TEACHING STUDENTS TO ESTIMATE PROBABILITIES: THE FREQUENTIST APPROACH AND ITS RELATIONSHIP WITH STATISTICAL UNDERSTANDING.

Steven J. Boyce
Virginia Tech
sboyce@vt.edu

Jesse L.M. Wilkins
Virginia Tech
wilkins@vt.edu

Beth Loveday MacDonald
Virginia Tech
bmacdona@vt.edu

An interview with a sixth-grade student illustrates how her number sense and understanding of variability relate to her ability and proclivity to apply a frequentist (statistical) approach to probability tasks. A general suggestion for teaching about mathematics of uncertainty through the gradual strengthening of estimation, as per the historical development, is also discussed.

Theoretical Background

The Three Views of Probability

There are three main views of probability: classical, frequentist, and subjective (Shaughnessy, 1992). Using the classical view, one first partitions a sample space into equally likely outcomes. The probability of an event is simply the ratio of the number of outcomes in which that event occurs to the total number of outcomes. In contrast, the frequentist approach to probability involves repeated trials. A person using a frequentist approach might conduct a simulation with a large number of trials, examine the data, and assert probabilities based on the observations. If the number of trials is large enough, and if the results are repeated in other contexts, then the probability is judged reliable. A classical approach examines *a priori* how different arrangements of events could happen in order to develop a uniform distribution model. The frequentist approach is mathematically more related to statistics, since it involves the search for a distribution and subsequent application of the distribution's properties. Thus the mathematics behind the frequentist approach tend more to the notions of limits and convergence, as relating to the law of large numbers (Shaughnessy, 1992). The third view of probability, the subjective view, also takes into account an individual's own knowledge, opinions, or feelings. Reliance upon subjective reasoning may signal misconception, lack of confidence, or uncertainty of the relevant mathematics, but its use is not necessarily irrational. A child might always express a favorite color to be most probable on a spinner, while some situations, such as the probability of a Mars landing in the next century, can only be estimated by a subjective approach (Jones, Langrall, & Mooney, 2007).

Development of the Mathematization of Statistics and Probability

The oldest examples of statistical thought each related to the concept of estimation (Bakker, 2003). Examples from Indian, Egyptian, and Greek stories contained phenomena similar to the *mode*, *mean*, and a measure called the *midrange.* For data that has a symmetric distribution, the mean, median, mode, and midrange all coincide, so there is no need for their distinction. Bakker (2003) found, using classroom teaching experiments, that modern-day students also benefited from beginning with estimation while learning measures of center. It was not until students were faced with the task of computing an average with non-symmetric data that they felt the need to develop and formalize other methods of average.

The parallel between the historical development of average and the historical development of probability is the original expectation of symmetry (or uniformity), the subsequent adjustment to increase the accuracy (or number of successes), and the reliance upon estimation. The oldest manuscript describing observed frequencies and non-uniform distribution was written in the 13th

Wiest, L. R., & Lamberg, T. (Eds.). (2011). Proceedings of the 33rd Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education. Reno, NV: University of Nevada, Reno.

century, while the first known and solved probability problem (by Galileo) took place about 1620 (Batanero, Henry & Parzyz, 2005). The conflict between theoretical calculation and observed frequencies is what led to the development of more rigorous combinatorics methods (Batanero et.al., 2005). A teaching experiment using non-uniform dice by Nilsson (2009) showed that students went through a similar process when their previous conception of uniformity was challenged by observation. When the amount of variation was too great, the students reexamined their assumptions based on empirical data and modified their theoretical model.

Probability and Statistics Frameworks

Jones, Langrall, Thornton, and Mogill (1999) developed a framework for the development of probabilistic thinking through the middle grades by observing students' responses to tasks that could be solved using classical probability. The framework consists of four levels: subjective, transitional, informal quantitative, and numerical. At level one, a student might only be able to recognize certain or impossible events; at level two, most or least likely events. By level three, students' quantitative reasoning and measures are used to describe likelihood. By level four, students are able to assign numerical probabilities (Jones et al., 1999). As a student's level increases, the tendency to use subjective judgments decreases. Polaki (2002) validated this framework in a study in South Africa, and found that the highest levels of probability thinking require part-whole reasoning, while students without even part-part reasoning generally operate in the subjective level.

Watson, Collis, and Moritz (1997) describe a complementary framework of probabilistic thought. There is a hierarchy of four levels: prestructural, unistructural, multistructural, and relational, and students must pass through two cycles within the levels to achieve the highest realm of probabilistic thinking. The first cycle involves the development of probability as a measure, while the second cycle involves the development of that measure. Considering that the second cycle requires part-part reasoning for the multistructural level and part-whole for the relational level (Watson et al., 1997), this view is compatible with that of Jones et al. (1999). The additional levels of the framework more clearly describe the phenomena of students' demonstrating a relational level of reasoning for a single task, while a subjective level for a multiple-part task.

Both of these frameworks support the view that probabilistic instruction should begin with part-part comparisons instead of part-whole relationships. However, the classical view of probability necessitates part-whole comparisons in order to declare a uniform distribution (at least implicitly). Since comparing frequencies of outcomes is essentially performing a statistical task, perhaps further insight into the requisite knowledge for learning probability can be obtained by studying existing frameworks for the learning of statistics. According to Shaughnessy, "there are important connections between probability and statistics, particularly when repeated trials of probability experiments generate a distribution of possible outcomes" (2007, p. 981).

A relevant connection is the interaction between students' understanding of expectation and variability. Watson, Callingham, and Kelly (2007) developed a framework for the understanding of expectation and variation with six levels ranging from idiosyncratic, with little or no appreciation of either variation or expectation, to comparative distributional, in which links between variation and expectation are established in comparative settings with proportional reasoning. To compute simple probabilities with the frequentist approach, only the fifth level of statistical reasoning is needed: understanding of the relationship between variation and expectation within a single context. It is at this level that students can articulate an expectation; at levels four and below, they are able to articulate only comparative aspects such as more or less likely. Thus the higher levels of this framework are distinguished by the use of part-whole reasoning, as was the case with the probability frameworks. In order to use the frequentist approach to *develop* a theoretical distribution, one must determine a permissible amount of variance between observed frequencies and a theoretical distribution, i.e., a search for a signal of variability (Shaughnessy 2007). This requires the highest level of statistical reasoning, for it necessitates the comparison of distributions across two groups: the observed frequency and the expected frequency (Watson et al., 2007).

Purpose

The purpose of this study was to examine the relationships between students' understanding of probability and statistics. The frequentist approach to computing probabilities relies on an understanding of what constitutes an acceptable level of variance. On the other hand, students' expectations for variance may be influenced by an assumption of uniform probability. Thus we sought to ascertain whether the students' level of statistical understanding could help explain their level of probabilistic understanding, and consequently further understand the role of the teaching of statistics and variability in teaching probability.

Methods

An extended clinical interview was conducted with one sixth-grade female from Virginia. The interview was divided into two sessions, each lasting approximately thirty minutes, involving a total of nine tasks. The first session, consisting of the first seven tasks, was designed to gauge the student's level of probabilistic thinking as per the frameworks of Jones et al. (1999) and Watson et al. (1997); questions and activities similar to their released items were used to assess the student's probability thinking. In order to ascertain the student's level of statistical thinking, tasks similar to those described by Watson et al. (2007) and Bakker (2003) were used. Each task was presented orally by the first author who served as the teacher researcher throughout the study. Manipulatives available to the student included physical dice, pencil and paper, and virtual spinners. Following are the tasks presented to the student in the first session.

1. If you were to roll this die (student is presented with a six-sided die), do you think it's easier to roll a one or a six?
2. In mathematics class, there are 13 boys and 16 girls. If a teacher were to write the names of the students on slips of paper and draw one out of a hat, would it be more likely that the name would be that of a boy, that of a girl, or is it equally likely?
3. There are two boxes, Box A and Box B. Box A has six red marbles in it. Box B has 60 red marbles. Box A has four blue marbles, while Box B has 40 blue marbles. If you want to pick a blue marble, from which box should you choose?
4. Create a possible graph depicting the monthly high temperatures in your hometown, given that the average yearly high temperature was 69 degrees.
5. A game is played where two spinners, each 50% black and 50% white, are spun. If they are both black, then person A wins; if they are different colors, person B wins. Which person would be more likely to win?
6. Estimate the number of penguins in a photograph in which the penguins are not of uniform size.
7. Ten Twizzlers of different colors are placed in a bag, and the student is asked to guess which color is most likely to be drawn out.

Each session was videotaped, and the authors planned additional tasks after discussing their individual interpretations of the video. The second session consisted of an extension of Task 4 and a new task that merged two dice throwing tasks, the first by Watson and Kelly (2007) and the second by Nilsson (2009). These were designed to explore the role of variance in the student's perception of the likelihood of outcomes.

The focus of the first session was that of a clinical interview: both the questions and the sequence of tasks were chosen in advance, and no attempt to teach the student was made. The second session was more of a teaching experiment (Steffe, 1991), as the student was encouraged to examine her notions of variance and expectation in the following tasks.

8. Create and critique possible graphs depicting the average yearly temperature.
9. Predict the results of rolling a six-sided die 60 times. The dice in question were (a) fair, (b) had two ones and no sixes, (c) were "loaded" in favor of ones.

Results

Probability Tasks Results

The student response to the first task was that it was equally likely to roll a one or a six: "because there are the same number of sides, so I don't think it really matters the number."

The second task response was "a girl, because there's more girls than boys." This correct response using part-part reasoning indicates achievement of the multi-structural level of Watson's framework (Watson et al., 1997) and the transitional level of Jones' framework (Jones et al., 1999).

According to Watson et al. (1997), a correct response to the third task is associated with the relational level of reasoning. The conversation associated with this task proceeded as follows:

Student: I would think they would be the same, but, ... maybe Box B just because it has a bigger number?

Interviewer: What makes you think they might be the same?

Student: Because that has 10 and that has 100, so out of, like, 100 percent would be 60 percent for both of them, to 40 percent.

Interviewer: What made you think that maybe Box B would be the one to pick?

Student: Ummm...maybe because it has the bigger number. I don't know.

The student's initial correct response to the task was based on proportional reasoning, which is requisite for the higher levels of probabilistic thinking in both Jones' and Watson's frameworks. The fact that she thought that Box B having a bigger number might make it the one to pick suggested that there might be some confusion over the law of large numbers. It was suspected that the student might believe that having "large numbers" is desirable when computing probabilities, but that she had not yet conceptualized a justification.

The teacher researcher then presented her with the option of simulating the task using Probability Explorer (Lee, 2005) because he wanted her to become familiar with the program for a future task and also wanted to see her reaction to the result of an experiment. Specifically, he wanted to see if the result of the experiment would help her make a decision. The conversation continued:

Interviewer: Which do you think is more likely to come up, red or blue?

Student: Red.

Interviewer: Why?

Student: Because there are more reds than blues.

Interviewer: [Clicking the button to simulate a grab, which was blue.] Why do you think it came up blue instead of red?

Student: I don't know, I'm not sure.

Interviewer: What if I click it again, do you think it will come up blue again, or do you think it will come up red?

Student: I'm gonna guess red. I guess it's a fifty-fifty chance because there's two colors. I don't know.

The inconsistency in the responses here is indicative of the transitional level in Jones' framework. The simulation didn't help her decide an answer - in fact, it encouraged her to even question the part-whole reasoning with which she earlier seemed comfortable. Here the term “fifty-fifty” seemed to mean that it could be either of two outcomes rather than a rejection of the part-part reasoning earlier displayed. This language pattern has often been noted by researchers (e.g., Jones et al, 1999; Shaugnessy & Ciancetta, 2002). The uncertainty of her response seems to indicate a lack of knowledge or confidence in the relationship between her theoretical model and empirical trials.

The student again said that there was a “fifty-fifty” chance that each person would win when faced with the fifth task on spinners. This was not surprising, since it involves a compound event and hence would require the higher levels of probabilistic thinking (Jones et al., 1999). When presented with a simulation in which the spinners came up differently seven times out of ten, the student decided that choosing different colors probably had an advantage. This aligns with Shaughnessy and Ciancetta’s (2002) results which indicated that playing this game and seeing variation could help students reject their equiprobable hypothesis.

The seventh task was also aimed at understanding the student’s level of probabilistic thinking. She estimated that blue would be most likely to be drawn out since there were more blues than any other color. A blue was drawn and not replaced, leaving 2 green, 2 blue, and 2 yellow. She was asked again what color was most likely to be drawn, and she said "blue, green, or yellow" because they had the highest number. When the blues were replaced, she said that the chance of a blue would increase because the number of blues increased. This is evident of a relational understanding (comparing the possibility across two sample spaces).

Overall, the student's level of probabilistic reasoning appeared to be “informal quantitative” in the framework of Jones et al. (1999). She displayed use of part-part reasoning throughout and at times exhibited part-whole reasoning such as percentages. She was able to compare sample spaces and make relational judgments such as “more or less likely,” while at the same time used subjective judgments for compound events. This indicated that she was in the first-cycle relational stage or second-cycle idiosyncratic stage in the framework of Watson et. al. (1997).

Statistics Tasks Results

When creating a scale for her graph in Task 4, the student made the lowest value 30 degrees and the highest 80 degrees “because I think the coldest it would get is around thirty” and “eighty is probably about the highest it gets around here.” Bakker (2003) observed similar behavior when asking students to explain their understanding of “average,” in which several students gave a response indicative of a mid-range concept.

The student seemed to display an expectation for variation in temperatures, as the temperature trends cooler in the winter months and warmer in the summer, but the overall mean of the temperatures she displayed was much lower than 69 degrees. It appeared that she was focused on the variation and the range but not the mean. Based on these initial findings, it was decided to follow up in the next session with other graphs to compare to hers in order to determine her level of understanding of variability and expectation via the framework of Watson et al (2007).

When judging the number of penguins in the sixth task, she originally did not want to guess, but said “two million” upon encouragement. I believe that the hesitation toward calculation was due to the fact that the penguins were not uniform in size and were in rows of non-uniform width. She was offered a ruler, and encouraged to create a way to estimate using proportional reasoning. She counted the number of penguins in the bottom row, estimated visually the number of rows, and multiplied to get an estimate of 1,200 penguins. This was similar to the method that Bakker (2003) found children using to estimate the number of elephants in a picture, although the students in his study first found an “average” block. She made no mention of the fact that the penguins in the row she chose were larger than the penguins in the other rows, even after that error led her to a much smaller estimate than she had originally guessed. Her behavior in this task was opposite of her behavior in the temperature-graphing in which she initially focused on the mean and subsequently on the variation. In the penguin counting she was initially focused on the variation (causing the difficulty in counting) and subsequently ignored the variation when performing the calculation.

In preparation for Task 8, the student was introduced to four graphs of supposed student work and asked to critique them as possible graphs for the average high temperature. The first graph depicted a uniform temperature of 69 degrees, which she decreed unlikely since “February was too warm…and August is too cold.” The second graph was also unlikely, for although the temperatures varied, they did so linearly, which was “too perfect.” However, she did express that the first two graphs were possibilities. The third and fourth graphs both had varying temperatures, and neither graph increased or decreased linearly. However, by putting a pencil across the graph at 69 degrees, the researcher showed her that graph 3 had substantially more data below the line than above the line (this would technically be comparing to the median, but the data is nearly symmetric and hence the mean and median are close). It was thought that this would have led her to believe that it was not a possible graph of the data with mean 69 degrees, and she did make the observation that “it was too low.” She drew a similar line using the fourth graph, and found about the same number of temperatures above and below the line. But when asked whether the third or fourth graphs were more likely, she chose the third because in the fourth graph the change in temperatures was closer to linear. She seemed to value the observed variance and high probability of “randomness” from month to month more than the presence or absence of the desired overall mean.

When faced with the die rolling tasks, the student was first asked to write down her prediction of how many of each number would come up. She said that she expected an equal number, 10, of each of the six values on the dice, which is indicative of level 2 on the scale of variation used by Watson and Kelly (2007), a “strict probabilistic prediction”(p. 3). When she rolled a standard die, six came up 17 times, which she attributed to chance. When asked if she would like to revise her prediction, she declined and indicated that she still thought each number would come up ten times if she rolled again.

The results of sixty throws of the second die (biased in favor of one) included a total of 21 ones and zero sixes. It was expected that the student would think that having no sixes was very unusual, and that she might wonder about the fairness of the die. She said that she thought it was “weird,” but once again attributed the outcome to “chance” and declined to revise her prediction.

The roll of the third die (weighted) resulted in 29 ones, this time with two sixes. Once again, the student was not surprised by the result, attributing it to chance. It was expected that she would immediately question the die, for Nilsson (2009) found that students noticed unexpected frequencies with non-uniform dice and resolved the conflict between theoretical and empirical probability by developing an empirical model.

The teacher researcher felt that it might not occur to the student that some of the dice might be unfair, so she was encouraged to examine the second die more closely. After she noticed that there were two ones and zero sixes, she explained that she was no longer surprised about the presence of 21 ones. “The ten from the sixes went to the ones and 21 is near 20.” She was then asked what might explain the 29 ones in the third trial. She ruled out the possibility that there were any numbers missing with the third die, because she had rolled at least one of each number. The teacher researcher encouraged her to watch the die as it spun, and she noticed that it was “loaded” by the way it landed. When asked to predict the way it was loaded and determine how many ones would be expected, she first suggested that we should expect to get 20 ones, just as in the second problem, but was unable to articulate a reason for this prediction. She was then asked, “If I were to roll this die a million times, how could you predict how many ones would come up?” She suggested that we had rolled “almost 50 percent ones,” in the first 60 rolls, so we could “roll it another 100 times, and see if it came up about half ones.” When asked why she chose 100 times instead of more or less than 100, she said “maybe do more, like 300.” She had suggested the use of a frequentist approach to compute probability.

Discussion

The student's performance on the probability tasks showed initial understanding and preference for part-part comparisons and the occasional use of part-whole reasoning when faced with classical probability tasks. However, when engaged with actual events, she showed a lack of confidence in the application of such comparisons to make predictions by rejecting the results of her comparisons and relying instead on subjective judgments. While her performance fell into the “informal quantitative” stage (Jones et al. 1999), the reliance on subjective probability in simulation showed the value of also considering the student's statistical understanding. Her initial demonstration of a “purely probabilistic” approach to variance shows her view of the calculation of probability as a deterministic exercise. Without further statistical understanding of variance, average, and the law of large numbers, she was inconsistent with both the application of probabilities to make predictions beyond a single event and the reconciliation of her prediction with a contrary outcome.

The interactions with this student led us to believe that after the second session, she became more likely to consider a frequentist approach to calculating probabilities of events that can be simulated. This mirrors the historical development of probability, in which theoretical notions of expectation are strengthened or rejected based on the observation of the frequency of outcomes. Because the student had been exposed to classical probability, she originally focused on the mathematical task of comparing outcomes in a uniform sample space. The performance on the tasks showed consistently that her probabilistic reasoning was in transition between focusing on part-part relationships and focusing on part-whole relationships. The statistical tasks also showed that she had difficulty gauging what was a reasonable level of variation, resulting in her subjectivity in and hesitation toward the rejection of a theoretical model based on trial outcomes. When tasks involved calculations, she ignored variation, and when tasks did not involve calculations, she focused on the variation.

Conclusions

Statistics and probability understanding are connected, for the evaluation of probabilistic claims with the frequentist approach requires one to both understand and expect variance in

Wiest, L. R., & Lamberg, T. (Eds.). (2011). Proceedings of the 33rd Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education. Reno, NV: University of Nevada, Reno.

situations of uncertainty. Students who are taught to calculate probabilities using the classical approach may have difficulty reconciling empirical evidence that differs from their calculations. By exposing children to probabilistic situations in which their intuitions and theories are challenged, teachers can encourage them to evaluate their own and then others' claims. Future research may explore in more detail the relationship between the transitions that students make between levels of probabilistic understanding and understanding of variation and expectation.

References

Bakker, A. (2003). The early history of average values and implications for education. *Journal of Statistics Education, 11* (1), 12.

Batanero, C., Henry, M., & Parzyz, B. (2005). The nature of chance and probability. In Jones, G.A. (Ed.) *Exploring probability in school: Challenges for teaching and learning* (pp. 15-32). New York: Springer.

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: Springer.

Jones, G., Langrall, C., & Mooney, E. (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, N.C. : Information Age Publishing.

Jones, G., Langrall, C., Thornton, C., & Mogill, A. (1999). Students' probabilistic thinking in instruction. *Journal for Research in Mathematics Education, 30* (5), 487-519.

Lee, H. S. (2005). *Probability explorer*. [Software]. Available from http://www.probexplorer.com

Nilsson, P. (2009). Conceptual variation and coordination in probability reasoning. *The Journal of Mathematical Behavior, 28* (4), 247-261.

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

Shaughnessy, J.M. (1981). Misconceptions of probability: From systematic errors to systematic experiments and decisions. In A. P. Shulte (Ed.), *Teaching statistics and probability: 1981 yearbook of the National Council of Teachers of Mathematics* (pp. 90-99). Reston, VA: National Council of Teachers of Mathematics.

Shaughnessy, J.M. (1992). Research in probability and statistics: Reflections and directions. In. D. Grouws (Ed.). *Handbook of research on mathematics teaching and learning* (pp. 465-494). New York: Macmillan.

Shaughnessy, J. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.) *Second handbook of research on mathematics teaching and learning* (pp. 957-1010). Charlotte, N.C. : Information Age Publishing.

Shaughnessy, J.M., & Ciancetta, M. (2002). *Students' understanding of variability in a probability environment*. In Proceedings of the sixth international conference on teaching statistics: Developing a statistically literate society. Cape Town, South Africa.

Spinillo, A. (2002). Children's use of part-part comparisons to estimate probability. *Journal of Mathematical Behavior, 21* (3), 357-369.

Steffe, L. (1991). The constructivist teaching experiment: Illustrations and implications. In E. von Glasersfeld (Ed.) *Radical constructivism in mathematics education* (pp. 177-194). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Wiest, L. R., & Lamberg, T. (Eds.). (2011). Proceedings of the 33rd Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education. Reno, NV: University of Nevada, Reno.

Watson, J., Callingham, R., & Kelly, B. (2007). Students' appreciation of expectation and variation as a foundation for statistical understanding. *Mathematical thinking and learning, 9* (2), 83-130.
Watson, J., Collis, K., & Moritz, J. (1997). The development of chance measurement. *Mathematics Education Research Journal, 9* (1), 60-82.
Watson, J., & Kelly, B. (2007). Assessment of students' understanding of variation. *Teaching Statistics, 29* (3), 80-88.

AN INFORMAL FALLACY IN TEACHERS' REASONING ABOUT PROBABILITY

Egan J. Chernoff
University of Saskatchewan
egan.chernoff@usask.ca

Gale L. Russell
University of Saskatchewan
gale.russell@.usask.ca

The main objective of this article is to contribute to the limited research on teachers' knowledge of probability. In order to meet this objective, we presented prospective mathematics teachers with a variation of a well known task and asked them to determine which of five possible coin flip sequences was least likely to occur. To analyze particular normatively incorrect responses we utilized a brand new lens – the composition fallacy – instead of the traditional lenses and models associated with heuristic and informal reasoning about probability. In our application of the new lens we were able to determine that fallacious reasoning, not just heuristic reasoning, can account for normatively incorrect responses to the task. Given the success of the new lens, we contend that logical fallacies are a potential avenue for future investigations in comparisons of relative likelihood and research in probability in general.

The general purpose of this article is to contribute to the paucity of research on (prospective) teachers' knowledge of probability (Jones, Langrall & Mooney, 2007; Stohl, 2005). More specifically, the purpose of this article is to merge the established thread of investigations into comparisons of relative likelihood (e.g., Borovcnik & Bentz, 1991; Cox & Mouw, 1992; Hirsch & O'Donnell, 2001; Kahneman & Tversky, 1972; Konold, Pollatsek, Well, & Lohmeier, & Lipson, 1993; Rubel, 2006; Shaughnessy, 1977; Tversky & Kahneman, 1974; Watson, Collis, & Moritz, 1997) with a developing thread of investigations into prospective teachers' comparisons of relative likelihood (e.g., Chernoff, 2009, 2009a, 2009b).

In order to achieve the general and specific goals detailed above, prospective teachers, as has been the case in past research, were presented with five different sequence of heads and tails – derived from flipping a fair coin five times – and were asked to declare which sequence was least likely to occur. However, unlike previous research, we utilize a brand new lens to account for certain responses; we demonstrate that certain responses fall prey to the fallacy of composition (i.e., because parts of a whole have a certain property, it is argued that the whole has that property). Further, we contend that informal fallacies, in general, create a new research opportunity for those investigating comparisons of relative likelihood.

A Review of the Literature

In mathematics education (and psychology) research, comparative likelihood responses are categorized, in a broad sense, into two particular categories: correct responses and incorrect responses. While correct responses are, for the most part, associated with normative reasoning,