next shot goes down as the basketball player makes more shots is further proof that S9 believes the next outcome should be different from the previous or current outcomes. S6 noted that the basketball player will miss the fourth shot because the shooter made the previous three shots. S6 later changed his reasoning on this task. S6 stated that the basketball player will miss the next shot because he is a fifty percent shooter and he is already over 50% with three out of four made shots. S6 departed from recency in this situation for reasoning purposes but not in outcomes. S2 was like S6 and changed her answers slightly when pressed. S2 initially said that the basketball player was more likely to make the fourth shot because the player had made the previous three shots. S2 and I had the following interactions (R – Researcher):

S2: "He will make it...because he has already made it three times..."

R: "why is he more likely to make the next three?"

S2: "...there is always that probability that he will…"

R: "What if he had missed the last three shots?"

S2: "He would miss the next one."

S2 was unique in that she was the only student to display any positive recency. However, once pressed about the fact that the player was a 50% shooter, S2 changed her answer and reasoning. S2 and I then had the following interaction:

> R: "How does the 50% shooting effect the fourth shot?"
>
> S2: "…the next three shots will be misses."
>
> R: "Why did you change your answer?"
>
> S2: "…he is only a 50% shooter. He can only make half of them. Or he only makes half of them from that spot."

On the surface, this interaction implied that the student was using positive recency, or the belief that the same event will happen again because it has already happened (Bryant & Nunes, 2012) to decide the outcome of the next shot. Beyond the recency that appears, both positive and negative, S2 also appeared to be developing her independence understanding as she did not view the events as separate from each other, rather she viewed the events as a collection.

The interchange with S2 was somewhat common with the Beginner group. All the students at some point noted the percentage or ratios of success as being a determining factor in the likelihood of the next outcome. Although recency played a role of the initial decision, deeper probing led to discovering frequency played a greater role and appeared to have a larger influence than recency on decisions about probability as noted with S2. S6 did something similar when pressed about the fourth shot. S6 noted that "he is a 50% shooter…" explaining that the shooter would miss the next three shots. When pressed about the shooter missing four shots, then making three in row, what would the eighth shot be, S6 noted that the basketball player would hit the eighth shot because he has already missed four, so the basketball player needs to make the eighth shot to bring the average back up. S9 was unique in that he started in this spot of the shooter missing the fourth shot because he was already over 50%.

S10 displayed a strong sense of making decisions based on personal, non-mathematical situations. For instances, I explained a corollary situation that S3 created during her explanation of the Hot Hand task. S3 suggested that the basketball task was similar to the birth of four children with the sex of the first three being the same, say boys. Then one predicts the sex of the fourth child. S10 decided that the fourth child must be a girl. S10 stated that was his family's situation and therefore what he believes will happen. S10 does something striking after that. He says that this was not the most likely outcome, a boy was the most likely outcome because there have already been three boys. The interaction seemed confusing at best and only informed me that S10 does not view random events as unordered nor unpredictable.

Understandings of independence. Students understanding of independence among the Beginner group was centered around their use of representativeness. Representativeness means that someone views a small sample of the events as being an accurate representation the all possible events for a given situation. For instance, a series of four flips of a coin has the same appearance or make up of 1000 flips. The use of representativeness can be viewed as either being an indicator of a student's understanding of independences or an inhibitor of students' understanding of independence. During this study, the representativeness was viewed as an indicator of current understanding, the latter was not a point of concern during the interviews. For instance, when completing the Hot Hand task, S2 stated that the basketball player would miss the next three shots because he already made three shots which was changed from her initial answer of the basketball player would make the fourth shot after making three in a row. S2 was asked to explain how she came to this conclusion. S2's initial answer seemed to be an issue with recency; in particular with a randomness issue but her reason for the change appeared to focus on independence. As discussed before, S2 stated that, "...he is only a 50% shooter. He can only make half of them." S2 failed to further explain her reasoning about either the original statement or the adjusted statement just presented. However, based on the adjusted statement alone, S2 suggested that she viewed the sample of 6 shots as needing to be an accurate representation of all the shots the basketball player will take. A similar interaction with S6 was noted before where he said, "…he's only a 50% shooter…" The interaction with S2 and S6 was typical of the beginner group and the students used representativeness to make decisions about outcomes of probability events. All members of the Beginner group appeared, on some level, to view a small sample as representative of the collection of all events.

S9 had a similar interaction with the Hot Hand task. His level of understanding of independence was deemed to be a L2 due to his occasional reliance on representativeness. While working the Hot Hand task, S9 noted that the basketball player would miss the fourth shot after making three in a row. However, when presented with the idea of missing four, then making three, and being asked the result of the eighth shot S9 noted that the basketball player would make the eighth shot. The answer for the first scenario of four total shots appears to be the result of negative recency, which was discussed earlier. The alternative scenario of eight total shots appears to focus on representativeness. This brings into question whether S9 was using negative recency for the first scenario for the Hot Hand task. Perhaps S9 was actually thinking of the four-shot sequence mirroring the entire data set for the basketball shooter while working the Hot Hand task. S9 appeared to believe that the eight-shot sequence that he was asked to consider needed to the have the same characteristics as the entire data set. Each scenario, the four-shot sequence and eight-shot sequence, needed to be half misses and half hits. S9's strategy was to match the frequency of the entire set with the frequency of a subset. Further, it was interesting that S9's reliance on representativeness appeared while S9 was working the Hot Hand task but not during the Black and White task. The Black and White task did not appear to cause S9 any conflict with representativeness. S9 never alluded to the idea that six rolls had to look like any set of events.

The role of skill was a point that was discussed during the Black and White task with S9. He said, "I'm not rolling it differently because I get black or white, but the odds are the same each time" because the color of the sides does not change. S9 was referencing how skill would affect the basketball player but not the die roll. The skill involved in shooting the basketball improved with successive shots over time and there would be fewer missed shots. This idea seemed to dictate S9's thinking about randomness, which contrasts with what was initially found when discussion of the Hot Hand task with S9. S9 displayed negative recency during the Hot Hand task by saying, "when he made it three times in a row, it made it seem less likely that he makes the next one." Here, S9 was responding to the idea of three made shots in a row. Next, S9 was asked to decide the eighth shot after the basketball player missed four shots then made three shots. S9 thought that the eighth shot in the four miss, three hit sequence stated previously seemed more likely to be a hit or "it seemed more realistic." The follow up question about eight shots demonstrates positive recency, causing a conflict about recency. It was possible that S9 was not thinking about recency, but rather was trying to maintain a particular frequency. Frequency was more closely related to a student's understanding of independence than randomness, therefore it might be that S9's understanding of randomness was tied to his understanding of independence.

One noteworthy aspect of representativeness was that at some point all four members of the Beginner group made a statement like "...he can only make half of them…" This one statement pushed all four to have a L2 or L1 rating on the LUM for independence because the statement implies a reliance on representativeness. This suggest that all four are considered a subset of the events rather than on the entire data set of shooting.

Separate events. S9 and S10 achieved a L2 for independence based on their abilities to view independent events as separate events, but only for the Black and White task. In the Hot Hand task and Fake Flip task, S9 and S10 appeared to view the separate events as a collection of events. S2 and S6 did not at any point appear to separate events as individual events, but rather saw them as a collection of events.

While completing the Black and White task, S9 noted that the outcome of one roll did not affect the outcome of the next roll. S9 noted that, "I'm not rolling it differently because I get white or black, but the odds are the same each time" because the sides do not change was his reasoning for why the outcomes are unaffected by each other. S9's reasoning in the Black and White task indicates that he viewed the rolls as separate events although S9's reasoning was not direct nor was it free of interpretation. However, on the Hot Hand task S9 made comments that led one to believe the opposite. For instance, S9 and I had the following interaction during the Hot Hand task when presented with the idea of missing four shots in a row, making three shots and now we are talking about the outcome of the eighth shot:

> R: …it's a little tricky…how's that a little tricky?
>
> S9: Well probability is tricky for me. I'm not good at basketball…it just seems different…when you put it in different scenarios, it just seems to have like…different likelihood…
>
> R: Does making or missing this shot influence making or missing the next shot?
>
> S9: I think it does. In my mind it does. It seems, I am still thinking about that. I am still thinking about how you put it in a different perspective.

S9 went on to discuss how originally, he felt the task was built on the idea of a pattern, but, when presented with an alternate progression, it was more random. The two big ideas that arose in the discussion were 1) S9 did not see individual shots as separate events and 2) the context of the task impacted S9's perception of the independence and randomness.

S10 held similar views to S9 for both the Black and White task and the Hot Hand task in reference to seeing independent events as separate events. Namely, on the Black and White task, S10 stated "the dice decide" or it was "luck". S10 pointed out that all the events are separate from each other, the roll of one die does not affect the roll of another die. However, just like S9, S10 noted that the next shot would be affected by the previous shot for the Hot Hand task. Again, it appears that the context of the problem influenced the level of independence that was perceived by the student. Additionally, it appears that when skill was perceived as a factor in an event, the perceptions of independence and randomness are affected.

S2 and S6 did not see any separation between outcomes on any of the tasks. As a part of the interview, S2 rolled a die with five black sides and one white side six times to mimic a possible answer choice for the Black and White task. S2 was surprised to have more black sides appear than white sides. S2's reasoning for being surprised by this was that "white could appear but does not have to…" Although one could talk about the understanding of randomness that this statement implies, another interpretation was that S2 was looking at the six rolls as an event rather than a collection of trials. S2 was surprised that one roll could be different from the others and therefore was not separate. S2 showed similar thinking on Hot Hand and Fake Flip. During the Fake Flip S2 noted that "it's not possible for it to land on heads or tails five times each." It seems that S2 looked at the five flips as an event rather a collection of trials since a true independent event would not have any connection between trials and 5 heads or tails in a row would be no different than 40 of the same type.

S6 presented similar understandings to S9, S10, and S2 with regards to events being separated in references to the Hot Hand task. However, like S2, S6's thinking was constant for both the Black and White and the Hot Hand tasks. S6 choose D on the Black and White task, which was six black results, as his answer initially. S6 gave little reason for this answer other than noting that black will occur the most. S6 later stated that the black side has a "90% chance of occurring" but since the likelihood was "not 100% so it has to land on white at least once..." S6 knew that black was more likely to occur but believed that since there was a chance for white to occur, then white must occur at least once in the listing of six rolls. S6 appears to be mistaking expected value for likelihood. S6 was asked why the answer was not C which was six white rolls. S6 stated "if you picked it (the die) up and it was white and you dropped it, it's probably not going to be white again." The word "again" implies that he views the next roll, at least on some level, as related to the previous roll. S6's reply implied that he viewed the events as related and that there was some skill involved in the activity, he did this a couple of times on different task. Furthermore, S6 was referencing the skill involved in rolling a die. S6 notes something similar for the Hot Hand task referencing the basketball players ability to consistently recreate the same movement every time while shooting the ball. The skill involved in shooting the basketball was ultimately a factor in S6 not viewing the shots as separate events. S6 mentioned that the basketball player must "remember" how he shot the ball last time to make the next shot. Thus, S6 is linking the next and previous shots together.

Interplay between randomness and independence. The Beginner group appeared to have an association between representativeness and recency. All the students in the group demonstrated parallel levels of understanding and usage of representativeness and recency in reference to independence and randomness. Alternately, it appears that S2 and S6 may have an additional issue with randomness as both students appear to use frequency to make predictions about the next event. The frequency issue referred to here is centered around the student's desire to maintain a given or known percent for an event while using a given subset or representativeness. Thus, it seems that S2 and S6's understanding of independence prevents their understanding of randomness from being truly perceived. S9 and S10 are separated from this discussion, while they both displayed representativeness, S9 and S10 both routinely alluded to past events when making predictions about future events. Thus, an indication of recency. The play between recency and representativeness is being discussed because, with the Beginner group, the two topics of independence and randomness appear to be intertwined at all levels of understanding.

S2 noted that “there was a possibility that he will make it, but there is also a possibility that he will not make it” in reference to the fourth shot taken by the basketball player. The discussion with S2 noted here was discussed earlier, however, it is being revisited to examine the interplay between independence and randomness. S2 discussed multiple aspects of her understanding in a single description. First, S2 notes that the event is random by mentioning that there was a possibility of either event happening. However, S2 also notes that the event of the next shot was connected to the previous shot by mentioning that the basketball player has already made a shot and this outcome influences the next outcome. S2’s understanding of independence is based on her understanding of randomness. She views the event as a random event with some order, a random event with order to it which is insightful as well. But, S2 also noted in the Black and White task that getting all of one color is not a random event. S2’s views six rolls of a die as a single event, not six separate events. S2’s understanding of independence has an influence on her understanding and decisions about random events.

S6, S9, and S10 all had similar episodes where their understanding of randomness was tied to their understanding of independence. S6, for instance, stated on the Hot Hand task that the basketball player would miss the next shot because the shooter made the last three shots. S6 is viewing the four shots as one event rather than four events. Simultaneously, S6 notes that the next shot “…is unpredictable…you don’t know if he is going to be able to make it or not…” Thus, S6 is aware that the events are random events, but his views on the events not being independent are shaping his view about the random event. S6 does not appear to believe that the outcome is completely predicable at this point, but rather that one outcome is more likely thereby making the fourth shot for the basketball player “less random.”

S9 and S10 are mildly separate from S2 and S6 in their understanding of independent events but not in their observed connection between independence and randomness. S9 and S10 both appear to have slightly more developed understandings of randomness and independence, although it was not substantial enough to merit the two pairs of students being placed into different groups. S9, for instance, noted during the Black and White task that the rolls were independent of each other although he still had a reliance on representativeness. S9 viewed the shots of the basketball player as a single event just like S2 and S6. However, on both the Hot Hand and Black and White tasks, S6 noted conflict that he had about the outcome of the shot. S2 and S6 never noted any such conflict. S9 noted “…that makes it seem like it’s a little bit more random” when he had completed rolling the die six times. S6 was referring to the fact that I informed him that he was the only student that got their predicted list of outcomes. S9 viewed the outcomes of the die as independent and random, something S2 and S6 did not relay in their discussion during the interview. S9 noted that the randomness could change based on the situation, a belief held by S6 as well who viewed the Hot Hand task as being non-independent events. S10 also did not view the shots in the Hot Hand task as being independent. However, when completing the Black and White task S10 noted, “I don’t think anyone could predict what it would land on all six times unless you get lucky…it’s like winning the lottery…no one is that good.” S10 is implying that no one can win the lottery twice. Furthermore, S10 is implying that the random event of winning the lottery is less random after you win it once. The "good" statement that S10 makes implies that he views that there is some level of skill involved in winning the lottery, although S10 might have been using lucky and good interchangeability.

Each of the four students of the Beginning Group indirectly noted a connection between independence and randomness. The understanding of randomness can cause a student to view an event as independent or not. Conversely, it appears that not viewing a set of events as independent effects the level to which a student perceives an event as being random with a lack of independence associated with less random events and vice versa.

Sample space. Another common characteristic that emerged in the interviews with the students in the Beginning group was their abilities to use sample space. S2 was the only student from the Beginner group who listed a complete sample space in the tasks. The ultimate objective of including tasks that bring out discussions of sample space in the study was to determine if students could use sample space as a tool to access and gather more information to better deal with a probability task. None of the members of the Beginner group used sample space as a tool. The primary reason might be because they could not build the entire sample space. However, none of the Beginner group students used their incomplete sample spaces to answer or modify solutions or supply justification to previous presented solutions.

All students in the Beginning group, S2, S6, S9, and S10, struggled to build the Fake Flip sample space with any detectable method beyond an alternating series. The set created in Fake Flip is not a true sample space in that it is one trial of forty flips rather than all trials for forty flips. The Beginning group students were mostly unique in that none of them adopted a second method to build another sample set while the other two groups, Intermediate group and Advanced group, did deploy additional methods. The alternating series idea is important to mention as it is the only method detectable from the interviews and it points out that the students can create a sample space, but their methodology is limited and still developing. The ability to create a sample space is important as to there being a deeper issue with building sample space with other tasks presented. For instances, S9 built the sample space represented in Figure 7.

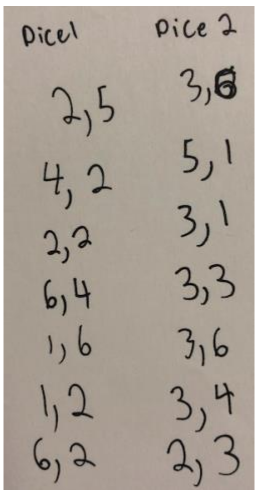

Figure 7: S9 Double 6 Solution

The sample space that S9 built was done so seemingly without any rules or process. When questioned about it, S9 stated that he created the sample space based on how he thought the rolling of two dice would appear. S9 was creating a list of random rolls of dice rather listing of all possible rolls. This was interesting insight into S9's perception of what constitutes a sample space. To have created a complete list using S9's method, he would have to create a list with hundreds of rolls and use actual dice. S9 had some background in what a sample space means, but his application for the context of Double 6 signifies that there is still much development needed for S9's sample space understanding. Additionally, S9's sample space only has 14 elements in it. When questioned about this, he stated that there were more elements in the list, but "I don't know them." The statement further reiterates that S9 does not have a solid or transferable methodology to build a sample space. S9 did not use the sample space to modify or further back up his explanation for the Double 6 task. It might not be surprising given that the list S9 created was incomplete and a list based on his guess of the outcome of the dice rolls. Additionally, S9's list contains the pairing 3,6 twice, in that order claiming that it was an "accident." S9 goes on to say that 3,6 and 6,3 are different pairings. Thus, S9 created a list with one pairing occurring twice. Any conclusions that S9 could have made from the sample space would be misguided, not only due to the incomplete list, but also due to the fact that pairs reoccur in the list. S10 created an incomplete list for Double 6 as well. It was unordered and had values that did not belong such as the pairings 10,1; 9,3; and 11,1 (see Figure 8).

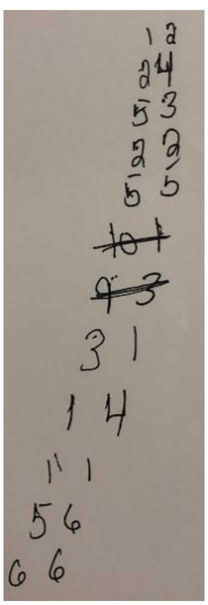

Figure 8: S10 Double 6 Solution

S10 created the pairings with 9, 10, and 11 in spite of being handed two six-sided dice to demonstrate the nature of the problem. When asked about the 10 and the 9, S10 noted that he thought the dice were 10-sided. S10 can understand what makes something a part of the sample space but as a rule cannot or will not adhere to the rules of the system. S10 ask if he can put the pairings from the task in the list: 5,6 and 6,6. S10's question about those two pairings implies that he is still developing his understanding of what belongs in a sample space. S10 did not use the sample space to modify or validate any of his previous answers possibly meaning that he did not fully understand the purpose of a sample space.

All four members of the Beginner group completed the Fake Flip task with all four displaying similar behaviors in the development of the list. S9 and S10 were the only two that completed the Double 6 task so the comparison between the two task is only applicable to them. One major difference for S9 and S10 was that both students relied on the use of patterns to create their list of fake flips for the Fake Flip task. The most interesting aspect of this behavior is that neither deployed any such behavior for the development of the Double 6 sample space. The use of patterns would have been useful in the Double 6 sample space, but is actually counter to what one might expect for the flipping of a coin 40 times. This is not to say that a list of coin flips cannot result in a list that appears to have a pattern, only that the students used this idea almost exclusively to build a list that should be random. None of the Beginner students built a list for the Double 6 task using any patterns, but all of them did for the Fake Flip task. The idea that the students engage patterns is an indicator that the capacity to build a complete sample space is possible. However, the fact that students did not deploy the use of patterns in the creation of the Fake Flip task indicates that students are struggling with either the context of the Fake Flip task or are still developing the ability to apply and transfer strategies into new scenarios. The transfer may not be taking place since the methods being deployed by the students is not robust enough to produce entire sample spaces. S2 for instance created two separate listing of fake flips (list 1 and list 3) that were negative images of each other, see Figure 3.

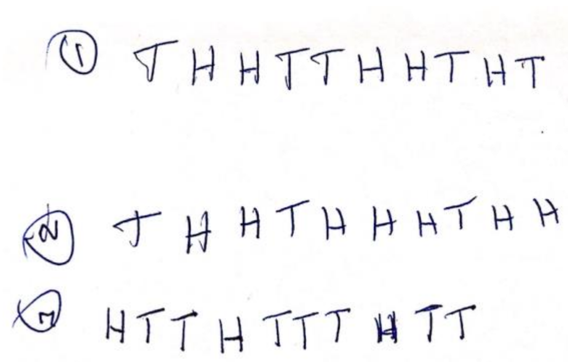

Figure 3: S2 Fake Flip Solution

S2 might have been thinking systematically, but she is not using a system that will yield all possible list of flips. S6 deployed a similar idea, see Figure 4 where his first listing was a list of alternation between heads and tails. S6 did not complete the Double 6

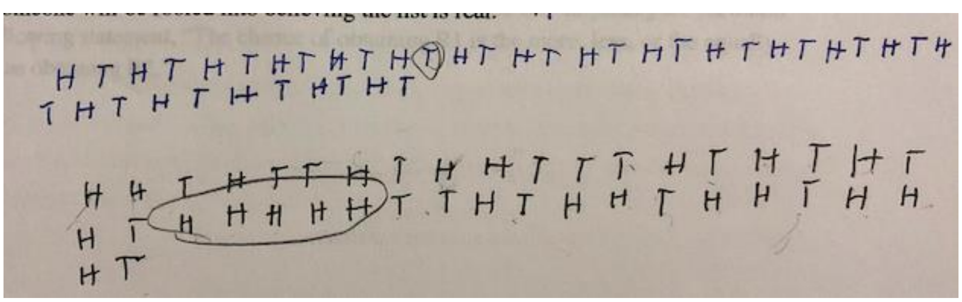

Figure 4: S6 Fake Flip Solution

task, thus a comparison cannot be drawn to for insight to S6's approach to build a sample space.

One of the defining differences between the Beginner group and the Intermediate group was the ability to build a sample space completely and systematically. This was ultimately the reason why S9 was placed in the Beginner group rather than the Intermediate group.

The Intermediate Group

The Intermediate group consisted of S1, S5, S7, and S8. Like the students in the Beginner group, the Intermediate group's understanding about randomness was often influenced by the students' understanding of independence and the reverse was true as well. The connection appears to exist to a less noticeable degree and the abilities among independence and randomness appear to be more distinct in that the students appear to be able to separate the abilities with greater ease.

Table 11 outlines the characteristics for each student in the Intermediate group. The skills listed in Table 11 are parallel to the skills listed in Table 10 for the Beginners Group. The analysis created a list of characteristics for the Intermediate group isolated from the LUM. The levels of understanding of each concept or topic was based on my interpretation of their responses as they explained their thinking in the interviews.

Table 11: Intermediate Group	S1	S5	S7	S8
Randomness				
Positive recency	No	No	No	No
Negative recency	No	No	No	No
Makes predictions based on patterns or haphazard models	No	Yes	Yes	Yes
Provides justification for random events	Yes	Yes	Yes	Yes

Communicates randomness as an unordered listing	Yes	Yes	No	No
Uses qualitative intuitions to make decisions about randomness	Yes	Yes	Yes	Yes
Uses frequencies to make decisions about randomness	Yes	Yes	Yes	No
Believes that skill effects random events	No	No	Yes	No
Independence				
Displays representativeness	Yes	Yes	Yes	Yes
Identifies independent events	No	No	No	Yes
Provides justification for independent events	No	Yes	Yes	Yes
Sample Space				
Builds sample space methodically	Yes	Yes	Yes	Yes
Builds complete sample space	Yes	Yes	No	Yes
Makes use of sample space	No	Yes	Yes	Yes
Independence and Randomness				
Allows non-mathematical beliefs to sway decisions about randomness or independence	Yes	Yes	Yes	No

Generally, the Intermediate group demonstrated elevated levels of understanding of randomness and independence with significant understanding of sample space compared to the Beginner group. The group was characterized by not using recency, but they did use patterns to make predictions about randomness. Two students from the Intermediate group were able to consistently communicate randomness as an unordered list and the other two were able to convey the idea, however, it was not direct or consistent in the interview. Only one of the four Intermediate students held a reliance on representativeness, although her reliance was mild. Only one student from the

Intermediate group was able to identify independent events as separate events. A departure from the abilities of the Beginner group, the Intermediate group students possessed the ability to justify their answers related to randomness and independence. Additionally, the students had the ability to build and use sample spaces. The sample spaces were not always correctly constructed, or were incomplete, and sometimes the sample spaces were misinterpreted by the students. However, the misuse of the sample space is secondary to the fact that the students used the sample space to make predictions or decisions about the events.

Understanding of Randomness. The Intermediate group had a more robust understanding of randomness. The Intermediate student S1 displayed possible positive recency as evident by her deciding that the basketball shooter would make the next shot because he made the last one. S1 grappled with the result of the fourth shot for some time. It appeared that she wanted to say the shooter would make the next shot, but she decided that he would possibly miss the fourth shot to maintain the percentage. S1 was referring to the fact that the basketball player was a 50% shooter and was already over the 50% mark with the third shot. S1 went on to state, "I don't know how many shots he is taking." S1’s statement about the number of shots the basketball player will, or has taken, has multiple implications. For now, I will focus on the randomness aspect of the statement. With this statement, S1 is relying on the frequency of made shots in order to make a decision about the outcome of future shots. S1 felt the need to engineer the result to match the shooting percentage of 50%. Thus, if the basketball player made the first three shots then the basketball player would miss the next three to maintain the percentage. The idea that S1 feels the need to build the outcomes in such a manner allows for a fairly solid assessment of her understanding about randomness. Namely, S1 is forsaking the possibility of random events for the promise of maintaining the 50% shooting average. S1 went on to explain, "…he (the basketball player) would make four (shots) and miss four…" when S1 was given the parallel event of the basketball player taking eight shots. Therefore, S1 demonstrates a desire, either consciously or unconsciously, to maintain the ratio of the event on subsets of the events which is an independence issue and S1 is demonstrating representativeness. However, S1 was able to correctly identify the event as random and define randomness. S1 does not necessarily specify a pattern of the afore mentioned hits and misses of the eight-shot sequence but rather simply states that half of each would be hits and misses. S1 allowed for the possibility that the shots were a mixture of hits and misses in an unspecified order. The idea of randomness is there, but S1 is limiting the degree the event can be random. Thus, S1 was marked as an L2 for randomness.

The concepts of independence and randomness appear in S5's explanations. S5's understanding of independence is demonstrated when S5 explained how the three previous shots the basketball player took impacted the next shot that was taken. S5 noted this by stating that the shooter was less likely to make the next shot because the basketball player had made three in a row. S5 pointed out that the basketball player had already made three out of four shots which was already over the 50% shooting average the player had from a location. S5 noted that the basketball player should have only made two out of four, but the three out of four is over the basketball player's average. Therefore, to maintain the average, the basketball player was more likely to miss the next shot to help bring the average back to 50%. S5 was directly connecting the three previous shots with the possible outcome of the next shot. S5 was making predictions for future events based on previous events. Additionally, S5 stated, “he is only supposed to make half the shots…” implying S5 was referencing representativeness in her discussion. S5 believed that the small subset of shots should look like the complete set of shots.

S5’s also thought in terms of patterns, viewing the three shots as a cluster, and thus her randomness understanding could be examined through the Hot Hand task. S5 displayed recency when she was asked the following, “what if the basketball player has taken ten shots so far in the game and these are the first three he has hit?” The probing question was designed to help determine if S5 was accessing recency or was truly trying to force the percentage to be 50% for the shooter. S5 replied with, “he is still more likely to miss the next one…” S5 appeared to be using negative recency in that S5 believed the fourth shot will be different from the three previous shots. However, S5 talked about the three hits and the unknown fourth shot as being a subset of the overall trend. S5 is grappling with randomness, but S5’s understanding of independence may have been overshadowing her usage of randomness. The follow up questions to noticing her viewing the four shots as a subset yielded S5 stating that if the basketball player had toggled between hits and misses, with the last shot being a miss, then the next shot would be a hit. Thus, reiterating that negative recency was having a strong influence on her thinking. The internal play that was apparent between representativeness and recency came to a head when I asked her to consider the following sequence of shots: miss, miss, miss, miss, hit, hit, hit. S5 thought for some time about this question, beginning to speak multiple times only to stop mid-word. It appeared that S5 was trying to decide if it would be a miss because of the three previous shots or a hit because of the first four misses. S5 ultimately decided that the eighth shot would be a hit to maintain the percentage or "…to make the 50% shooter range to still fit." S5 went on to state that whether the basketball player hits or misses, the previous shot will have an effect on the next shot. Either way, S5 scored a L1 for both randomness and independence as a result.

In the Hot Hand task, S5 noted that the basketball player was a "professional" and thus there was an element of skill involved. S5 believed that the skill of the player could affect the outcome for the Hot Hand task. S5 noted that the event was not random because the basketball player was a "professional." If the basketball player were a high school player his shooting would be random, but the skill level involved with the professional player prevents the event from being random. S5 held this belief despite the 50% shooting. The following discussion arose from our conversation which lead to the conclusion that she was not accessing recency but rather was trying to maintain a certain frequency on a small subset of data:

> R: …he (the basketball player) has taken ten shots in game so far and those are the first three he hit.
>
> S5: I would still say less likely because they (the made shots) are three straight in a row… when he's there he will shoot some of them (make the shot) and won't shoot some of them (miss the shot). He has already shot three of them straight in a row. He is less likely to shoot the next one because when he is doing it, its …um…it's like he is making three but he is only supposed to make 50% and it's very unlikely for him to make five straight in a row and then miss the next five.
>
> R: …what if he had taken six shots, made the first, missed the next, made the next, kinda teetered back and forth. So, now we are to the point where he had just

missed the last one. Is he (the basketball player) going to make the next one (shot)?

S5: Okay…I see… (long pause) … so he made six…

R: …yeah, he got a hit, then a miss, then a hit, then a miss

S5: I am just writing it out…(long pause while S5 writes the pattern down on paper)…so he would be able to hit the other one. Cause, the reason I am saying he would miss is the wording it says, “his last three shots”. So that means he has made three consecutive ones… but this one (the alternating hit misses) is showcasing his 50% more…

The dialog emphasized the frequency thinking that S5 was using. S5 was trying to make the series of shots fit into the known quantity of making 50% of the shots. S5 was displaying no recency in this interchange but her views of randomness in this situation are being ruled by her views that the shots are not separate events, which will be discussed in more detail later.

S5 was presented with the idea of flipping a coin and achieving three heads in a row as a parallel discussion to the Hot Hand task. S5 stated that the most likely next outcome would be tails initially but later started to say that we do not know the outcome. Then, S5 changed her mind again and reaffirmed that the outcome would be tails.

R: If I get three heads in a row, what is the most likely next outcome, if there is one?

S5: That would be tails.

R: Why?

S5: …(student starts to writing down something to help her think)…Yeah I would say it would be…so you said the last three where heads right?

R: What would happen next? Or can you even predict what would happen next?

S5: No, it's just like (long pause) I kinda use the same logic for this one so it would be tails because of the consecutively…

R: Because we have too many heads in a row?

S5: Mmmhm (head nodded yes).

S5 ultimately went with tails because of the three in a row because only 50% can be heads. S5 did not appear to be accessing recency. Rather, S5 was trying to make the subset event match the appearance of a set of many trials where the actual probability was approaching the theoretical outcome. Therefore, S5's understanding of independence was guiding her thinking and decisions about probability events more than randomness.

S5 stated that a coin flip would be random but a professional basketball player with a 50% chance of making the shot was not random, the idea of skill involved in the activity, in spite of the associated percentage, might have been underscoring S5's thoughts. S5 was asked what type of affect would a highly skilled player, such as LeBron James, have on the outcome. An increase in the percentage was implied with the highly skilled player but not directly stated. S5 noted that the outcome would be the same for a highly skilled player as it would be for any other professional basketball player.

S7 shared S5's view that skill will affect a random event outcome. S7 states that the shooter has a higher percentage of making the next shot because he made the last three. S7 stated that the shooter's 50% shooting is based on skill as noted in the following dialog:

S7: He shot 100 shoots, that got him his 50%. But now the shooter has made three in a row, that brings his percentage up to 53%. So, he has a higher percentage.

R: The shooter is more likely to make the 4th shot for no other reason than he has made three in a row?

S7: It stacks the statistics in his favor.

S7 displayed some evidence of positive recency but was justifying it with the belief that the frequency (or probability of the event) was changing with time. The interaction provided insight to possibly discount recency, but to have done so completely at that point would be inappropriate. S7 appeared to make decisions about outcomes based on frequency rather than recency. S7 appeared to be alluding to the idea that the skill level of the basketball player was increasing his skill level and this would have affected the likelihood of making the next shot. S7 provided insight at this point that recency was not a consideration of his thought process rather he was relying on frequency. S7 did not discuss the possibility that the shots could appear in any order of hits and misses or randomly, however S7 also did not discount the possibility that the shots could have been in any order. The collective reasons above led to S7 being assigned a L3 for randomness.

S8 appeared to see the basketball player's shooting ability as separate from randomness in the Hot Hand task. S8 stated that all four shots were "independent" of each other. Thus, "…it shouldn't affect his chance of making the next (shot)." S8 did not make a prediction for the fourth shot, he only stated that the next shot was unaffected by any previous outcomes. A major insight about S8's understanding of randomness could be gathered from S8's lack of prediction about the basketball player's next shot. S8 indirectly communicated that the basketball shot was not predictable S8 appears to have a strong understanding of randomness based on his response to the Hot Hand task.

S7 responded to S8's explanation of the Hot Hand task by spending some time explaining kill-death (KD) ratios for first person shooter video games. S7 used the KD ratios to justify why he felt the fourth shot would be a hit to S8. S7 noted that in a video game, that if a player went on a streak of several "kills" without "dying" it would have increased one's KD ratio because "your skill increases." The situation S7 was referencing was based on the player learning where the attacks on the video game would come from or learning the procedures and other skills that would alter the outcome of particular events in the video game. S7 was building a parallel to the Hot Hand task with his video game discussion with the following argument:

> S7: …say he died 100 and got 50 kills, but one time he got a three-player kill streak and killed three people in a row instead of dying immediately. So that means he has a point five three KD ratio. Which means he (the basketball player) is getting "better" in some sense. So, I think it will increase his chances. So, he won't keep dying immediately, cause there is more of a chance of him getting a kill and then dying immediately. So, I think there is some correlation between all of the shots.
>
> R: What about you (S8)? You disagree with this?

S7 believed that the basketball player, just like video game player, would have increased in skill which would have changed the shooting percentage for the basketball player. S8 refuted with:

S8: If he (the basketball player) has a shot ratio, it should stay the same no matter what.

R: What about the KD ratio thing?

S8: I don't understand that.

S7 then precedes to explain KD ratios to S8. S7 does by introducing an example of two players playing a game such as *Halo* or *Call of Duty*.

S8: But in the long run, if you do that forever, it's eventually going to get closer to the actual…original ratio.

S7: No! I disagree. You get better at whatever you practice at. This guy is obviously practicing, to make three shots in a row. He (the basketball player) is getting better.

R: We are looking at small subset…we are looking at a four-shot event. Are we going to see a skill increase in a four-shot event?

S7: I should say so.

R: What you think? (to S8)

S8: (slightly shakes his head no)

The disagreement was about whether the skill level involved in the activity would have changed the randomness of the outcomes. S7 argued that the frequency of the events would change as the skill level of the basketball player increased. S8 suggested that the frequency was already predetermined and successive attempts would only reinforce the established ratio of 50%. S8's discussion was centered around the result of many trials without any reference of skill building. S7's thinking was the skill would increase and that would change the outcomes. It is interesting to see that S7 uses a frequentist probability approach where the experimental probability approaches the theoretical probability.

Frequency and randomness. The discussions with all three groups of students revealed some level of frequentist thinking that might have interacted with students understanding of randomness. S7, for instance, noted that frequency of recent events would change the skill involved in an activity and influence the subsequent outcomes when he discussed the Hot Hand task and the parallel of the video game kill-death ratio. S8 noted that the long term, combined frequency of long-term trials would yield a percentage that would be in line with theoretical probabilities while discussing the Hot Hand task. S5 noted that tails would be the next outcome on the Fake Flip task, thereby bringing the percentage back to 50%. S1 had trouble making a prediction since the total number of shots that the basketball player had taken, or still had to take, was unknown. The concern for S1 was that the total number of shots was needed information in order to determine the outcome of next shot to ensure that the overall percentage for all shots was maintained at 50%. All three students were highly concerned about maintaining the frequency of 50% shooting or an outcome of half heads and half tails. The idea of frequency reappeared several times with different tasks with different students. The manner in which the students in the Intermediate group discussed outcomes in terms of frequency did not match with any recency thinking. However, the reason that representativeness is not immediately clear in these instances was due to the interaction between the randomness and frequentist thinking of the students. Students were not necessarily trying to match the global average with the subset of outcomes. However, they were looking locally at the current set of trials and trying to match a given frequency to the subset. Representativeness was observed with the Intermediate group and will be discussed in the next section. Overall, the students for the intermediate group had a reliance on frequency when making decisions about random events.

Understanding of Independence. All four students, S1, S5, S7, and S8 demonstrated some level of reliance on representativeness. S7 scored a L2 for independence on the LUM while the other members of the group all scored a L1. Collectively, the Intermediate group scored lower on independence than the Beginner group but higher in sample space and randomness. The first reaction to this realization is that the students in the Intermediate group might have been misplaced. However, after carefully analyzing the data, it appears that a deeper understanding of randomness and sample space might have impacted the students' understandings of independence. A drop in understanding of one knowledge element was not shocking, but is quite interesting. The drop implies the understanding of independence and randomness are separate in some way. Later discussion will show some relationship between independence and randomness with the Advanced group. The Intermediate group's understanding of sample space may have strongly encouraged the students' views on representativeness. The Intermediate group, unlike the Beginner group, accessed sample space both for justifications and processing the task. It is possible that the students' efforts to use sample space to justify answers might have caused some perceived lower score on the LUM for independence. Students' dominant use of representativeness combined with their thinking about the sample space came to the fore and independence was less frequently called upon to separate trials. The sample space was not an inhibitor of independence understanding but rather independence is a factor to consider when dealing with sample space.

S1's understanding of independence was influenced by the use of representativeness. S1 used representativeness in such a way that it overshadowed some of her other understandings such as her tendencies towards recency. S1 stated, "I don't know how many shots he is taking." This statement moves away from the idea that S1 is displaying either positive or negative recency and moves towards making predictions based on the idea that a small subsample of the data should maintain the behavior of the complete data set. This is representativeness. S1 was thinking that the small subset of shots would be 50% misses and hits just like the much larger data set of the basketball players cumulative shooting statistics. Representativeness, based on literature, focuses on a student's understanding of independence rather than randomness. In the current context of the Hot Hand task, S1 appeared to display a need to maintain a particular frequency. Representativeness as evident by S1 noting her need to know how many shots were taken as was seen in the interview where she was told to assume eight shots:

> R: Can you make a listing of eight shots? What would that listing look like if he (basketball player) took eight shots?
>
> S1: He (the basketball player) makes four and misses four.
>
> R: And we already know that these are the ones he's got, right? (see Figure 16), (underlining third made on List 2 of the student's work). Why did you go made, miss, miss, miss?

S1: Because he has…he only shoots…he hits half the shots he makes. So, half of eight is four, he will make four (counts out the mades) one, two, three, four. Oh, I missed one, (counts out List 2), one, two, three, four, five, six, seven…miss.

R: Okay.

S1: So, he will make four and miss four…but you really don't know if it be like… it could be made, miss, and made again and miss, and then made and then miss (see List 3 of Figure 9). So, I don't really know…either way he is still making half of them.

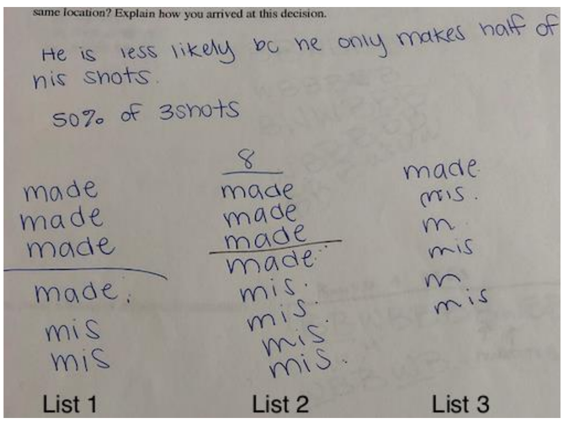

Figure 9: S1 Hot Hand Solution

S1 displays reliance on representativeness in this interchange by maintaining a 50% shooting on List 2 and List 3 (see Figure 9). S1 is not viewing each shot as a separate event. In our interaction, S1 reveals that she believes that eight shots are enough to representative the entire collection of shots taken by the basketball player.

S5 noted that the shooter is less likely to make the next shot because they have made three in a row. On the surface it seems that the response is guided by recency, S5 went on to explain that the 50% shooting average had to be maintained for the shooter. S5 noted that three of the four shots were made, so the fourth shot would be a miss. S5 was referring to the likelihood that the basketball player should only make two out of four, but is already over this average with three out of four. Therefore, to maintain the average, the basketball player is more likely to miss the next shot. Just like S1, S5 is operating under the belief that four shots should have the same appearance of the collection of shots over a long time period.

The Fake Flip task yielded different results for S1 and S5. S5 pushed back on representativeness on the Fake Flip task. She believed that streaks of heads or tails were possible, however, only mild imbalances were possible between the number of heads and tails. S5 was indirectly stating that a forty-flip sequence did not have to have the same appearance as 1000 flips in that there might be slight imbalances in forty flips that would not exist for 1000. S5 had the expectation that 1000 flips would be a fairly even distribution of heads and tails, if not a perfect distribution. S1, in contrast to S5, noted that there should be an equal number of heads and tails without accounting for the number of flips in the listing. S1 added that she believed that equal numbers of heads and tails is the most believable outcome because it is the most likely outcome. The idea that a half and half split was only justified by the belief that it is “most likely” with no additional support given. S1 created five lists as seen in Figure 10. The five list are listed in the order that S1 believed them to be the most likely starting with L1.

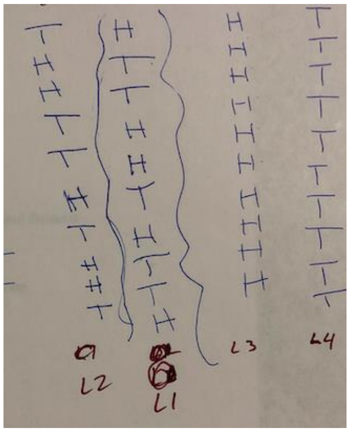

Figure 10: S1 Fake Flip Solution

However, S1 created the list in the following order, L1, L3, L4, and L2 and toggled between L1 and L2 as the most believable noting that L3 and L4 were not believable and "probably does not happen." The three interesting aspects of her list were: 1) L1 and L2 are negative images of each other, 2) S1 ended up choosing L1 because it started with heads, and 3) L1 and L2 are perfectly balanced between heads and tails. The third item here, the balance between heads and tails, was a demonstration of S1's reliance on representativeness. S1 was not asked to flip a coin for comparison.

S5 created the list in Figure 11. S5 was then asked to flip a fair coin ten times for comparison to her created list.

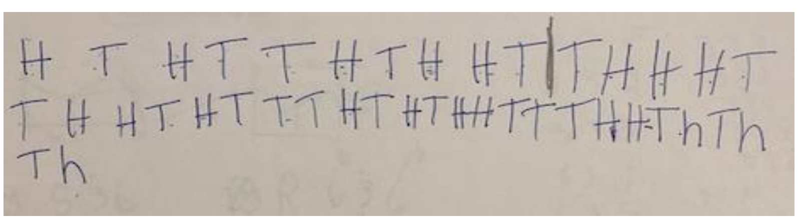

Figure 11: S5 Fake Flip Solution Part 1

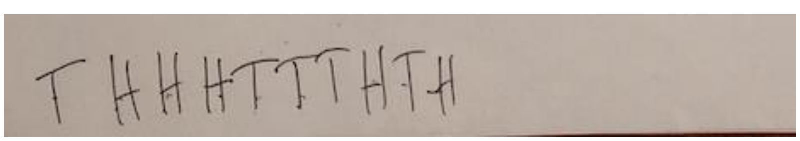

Figure 12: S5 Fake Flip Solution Part 2

S5 believed that her fake flips list had more of a pattern than the actual flips, hence the fake list was less believable. It was not clear what she meant by “pattern”. I interpreted it to be an even distribution of heads and tails. S5 used the reasoning to pick the actual flips as the list her friends would believe as the real list of flips. S5 justified this with noting that it was such a short list (of ten flips) that one should see less of a pattern. The pattern that would not be seen within the first ten flips would be an equal number of heads and tails. S5’s fake flips list had an equal number of heads and tails but the actual flips were unequal. It was interesting that S5 was easily swayed by a single trial. Overall, in terms of independence, she believed that small sets of events did not need to appear like the larger set. S5 appeared to have a reliance on representativeness that affected her decision making on some level.

The discussion between S7 and S8 for the Hot Hand task led to little insight into their understanding of independence, so the discussion about S7 and S8 will mainly focus on the Fake Flip task. One aspect that came to the fore during the Hot Hand task for S7 was the belief that the probability to make the next shot was always in flux. The result for representativeness is that S7 did not demonstrate any for the Hot Hand task because of his perception of the task. S7’s belief that skill will alter the next shot makes the discussion about representativeness irrelevant as he views the data set as continuously growing. S8 believed the events were completely independent in that they were separate events. Thus, S8 was not trying to match the four-shot sequences to the overall data set rather each shot was a new event. The context of the Hot Hand task appears to contribute to both students’ ignoring the comparison of the four-shot sequence to the cumulative data set. S7 and S8’s perception about the Hot Hand task does not discount their ability, rather it provides greater insight into their understanding of independence in that they demonstrated conflicting views between each other and between the Hot Hand task and the Fake Flip task.

S8 demonstrated representativeness while completing the Fake Flip task by creating a list of fake flips that had exactly half heads and half tails through intentional design (see Figure 13).

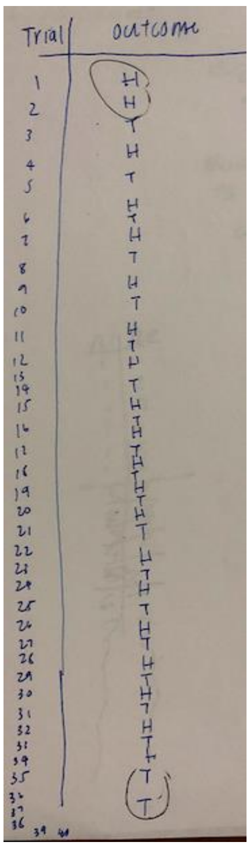

Figure 13: S8 Fake Flip Solution

S8 noted that since the coin has a 50% chance of landing on heads or tails, the list of flips needs to reflect that outcome. When pressed about this outcome, S8 notes that the list of 40 flips does not have to be 20 heads and 20 tails, but it can't deviate too much from the 50% mark. S8 seemed to believe that 27 to 13 would be the highest the ratio could go for a small sample set like 40 flips. S8 continues that 54 to 26 is "not legit" as there are too many of one type. It was interesting that S8 was okay with the same ratio when the set was smaller but not as the set grows. S8 appears to be thinking in terms of frequency. Thus, the frequency would approach the ideal as the trials go on. The discussion was not centered around whether S8 was wrong in thinking that the larger the number of trials yields something more closely resembling 50% heads and 50% tails, only that the larger set has to self-balance. S8 appeared to believe that 80 flips were a large enough data set to see the "correct" ratio of 50% of heads and tails. Conversely, S7 did not seem to believe that the size of the set mattered. Furthermore, S7 appeared to believe that a half and half split would be unlikely to the point of not possible at all, seemly at any data set size. S7 backed this up by noting that S8's list looked unbelievable because of the lack of variation. S7's argument was mostly centered around S8's list had "perfect" alternations of heads and tails and the set was "too" balanced between heads and tails. S7's list in contrast listed more outcomes of heads (see Figure 14).

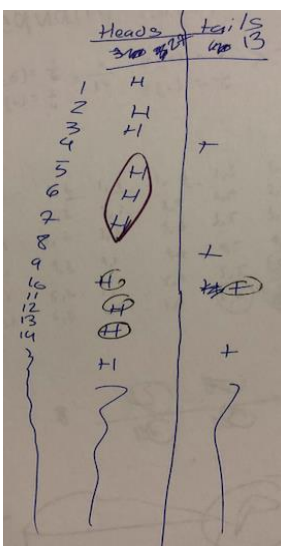

Figure 14: S7 Fake Flip Solution

Figure 14 shows that S7 built a list that had ten heads and three tails. Additionally, almost all the heads came in groups of three. Then, S7 stopped building the list at the 14th flip noting that the list would continue according to a similar pattern. S7 was critical of S8's pattern of heads to tails throughout his list but S7 appeared to be unaware of the pattern that he created in his own list. The pattern that S7 created indicated that S7 relied on representativeness. S7's reliance on representativeness was more evident in S7's response to S8. S7 believed that a small subset of flips could not look like a sequence of alternating heads and tails because a large set of flips would not have an alternating appearance. This indicates that S7 viewed a random event, like flipping a coin, as not having an alternating sequence. The implication for representativeness is that S7 views an alternating sequence not possible for the subset because it was not possible for the larger set. A larger set of flips was referring to a collection of flips that would reflect a 50% for heads and tails. S7 implied that the larger set of flips would need to look more like S7's set that was both head heavy and had multiple heads in a row. Therefore, S7 was using representativeness when he decided that S8's fake list was not believable.

Separate events. A student's ability to identify separate events is both related to representativeness and a stand-alone characteristic. Students of the Intermediate group all displayed different levels of understanding separate events implying that some had slightly more developed understandings of independence. Thus, the need to consider separate events set apart from representativeness is required. S1, S5, and S7 did not consistently identify independent events while completing the Hot Hand, Black and White, Double 6, and Fake Flip tasks. S8 was the exception in the group with respect to separate events in that he noted that the basketball players shots, in the Hot Hand task, were unaffected by each other. However, S8 appeared to view the coin flips of the Fake

Flip task differently from the basketball players shots. All the students' responses to the Hot Hand task will be discussed here in terms of separate events. Additional insight from the responses of other tasks will be provided to either provide contrast to perceived understanding of students or reinforce it.

S1 was asked what would happen if a basketball player took eight shots, missing four then making three and the eighth shot was in question. The purpose of asking this extension question was to determine if S1 would change her answer from the previous question based on only three made shots to which S1 responded that the fourth shot would be a miss. S1 decided that the eighth shot would be made in the extension question. S1 maintained a desire, either conscious or unconscious, to maintain a 50% ratio of hits and misses on subsets of basketball shots. Representativeness was being displayed, but more importantly S1 was not separating the shots into separate events on the Hot Hand task. Rather S1 was considering all four shots or all eight shots as a single event. S1 viewed the previous three shots as affecting the fourth shot. S1's thinking was evident when considering her responses, "I don't know how many shots he is taking" and "…he would make four and miss four...". The statements point to how S1 does not view each shot as a separate event.

S1 displayed some ability to identify separate events while completing the Black and White task. S1 stated, "you could get white every time you roll or you could get black every time you roll." There was possibly more information in this statement than just separate events but that was one item indirectly discussed here. S1 was noting that each roll is a separate event in that each outcome was unknown as reinforced by the statement, "they are likely because you honestly don't know what you are going to roll…I would choose all blacks because there are more than whites.” The admission that white could reappear endlessly was a good indicator that S1 can identify separate events.

The context of the Hot Hand task was altering S1’s view of the independence of the event(s) whereas the context of the Black and White task was not. There was the possibility that the skill factor involved in basketball was leading her to believe that the events are not separate. S1 noted that coins are “funky” when completing the Fake Flip task, suggesting that events involving coins are subject to outside influences such as how the coin was flipped. The Black and White task saw S1 mention that the die was an uncontrollable item and thus subject to outside influences.

S5 held similar views to S1 about separate events while completing the Hot Hand task. S5 also noted that making the previous three shots would affect the outcome of the fourth shot. S5 was basing this on the fact that the basketball player was a 50% shooter. Thus, the basketball player should have only made two of the four shots, but he had three of the four already. Therefore, the basketball player had to miss the fourth shot to help bring the percentage back to 50%. S5 was combining the events into one collective outcome that has the overall average of 50%.

S7 was different from S1 and S5 in that, although there was some representativeness displayed while working the Double 6 task, there was no representativeness displayed while working the Hot Hand task. S7 viewed skill as being involved in basketball players’ averages and thus influenced the outcome of the event. S7 stated that the shooter has a higher percentage of making the fourth shot because he made the last three. Although S7 was displaying representativeness here, the concern during the discussion was separation of events. S7 noted that the basketball player's 50% shooting average was based on previous attempts. S7 explained this as follows:

> "Say he shot 100 shots, that got him his 50%" but now the basketball player made three shots in a row, that "brings his percentage up to 53%. So, he (basketball player) has a higher percentage."

S7 stated the skill level involved in basketball can be developed thereby negating the historical data. However, the implication here was that S7 believed that the events are related, and one event influences the next event. Therefore, from any perspective S7 did not view the Hot Hand task as being made up of separate events but as one single event. The context of the problem might have played a role in how S7 viewed the situation, possibly even encouraged the idea that the events were related.

S8's response to the Hot Hand task was different from the other three students. He stated, "there is an equal probability so whether he makes the shot or not it is independent." S8 was asked to explain what he meant by independent to which S8 replied with, "if he (basketball player) makes the first one (shot), it shouldn't affect his chances of making the second one." S8 was the only one in the Intermediate group to identify separate events. S7 strongly disagreed with S8 on this stance but most of the resulting conversation centered around S7's belief that skill was involved in the task compared to S8's view of the task as a statistical situation. Therefore, it appeared that the context of the situation can have an impact on the which skills students access to complete the task.

A contrast about separate events was painted with S8 when examining the Fake Flip task. S8 made a list of flips that alternated from heads to tails (see Figure 13). Although S8 did not talk directly about getting multiples of either heads or tails in a row, he did reference the idea of maintaining a ratio of heads and tails within a forty-flip trial, namely roughly half of each heads and tails. The fact that S8 believes that too many of either heads or tails would make the list “not possible” points his understanding of separate events was still developing as S8 does not view each flip as a separate event. Rather, it appears that S8 views the forty flips as a collection of events. S1’s experience with the Fake Flip task revealed that she viewed each flip as separate event. Interestingly, S1 viewed the Fake Flip task as separate events but not the Hot Hand task and S8 was the opposite. S1 and S7 appeared to hold similar views on the Hot Hand task although S1 never mentioned skill as a factor in the discussion. It was possible that S1 was thinking that skill was a factor for the Hot Hand task. But, S1 was the first interview and the idea of skill being a factor or an aspect of context had not been brought up yet and therefore it was not a point of discussion. The context of the problem appears to affect views of separate events, pointing to the fact that context of a task can change what knowledge and which skills a student will access to complete a task in spite of the intention of the educator or researcher.

Skill and Independence. S7 noted that the player’s skill is increasing and therefore the basketball player was more likely to make the next shot. This interaction between S7 and S8 as noted before on the Hot Hand task is a good example of how skill involved in an activity can affect perceptions for students. S7 did not appear to access representativeness but S7 was concerned with the number of shots made before the three presented in the task. S7 was accessing the frequency of the situation in a way to include the skill of the activity. The context of the problem led S7 to believe that situation is in constant flux and therefore the next shot is affected by previous outcomes. This perspective is not necessarily wrong for this problem as basketball is an activity of skill. Thus, the debate here is not whether S7 had a good understanding of independence but rather it is an example of how students can interpret the context of a task that negates any discussion about independence. S7's independence understanding is not visible and any discussion about independence is irrelevant. The context of the situation can have an impact on which skills students access to complete the task. Even though the Hot Hand task was picked with the intention of investigating students' understanding of randomness and independence, such information was sidetracked by S7's belief of skill on the activity. The discussion was still rich and I gathered great insight into the student's understanding. However, the insight was more about S7's understanding of populations and frequency which were not focuses of this study.

Interplay between randomness and independence. The interplay between randomness and independence was further explored while analyzing the Intermediate group. S7 and S8, for instance, appeared to blur the lines between their understandings of both independence and randomness while completing the Hot Hand task and the Fake Flip task. The Fake Flip task revealed that both students created "random" list that had easily definable patterns. S7's list was three heads, then one tail and the pattern repeated. S8 alternated between heads and tails for the entire list. Both students' ideas of randomness drove the creation of their respective list. Additionally, both students were still developing their understanding of independence as evident in their use of representativeness and what they identified as separate events. The students' understanding of randomness and independence appear to be connected.

Parallels between the understandings of independence and randomness became apparent while comparing the Fake Flip task and the Hot Hand task. S8 noted that, with Hot Hand task which also had a 50% probability, either a hit or miss was possible and it was not possible to determine the next shot. S7's views of the skill involved simultaneously removing randomness from the situation and making the next shot dependent to the previous shot. S1 noted on the Fake Flip task that "you really don't know what you are going to get." Here, she was speaking to the way she created her fake flips lists. While discussing flipping a coin, S1 admitted that each was a separate event with an unknown outcome. However, S1 did not view the Hot Hand task as containing separate events as evident when she took issue with the fact that she did not know the total number of shots that the basketball player took or was going to take in the game. The Hot Hand task revealed that S1 was developing her understanding of randomness and independence. I noticed a trend: students were developing their understanding of randomness while also developing their understanding of independence and vice versa. The correlation between the understanding of randomness and independence is noteworthy because the Beginner group had consistent understanding of randomness and independence while completing all the task. Thus, the Intermediate group's variations of understandings of randomness and independence was useful in that it showed that both understandings of randomness and independence appear to grow parallel.

Sample Space. A primary difference between the Beginner group and the Intermediate group was the more developed level of understanding that the Intermediate group held with sample space. All but one member of the Intermediate group scored a level 3 on the LUM with S7 scoring a level 2 on the LUM. The students in the group could systematically build a sample space. S7 was the only student of the Intermediate group to not build a complete sample space but appeared he did not do so out of apathy rather than ability. All the students of the Intermediate group understood sample space well enough to either use a sample space to justify an answer or to answer the task, although they might have required prompting to do so.

S1 scored a level 3 on the LUM based on her abilities to build and use sample space. S1's primary reason for not having a level 4 for sample space was the fact that she did not use the sample space on Fake Flip task to answer the question or justify her answers. The sample space was more of an extra step that S1 needed to complete for the task rather than a tool to be used as a part of the process of completing the Fake Flip task. S1 was able to build samples spaces completely without prompting and did so with great attention to detail. The Fake Flip task found S1 building four primary list of fake flips of 10 trials each. The manner in which S1 builds the first list is interesting in that S1 stopped building the first list about halfway through to build list 3 and list 4. After list 4, S1 finished list 1 and then S1 built list 2. List 3 and 4 were homogeneous list of all heads or all tails. S1 noted that she did this because "anything could happen.", meaning that one did not know the outcome, she was accounting for the randomness of the outcomes and the possibility of such outcomes. Thus, S1 demonstrated a significant method to build sample space. Recall that S1 created her lists systematically with list 1 and 2 as alternating heads and tails with pairs of heads and tails appearing periodically with the unique trait that list 1 and 2 were negative images of each other (see Figure 8).

S1 appeared as though she might have been making lists that others would believe but she was not entirely sure herself that the lists where believable. This thought is based on her statement, “you really don’t know what you are going to get.” However, she chose a list with equal numbers of heads and tails as her list that would fool her friends. S1 noted that she did not intend the list to be balanced with heads and tails rather S1 was trying to make a list that was “believable.” Ultimately, S1 had a method of creating her sample space although there was little evidence that she used a sample space to solve a problem or justify a solution.

S5 scored a L3 on the LUM based on her abilities to build and use sample space. S5 would have received a L4 on the LUM for sample space but needed prompting to correct the sample space, although S5 did not need prompting to use the sample space to answer the Double 6 task. Once prompted, S5 did use the sample space on the Fake Flip task and the Double 6 task to answer the question or justify her answers. S5 was able to create a completed sample space methodically and use the sample space to justify her reasonings as was demonstrated on the Fake Flip task and the Double 6 task. S5 spent a great deal of time and thought to create her list of fake flips. S5 had at most three heads or tails in a row with the list almost alternating between heads and tails on each flip. At times, it appeared S5 might have been virtually flipping a coin in her head. S5 appeared to attempt to construct a list that contained few patterns because she felt that a real coin flip would contain no patterns (see Figure 11).

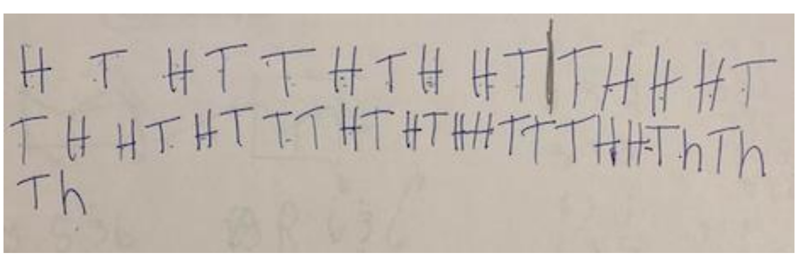

Figure 11: S5 Fake Flip Solution Part 1

S5 latter stated that she believed a real list of coin flips would not have any patterns. After constructing her list, S5 was asked to flip a coin ten times and record her results. S5's actual flip list has streaks of three heads and three tails (see Figure 12).

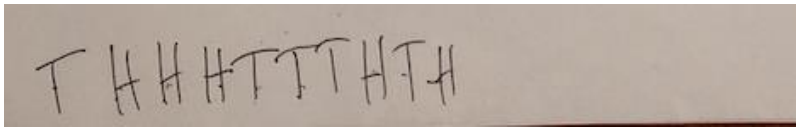

Figure 12: S5 Fake Flip Solution Part 2

S5 was asked follow up questions after she flipped the coin ten times:

> R: Did you get what you expected?
>
> S5: Well…um…not…there, there is that consecutively but at the same time there is like one, two, three, four, five, one two, three, four, five, (as she counts the number of heads and tails). Even with how am I going to flip this coin…just throw it (the coin) up and it lands on the same one I throw it up…but even with the randomness of how do I flip this it's still the fifty-fifty because no matter what you still only have the two options.
>
> R: …you did say that as it went on that a pattern would emerge…but we got heads, heads, heads, tails, tails, tails…which would your friends believe?
>
> S5: They will believe the randomness (pointing to the real list). Because they would say you can't exactly have pattern. I said there would be a pattern because…when you have a small number of them…you will have a pattern because there are only so many you are going to do.

At first, it appeared that S5 believed that her fake flips, the first ten at least, have more of a pattern than her actual flips. This assessment was based on the, "they will believe the randomness," comment. However, "the randomness" S5 is referencing appears to be the actual flipping of the coin. S5 noted that the coin is the random part of the activity. S5 was interestingly unconcerned about the groupings of heads and tails in the actual list of flips. S5's unconcerned nature about the groupings of heads and tails combined with fake list she created implies an understanding of randomness that allows for the possibility of multiple repeated events.

S5 was asked to create the sample space for the Double 6 task (see Figure 15). S5 starts with all the identical pairings, such as 1,1; 2,2…6,6, as can be seen in the second column from the right going down the page. Then, S5 list the pairs for the first number being one, then the pairings starting with two. The pairings starting with three begin at the top right column and the list continues through the pairings starting with five. The pairings starting with six are in middle of the page. S5 skipped the identical pairs in this process as she had already listed those.

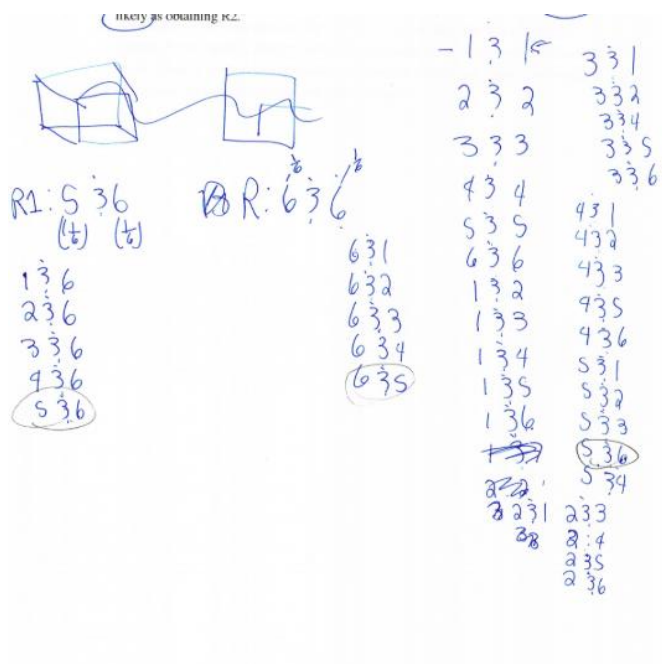

Figure 15: S5 Double 6 Solution

S5 misses a few pairs and repeats a few but finds her mistakes relatively quickly. Then she decides that the list needs to be reversed as she starts to do on the far left starting with the pairs that begin with six. S5 does this to account for the other order than the dice occur in. S5 is pushed about the pairings not already appearing and ones that are appearing to determine if she intended to list the pairs multiple times or if she lost track of the pairings she listed. S5 insist that the reversal pairings she added on the far left are necessary. S5 states, "the first is the six and the second die is the three...but now the first die has the three and the second die is the six." The pairing of 5,6 occurs twice plus the pairing 6,5. When asked about these occurrences, S5 creates a circular discussion that involves the need to "flip" the values so that the "big number comes first then it comes last". S5 adds that she did not realize that the pairing 5,6 appeared twice in the list and was a mistake. However, S5 maintains that the pairing 6,5 is different from the pairing 5,6 and needs to be in the list. S5 then states that she listed all the possible outcomes and marks out the ones that are the same. S5 then used the list she created to modify and justify her answers to the Double 6 task.

S7 scored a L2 on the LUM based on his abilities to build and use sample space. S7 might have received a L3 on the LUM for sample space but he did not create a complete sample space for the Fake Flip task or the Double 6 task. However, the sample spaces that S7 created appears to be incomplete by choice rather than ability. S7 also used the sample space to answer the Fake Flip and Double 6 tasks. S7 also used the created sample spaces to justify his answer to his co-interviewee, S8. S7 built his sample space using three heads, then one tail, repeated this pattern a few times, then drew a line down the page to signify the pattern continues (see Figure 14).

The Double 6 task further demonstrates that his understanding of sample space is a L2 on the LUM. S7 created an incomplete sample space but does it systematically although S7 deployed multiple methods to create the sample space. S7's multiple approaches to building the sample space might have contributed to the sample space containing some mistakes. S7 kept the first die constant for the first six rolls, then kept the second die constant for the next six rolls, repeating one pairing of numbers. S7 stopped after the two-iteration leaving the sample space incomplete. S7 used this incomplete sample space to answer questions related to the Double 6 task and justify his answers about the task although S7 had to be prompted to use the sample space. S7 also created outside scenarios to discuss his reasoning for the Double 6 task. For instances, S7 stated that there might be three ways to get to your grandmother's house but, since there was only one house to arrive at, the three paths still represent one event. S7 was using the grandmother discussion to explain why the pairings 5,6 and 6,5 are one event for the purposes of probability. S7's reliance on outside scenarios is an indicator that he does not understand sample space at the level required to apply the sample space to a problem. S7 might be still be developing how sample space and the task itself connect with each other.

S8 scored a L3 on the LUM based on his abilities to build and use sample space. S8 might have received a L4 on the LUM for sample space but he did not access the sample space without prompting to justify his reasons or answer the Fake Flip task or the Double 6 task. S8 built a complete sample space of forty flips as requested. His sample space from the Fake Flip task consisted of alternation of one head then one tail but began with two heads and ended with two tails. S8 built a complete sample space systematically. S8 built his sample space to be representative of what he felt would need to be true about a large sample of flips, namely there should be approximately an equal number of heads and tails. S8 stated that you would have a "50/50" chance of getting a head or tail so the sample space needed to reflect that. S8 used the sample space as means to justify his views of a coin flip. However, S8 also used his belief of a half and half split to construct the sample space which implies that S8 was thoughtful and deliberate about his constructions of the samples space.

S8 demonstrated ability to build and use sample space while completing the Double 6 task. S8 can create a complete sample space (see Figure 16).

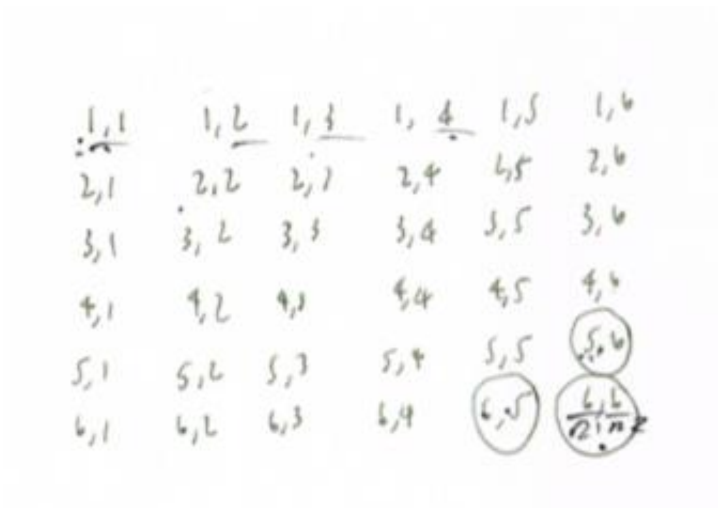

Figure 16: S8 Double 6 Solution

S8 then used the sample space to answer questions about the Double 6 task and justify those answers. S8 used the sample space to argue his perspective about which outcome was correct. S8 engaged in discussion with S7 about the Double 6 task. S7 did not agree with S8's decision about the pairings 5,6 and 6,5 being different pairings, as noted earlier. S8 referenced his sample space in his rebuttal to S7 further suggesting that S8 held a deep enough understanding of sample space to use the sample space as a tool. S8 had to be prompted to use the sample space to reevaluate the task in the initial part of the debate between S7 and S8 which is the primary reason why S8 is a L3 and not a L4 for sample space.

Justification provided by students. The Intermediate group was able to provide some justification for their reasonings. The justifications were often incomplete or used erroneous logic. However, a major distinction between the Intermediate group and the Beginning group was the ability to use sample space to validate answers. S8 for example noted that pairings 5,6 and 6,5 were different because the order matter. S8 used the information to decide that his original answer to the task was incorrect and changed his answer. S8 went on to argue this point with S7. S7 used the two pairings of 5,6 and 6,5 in the opposite way declaring they were the same event thus stuck to his original answer on the Double 6 task. Which was correct was secondary to the fact that both S7 and S8, as well as the other members of the Intermediate group, used the sample space to answer, change answers, and justify answers. The Intermediate group also processed the ability to create examples as parallels to the problem at hand to explain their thinking. The interaction between S7 and S8 did this when they discussed going to grandma's house using different routes as a parallel event to the pairings of 5,6 and 6,5. The combination of these two types of justification lead to deeper understanding of the Intermediate students' thinking.

One major contrast between the Intermediate group and the Advanced group was the members of the Advanced group did all of this unprompted. As will be discussed later, the Advanced group's use of sample space was set apart from the Intermediate group by their ability to do everything the Intermediate group did unprompted.

Advanced Group

The Advanced group was characterized by their ability to build and use sample space to answer tasks unprompted, which was a major departure from the students in the Intermediate group. S3 held the ability to build a complete sample space but was marked as a “No” in Table 12 due to a small error that S3 did not change in one sample space. However, the error could easily be attributed as a small oversight or inattention to detail that has no reflection on the understanding of the task or sample space. Additionally, the Advanced group displayed no evidence of any recency and could communicated that random events were unordered list. The students in the Advanced group displayed little evidence of representativeness but were able to identify separate events and effectively justified their solutions. The Advanced group consisted of two students - S3 and S4; based on the scores, these students had the highest collective understanding as can be seen in Table 6. Table 12 displays the major findings among S3 and S4. It is important to note that collection of perceived understandings allowed for the classifications. This point is especially important to remember since S3 and S4 did not vary greatly in their understanding of randomness from S1 or S7. S3 and S4 were distinct with their perceived understandings of independence and sample space from the students in the Intermediate group.

S3 and S4 were the only students to have an L3 understanding of independence and a L4 understanding for sample space. These two combined factors are underlining the reason why S3 and S4 are separated from the Intermediate group. An additional component that set them apart from the Intermediate group was S3 and S4’s understanding of separate events consistently. Their average understanding of randomness was also higher than the Intermediate group but this fact seems secondary to S3 and S4's understanding of independence being stronger than any other student. It should also be noted that the S4 displayed representativeness once in one task, the Black and White task. The display was noted but faded with further discussion. The emphases on independence is made here because independence appeared to be a contributing factor to students' overall strength in probability. This section of the findings will attempt to demonstrate that S3 and S4 held a noticeable difference in understanding of probability over the other eight students.

Table 12: Advanced Group	S3	S4
Randomness		
Positive recency	No	No
Negative recency	No	No
Makes predictions based on patterns or haphazard models	No	No
Provides justification for random events	Yes	Yes
Communicates randomness is an unordered listing	Yes	Yes
Uses qualitative intuitions to make decisions about randomness	Yes	No
Uses frequencies to make decisions about randomness	Yes	No
Believes that skill effects random events	Yes	Yes
Independence		
Displays representativeness	No	Yes*
Identify independent events	Yes	Yes
Provides justification for independent events	Yes	Yes
Sample Space		

Builds sample space methodically	Yes	Yes
Builds complete sample space	No**	Yes
Makes use of sample space	Yes	Yes
Independence and Randomness		
Allows non-mathematical beliefs to sway decisions about randomness or independence	Yes	Yes

Randomness. S4 achieved a L4 understanding of randomness while S3 scored a L3. Neither student displayed any recency while completing any tasks. Both students could routinely express a random event as an unordered list of outcomes although they may have communicated this indirectly through the explanation of a solution to a task or in the explanation using a student generated example. While completing the Hot Hand task, S3 stated, “well anything can happen...it depends on the situation.” S3 initially provided evidence that she views random events as unknow outcomes or unpredictable. The second part of the statement is interesting as well because S3 is alluding to one of two possibilities as the “situation.” The first possibility is that S3 believes that the context would affect the randomness of the event. The second possibility is that the term “situation” here references the unknown outcome or does the basketball player make or miss the shot. The second idea appears to be S3’s usage. S3 goes on to note that the situation will determine the likelihood of an event, but the events do not become more likely just because a particular event has occurred repeatedly, this is what I gathered S3 meant by “situation.” S3’s “situation” statement does not discount that she understands randomness, rather appears to strengthen the case that she understands randomness quite well.

Through both the Hot Hand and Black and White tasks, S3 displayed no recency. The Hot Hand task revealed that S3 put no weight on the previous events affecting the outcome of the next event. The surface of the previous statement seems rooted in independence but here the focus is S3 did not make a decision about future events based on previous events. Rather, S3 is determining outcomes based on a known or given probability. S3 engaged in a short discussion about the Hot Hand task where she stated, "it should not matter...it's not going to change." The "it" here is the previous and next shot by the basketball player. S3 is affirming that she believes that the 50% shooting average is the sole determining factor in the likelihood of the next shot. S3 creates a baby example corollary to the Hot Hand task. S3 notes that three boys born sequentially to the same mother do not make it more likely to have a fourth boy. While S3 swayed slightly about whether the basketball player would make the next shot, she ultimately decided that the outcome was unknown, and the three previous outcomes do not influence the fourth outcome just like three boys born in sequence do not affect the sex of the fourth child.

S3's understanding of randomness is further backed up when she completed the Black and White task. Initially, S3 choose 'D' (six black rolls) as the most likely outcome. S3 noted that this was the most likely outcome because there are more black sides of the dice. However, as soon as S3 was pressed about her choice, she changed it to 'B' (five blacks and one white at the end). S3 could not provide any reasoning as to why she chose 'B' over 'A' (one white at the beginning followed by five blacks). Then once S3 is pressed about this choice, she changes her mind again back to 'D'. S3's explanation about her change to 'D' is centered around the idea that there are more black sides than white sides and thus black will occur more often. Here again, S3 is noting that the subsequent events are unaffected by previous events, further providing evidence that recency is not being used to make decisions.

The discussion during the Black and White task also established that S3 views random events as unordered lists. The unordered aspect was evident when S3 noted during the brief time she choose ‘B’ as her answer for the Black and White task. The fact that S3 could not provide a reason for choosing ‘B’ rather than ‘A’ implies that S3 views both lists as equally likely. S3 when on to noted that the white “could come anywhere.” Therefore, S3 views the list of outcomes as an unordered list because the white side can appear anywhere in the listing.

S3 further demonstrated a strong understanding of randomness while completing the Fake Flip task. S3 created three list of fake flips and was asked to pick which one would fool her friends. S3 admitted that she believed that all three were equally likely and therefore all three list were equally believable. S3 went on to say, “it doesn't really matter what you write...you always have a 50/50 chance with a coin, so you don't know what is going to come out.” S3 was noting that the list itself is random and therefore had no definable structure to it and thus not predicable in the outcomes.

S4 appears to have well develop understanding of randomness as well as noted by the fact she was the only student to achieve a L4 for randomness. S4 did not appear to display any recency and appeared to understand randomness was the unordered. She provided justification for all her decisions and rarely waivered once a decision had been made. This final point was consideration for the level of difference between S3’s L3 and S4’s L4. S3 would waiver while completing a task, granted this could have been the result of out loud thinking. However, there were times when S4 appeared to be thinking out loud but S4 did not change her mind about answers. S4 was asked to compare the Hot Hand task to flipping a coin. The following interaction was the basis for much of the afore mentioned discussion about S4's understanding of randomness:

R: Are both situations random?

S4: Yes.

R: Why?

S4: You don't get to pick one it is, it's just by chance.

R: What do you mean by chance?

S4: You don't know which it will be until it happens.

R: So, there is no way of predicting it?

S4: There is if you do it a certain amount of times, like you do it a lot. Then you can probability predict what will happen.

R: What do you mean if you do it a lot?

S4: Like if we did it like 500 times, we could probability predict how often it would happen.

R: Oh, so, you are saying we would know how many of the 500 or you saying that we would know what the 501st is?

S4: (long pause) We wouldn't know what the 501st was but we would know how many of the 500 he would make and that would lead us to make a guess or an educated guess on what the 501st will be.

The afore mentioned discussion demonstrated that S4 did not use recency to make decisions about random events. Although S4 did not answer this directly, also imbedded in the discussion was the idea that the collections of 500 shots are unknown in their outcomes. More specifically, the order of hits and misses or the length of streaks of either is unknown and not predictable. I proposed a different scenario; the basketball player misses four shots, makes three shots, and is now on the eighth shot. S4 did not appear to use any form of recency when formulating her responses. S4 stated the result is not known because the shooter has a 50% chance of making each shot. Thus, three made shots in a row have no positive nor negative affect on the result of the next shot. S4 admitted that the basketball player's 50% shooting average is no different from flipping a coin and just as random. It was noted that flipping a coin with heads equivalent to making the shot and tails equivalent to missing the shot could be used in place of the basketball player.

S4's lack of recency was further confirmed when she noted that the seventh roll would not be predictable while completing the Black and White task. S4 also noted that she was not choosing 'A' or 'B' because both listings were equally likely. S4 considered this an issue because there was no way of knowing if the white roll comes first or last and therefore it was impossible to predict which event was more likely to occur thereby demonstrating that S4 believed the event to be random. The fact that S4 did not use the term random to discuss the task is secondary to the fact she described a random situation as it applied to the Black and White task.

Independence. S3 and S4 displayed well developed understanding of independence both scoring a L3, which was the highest among all the students for independence. While S4 demonstrated some representativeness on one task, S3 did not demonstrate any representativeness. Furthermore, what appeared as representativeness for S4 may have been S4 superimposing beliefs on others but not herself. S3 and S4 were able to consistently determine that events were separate.

S3 showed that she could identify separate events clearly when completing the Hot Hand task. Much of the information that appears here are repetitive statements from the randomness section. The statements from S3 was repeated here because they touched on multiple aspects of S3's thinking. Namely, S3's understanding of randomness and independence appeared to exist concurrently and influence one another. The concurrent understanding was displayed when S3 stated, "it should not matter..." in refence to how previously made basketball shots affect the outcome for the next shot. S3 went on to say, "...it's not going to change." In addition to the implications for randomness thinking, S3 was also noting that the 50% shooting of the basketball player was predetermine and making three shots in a row would not affect the 50% shooting average. S3 also said, "well anything can happen...it depends on the situation." The first part of the statement provided great insight into S3's understanding of randomness, but the second part of the phrase is of interest to independence, namely the term, "situation." S3 points out that the "situation" of making the basketball shot or not was the only thing at play. As mentioned before, S3 was using the term "situation" to describe the event of interest or the shot itself. Thus, S3 was noting that the only thing affecting the outcome was the event of shooting the fourth shot and was removed from all other considerations. S3 was not distracted by the previous shots and was solely focused on the fact that the shooter had a 50% shooting percentage. The fourth shot was dependent on the known shooting percent, a streak of previous shots.

The previous interaction involving the "situation" also showed how S3 was also not assuming the fourth shot needed to be a miss to bring the streak of shots back to 50% to match the basketball players shooting average. S3 is unconcerned with the streak that is present in the current situation, thus displaying no reliance on representativeness.

S3 further backs up her lack of use representativeness while completing the Fake Flip task. S3 was asked to create a list of fake flips of a fair coin (see Figure 17).

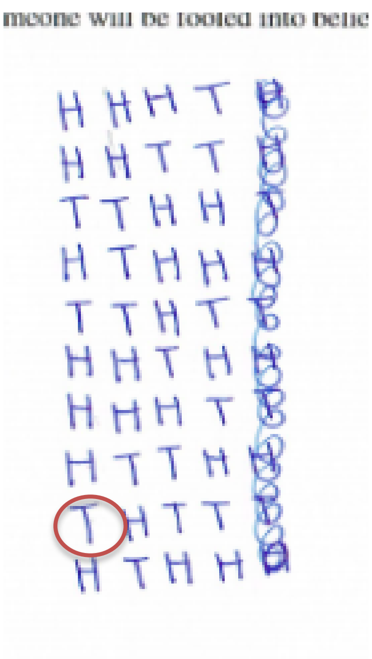

Figure 17: S3 Fake Flip Solution

S3 made the list starting in the top left, made the first column, then moved across for the third and fourth columns. S3 was asked would her friends be suspicious of her list if one of the tails, the tail with the red circle, was changed to a head so that there seven heads in a row. S3 responded with, "I think it depends on the person" going onto reference taking a multiple-choice test. S3 pointed out that getting several A's in a row would make some people worry that they were wrong. Ultimately, though, the answers are "separate" so it does not matter. While S3 point out that she believes the events are "random", the discussion further demonstrates that S3 can identify separate events. The discussion also demonstrates S3's understanding of randomness was linked to her understanding of separate events. Additionally, she makes allowances or alludes to the fact that some people will not see the events as independent. Therefore, creating a fake list of flips that looks believable involves creating a list where the values are not independent of each other.

S4 demonstrated no representativeness while completing the Hot Hand task and minimal representativeness while completing the Black and White task. The Hot Hand task revealed that S4 does not believe that the four-shot section needs to have the same characteristics as the full history of the basketball player, or that historically he makes 50% of his shots. S4 backed up this idea with the statement, "...it will always be 50%..."; thereby showing that S4 is making decisions based on a known probability rather a recent history. Furthermore, S4 was also noting that the previous three shots did not an affect the fourth shot. S4 went on to say, "Some might say that as he makes more his confidence will go but..." this will not affect the outcome. Again, S4 is noting that the 50% shooting average of the basketball player is predetermine and is unaffected by the current series of events.

The Black and White task revealed mild representativeness for S4, but the demonstration might have been S4 superimposing expected value onto a likelihood question or the interviewer's interpretation of S4's explanation. S4 initially chooses 'E' as her answer. She noted that all answer choices were possible but 'E' had "success and failures" in the list. S4 was using "success and failures" to describe possibility that all the sides could appear. S4 went on to talked about the idea that one white side might appear when discussing how choice 'D' is less likely than the trial she created with a die that had five black sides and one white side. However, after some discussion S4 changed her answer to 'D', which was all the outcomes were black. It became clear that S4 understood that the frequency of the white side was one-sixth and therefore less likely to occur in the sequence. But, S4 appeared to discount this possibility due to the fact she viewed each roll as a separate event. S4 stated that there was, "probably going to be at least one white," but a white side does not have to appear at all. S4 said, "probability" but through the rest of the discussion, it appeared that S4 was referencing expected value or that one out of every six rolls of the die would yield a white. Furthermore, S4 noted that, "I could have gotten all white or all black," after she rolled a die six times. Representativeness appears to be present on the surface perspective of S4's comment, but a deeper investigation revealed that S4 was trying to communicate the need to account for all possibilities as S4 stated multiple times, "all the outcomes (listed) are possible."

Although the surface discussion was about likelihood, the deeper discussion led to S4 trying to track which black side appeared to determine if all the black sides appeared in each six-roll sequence. S4 used this reasoning to decide that the white side could appear in a six-roll sequence. There was not a point where S4 referred to the idea that, in a large set of rolls, one-sixth of the rolls would be white, only that she expected one or some of her six rolls to be white. An argument could be made that expected value and representativeness go hand in hand, but, in the context of the discussion that S4 presented while working the Black and White task, it appeared to be less representativeness and more expected value. S4 appeared to be basing her responses on the expected value of the six-roll set but also considered the six-roll set as “too small” to make any decisions about overall behavior of the rolls or to be able to make predictions about a seventh roll of the die. S4 did not note that one-sixth of the rolls would be the white side but rather she noted that not all of the six rolls would be black because not all of the six sides were black.

The Black and White task revealed that S4 appears to have developed understanding of separate events. S4’s understanding of separate events in the context of the Black and White task may have had more impact on her belief that a white would appear in a six-roll sequence than an expected value. S4 did mention that the rolls did not affect each other when she discussed why she was staying away from choices ‘A’ or ‘B’. S4 felt that both ‘A’ and ‘B’ are equally likely and there is no way of knowing if the white side comes first in the sequence or last in the sequence. Underlining S4’s thinking in this context was that she could not choose either as the most likely because both were equally likely as would any sequence with one white appearing in it. S4 displayed that she saw each roll of the die as separate event.

Interplay between independence and randomness. The interplay between independence and randomness was still evident with the advanced group. S3 and S4 both appeared to have the strongest level of understanding of both randomness and independence. This fact was not surprising, but it was noteworthy. Neither student displayed any recency and the presentences of representativeness was questionable at best with S4 and non-existent with S3. There were many instances were a student’s understanding of one topic appeared to influence their understanding of another; a connected approach. For instances, S3 alluded to her randomness and independence understanding while working the Fake Flip task. After she had created her fake list of heads and tails, I asked S3 if it would raise concerns with her friends if one tail was changed to a head so that there seven heads in a row. S3 responded with, "I think it depends on the person." S3 continued by referencing an example of taking a multiple-choice test and getting several 'A's' in a row. S3 went to explain that several A's in a row would have made some people worry but ultimately the answers are "separate" so the length of repetition of the same outcome did not matter. S3 believes this because the events are "random."

Sample space. Both S3 and S4 achieved a L4 on the LUM for sample space. Both students were consistently able to build complete sample spaces and used those samples spaces to adjust and justify answers to presented task. The Double 6 task, for instance, yielded a great deal of insight into sample space for S3 and S4. Both S3 and S4 built a complete sample space systematically. S3 started by creating the sample space for rolling two dice (see Figure 18).

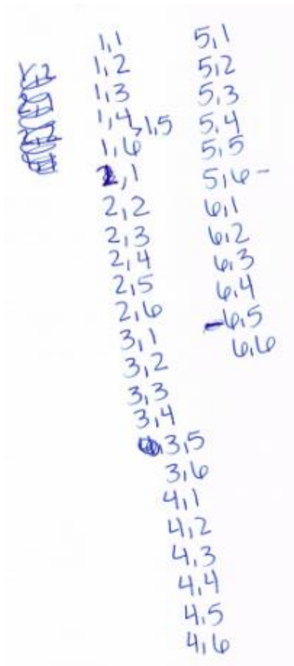

Figure 18: S3 Double 6 Solution

S3 did this by creating all the permutations for one and two but marked it out. S3 then began a second time with the ones and its pairs then two and its pairs, then completed the list in this manner. Both methods were systematic in manner, but S3 appeared to believe that the first method was limiting her ability to complete the list, thus she adopted a different method. S3 missed the pair 1,5 but caught it in the discussion and filled in the missing pair. The fact that S3 found her own mistake in the list is further proof that she understands how to build a good sample space. I was less concerned with the missing pair and more interested about the fact that S3 found it without prompting as it implies that S3 understands what makes a sample space complete. As a follow up question, S3 was asked if the sum of 11 was more likely than the sum of 12. S3 stated, "No...it would be the same...it would be more likely to get 11...because you have two choices the two chooses of five and six rather than just the six and six..." S3 was able to use the sample space to answer a follow up question. S3 did this without being prompted and used the sample space to correct her own thinking about the problem.

S3's use sample space in the Double 6 task was of further importance as it provided the means to collect the information that was used to designate her as a L4 for sample space. There were variations in her abilities with sample space (see Table 5). For instance, on Black and White task, S3 had difficulty creating the sample space. The difficulties might have been due to the context of the problem. S3 knew the sample space for six rolls of a die would be larger than two rolls. S3 may have been struggling with the mindset of building such a large set, she might have thought that the sample space may have seemed too large to manage. For this reason, S3 was marked as a L2 on Black and White for sample space. Therefore, the decision was made to focus determining S3's understanding about sample space chiefly on the Double 6 task.

S4's experience with sample space during the Double 6 task was similar to S3 in that she created a complete list systematically and completely (see Figure 19).

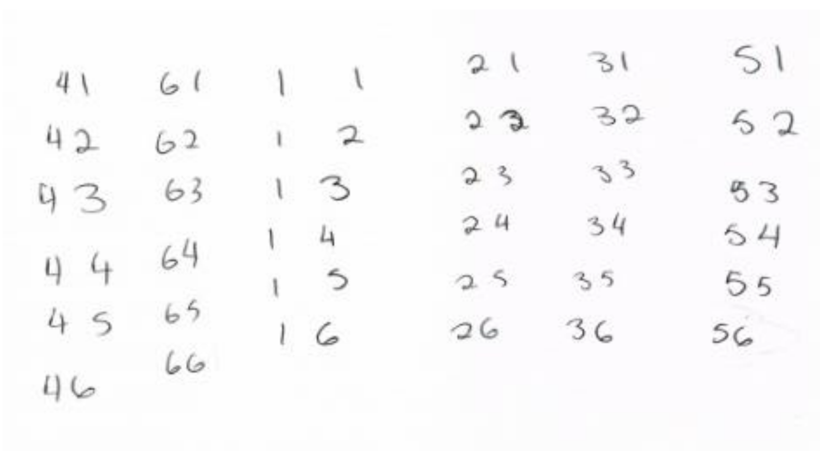

Figure 19: S4 Double 6 Solution

S4 did not make any mistakes in the creation of the list the first time so no corrections were necessary. Just like S3, S4 changed her answer about the Double 6 task after building the sample space. Initially S4 said the result was one-third for R1 and R2. S4 initially thought that there were 12 possible outcomes for the rolling of two dice. Without prompting, S4 changed her answer for the Double 6 task once she had created the sample space noting that “there are more than 12 outcomes.” S4 was referring to the fact that she now had 36 outcomes to choose from. The sample space enabled S4 to change her answer to the task and back it up with justification.

Ultimately, both S3 and S4 had the ability to use the sample space to change and justify their answers. The adjustment process that S3 and S4 engaged in was important to evaluating the students’ understanding of sample space and separating S3 and S4 from the rest of the students and they were the only students that did so unprompted which further separated them from the group. The use of sample space as a tool to solve probability task was noted by Watson and Callingham (2014) as an indicator of the student’s understanding of deeper probability concepts such as independence. Additionally, English and Watson (2016) noted that building the sample space can equip students with the ability to properly justify their answer as was seen with both S3 and S4.

Skill affects randomness. S3 and S4 were asked about the skill level involved in shooting the basketball. S4 noted that the basketball player hitting of the fourth shot was unaffected by his skill because he was a 50% shooter. S3 initially answered it will influence the basketball players fourth shot, but in the same sentenced changed her answer to it will not have an effect. It is worth noting that S5 and S7 held similar viewpoints about the basketball player and his skill level in that S7 believed that the basketball player’s skill would have increased with every shot and S5 believed that the skill level would matter if the basketball player was a professional but it would not matter if the basketball player was a high school player. This is contrasted to S4 and S8 whom both viewed the 50% average as an established average that already considers the shooter's skill. S4 stated, "some might say that, as he makes more, his confidence will go up...this will not affect the outcome - it will remain 50%" when discussing the shooter achieving a string of three hits in a row. It should be noted that representativeness, by Jones (2005), is not a factor during this part of the discussion in reference to independence for S4. S4 stated "...if he shoots 500 we might know something about the outcomes..." S4 was referring to that if the shooter takes 500 shots we will know that half of them will be hits and half will be misses. S4's observation about 500 shots leads me to think that she did not view the string of three hits in a row as a representation of the overall set of shots. S8 noted similar understanding in that he stated that over time the shooting percentage will approach 50% as more shots are taken. It is worth noting that several students, such as S1, S5, S9, S10, all made note that they were not good at flipping a coin. Coin flipping may not be something that is considered to be a skilled activity such as basketball or video games, but many of the students appear to believe that the skill involved in all three activities is important enough to mentioned as they appeared to believe that their skill at flipping a coin has some type of effect on the outcome.

Conclusion

Ten students completed the tasked based interviews and collectively provided large qualities of data about the probability understandings of high school students. Overall, the students were successful in completing the tasks and providing answers and justification. The primary purpose of the tasks was to gather an understanding of how students think about randomness, independence, and sample space. However, along the way, students also provided small insights into context and beliefs about probability. Those aspects were stated in the presentation of findings as appropriate to the description at hand.

The findings presented above are a small collection of behaviors and understandings of ten high school students. The findings are not generalizable but are insightful in that the findings display a collection of student understandings of probability. Major themes and any implications to teaching practice will be discussed in Chapter 5. Some of the major points of interest were recency and it being almost totally non-existent in this sample, the affect sample space had on other aspects of probability, and the use of sample space with an increase in understanding of randomness and independence. The use of sample space as a tool was noted during Chapter 4 and will be discussed in greater detail in Chapter 5.

CHAPTER FIVE: DISCUSSION

The purpose of this study is to understand the probability literacy of high-school students. There were few studies in prior research that examine the probability literacy of high school students. There is a need for students to have a robust grounding in probability to be able to function effectively in a highly technological and data driven world. The research question guiding the study was:

> What is the probability literacy of high-school students? In particular, what do high school students understand about the probability concepts of chance, randomness, independence and sample space?

Overview

This study examined the understanding of ten high school students' probability literacy through task-based interviews. The students were asked to solve and discuss their thinking about five specially designed tasks adapted from prior research. The Levels of Understanding Matrix (LUM), which was designed based on prior research and was build based on pilot study student interviews, was the primary tool used to understand and classify students' thinking in the three major knowledge elements of probability – randomness, independence, and sample space.

The study showed that the students demonstrated a complex and intertwined understanding in the areas of randomness, independence, and sample space. Special aspects of the three areas included recency, understanding randomness as an unordered list, representativeness, separate events, systematic development of sample space, and using sample space as a tool to answer questions about randomness and sample space.

Each student demonstrated some aspect of one or many of these traits at some level during the interview while completing one of the five presented tasks.

Based on the LUM, four students were classified as Beginners, four were classified as Intermediate, and two were classified as Advanced. The classifications were based on a student's collective scores from each category of randomness, independence, and sample space. The Beginner group scored one's and two's on all three categories, the Intermediate group scored one's and two's in randomness and mostly three's in sample space. Sample space understanding was the primary separation in student understanding between the Beginners and Intermediates. The Advanced group scored three's and four's in randomness and independence and four's in sample space. The sample space scores, which were four's for both the Advanced students, along with their independence scores were the deciding factor in classifying those students as Advanced. Both scores were noticeably different from the other two groups.

The Beginners were all in Math 1, Math 2, or Math 3. The Intermediates and Advanced had or were currently taking a statistics or pre-calculus courses except for one student who was taking Math 2 but was enrolled in a Saturday academic enrichment program. Based on observational data, courses that students have taken appear to have an impact on the students' understanding of randomness, independences, and sample space.

My study revealed, through the framework of the LUM, that students have various levels of understanding of randomness, independence, and sample space. More specifically, high school students do not appear to access recency in relation to randomness but do access representativeness in relation to independence. Both recency and representativeness appear in the LUM as factors to gauge levels of understanding for their respective topics. The LUM and its implications will be discussed in greater detail in a later section. High school students appeared to hold some understanding of sample space and many held the ability to use sample space, on some level, to assist them with answering probability tasks. Moreover, understanding of randomness and independence appear to develop in parallel for most students. Although there was a dip in an understanding of independence for the Intermediate group, their understanding of sample space appeared to increase dramatically over the Beginners. In general, a student's abilities appear to complement each other in that students used their understanding of randomness to understand independence and vice versa. Students appeared to access sample space while dealing with randomness in that, if a student had a higher level of understanding of sample space, they had a stronger understanding of randomness. Sample space appears to aid students at a lower level of development in understanding randomness but might inhibit the same student's understanding of independence. The students' understanding of independence appears to level out with randomness once the students' understanding of sample space is solidified.

Comparisons to Past Studies

There have been several studies that have gained valuable insight into students' understandings of probability and probability literacy. Those studies have revealed behaviors such as recency, representativeness, and the use of sample space as a tool. Additionally, studies have found that some students can correctly answer questions but cannot justify their answers using probability.

Most studies conducted in researching probability have centered around younger children; elementary and middle school age students (Bryant & Nunes, 2012; English &

Watson, 2016; Fischbein & Gazit, 1984; Gilovich, Vallone, & Tversky, 1985; Gal, 2005; Kerslake, 1974; Piaget & Inhelder, 1975; Watson & Mortiz, 2017). This section will discuss past studies and how my findings relate to those students. This will include times where my study supported the findings of past studies as well as instances where my study found disagreement with past studies.

One major aspect that should be displayed throughout this discussion is how most of the other studies in probability focused on elementary and middle school age students. When considering probability literacy, few studies have been conducted at the high school level. Bryant and Nunes (2012), for instances, completed a study of young adults and older students at the middle school level but spoke little of high school students. Rubel (2007) was a notable exception where she examined the probabilistic thinking of both high school and middle school students.

The Absences of Recency

Bryant and Nunes (2012) observed a behavior of recency as middle school students and young adults made decisions about random events. Bryant and Nunes (2012) described recency as an individual making decisions about subsequent outcomes based on the repetitive nature of previous outcomes. Meaning, if one decides that the next basketball shot will be made because the last three were made, that is positive recency or, if one decides the next basketball shot will be missed because the previous shots were made, that is negative recency. This phenomenon was observed in this study. However, only one of the students in this study displayed documentable recency. The other students in this study displayed recency-like behaviors but it was decided that it was not recency but instead the behavior being noticed was based frequency of the event.

The research completed by Bryant and Nunes (2012) suggested contrary findings to this finds. Bryant and Nunes (2012) found that younger students would display positive recency and older students and adults would display negative recency (Bryant & Nunes, 2012). My study found little evidence of recency present for the youngest or oldest students in the study. Additionally, my study found little evidence of recency with students that scored L1 on randomness and those that were still developing their understanding of randomness. Namely, it was found that students thought in terms of frequencies. Students made predictions based the values or outcomes that would have brought the events closest to a predetermined or an anticipated statistical outcome.

The Hot Hand task provided further evidence about the use of recency as many students would answer the opposite for the fourth basketball shot; a miss if there had been three hits or a hit if there had been three misses. The Hot Hand task was based on a study completed by Gilovich, Vallone, and Tversky (1985). Gilovich, Vallone, and Tversky (1985) asked adult fans about the phenomenon known as “hot hand” or a belief that a basketball player will go on a streak of making several shots in a row. Their study was not looking to prove recency but rather, in the context of college or professional basketball, do people believe making one shot increases or decreases the likelihood of making the next shot. Their study found that people in general do believe the chances are increased with every successful shot made, but the statistics of the basketball players do not support this idea of going on streaks of making several shots in a row. My study found that, even in the instances when students believed the next shot was more likely, it was not because they made the last three shots. Most of the students in this study either believed that the next shot was more likely to be a miss or they could not make the prediction because the event was random. The students that believed that the next shot would be missed based their decision on a desire to maintain the 50% shooting average. This point was further reiterated on the Fake Flip task when students would try to keep the series of forty flips balanced. My study is not denying the presences of recency among some populations, only that it did not appear among the population that was used for this study.

The lack of reliance on recency by students does not imply that students held a strong understanding of randomness, although some did. Rather, it provided evidence that most of the students were basing their understanding of randomness on other aspects of the phenomena of shooting a basketball. For instance, items such as frequency and skill arose in the course of the interviews. The results demonstrated that the students appeared to access frequency with some regularity when discussing randomness.

Randomness understanding was still developing for most of the high school students in this study and that fact was emphasized by four of the ten students scoring a L3 or L4 on the LUM for randomness. If proficiency standards were applied to the LUM, which would be grossly inappropriate, one might find one student was proficient in randomness with three others closely behind.

Other Findings

Piaget and Inhelder (1975) found that high school students should have a robust view of randomness that should be grounded in some mathematical reasoning. Younger students might be primarily relying on intuition, but high school students should not be as their understanding of randomness should have progressed beyond using intuitions. It was found that many of the students interviewed in this study still relied on intuitions or other self-constructed regulations to make decisions about random events. The finding of high school students' reliance on intuitions was a primary reason for including intuitions in the LUM for L1 and L2 randomness as it points to a still developing understanding of randomness. Furthermore, despite Piaget & Inhelder's (1975) findings that high school students should have a robust understanding of randomness, the findings in this study show that their understandings where influenced by peripheral topics such as skill or the context of the tasks being used. The implications are that some students are still developing a robust understanding of randomness that should be supported and developed further through proper instruction.

Fishbein and Gazit (1984) found that students possibly link randomness and unpredictability meaning that students might believe an event is random if they cannot predict the outcome. My study did not refute this finding but rather found that high school students may apply levels to randomness with students having noted that the levels could have been affected by previous outcomes or skill. Therefore, shooting a basketball could be less random for a highly skilled player than a player with low skills even if both maintained a 50% shooting average. The students did not seem to consider that a highly skilled player would have a shooting average more than 50%. The Fake Flip task reinforced this idea of changeable randomness when students noted that a person can learn to control a flip of a coin.

Representativeness and Frequency

Representativeness is a mindset that a subset of trials has the same characteristics of the entire, much larger set of trials (Jones, 2005). As stated before, the implication of representativeness is that students struggle with separating an event characteristic from the population characteristics (English & Watson, 2016). In this study, the students used what they knew about the event to make predictions about outcomes based on previous outcomes. All but one student (S3) displayed representativeness in this manner on most tasks. It was found in this study, on the Hot Hand task, that students would decide that the fourth basketball shot in the sequences would be a miss not because the three previous shots were hits but because the shooting average was 50% and three out of four was already over the predetermined 50% level. Many of the students made note that half the shots would be misses, thereby transferring event characteristics to a sample.

Some students engaged in discussions about the Fake Flip task revealed that they believed a smaller subset should hold the overall theoretical ratio of outcomes. As noted, some students built lists of flips that were essentially alternations of heads and tails designed to maintain a 50% distribution of each. Many of the students that participated in this study believed that the distribution of heads and tails needed to be maintained at roughly 50%, even for small sets of trials where a bit more variation would be allowed but not to extreme levels. The students held mildly different understandings of independence as separate events but most students held to the belief that the variation would need to be small to be believable on the Fake Flip task. The major conclusion from this study was that students understood that the events were separate but the relationship that existed between sets of trials and the theoretical probability of an outcome can be unrelated. Other studies found the use of data sets can help students develop their understanding of independence (Jones, 2005; Watson & Callingham, 2014). Based on my interactions, most of the students benefited from using data sets in connection to independence. This will be reexamined in the sample space section later in this chapter.

Independences and Separate Events

Many students appeared to struggle with viewing separate events which supports findings from Konold (1993). Konold (1993) found that middle school and college students alike struggle to understand independence. This study revealed that students viewed the basketball shots on the Hot Hand task as not being independent. One student noted that the basketball player was getting better and therefore one shot would affect another shot. There was a belief skill played a role in the outcome of the trials for the basketball player, believing that the player was increasing his skill and thus the trials were not separate from each other. Therefore, the students were not wrong from this perspective to view the trials as a collective. However, the task was discussing a four-shot sample with a predetermined shooting percentage. Therefore, the students might have been taking liberties with the task. Another student noted that, "I don't know how many shots he is taking…", suggesting that all the shots the player takes are related to each other, even future shots. The second case mentioned here was repeated in different forms by six of the students in my study. The students, regardless of skill involved in basketball, spoke of shots not being separate events. It is important to distinguish that the task of pulling marbles from an urn without replacement would not be separate events in that the count in the urn changes with every pull thereby changing the probability for each remaining marble. The Hot Hand task did not have such criteria thereby allowing students to superimpose skill onto the task, negating the independence of the event and making it a compound event. The discussion here was not whether skill would have affected the outcome or even whether hot hand is a real phenomenon. The point of discussion was about how did students view separate events; as a collection or a series of separate trials.

The discussion about skill in relation to the Hot Hand task pointed out that sometimes students have created rules in their minds that change the way they view a situation. Piaget and Inhelder (1975) found that students will indeed create arbitrary rules or justification when asked to make predictions about probability outcomes. My study found this to be true as it was observed that students would use self-created rules as a method to justify solutions when they cannot reconcile what is observed with what they believe to be the outcome.

The student-created rules appear to be confined to some contexts. For instance, some students applied different perspectives to the Hot Hand task compared to the Fake Flip task even when skill was not a factor. There was the belief that the list of flips would be exactly half heads and half tails and “too much” of a deviation from half and half would be suspicious. Ignoring the definition of “too much”, the primary point of concern in the present discussion was the view that each flip was not separate which is a sharp contrast to the Hot Hand task where students viewed the basketball shots as separate and thus the next shot was unpredictable. The reverse situation was true with students where Hot Hand task contained related trials but the Fake Flip task did not. This supports findings from another study were students conducted multiple independent trials for an experiment to develop their understanding of independence with particular emphasis on separate events (Batanero et al., 2016). In Batanero et al. (2016) students conducted large numbers of trials and demonstrated independence. Batanero et al. (2016) found that early in the experiment students did not always view the flips as separate but, as the experiment list was built, the concept of independence emerged and students gathered a greater understanding of separate events. Helping students to better understand any topics was outside the scope of my study, as I was only seeking their current understandings.

Another aspect to consider was studied by Fishbein, Nello, and Marion (1991) when they found that students believed one could learn to control the flip of a coin. It is possible that some of the students were giving answers about the Fake Flip task with the background idea that a coin flip can be controlled. Furthermore, many students made note that their ability to flip a coin was poor and thus the outcomes were possibility varied as a result. An assumption would be that students believe that the Hot Hand task would have included a similar skill component. However, as previously stated, few students viewed the events as being truly parallel in their nature. This inconsistency adds to Piaget and Inhelder's (1975) findings about arbitrary rules and this study shows that they are a strong governing force for students, even at the high school level.

The Interplay of Independence and Randomness

This study purposely attempted to gather an understanding of randomness and independence of students as separate concepts. However, in the interview interactions the students would intertwine notions of independence and randomness in their responses. The association was evident at every stage of data collection. Students were asked probing questions with the intention to reveal more about their understanding of randomness, but they spoke about their understanding of independence. For instance, one student noted that the black side of the die on the Black and White task will occur more often while stating that white will have to appear in a six-roll sequence. This student noted independence and randomness in the same discussion but they were mildly conflicting with each other. The discussion alluded to the view that the events were random but also viewed the six rolls as one event rather than a collection of trials. The implications described here about the association between independence and randomness suggest that the teaching of independence and randomness should occur concurrently. Furthermore, one skill may act as scaffolding for the development of the other skill. A recommendation would be to use tasks to teach randomness and independence that involve both randomness and independence.

Discussion of the Methods

My study was centered on determining the current understanding of high school students through the use of tasks and a research-created LUM as a framework to gauge the student responses. The LUM was rooted in other studies and other frameworks, however, the final form was entirely generated based on my study and the responses given by my students. The tasks used in this study were adaptations from previous studies picked for their potential to generate discussion and to focus students on the particular topics of randomness, independence, and sample space. The discussion that follows was constructed to reveal the details of the LUM and the task usage.

The LUM

The Levels of Understanding Matrix (LUM) evolved during the study and served as a means to analyze students' understanding of probability. The evolution during the study occurred as new data was introduced via an interview but did not fit any existing part of the LUM. The LUM was refined, then the LUM was reapplied to all the interviews that had taken place up to that point in order to ensure that all students were measured with the same standard. The LUM provided criterion to effectively understand student responses for each task which had the potential to be vast and seemingly unrelated. While the responses were generally aligned, albeit at different levels of understanding, the LUM help to remove much of the guesswork associated with the process of analyzing student data. The LUM was created partially for this purpose as there was not an existing framework to gauge student understanding for the concepts and grade level in question. The LUM also assisted in focusing the analyses around several anticipated themes such as recency or representativeness. The LUM also assisted in topics that were new or unanticipated such as the affect that skill might on have on a student's response to a task.

The LUM indirectly provided a lens to view the value of each task. One task, the Straight Lottery task, was removed from use after the development of the first version of the LUM. The LUM displayed that the data being gathered during the first few interviews on the Straight Lottery task as being parallel in purpose for data about independence and randomness.

The LUM was based on Gal's (2005) definition of probability literacy and Jones, Langrall, and Mooney's (2007) framework of particular probability components such as independence. Gal's (2005) framework focused my study on components that collectively define being literate in probability. Jones, Langrall, and Mooney's (2007) framework served as the basis for certain structural aspects of the LUM. The LUM also reflected some of the standards in the CCSSM (NGACBP, CCSSO, 2010). For instance, both Jones, Langrall, and Mooney (2007) and the CCSSM outline the use of probability to solve problems. The LUM included this idea under sample space as major factor for gauging how students used sample space to approach the solutions to the interview tasks.

The other components of Jones, Langrall, and Mooney's (2007) framework were similarly included, although loosely at times, such as the use of random variables to interpret distributions. The LUM contained components about regular and uniform distributions and how students access these distributions in references to their understanding of randomness. The remaining component of Jones, Langrall, and Mooney's (2007) framework included the use of simulations. The use of simulations was not directly included in the LUM but the LUM made particular notice of using students' abilities to make predictions based on previous outcomes. These outcomes can be theoretical, as was the case on the Black and White task where students discussed the likelihood of one outcome compared to four other given outcomes. Contrasted to the theoretical scenarios were the trials that students completed such as the Fake Flip task were students flip a coin up to forty times for comparison to a fake list of flips.

Gal's (2005) definition of probability was more directly translated into the LUM by use of three major knowledge elements of randomness, independence, and sample space. The big ideas also served as the foundation for choosing the tasks that were used in the study. Additionally, my study was concerned with the knowledge element of figuring probabilities. Other components of probability literacy such as context and beliefs, though noted in the findings of the study at a limited quantity, were not a focus and were not directly included in the LUM. Rather, the components such as context and belief were used as tools to better understand and discuss students' understanding of randomness, independence, and sample space.

The LUM went through multiple iterations that included additions to the independence and randomness components. The independence levels were changed based on a couple of interviews that revealed students held beliefs that appeared to affect their views on independence. A major concern was to determine if students were making decisions about independent events while ignoring the "noise" of a tasks or situation. The term "noise" here refers to the unrelated context or scenario of the task. For instances, a student might decide the outcome of the next coin flip was based on what the previous 20 flips had yielded rather known theoretical probability of 50%. A similar concern as based on the Hot Hand task and what the previous shots resulted in, hits or misses. This touches both on recency of randomness and of separate events of independences. Therefore, the LUM needed to reflect this idea for both independence and randomness. The place where it was included for randomness was under the "makes predictions based on pattern or haphazard model" and was primarily used to denote students as being a L1 or, if it was limited in usage, a L2. The independence connection was the inclusion of "makes prediction for subsequent outcomes on previous (or future) outcomes…" which was used to place students at a L1 for independence.

Randomness continued to be adjusted through the tenth interview. The "personal experiences, non-mathematical, beliefs…" aspect was added to L1 after the tenth interview. The student in the tenth interview noted multiple times that his rational for an answer choice was based on his own experiences. For instance, he noted that if three boys are born in a row, then the third child will be a girl because that is what his family unit looks like. Additionally, there were several instances where it was noticed upon second analysis of the data that many of the students would make similar assessments such as deciding that heads appeared more often because that is what he or she sees happen. The concern about personal belief appeared to have an effect on how students viewed randomness, thus it seemed prudent to include it under randomness at L1.

The personal belief aspect should have been noticed earlier, but it was not until the tenth interview when the student mentioned that his household had three boys and a girl that the connection was noticed clearly. Through the process of re-examining the data, it was noticed that personal beliefs were scattered throughout most students' discussion of randomness. The discovery further emphasized to the researcher how the disposition elements blend into the knowledge elements and affect student understanding and decisions. The result is that even when the focus is on randomness, independence, or sample space, many other aspects of probability literacy will be present. A focus on a single element of probability literacy was maintained by centering interview questions around topics of concern. For instances, when the concern was independence, probing questions would have been centered around separate events and representativeness to determine a student's understanding of these ideas.

The LUM has the potential to be built to include all aspects of probability literacy. Future research can continue to provide data that will drive the development of the other components of probability literacy such as context, language, beliefs and attitudes, and personal sentiments. Many of the tasks used for this study could be used to research additional components of the LUM by adjusting the probing questions. For instance, asking students to build parallel situations to the task being discussed would provide data about the student's understanding of language being used or asking students if they would bet money on the fourth shot from the Hot Hand task would provide insight into personal sentiments about risk. The LUM, tasks, and the probing questions work in conjunction with each other to provide a detailed perception of a student's understanding.

A completed LUM that includes all the probability literacy components would provide researchers and practitioners alike with a solid tool to assess student understanding of probability. The LUM provides the ability to group students by ability in probability literacy, an ability previously not available at the high school level. In the classroom setting, it would allow for teachers to better understand what students need to become probability literate. This would help the teacher in choosing tasks for classroom lessons that would build the needed probability literacy skills. It should be noted that the LUM has been designed with the use of open-ended task that required the researcher to interact and discuss the task with the student. Optimal use of the LUM by researchers or practitioners would include a similar format. The task-based interview format would not be a concern for researchers but the format might require some adjustments for classroom use by teachers.

The Tasks Used

Recall from Chapter 3 that the tasks were used as the primary tool to gather information about students' understanding regarding randomness, independence, and sample space. There were several tasks initially considered for use in the study. Ultimately, tasks that were chosen because of their perceived familiarity to students. One task used in the pilot and with one student early in the study was the straight lottery task. The task was chosen because of its connection to randomness and sample space. The task was discounted from use later in the study because students in the pilot and the early in the study struggled with the context of the problem. There is the possibility of including it in subsequent studies that have the goal of discovering more about students' understanding of probability in reference to context.

Each task was used to gather an aspect of understanding. For instance, the Hot Hand, Black and White, and Double 6 tasks were chosen for their ability to gather data about students' understanding of independence. The Hot Hand, Black and White, and Fake Flip tasks provided insight into students' understanding of randomness. The Fake Flip, Double 6, and Black and White tasks also provided insight into sample space. One task provided large quantities of data about at least two aspects of the study if not all three aspects. Frequently, short interactions between me and the student would provide large amounts of data about randomness, independence, and sample space simultaneously.

The tasks that were used and contributed the greatest quality of data were tasks with topics that were easily accessible to students such as basketball, dice, and coins. The Hot Hand task was one of the best tasks at revealing students' understanding of both randomness and independence. The three previous made shots presented in the problem revealed where students were in their understanding of independence by forcing students to separate the fourth shot from the three previous shots. The task also displayed where students were in their understanding of randomness in that they were asked to make a prediction about the fourth shot. In the context of the problem, with skill of the basketball player not being a factor, the fourth shot was meant to be random and thus unpredictable. Students' responses to this situation revealed their understanding of randomness. The exception to this situation were the students that factored skill into the task, an inclusion that changed the randomness aspect of the problem. However, when skill was included as a factor, further discussions about context were still possible. Therefore, the Hot Hand task will be maintained in its current format for subsequent studies but follow up questions will be tooled to bring about discussion of skill involved. The skill aspect was an interesting and unexpected outcome of the task which will be discuss later in this chapter. Changing the context of the problem slightly with discussion about skill will hopefully reveal new insight into students' understanding of probability in different context.

The Fake Flip and Double 6 tasks provided extensive data about students' understanding about sample space which also allowed for rich discussion about randomness and independence. The Fake Flip task required that students build a collection of trials of a flip of a coin that revealed how students viewed randomness based on how they built the list and explained their reasoning for the list. However, the Fake Flip task provided limited data about how students build samples spaces and it revealed that listing all possible sample space variations for the Fake Flip task proved to be extensive for any of the students to attempt. The Double 6 task required students to build the entire sample space for the rolling of two dice. The Double 6 task was significantly more accessible for students in terms of building the entire sample space. The Fake Flip task was excellent for discussion based on what was created but the Double 6 task allowed for more insightful discussions about sample space. Both tasks forced students to consider all possible outcomes of each situation and displayed students' understanding of various topics.

The Black and White task proved to provide useful data about students' understandings of randomness and independence by requiring students to contrast different possible outcomes from six rolls of the die. Students had to consider the relationship, or lack of a relationship, between rolls of the die. Discussion rich answer choices included outcomes, one with the white side at the beginning of the list and another with the white side appearing at the end of the list. Engaging students in a discussion about similarities and differences between these two outcomes was insightful, especially when a student picked one of the two as their “most likely to occur” choice. The other answer choice was the all black outcome. Some students appeared to access expected value in reference to not choosing all black as their choice because one white “was possible” as many students sited. Overall, the task was relevant and insightful for randomness and independence. Researchers and teachers could make use of this task to compare and contrast the expected value and likelihood. This suggestion is based on the fact that multiple students made note of one white appearing in a six-roll sequence being the most “likely” occurrences because there was one white side on the dice. Students where referencing expected value, not likelihood which provides an entry point to discuss expected value with students as a lesson rather than an assessment.

Addressing High School Probability Literacy

An argument could be made that this was not a true probability literacy study since it did not examine all components of probability literacy overtly. Furthermore, the researcher-developed LUM, which was used as a framework to view and inform the data analysis, only accounts for certain knowledge elements of big ideas and figuring probabilities. However, the topics examined in this study are viewed as foundational to the understanding of probability. It was key to exam these topics to begin the examination of the other components of probability literacy.

The probability literacy title was maintained on this study for two additional reasons. First, the development of the LUM during the project was a major objective for the study and served as a framework to analyze the data. The LUM is comparable to other frameworks developed by other studies focused on aspects of probability literacy. These other frameworks exist at other levels of schooling but none that pertain particularly to high school students. Thus, it was important to focus the study on the development of the LUM to make it reliable and robust for randomness, independence, and sample space. The LUM in its current state will serve as the foundation for analyzing student understanding of these topics. The intent was to extend the LUM to include all aspects of probability literacy for high school students in future studies, as will be discussed later.

It should be noted that the LUM does contain aspects of the disposition's elements. For instance, it was found while looking at independence, that personally held beliefs by the student had an effect on the student's understanding of independence and therefore a personal belief component was included in the rubric for independence. The randomness topic LUM included, "allows personal, non-mathematical, beliefs or experiences to sway decisions…" Therefore, additional disposition elements were included in the LUM for randomness. These items were included in the LUM as the result of the primary findings from the three students from the pilot study and the early stages of the main study. These students revealed that randomness and independence understanding include these components of disposition elements in the manner in which they interacted with the presented task. For instance, some students believed that heads appeared more often either as the first outcome or more often in the listing of flips which was contrary to their stated understanding of each side of the coin having a 50% chance of coming up on each flip. The contrast painted here pointed to the need to include some level of disposition elements in the LUM for randomness and independence in spite of the focus for this study being on knowledge elements. The reasoning was that the students' beliefs influenced their understanding of randomness and independence, even when students were aware of the contradictions that they were making. Therefore, it was important to document beliefs to properly interpret student understanding of the knowledge elements.

The second reason to maintain the probability literacy title was the desire to extend this study in the future to collect data about other probability literacy elements. The tasks used for this study have already been shown to be able to elicit responses about disposition elements. However, collecting or creating additional tasks that have a primary purpose of collecting data about the disposition elements would be ideal. The additional data would aid in the development of the extended LUM that includes the other components of probability literacy. However, the study in its current state could easily be repeated to gather more data to add validation to the current findings. With the right follow up questions, the same tasks could be used to gather information about the disposition elements.

Implications to Teaching and Learning Probability

While analyzing the students' responses for each group, independence and randomness appeared to develop in conjunction with each other. All three groups appeared to display this correlation with the Beginner group and the Advanced group displaying the strongest relationship. The data here is far from conclusive, but the occurrences do provide insight for teaching and learning. Namely, independence and randomness should not be taught in separation from each other. Rather, independence and randomness should be taught as compliments to each other to help students develop a more holistic understanding of probability.

There was some literature that expressed the idea that students viewed independent, random events as compound events (Bryant and Nudes, 2012). Fischbein, Nello, and Marino (1991) completed a study where it was found that independence and compound events could not be separated in student thought or learning. Thus, it seems logical that randomness needs to maintain a relationship with independence in student learning. The findings from my study imply that students relying on their understanding of independence to better understanding randomness and the reverse was true, too.

Students who did not understanding how to build and use sample space struggled with further development of randomness and independence. Students of the Beginner group did not hold the ability to build a complete sample space or use the sample space with the task at hand. The Beginner group also could not communicate that randomness was an unordered list compared to students in the Intermediate group and the Advanced group who all could build sample spaces and communicated randomness as an unordered list. Further investigation is required to determine if students that deploy simple patterns to build the sample space for the Fake Flip task would develop similar methods for the Double 6 task or if they would behavior similar.

The ability to build a complete sample space appears to be associated with better developed understandings of randomness and independence as evident with the Intermediate and Advanced groups. The Advanced group held the highest level of understanding for sample as determined by the LUM of any group. The implication was that sample space should be a fundamental topic of discussion for students to development deeper understandings of randomness and independence. The function of the sample space here was not the point of discussion, only that well-developed understandings of sample space is associated with a well-developed understanding of randomness and independence. This study did not purposefully look for causality between sample space and randomness or sample space and independence or sample space and both randomness and independence. Currently, the study has only noted that there appeared to be an association. However, based on the literature available about sample space it appears that sample space is a tool to understanding other probability topics. Thus, the association between sample space and other big ideas seemed logical. It remains to be seen if the process of learning how to build a complete sample space can cause an increase in the understanding of randomness or independence or if the ability to build a complete sample space simply equips students with tools to continue investigating randomness and independence at a deeper level.

This study found that randomness, independence, and sample space should not be taught in isolation because they are not understood in isolation. The translation to classroom instruction is that randomness and independence should be discussed when students are learning about sample space and vice versa. The interconnectedness of the topics based on the way students responded emphasis the importance of the teaching and learning of the topics collectively rather than in isolation. Teaching the topics appears to play to the students' natural understandings of randomness, independence, and sample space. Thus, teachers should look for tasks that utilize two or three of the big ideas at a time to help students build a collective, deeper understanding of randomness, independence, and sample space. Furthermore, the grouping ability of the LUM provides teachers the ability to better predict what aspects of probability students still need in order to deepen their probability literacy. The ability to group students is referencing the ideas that the beginner students all appeared to still be developing their understanding of sample space. Namely, the Beginner group still needs to learn methods to build complete sample spaces. The Intermediate group saw a dip in their independence understanding and teachers can respond to this dip with a focus on supporting independence more in lessons. These are two examples of how the grouping ability that is possible because of the LUM can positively impact classroom teachers.

Students reliance on frequency rather than recency would imply that classroom teachers would need to use tasks, such as the ones used in this study, activities, and discussions around random events that force students to look at the frequency of the event. Any activities and discussions should mimic the tasks used in this study in that the activities should access multiple big ideas of probability rather than focus on a single component. Students would need training and support to learn the balance between the meaning of frequency of an event and the understanding of what it means for the event to be random. The findings from this study imply that running trials and pressing students about justifying their answers forces students to further develop their thinking. As a part of this study, when students were pressed about their understanding they would provide feedback that discounted recency or lead the student to change their response. Therefore, requiring students to justify answers or explain their reasoning should be a requirement of all the learning. It provides an avenue to assess student understanding. Additionally, it forces students to align their thinking with their results and then adjust for any misalignment.

Justifications

As stated before, the Advanced group of students held the ability to build student generated examples for justification which demonstrated a deeper understanding of randomness, independence, and sample space. The perceived deeper understanding was connected to the ability to transfer understanding between multiple contexts. For instance, one student created a parallel situation to the Hot Hand task using the birth of babies and not knowing their gender. The ability to created good parallel examples appears to coexist with the deeper understandings of the knowledge components of probability. It should be noted that in some of the interviews that followed the introduction of the baby example, students would be asked about the baby example as a follow up question to the Hot Hand task. It was observed that the baby example could be developed into a good task as it has the same premise of the Hot Hand task but removes any possible discussion of skill.

Using parallel scenarios in the classroom as a teaching methodology would be twofold. Primarily, the ability to create a parallel scenario provides insight into a student's understanding of the material. Several studies, including Fischbein, Nello, and Marion (1991) have shown that students struggle justifying their solutions but when students do provide parallel scenarios it is indicative of a deeper understanding of the topic at hand. Therefore, the ability to create a parallel example provides proof that student understandings the topic and teaching students to build parallel scenarios might also help move students from the Intermediate group to the Advanced group.

At a secondary level, the parallel scenario provides additional discussion between the teacher and student. For instances, the introduction of the baby example as a parallel to the Hot Hand task provided further discussion with the student about the same topic. The discussion about the Hot Hand task was exhausted as the student was not providing additional information upon further questioning. Then, the baby example was introduced by the student; three boys born in a row, what would the fourth child be, boy or girl? The student in question here was enrolled in a nursing program at her high school, so biological examples might have been more accessible to her. The parallel scenario elicited more discussion with the student that led to greater insight into her understanding of both randomness and independence. An interesting aspect for consideration for future research would be finding good methods to elicit parallel scenarios to engage students in deeper discussions. The situation being described here appears to be problem posing, which could open another area into studying probability literacy. Alternatively, it would be equally useful to determine if parallel scenarios are only a response from a student that understands the content. Thus, the implication would be that a student's ability to present a parallel scenario is an indicator of understanding.

Language and Context of Probability

Language and context were outside the scope of my study, however there were some noteworthy impressions that were noticed throughout the interviews. These impressions include a variety of unexpected discussions and outcomes that took place when interviewing students. One student initially had great trouble with the Black and White task due to a misinterpretation of the word "die". He thought that the task was referencing the dyeing of cloth and thus was confused how a shirt could be randomly dyed white or black. The situation reveals that language literacy of the student changed the context of the problem in ways that were not anticipated. Furthermore, it took a couple of minutes to resolve the concern as it was not immediately obvious to the researcher what the student was confused about.

Another unexpected outcome was the result of the presentation of alternate scenarios for the Hot Hand task where the basketball player had alternated between hits and misses and another where the basketball player had missed four, made three and the eighth shot was in question. A student's answer to the scenarios was not the unexpected part, rather her reaction to the idea of different scenarios where she noted, "…way it was worded it said he's made three consecutive ones…" The student's concern of the wording of the "last three" shots lead the student to create justification of why the answer was different for the different scenarios. The student appeared to be relying on the language and context of the problem to formulate a solution and to justify that solution.

The affect that language appeared to have on students' understanding of the different tasks was not a point of focus of the study but there are several examples were the language of the problem may have affected the response by the student. The effect appeared to be both internal, or how the students interpreted the task, and external, how students conveyed their understanding to the researcher. For instance, it was unclear at times which perspective students were speaking from, themselves or how other people would view an event.

Ultimately, all data gathered from the students about their understanding was a perspective issue. The examples discussed above are presented as a means to point out that the skills that students choose to deploy and their interpretation of the outcomes is at best partially effected by the language and context of the task and the situation. There were examples where students provided different answers to parallel task with essentially the same scenario but different context, such as the the Hot Hand task and the Fake Flip task. Both tasks were 50% outcome events, but the context lead several students to view them as very different and adjust the outcomes accordingly. Language and context were factors at all times in student understanding.

Sample Space Usage

Sample space appears to be a viable tool to assist students in their understanding of independence and randomness. Therefore, students should be taught how to build and use sample spaces. One major observation about sample is that the students that successfully used any sample space, generated or one presented by the task, held the ability to build said sample space systematically. Therefore, it should be a goal to equip students with the ability to build sample spaces systematically. The research does not go so far as to make recommendation of how this might best be done, as that was outside the scope of the study. But, the findings from this study are supported by other studies that encourage the use of sample space for developing understanding of independence (Jones, 2005; Watson & Callingham, 2014). Additionally, the study does not present any best practices on teaching students how to take advantage of samples spaces for the same reason as previously stated. However, both topics would be excellent topics for future research. It appears that coupling sample space in discussions with independence and randomness makes the conversations richer and might build stronger relational understanding for the students. Watson and Callingham (2014) noted similar findings in a study with seventh and eighth grade students they conducted using two-way tables. It was found that the skills used to dissect the data presented in a table help students with understanding independence (Watson & Callingham, 2014). Batanero, Chernoff, Engel, Lee, and Sánchez (2016) found that having students build a sample space can develop understanding of independence. Therefore, regardless how it is used, sample space is a useful tool in helping students develop their understanding of independence. Additionally, this study found that understandings of sample space and randomness appear to develop in parallel, thereby further emphasizing the link between the development of randomness and independence.

An argument could be made that the three Big Ideas of randomness, independence, and sample space appear to be in constant conflict with each other for student understanding. The Intermediate group's lower LUM scores in apparent independence understanding was an example of how the Big Ideas can interact with each other. As stated before, it was possible that the students' efforts to use sample space to justify answers might have caused some confusion about independence. The creation of confusion appears to cause a drop in understanding for students. However, the deeper issue may be that independence understanding needs to be better supported while introducing and initially using sample space to prevent a drop in understanding of independence. The Intermediate group did have higher LUM scores in perceived understanding of randomness and sample space which implies that sample space understanding progresses probability literacy, while simultaneously creating additional conflict for students to resolve. This fact does not mean that sample space is an inhibitor nor that sample space should not be taught to high school students. Rather, independence should be supported when students are receiving instruction about sample space.

Furthermore, independence should be reexamined within the context of sample space to better support the development of both.

Limitations of Study

There were ten students that took part in this study, all of whom volunteered to be interview. There was a gift card given as a part of their participation and most of the students also got to miss roughly an hour of their mathematics class as result of their participation. The students reported good grades in their mathematics classes. Thus, the understandings outlined here may not represent all high school students. The research appears to be exhaustive in that by the end of the study there was a great deal of repetition in types of student responses both in solutions and justification. However, due to the size and scope of the sample used, the study does not allow for extensive generalization. The study needs to be repeated with other populations to test the robustness of the LUM and to verify the findings.

Future Research

The next stages of research should look to verify the results found in this study as well as extend the findings to include other aspects of probability literacy that were observed passively or set aside entirely in this study. The LUM should be extended to include all aspects of probability literacy through repeating this study, or one similar to it, with an adjusted focus on context and critical questions to complete the knowledge elements and critical stance, belief and attitudes, and personal sentiments regarding uncertainty and risk to complete the disposition elements. The collection of remaining objects appears to be larger than what has already been completed, therefore it might be more effective to break the remaining pieces into multiple smaller studies to complete the LUM and have a complete collection of high school students' probability literacy.

The remaining knowledge elements of context and critical questions create other areas of special interest mildly departed from probability literacy on a whole such as what affect does the language have on context or the reverse. One student had difficulty with the Black and White task because the language of the problem led him to believe that the task was dying something black or white at random and he was unsure how to deal with dying something white after it had been dyed black. A simple language issue lead to a large context issue. This situation was not isolated to the one student although his was possibility the most notable example. Therefore, a study solely focused on the language aspect of the task might reveal large insight into a student's understanding with direct impact on gauging their contextual understanding of probability.

There appears to be a relationship between sample space and a student's understanding of independence or randomness. There are several aspects that need to be addressed still that could be addressed in conjunction with probability literacy and the LUM. Namely, how does a student's understanding of sample space affect the understanding of randomness and independence? This study found that the understandings increased in parallel but this study did not exam the interaction between sample space and randomness or independences. Deeper and more meaningful research should be done on this relationship at the high school level.

Conclusions

High school students appear to have a significant intuitive understanding of probability. The LUM showed that many students are well past primary understanding of probability by the time they are ready to leave high school. One suggestion that can be drawn from this study is that high school students are capable of having deeper exposure to probability. Therefore, probability literacy should have a more prominent role in schools and society as a whole. Despite their current understanding, students still have room for improvement in the concepts of randomness, independence, and sample space. Additionally, teachers and curriculum writers need to place greater importance on probability literacy.

As stated before, the current and an expanded form of the LUM combined with well-designed tasks could serve as formal assessments as well as teaching tools for high school teachers. The White and Black task that was discussed earlier contrasting likelihood to expected value would be an example of such teachable situations. One issue to be concerned with would be training for classroom teachers to properly make use of the LUM and tasks as they relate to their use as a formal assessment and a teaching tool in the classroom. When used properly, the LUM and corresponding tasks would provide great insight for teachers about how their students understand probability topics. This information would assist teachers in developing richer, more relevant lessons to engage students in probability education.

Lastly, the LUM in its current state allowed students to be classified into three groups: Beginners, Intermediate, and Advanced. The purpose of the classifications was intended to be able to group students, at least on paper, by their level of probability literacy. The ability to classify students by ability exist in other disciplines, such as reading levels, in high school but did not exist for probability literacy. The LUM filled this gap in the world of probability for the big ideas: randomness, independences, and sample space. The expanded LUM would do the same for all aspects of probability literacy.

References

Albert, J. H. (2003). College students' conceptions of probability. *The American Statistician*, *57*(1), 37– 45.

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

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

Batanero, C., Chernoff, E. J., Engel, J., Lee, H. S., & Sánchez, E. (2016). *Research on teaching and learning probability*. Retrieved from https://link-springer-com.librarylink.uncc.edu/book/10.1007/978-3-319-31625-3

Berenson, S., Dawkins, K., Blanton, M., Coulombe, W., Kolb, J., Norwood, K., & Stiff, L. (1999). *Proceedings of the Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education Volume 1 (20th, Raleigh, NC, October 31-November 3, 1998)*. Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Bossé, M. J. (1995). The NCTM standards in light of the new math movement: A warning! *The Journal of Mathematical Behavior, 14*, 171–201. doi: 10.1016/0732-3123(95)90004-7

Bryant, P., & Nunes, Terezinha. (2012). *Children's understanding of probability: A literature review (full report).* London, England: Nuffield Foundation.

Chernoff, E. J. (2009). An investigative lens. *The Journal of Mathematics Behavior, 28,* 19-29.

Chernoff, E. J. & Sriraman, B. (2014). *Probabilistic thinking: Presenting plural perspectives*. Dordrecht, Netherlands: Springer.

David, F. (1962). *Games, Gods and Gambling; The Origins and History of Probability and Statistical Ideas from the Earliest Times to the Newtonian era*. New York, NY: Hafner Pub. Co.

English, L. D., & Watson, J. M. (2016). Development of probabilistic understanding in fourth grade. *Journal for Research in Mathematics Education*, *47*(1), 28. doi: 10.5951/jresematheduc.47.1.0028

Falk, R. (1986). Conditional probabilities: insights and difficulties. *Proceedings of the Second International Conference on Teaching Statistics* (pp. 292–297).

Falk, R. (2014). Monty's dilemma with no formulas. *Teaching Statistics*, *36*(2), 58–61.

Falk, R. (n.d.). Conditional probabilities: Insights and difficulties.

Falk, R., & Wilkening, F. (1998). Children's construction of fair chances: Adjusting probabilities. *Developmental Psychology*, *34*(6), 1340.

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

Fischbein, E., Nello, M. S., & Marino, M. S. (1991). Factors affecting probabilistic judgements in children and adolescents. *Educational Studies in Mathematics*, *22*(6), 523–549.

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 70,* 1-25. doi: 10.2307/1403713

Gal, I. (2005). Towards" probability literacy" for all citizens: Building blocks and instructional dilemmas. In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 39–63). New York, NY: Springer.

Goldin, G. A. (2000). A scientific perspective on structured, task-based interviews in mathematics education research. In A. E. Kelly & R. A. Lesh (Eds.), *Research design in mathematics and science education* (pp. 517-546). Mahwah, NJ: Lawrence Erlbaum Associates.

Hald, A. (1990). *A history of probability and statistics and their applications before 1750*. New York, NY: Wiley.

Hirschberg, N. (1980). Differences in social judgement. In M. Fishbein (Ed.), *Progress in social psychology* (pp. 73-126). Hillsdale, NJ: Lawrence Erlbaum Associates.

Horvath, J., & Lehrer, R. (1998). A model-based perspective on the development of children's understanding of chance and uncertainty. In S. P. Lajoie (Ed.), *Reflections on statistics: agendas for learning, teaching, and assessment in K-12* (pp.121-148). New York, NY: Routledge.

Jones, G. A. (2005). *Exploring probability in school: challenges for teaching and learning*. New York, NY: Springer.

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 Publishing.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics*, *32*, 101–125. doi: 10.1023/A:1002981520728

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1999). Students' probabilistic thinking in instruction. *Journal for Research in Mathematics Education*, *30*, 487-519. doi: 10.2307/749771

Leavy, A., & Hourigan, M. (2014). Motivating inquiry in statistics and probability in the primary classroom: Motivating inquiry in statistics and probability in the primary classroom. *Teaching Statistics: An International Journal for Teachers*, *37*, 41-47. doi: 10.1111/test.12062

Lodge, J. M., Alhadad, S. S. J., Lewis, M. J., & Gašević, D. (2017). Inferring learning from big data: The importance of a transdisciplinary and multidimensional approach. *Technology, Knowledge and Learning*, *22*, 385–400. doi: 10.1007/s10758-017-9330-3

Kaplan, J. J., Rogness, N. T., & Fisher, D. G. (2014). Exploiting lexical ambiguity to help students understand the meaning of random. *Statistics Education Research Journal 13*, 9-24.

Kazak, S., Wegerif, R., & Fujita, T. (2015). Combining scaffolding for content and scaffolding for dialogue to support conceptual breakthroughs in understanding probability. *ZDM*, *47*, 1269–1283. doi: 10.1007/s11858-015-0720-5

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education*, *24*, 392. doi: 10.2307/749150

Metz, K. E. (1997). Dimensions in the assessment of students' understanding and application of chance. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 223–238). Washington, DC: IOS Press.

Mertens, D. M. (2015). Chapter 8: Qualitative methods. In H. Salmon (Ed.), *Research and evaluation in education and psychology* (pp. 235-277). Thousand Oaks, California: Sage Publications, Inc.

Nacarato, A. M., & Grando, R. C. (2014). The role of language in building probabilistic thinking. *Statistics Education Research Journal*, *13*, 93-103.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

National Governors Association Center for Best Practices, & Council of Chief State School Officers (NGACBP, CCSSO). (2010). *Common Core State Standards for mathematics*. Retrieved from http://www.corestandards.org/Math/Content/HSS/CP/

Patton, M. Q. (2002). Part 2: Qualitative design and data collection. In C. D. Laughton (Ed.), *Qualitative research and evaluation methods* (pp. 207-428). Thousand Oaks, California: Sage Publications, Inc.

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

Pollatsek, A., Well, A. D., Konold, C., Hardiman, P., & Cobb, G. (1987). Understanding conditional probabilities. *Organizational Behavior and Human Decision Processes*, *40*, 255–269.

Ross, S. M. (2000). *Introduction to probability models* (7th ed.). San Diego, CA: Academic Press.

Rubel, L. H. (2007). Middle school and high school students' probabilistic reasoning on coin tasks. *Journal for Research in Mathematics Education*, *38*, 531–556.

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

Skemp, R. R. (2006). Understanding and instrumental understanding. *Mathematics Teaching in the Middle School,12*, 88-95.

Wasserman, N. H. (2015). A random walk: Stumbling across connections. *The Mathematics Teacher*, *108*(9), 686. doi: 10.5951/mathteacher.108.9.0686

Watson, J. M. (1997). Assessing statistical literacy through the use of media surveys. In I. Gal & J. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 107-121). Amsterdam: IOS Press.

Watson, J., & Callingham, R. (2014). Two-way tables: issues at the heart of statistics and probability for students and teachers. *Mathematical Thinking and Learning*, *16*(4), 254–284. doi: 10.1080/10986065.2014.953019

Watson, J. M., & Kelly, B. A. (2007). The development of conditional probability reasoning. *International Journal of Mathematical Education in Science and Technology*, *38*(2), 213–235. doi: 10.1080/00207390601002880

Watson, J. M., & Moritz, J. B. (2002). School students' reasoning about conjunction and conditional events. *International Journal of Mathematical Education in Science and Technology*, *33*(1), 59–84. doi: 10.1080/00207390110087615

Zahner, D., & Corter, J. E. (2010). The process of probability problem solving: use of external visual representations. *Mathematical Thinking and Learning*, *12*(2), 177–204. doi: 10.1080/10986061003654240

Appendix A: Comparison of Gal and Jones Frameworks

	Jones, Langrall, and Mooney (2007)	**Using probability to solve problems, both theoretical and experimental**	**The use of simulations to approximate probabilities**	**The use of random variables and interpreting probability distributions, including the normal distribution**	**Apply random variables to generate and interpret distributions such as the binomial, uniform, normal, and chi-square**
Corresponding Gal (2005) items	**Gal (2005) - knowledge elements**	Big ideas: randomness, independence, variation, predictability/uncertainty, (sample space)	Figuring probabilities: ways to find or estimate the probability of events	Context: understanding the role and implications of probabilistic issues and messages in various contexts and in personal and public discourse	Context: understanding the role and implications of probabilistic issues and messages in various contexts and in personal and public discourse
Corresponding Gal (2005) items	**Gal (2005) - knowledge elements**	Figuring probabilities: ways to find or estimate the probability of events	Critical questions: Issues to reflect upon when dealing with probabilities.	Critical questions: Issues to reflect upon when dealing with probabilities.	Critical questions: Issues to reflect upon when dealing with probabilities.
Corresponding Gal (2005) items	**Gal (2005) - knowledge elements**	Context: understanding the role and implications of probabilistic issues and messages in various contexts and in personal and public discourse			Language: the terms and methods used to communicate about chance

Corresponding Gal (2005) items	**Gal (2005) - knowledge elements**	Language: the terms and methods used to communicate about chance			
Corresponding Gal (2005) items	**Gal (2005) - Dispositional elements**	Critical stance	Personal sentiments regarding uncertainty and risk	Personal sentiments regarding uncertainty and risk	Critical stance
Corresponding Gal (2005) items	**Gal (2005) - Dispositional elements**	Beliefs and attitudes	Critical stance	Critical stance	Personal sentiments regarding uncertainty and risk
Corresponding Gal (2005) items	**Gal (2005) - Dispositional elements**	Personal sentiments regarding uncertainty and risk			

Table 1 – The columns are Jones, Langrall, Thornton, and Morgill's (1997)'s framework. The rows are the instances where Gal's framework runs parallel with Jones, Langrall, Thornton, and Morgill's (1997).

Appendix B: Original List of Tasks Considered for the Pilot

1. A die is painted white on one side and black on the other five sides. If the painted die is rolled six times, which of the following two outcomes is most likely? a) Black side up on five rolls and white side up on the other role. b) Black side up on all six rolls. (Konold, 1993, pp. 393)
2. Which of the following is the most likely result of the five flips of a fair coin? a) HHHTT B) THHTH C) THTTT D) HTHTH E) All four sequences are equally likely. Which of the above sequences would be least likely to occur? (Konold, 1993, 397)
3. A chip colored red on one side, white on the other is flipped repeatedly, landing with the red side facing up twice in a row. Which outcome is most likely for the 3rd flip: red or white. 2007_Jones. This task could easily be completed as a hands on task
4. "R1 "a 5 and a 6 are obtained" and R2 "a 6 is obtained twice". The question asked is, "Do you think the chance of obtaining each of these results is equal? Or is there more chance of obtaining one of them, and if so, which, R1 or R2? Or is it impossible for you to give an answer, and if so, why?" (Lecoutre, 1992).
5. It is more likely to have the same number appear on a roll of two dice or different numbers? (1991_Fischbein)
6. Question AS: "Considering the sum of the points obtained when rolling a pair of dice, will you bet on 3 or on 6? Why?" (Fischbein et al, 1991)
7. Question B5: "Considering the sum of the points obtained when rolling a pair of dice, would you bet on 7 or on 10? Why?" (Fischbein et al, 1991)

8. Modified: Consider the sum of obtained when rolling a pair of dice, what is the sample space?
9. Can I do the fake vs. real coin flip list? Dice roll?
10. Gill has participated in a weekly lottery during the last two months. So far he has never won but he has decided to go on for the following reason: "Lottery is a game based on chance, sometimes you win sometimes you lose. I have already played many times and I have never won. Therefore, I am sure more than before that I shall win in one of the next games". What is your opinion with regard to Gill's explanation? (Fischbein et al, 1984)
11. Gilla is 10 years old. In her box, there are 40 white marbles and 20 black ones. Ronit is 8 years old. In her box there are 30 white marbles and 15 black ones. Each of them draws one marble from her own box, without looking. Ronit claims that Gilla has a greater chance of extracting a white marble because she is the older one, and therefore she is the cleverest of both of them. What is your opinion about this? (Fischbein et al, 1984
12. Shula once filled up a lottery form with the following numbers: 1; 7; 13; 21; 22; 36; and she won. Therefore, she claimed that she must always play the same group of numbers, because it turned out to be a lucky one. What is your opinion about this? (Fischbein et al, 1984)
13. Ruth prefers, when she participates in a lottery, to choose consecutive numbers like 1, 2, 3, 4, 5, 6. She claims that in this way she increases her chance of winning. On the other hand Jenny claims that the chance of getting six consecutive numbers like 1, 2, 3, 4, 5, 6 is smaller than that of getting a random

sequence of numbers. She says that a lottery is something chancy and therefore there is no chance of getting a sequence of consecutive numbers. What is your opinion with regards to the two attitudes, that of Ruth and that of Jenny?(Fischbein et al, 1984)

Appendix C: Probing Questions

General Questions:
Explain what you did to get this answer.
What was the process?
How did you arrive at that solution?
What strategies did you choose?
What lead you to choose that your strategies?
Explain the key to working this problem.
Why did you choose those processes?
What is another way you could have done it?
What are the different components of this task?
List all possible outcomes.
How does the order of the selections affect the outcome?
Explain what criteria you used to build the list of outcomes.
What happens to the probabilities as the items in the list of outcomes increases?
How does list influence outcomes?
Explain how the outcome of one event does or does not affect the outcome of another.
Explain how this situation is or is not random.
What is randomness?
What did you use to determine if this was a random situation or not?
What makes this a random event?
What would prevent it from being random?
How long will I need to wait to win the lottery?
What is unbiased?
How does consistency in play lottery affect winning?
What would be the next outcome if we were to extend the trial one more time?
What allows you to make that prediction?
What prevents you from making a prediction?
Explain related or unrelated the events are to each other.
What makes the events here related or unrelated?
What would need to change to prevent it from the events from being related?

Follow up questions:

Explain your current understanding of probability.
Please use one of the tasks we have worked to explain.
Please use a personal example to explain.
What type of exposure do you have in probability?
Explain how you use probability in your everyday life.
What is independence? (I might need to use common terminology in place of mathematical terminology.
What is randomness?
Explain the difference between independence and randomness.
What is sample space?

Appendix D: First Iteration of the LUM

	Level 1 (L1)	Leve 2 (L2)	Level 3 (L3)	Level 4 (L4)
Sample Space	Students makes incomplete list(s), does not use list(s) to make decisions	List set but not in a systematic manner, 1 & 2 stage sets*	Adopts a strategy to create complete list or uses an incomplete list to make decision	Adopts strategy to build list and uses it to solve problem make decisions
Independence	Student makes predication for subsequent outcomes on previous outcomes rather than current situation	Student can distinguish how probability does or does not change based on replacement	Student recognizes that probability does not change in replacement sets and does change in non-replacement sets	Student can identify an independent event and can assign numerical probability to independent events
Randomness	Student makes predictions based on: patterns, the most recent occurrences, or haphazard models, or created regular distribution, frequencies are not mentioned	Student piratically make predictions based on: patterns, the most recent occurrences, or haphazard models, or can create a mostly irregular distribution and "Qualitative intuitions of proportionality" (Piaget, 1977, pp. 54)	Student makes predictions mostly based on frequencies or can create distribution that mostly irregular or partially uniform distributions and sometime uses "Qualitative intuitions of proportionality" (Piaget, 1977, pp. 54)	Student makes predictions based on frequencies or can create an uniform distribution and can communicate that randomness is an unordered list of outcomes, and makes extensive use of "Qualitative intuitions of proportionality" (Piaget, 1977, pp. 54)
*stage sets refers to how students deal with compound events such as two dice being rolled at the same time				

Appendix E: Second Iteration of the LUM

	Level 1 (L1)	Level 2 (L2)	Level 3 (L3)	Level 4 (L4)
Sample Space	Students makes incomplete list and does not use list to make decisions regardless of prompting	Adopts a strategy to create list of outcomes but must be prompted to do so but cannot use it to solve problems.	Adopts a strategy to create complete list and uses the complete list to solve problems or to make decision, but must be prompted to use the list or creates an incomplete list but does not need prompting to create it or use it to solve problems.	Adopts a strategy to create complete list and uses list to solve problems or to make decision and student does so unprompted.
Independence	Makes prediction for subsequent outcomes on previous outcomes rather than current situation	Student can identify an independent event and can assign numerical probability with or without replacement, or can provide correct justifications for answers, or does not allow representativeness to sway decision or calculations about independent events (one must be met)	Student can identify an independent event and can assign numerical probability with or without replacement, or can provide correct justifications for answers, or does not allow representativeness to sway decision or calculations about independent events (two must be met)	Student can identify an independent event and can assign numerical probability with or without replacement, or can provide correct justifications for answers, and does not allow representativeness to sway decision or calculations about independent events. (all 3 must be met)
Randomness	Makes predictions based on: pattern or haphazard models. Regular disruptions or frequencies are not mentioned. Evidence that student makes decisions about randomness based on either	Partially make predictions based on: patterns, the most recent occurrences, or haphazard models, or can create a mostly irregular distribution and little use of "Qualitative intuitions of proportionality" (Piaget, 1977, pp. 54), some evidence that positive or negative recency influences decisions.	Make predictions mostly based on frequencies or can create distribution that mostly irregular or partially uniform distributions and sometime uses "Qualitative intuitions of proportionality" (Piaget, 1977, pp. 54), little evidence that positive or negative recency influences decisions.	Make predictions based on frequencies or can create regular or uniform distribution for decisions and can communicate that randomness is an unordered list of outcomes or "individual outcomes are uncertain" (Kaplan, 2014) and makes extensive use of "Qualitative intuitions of proportionality"

	positive or negative recency.			(Piaget, 1977, pp. 54), no evidence that positive or negative recency influences decisions.
*stage sets refers to how students deal with compound events such as two dice being rolled at the same time				