DIFFERENCES IN STUDENTS' USE OF COMPUTER SIMULATION TOOLS AND REASONING ABOUT EMPIRICAL DATA AND THEORETICAL DISTRIBUTIONS

Robin Rider
East Carolina University, United States
Hollylynne Stohl Lee
North Carolina State University, United States
riderr@ecu.edu

This paper reports a comparison of two separate studies using the same task and simulation software but with different age groups and abilities of students who have had different curricula experiences. One study examined how middle school students used computer simulation tools to reason between empirical data and theoretical probability. The second study replicated the first with secondary school students who had just completed an Advanced Placement statistics course. This comparison includes the similarities and the differences in the way each group approached the task and used the simulation software, given their background and prior knowledge.

BACKGROUND

The use of technology tools has and will continue to affect statistics curriculum in many pre-college and college-level courses. In a technologically-rich classroom where students are studying statistics and probability, the tools available should enable students to:

1. "Practice data analysis with an exploratory, interactive, open-ended working style and combine exploratory and inferential methods and graphical and numerical methods.
2. Extensively use multiple linked representations and simulations to construct meanings for statistical concepts and ideas.
3. Construct models for simple and multistage random experiments, and use computer simulation to study them" (Ben-Zvi, 2000, p. 151).

Over the past decade, technologies such as computer software and graphing calculators which allow the user to perform statistical calculations such as computing confidence intervals or finding lines of best fit have had a tremendous impact on the way statistics is taught. With such tools in hand, students no longer need to perform laborious algorithms to compute statistics. With a focus away from correctly computing, teachers can ask questions such as the appropriateness of a statistic or statistical procedures, the feasibility of an answer in a problem situation, or interpretations of results in the context of a problem.

The use of simulations has begun to become more mainstream, but is less apt to be used in many classrooms. At least in the United States, this is partly due to the overwhelming presence of graphing calculators and spreadsheets, coupled with teachers' overall unfamiliarity with the use of simulations in teaching (Stohl, 2005). *Computer simulation methods* (CSMs) (Mills, 2002) have the potential to change the way teachers approach instruction of probability and statistics in the next decade. Simulation software provides an opportunity for students to collect very large random samples and reason from empirical data to make inferences about populations or unknown theoretical probability distributions.

Our research is an effort to address the call for research on how CSMs enhance student learning (Mills, 2002) and on students' understanding of connections between relative frequency of empirical data and theoretical probability (Jones, 2005). Our analysis from two studies focuses on how students in two different age groups (ages 11-12 and 17-18) approached the same task using a computer simulation tool (*Probability Explorer*, Stohl, 2002). The students in the two studies had different prior experiences with simulation tools and different curricula experiences. We focus on the similarities and differences across the two groups with respect to students' attention to and use of sample size, variability, representations of data, and what the students consider as evidence to substantiate their findings.

IMPORTANCE FOR CURRICULUM

In recent years there has been a growing trend to implement stochastic concepts into the K-12 curriculum in the United States (Mills, 2002). The concept of probability has been traditionally taught with an emphasis on a classical Laplacean approach; however, a frequentist approach to probability, based on the law of large numbers, is becoming more common (Jones, 2005; Parzysz, 2003). With this approach, teachers would begin with an empirical introduction to probability by creating problem situations in which students, in order to solve a problem must utilize repeated trials of the same event, either with concrete materials or through computer simulations to estimate theoretical probabilities (e.g., Batanero, Henry and Parzysz, 2005; Parzysz, 2003). Unfortunately there is a lack of research on students' understanding of the connection between observations from empirical data and a theoretical model of probability (e.g., Jones, 2005; Parzysz, 2003).

The NCTM (2000) supports technology-based simulations when teaching probability because they "afford students access to relatively large samples that can be generated quickly and modified easily" (p. 254). Although research on the effect of simulation software on student learning is limited, there are studies which confirm that the use of CSMs increase the number of correct answers that students give to a variety of problems (Garfield and delMas, 1991).

Activities which allow students to explore the relationship between empirical data and theoretical probability are more consistent with what statisticians do as they develop a hypothesis, collect data to test the hypothesis, analyze the data, and reevaluate their hypothesis in light of that analysis. "Exploring data, designing data production, using diagnostic tools to ask whether a proposed method of inference is appropriate have a 'back and forth' flavor quite unlike the 'straight ahead' nature of traditional statistical calculations" (Moore, 1997, p. 126). Reasoning from a frequentist perspective requires this bi-directional movement between empirical data and theoretical probability (Lee, Rider, and Tarr, 2005; Stohl and Tarr, 2002).

PREVIOUS RESEARCH

There has been much research on students' understanding of probability concepts. Watson and Moritz (2003) established that many young children doubt that each outcome of a standard die is equally likely. However, when attempting to confirm or refute their beliefs about the fairness of a die few children used a data collection strategy to substantiate their claims. It is possible that this could be connected with the lack of emphasis in the curriculum on empirical data collection as a tool to answer questions about the probability of an event.

One of the critical aspects of collecting empirical data is determining how large a sample size to collect. When studying students in odd numbered grades from fifth to eleventh, Fischbein and Schnarch (1997) noticed an increased acceptance of small sample sizes in older students when asked to compare the likelihood of two events, one with a small sample size, and the other with with a large sample size (e.g., compare likelihood of 2 heads out of 3 fair coin tosses to 200 heads out of 300 fair coin tosses). However, this acceptance may have been due to an increased knowledge of and reliance on proportional reasoning to determine the likelihood of an event.

Although Fischbein and Schnarch (1997) and Watson and Moritz (2003) employed different situations in their research, the results found by each may give an indication of the effect of the context on students' consideration of sample size. None of the students in these studies was asked to determine an appropriate sample size and collect data themselves to examine the fairness of a die or to compare the probabilities of two events.

These studies give rise to two methods of considering variation in a probability context. Within a data set, variability in the frequency of outcomes can be compared to each other or to an expected distribution. Across data sets, variability in samples can be used to examine differences with a small number of trials, or similarities with large number of trials. From a frequentist perspective, reasoning across data sets (e.g., several samples of 50 trials of a die toss) is crucial for students to be able to make inferences about an unknown theoretical probability or distribution.

These findings suggest that students may benefit from a frequentist approach to teaching probability. However, in a typical classroom, time plays an integral factor in a teachers' ability to collect large amounts of data, a limitation that can be overcome with the use of computer simulations. The availability of computer simulations has given rise to more research regarding their use. Pratt (2000) reported that computer simulation software contributed to an increased understanding of connections between sample size and the distribution of data. Taylor (2001) found that the use of computer simulation software had a positive effect on elementary students' understanding of probability derived from empirical data when used in whole-class instruction with one computer display. Lee, Rider and Tarr (2005) observed how students use of large sample sizes and attention to variability when solving a task with computer simulation tools facilitated stronger connections between observations of empirical data and reasoning about an unknown theoretical distribution. These studies suggest that simulation tools may help students construct an understanding of the relationship between theoretical probability and empirical data.

CONTEXT AND TASK

This paper reports a comparison between results from two separately conducted studies. Although students in the two studies were from different age groups and different levels of experience, we examined similarities and differences in student reasoning across the two studies. The authors realize that this comparison confounds a number of important variables (i.e., age and academic experience) and future research is planned to examine differences between closer related groups. We considered each of these separate studies as part of a growing body of research on how students reason about theoretical probability from empirical data and as such, felt it was important to begin looking at differences across groups engaged in an identical task.

In this paper, we are comparing how middle school students (ages 11-12) reason between empirical data and theoretical probability (see Lee, Rider and Tarr, 2005; Stohl and Tarr, 2002) to how high school students (ages 17-18) approach the same task. The task (see Figure 1) used in our research challenges the students to collect and analyze data from a context with an unknown probability distribution and to make inferences about the distribution. Students must approach the task from a frequentist perspective, as the theoretical probabilities are unknown in the simulation.

Schoolopoly

Your school is planning to create a board game modeled on the classic game of *Monopoly*™. The game is to be called *Schoolopoly* and, like *Monopoly*™, will be played with dice. Because many copies of the game expect to be sold, companies are competing for the contract to supply dice for *Schoolopoly*. Some companies have been accused of making poor quality dice and these are to be avoided since players must believe the dice they are using are actually "fair." Each company has provided a sample die for analysis and you will be assigned one company to investigate:

Luckytown Dice Company	Dice, Dice, Baby!
Dice R' Us	Pips and Dots
High Rollers, Inc.	Slice n' Dice

<u>Your Assignment</u>

Working with your partner, investigate whether the die sent to you by the company is, in fact, fair. That is, are all six outcomes equally likely to occur? You will need to create a poster to present to the School Board. The following three questions should be answered on your poster:

1. Would you recommend that dice be purchased from the company you investigated?
2. What evidence do you have that the die you tested is fair or unfair?
3. Use your experimental results to estimate the theoretical probability of each outcome, 1-6, of the die you tested.

Use Probability Explorer to collect data from simulated rolls of the die. Copy any graphs and screen shots you want to use as evidence and paste them in a Word document.

Figure 1: Handout given to students describing the Schoolopoly task.

The students in each study had different curricula experiences with probability and statistics and use of technology tools. The middle school students (ages 11-12) had participated in a CSM-intensive unit of study on probability for 10 days before being given the Schoolopoly task. During the unit of study, students had many discussions where they reported what they noticed about sample size and variability when collecting empirical data from probability experiments such as coin tosses and drawing marbles from a bag with replacement. They also had experience in designing a computer simulation that they believed would accurately represent spinning real spinners with different size sectors. The high school students were at the end of an advanced placement statistics course (entry-level college material) with a heavy emphasis on the use of graphing calculators for computing various statistical tests. They had minimal experience with simulation techniques by hand or with technology. Their course work focused on theoretical aspects of probability and statistics and the use of technology to perform computations. Thus, their experience with effects of sample size and variation was in the context of calculations and learning about the law of large numbers and the central limit theorem. Although the high school students had no prior use of the specific computer simulation software program used in the study, they were given a brief introduction to the software and allowed to become familiar with important features that would be used in the Schoolopoly task. Since they were also very familiar with the graphing calculator, they were also given access to that technology to use in their work.

In this assignment, students in both studies worked in pairs and were given a substantial amount of time to collect and analyze empirical data. The pairs of students had to describe what evidence they used to arrive at their decision (see Questions 1-3 in Figure 1). Each pair of students presented their results to the class with a visual presentation in the form of a poster. The students then had to support their reasoning when asked questions by their classmates.

RESULTS

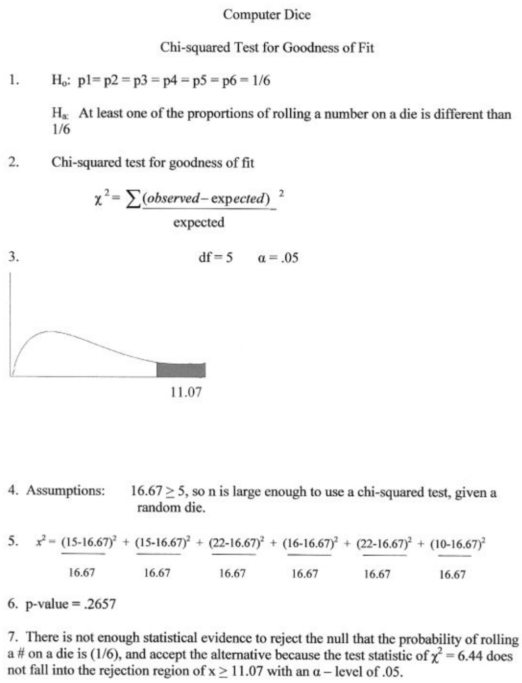
We compared the similarities and differences within each age group and across age groups to understand how students used the simulation software to collect and analyze the data. Preliminary results indicate that many of the middle school students reasoned from large samples ($n \geq 500$) and utilized the stability of the empirical results to support their reasoning. Many were successful in predicting whether their die was fair or not, and they made reasonable estimates of the underlying probability distribution for their die. In contrast, the high school students drew upon their understandings of appropriate statistical tests and computations. The high school students seemed to ignore the law of large numbers as a tool for helping them estimate the theoretical distribution. Thus in comparison to the younger students, they typically only collected one sample ranging in size from 30 – 500 and then performed statistical tests for goodness of fit with the null hypothesis that the die had equiprobable outcomes and an alternative that at least one outcome had a probability different from the rest. They tended to use the results of a single sample to generate an estimate of theoretical probability rather than using any application of the central limit theorem. A sample poster from each group of students is included in Figure 2.

Only one group of high school students utilized the underlying theory of the central limit theorem, although this particular group of students only took three samples of size 500 and then averaged the probabilities of each outcome across the three samples. The majority of the students had significant difficulty determining what to do to estimate the theoretical probabilities and wanted the teacher to give them "a hint of what formula to use." This reliance on procedures and formulas is indicative of a procedural emphasis in the curriculum. It is important to note that the majority of the high school students featured in this study was very successful in the class and on the Advanced Placement Statistics exam. Garfield (1995) has previously reported a reminder "that although students may be able to answer some test items correctly or perform calculations correctly, they may still misunderstand basic ideas and concepts" (p. 31). In most cases of the high school students, there was no connection between the empirical data they had collected, the theoretical probability they were trying to estimate, and the central limit theorem. Once the students had performed a statistical test to confirm or refute the hypothesis of equiprobable outcomes, they then moved on to estimating the theoretical probabilities using the relative frequencies of relatively small samples. The majority did not perform any additional data collection to confirm or refute the previously collected data.

In contrast, by performing many cycles of data collection, the middle school students were coordinating their mental image of the theoretical probability which initially was an assumption of equiprobable outcomes with empirical data that either confirmed or conflicted with that image. In each subsequent cycle of data collection it was apparent that students' images of the theoretical probability or distribution were informing their decisions on sample size, noticing properties such as stabilization of outcomes, and comparing variability within and among samples. Although all pairs of middle grades students demonstrated this bi-directional movement between an image of a theoretical model and the empirical data, the strength of the connections they were making varied vastly. Representation use and sample size were critical factors to the strength of the connections the students made and thus the overall success was different among the pairs of middle school students.

<table>
<tr><td colspan="2">Actual probability distribution (unknown to students) for Dice R Us company was
1: 12.5%, 2: 18.75%, 3: 18.75%, 4: 18.75%, 5: 18.75%, 6: 12.5%</td></tr>
<tr><td>

</td><td>



Computer Dice

Chi-squared Test for Goodness of Fit

1. H_o: p1= p2 = p3 = p4 = p5 = p6 = 1/6

   H_a: At least one of the proportions of rolling a number on a die is different than 1/6

2. Chi-squared test for goodness of fit

$$\chi^2 = \sum \frac{(observed - expected)^2}{expected}$$

3. df = 5 α = .05

11.07

4. Assumptions: 16.67 ≥ 5, so n is large enough to use a chi-squared test, given a random die.

5. $x^2 = \frac{(15-16.67)^2}{16.67} + \frac{(15-16.67)^2}{16.67} + \frac{(22-16.67)^2}{16.67} + \frac{(16-16.67)^2}{16.67} + \frac{(22-16.67)^2}{16.67} + \frac{(10-16.67)^2}{16.67}$

6. p-value = .2657

7. There is not enough statistical evidence to reject the null that the probability of rolling a # on a die is (1/6), and accept the alternative because the test statistic of $\chi^2 = 6.44$ does not fall into the rejection region of $x \geq 11.07$ with an α – level of .05.

</td></tr>
<tr><td>Middle School poster of results. Several sets of data, one with sample size of 500.</td><td>High School poster artifact of results. Sample size is 100. Estimate of the probabilities from their poster are 1: 15%, 2: 15%, 3: 22%, 4: 16%, 5: 22%, 6: 10%. They also included a pie graph and a frequency table.</td></tr>
</table>

Figure 2: Example work of middle and high school students

DISCUSSION

> If you use technology to simply carry [out] the same old thing, you get the same old results. To get different results, you must add new thinking to new technology …. Such a statistics curriculum, which takes advantage of the technology, can stress conceptual understanding, mathematical modeling and problem solving, real-world applications, and new methods of analyzing data. (Ben-Zvi, 2000, p. 151)

The middle school students featured in this study had no experience with formal hypothesis testing techniques. Thus to solve the problem of determining if the company produced fair dice, they used informal reasoning techniques such as taking large samples and watching for the stabilization of the bar graph and data tables to determine an estimate of the true probabilities of the outcomes of the die. They took advantage of the power of the simulation software to collect relatively large samples of data and made comparisons of the outcomes within a sample and across samples. Secondary school students, having studied formal hypothesis testing procedures, relied upon relatively small samples and typically did not take more than one sample. They would then apply what they knew about hypothesis testing techniques without examining whether those techniques were appropriate. They made no meaningful connections between the empirical data they had collected and theoretical probability distributions, and instead relied on procedural techniques made simpler by performing the calculations with a graphics calculator.

Identifying curriculum needs and the impact of various technological tools such as CSMs on student learning of statistical concepts is critical as the subject plays a more prominent role in the K-12 curriculum. The value of a frequentist perspective to probability in instruction is only beginning to be realized. Although the Schoolopoly task does not have real world importance, it does give an engaging context for students to go through the complete process of making an initial hypothesis, collecting data to confirm or refute that hypothesis, analyzing the data, collecting more data to confirm results, and making inferences regarding theoretical probability. Using a classical Laplacean *a priori* approach to probabilities does not allow students to experience the entire statistical inference process as a statistician would. Although students may be able to complete hypothesis tests and inference procedures on ready-made data sets, the experience of collecting data, including determining sample size, adds an additional dimension that requires students to go beyond statistical procedures. Tasks in which students must estimate unknown theoretical probabilities by utilizing empirical data combined with simulation technology may allow students to develop a more robust understanding of statistical concepts.

REFERENCES

Batanero, C., Henry, M., and Parzysz, B. (2005).The nature of chance and probability. In G. A. Jones (Ed.), *Exploring Probability in School: Challenges for Teaching and Learning,* (pp. 15-37). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. *Mathematical Thinking and Learning*, 2(1 and 2), 127-155.

Fischbein, E. and Schnarch, D. (1997). The evolution with age of probabilistic intuitively based misconceptions. *Journal of Research in Mathematics Education,* 28(1), 96–105.

Garfield, J. and delMas, R. (1991). Students' conceptions of probability. In D. Vere-Jones, S. Carlyle and B. P. Dawkins (Eds.), *Proceedings of the Third International Conference on Teaching Statistics*, Dunedin, New Zealand, Vol. 1, (pp. 340-349). Voorburg, The Netherlands: International Statistical Institute.

Garfield, J. (1995). How students learn statistics. *International Statistical Review*, 63(1), 25-34.

Jones, G. A. (2005). Reflections. In G. A. Jones (Ed.) *Exploring Probability in School: Challenges for Teaching and Learning,* (pp. 375-379). Dordrecht, Netherlands: Kluwer Academic Publishers.

Lee, H. S., Rider, R. L., and Tarr, J. E. (2005). Making connections between empirical data and theoretical probability: Students' generation and analysis of data in a technological environment. Manuscript currently under review.

Mills, J. D. (2002). Using computer simulation methods to teach statistics: A review of the literature. *Journal of Statistics Education*, 10(1), http://www.amstat.org/publications/jse/v10n1/mills.html.

Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review*, 65(2), 123-165.

Parzysz, B. (2003). From frequency to probability. Some questions posed by the new French senior high school curricula. Paper presented at the 54th International Statistical Institute, Berlin, Germany, http://www.stat.auckland.ac.nz/~iase/publications.php?show=3.

Pratt, D. (2000). Making sense of the total of two dice. *Journal of Research in Mathematics Education,* 31, 602-625.

Stohl, H. (2002). *Probability Explorer v. 2.01*, http://www.probexplorer.com.

Stohl, H. (2005). Probability in teacher education and development. In G. Jones (Ed.), *Exploring Probability in School: Challenges for Teaching and Learning*, (pp. 345-366). Dordrecht, Netherlands: Kluwer Academic Publishers.

Stohl, H. and Tarr, J. E. (2002). Developing notions of inference with probability simulation tools. *Journal of Mathematical Behavior,* 21(3), 319-337.

Taylor, F. M. (2001). *Effectiveness of Concrete and Computer Simulated Manipulatives on Elementary Students' Learning Skills and Concepts in Experimental Probability.* Unpublished doctoral dissertation, University of Florida.

Watson, J. M. and Moritz, J. M. (2003). Fairness of dice: A longitudinal study of students' beliefs and strategies for making judgments. *Journal for Research in Mathematics Education*, 34, 270-304.