International Electronic Journal of

Mathematics Education

Volume 4, Number 3, October 2009 www.iejme.com

Special issue on "Research and Developments in Probability Education"

Manfred Borovcnik & Ramesh Kapadia (Eds)

BUILDING A CONNECTION BETWEEN EXPERIMENTAL AND THEORETICAL ASPECTS OF PROBABILITY

Seth Ireland & Jane Watson

ABSTRACT. This paper addresses a question identified by Graham Jones: what are the connections made by students in the middle years of schooling between classical and frequentist orientations to probability? It does so based on two extended lessons with a class of Grade 5/6 students and in-depth interviews with eight students from the class. The Model 1 version of the software TinkerPlots was used in both settings to simulate increasingly large samples of random events. The aim was to document the students' understanding of probability on a continuum from experimental to theoretical, including consideration of the interaction of manipulatives, the simulator, and the law of large numbers. A cognitive developmental model was used to assess students' understanding and recommendations are made for classroom interventions.

KEYWORDS. Experimental estimates of probability, Theoretical probability, simulation, law of large numbers, *TinkerPlots*.

Copyright © 2009 by GOKKUSAGI

ISSN: 1306-3030

1. INTRODUCTION

The motivation for this study grew from the first author's interest in the contribution of technology to learning in the classroom. The narrowing of the focus resulted from the availability of the *TinkerPlots* (Konold & Miller, 2005) Sampler, which was under development, and a reading of *Exploring Probability in School* (Jones, 2005). Jones, reflecting on the current teaching of probability in classrooms, noted,

> "In spite of the apparent robustness of research on elementary school children's probabilistic reasoning, it is evident that there is a void in the research associated with the frequentist approach to probability; that is, research dealing with children's cognitions on experimental probability. In fact, there is almost no research on whether children can make connections between classical and frequentist orientations to probability even though teachers are encouraged to use these connections in the classroom. … such research needs to document effective classroom practices including those that use the technology and software that is becoming available for young children." (p. 368)

Only recently has research focused on this area. Stohl, Rider, & Tarr (2004) investigated how sixth grade students used computer simulated experimental data to make inferences about unknown probabilities of a loaded die. They examined "the ways in which students' understanding and use of sample size, independence, fairness, and variability interacted with their use of external resources such as the task context, multiple representations of data and social negotiations with a partner" (p. 1). Stohl et al. (2004) concluded that students who used larger sample sizes and multiple representations of data made appropriate inferences regarding fairness.

Abrahamson & Wilensky (2007) employed a problem-solving context based on the binomial distribution to implement their design-based research framework for mathematics learning with cognitive conflict. Incorporating learning axes with two complementary components of related statistical concepts at the ends and computer-based bridging tools embodying in an ambiguous fashion the two components, they explored student learning outcomes as they struggled to resolve the question of the proportion p in an unknown binomial distribution. Their examples of classroom learning episodes illustrate the use of sampling to resolve conflict, for example in relation to theoretical-versus-empirical probability.

The *TinkerPlots* Sampler is being developed as a probability simulation addition to the currently available software. Although other simulation software has been developed to assist middle school students' understanding of probability (e.g., Abrahamson & Wilensky, 2002; Stohl & Tarr, 2002), *TinkerPlots* Sampler provides a wide range of flexibility in its modeling capacity.

Recent research (Konold, Harradine, & Kazak, 2007; Konold & Kazak, 2008; Konold & Lehrer, 2008) is showing the potential of the software to enhance student understanding of variation, distribution, and the links between data and chance. As well, *TinkerPlots* itself has been recognized as making a valuable contribution to classroom practice and student understanding of other data handling topics (Ben-Zvi, Gil, & Apel, 2007; Paparistodemou & Meletiou-Mavrotheris, 2008).

The Sampler provides two ways of simulating a natural phenomenon: The Mixer and the Spinner. The Mixer in Figure 1 is similar to a bag of marbles, where the balls are the same size and shape and the sample space is altered by adding additional balls to the Mixer. The Mixer simulates randomness visually by bouncing the balls around the bag and releasing an outcome out of the bottom of the bag. The Spinner in Figure 1 incorporates an area model. The size and proportions of the elements of the sample space are altered by the percentage of area allocated to each outcome on the Spinner. The Spinner simulates randomness visually by spinning a pointer around, which eventually stops, giving the result of the trial.

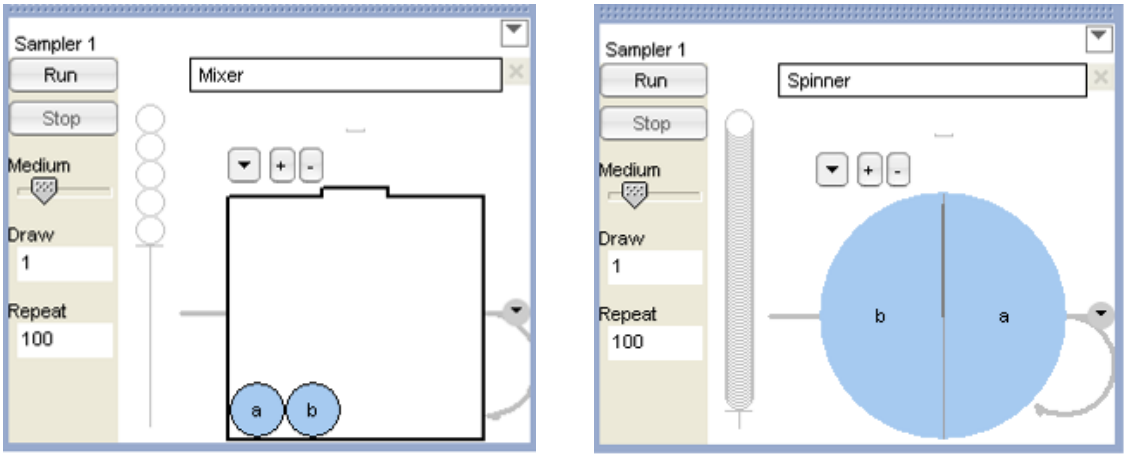


Figure 1. The Mixer and the Spinner in TinkerPlots Sampler (click to see the videos).

Computer simulations imitate the process of "real-world" phenomena on a computer (Malhotra, Hall, Shaw, & Crisp, 1996) and "should imitate the internal processes and not merely the results of the thing being simulated" (Wordnet, n.d., para. 1). Students, need not only make conceptual connections between the three-dimensional concrete manipulatives, such as coins, and the two dimensional abstract representation on the simulator, but also make the connection between the internal processes of the simulation and the process of obtaining outcomes using the coin.

Many questions arise. For example, what is the students' understanding of how the computer chooses outcomes randomly? Do the students make connections between the computer

simulation and the concrete manipulatives? What understandings are required for this transition? Much research has been carried out into the use of concrete models and computer simulations to develop other abstract mathematical concepts for students (e.g., Clements, 1999; Mills, 2002; Rider & Stohl Lee, 2006; Sarama & Clements, 1998). However, except for the work of Pratt (e.g., Pratt & Noss, 2002) on randomness based on the notion of situated abstraction and that of Abrahamson & Wilensky (2007) with their learning axes and bridging tools, there appears to be limited research concerning what is required to facilitate an effective transition from concrete models to abstract computer simulations in the field of probability. This realization has provided further impetus for the study.

Konold (2006), discussing the design of the *TinkerPlots* software, gives some insight into how the transition between concrete manipulatives and the simulation might be accomplished.

> We then worked to implement these operations in the software in a way that would allow students to see the computer operations as akin to what they do when physically arranging real-world objects. This sense–that one already knows what the primary software operators will do–becomes important in building up expectations about how the various operators will interact when they are combined. (p. 4)

It is the connection suggested by Konold, between the concepts considered concrete (real-world objects) and more abstract (computer simulation), which provides a starting point for this study. The literature reveals many definitions of concrete and abstract and their interaction (Erickson, 2006; James & James, 1959; Steen, 2007; Stein & Bovalino, 2001). The ideas, that the concrete leavens the abstract (Erickson, 2006) and that abstract concepts can become concrete (Basson, Krantz, & Thornton, 2006) suggest a continuum linking the two extremes with the two ideas reinforcing each other depending on context. In the area of probability the identification of experimentation using actual objects with the concrete end of the continuum and theoretical probability with the abstract end, appears feasible.

With manipulatives associated with experimental probability and the simulator providing one link from the manipulatives to the theoretical, the Law of Large Numbers fits also in a proposed model along the continuum (Stohl & Tarr, 2002) as shown in the framework in Figure 2. The Law of Large Numbers is placed in the center of the proposed continuum because of its underpinning explanatory power in relation to the increased sample sizes possible for trials associated with manipulatives and simulations. These three elements provide the links considered in this study in developing students' understanding of the relationship between experimental and theoretical probability.

Further discussion of the development of the model is presented in Ireland (2007). Wilensky (1991) offers an interesting insight into the development of relationships by learners of mathematics, by using the term concrete as the ultimate adjective to describe the successful process. Although this characterization might apply to students' successful construction of the relationships in Figure 2, the term is not used in that sense in this paper. Rather concrete is used as a complement to abstract as suggested by the fore-mentioned authors.

The key research question for this study is then, "What connections do students make between experimental and theoretical probability after work with manipulatives and with the aid of *TinkerPlots*?"

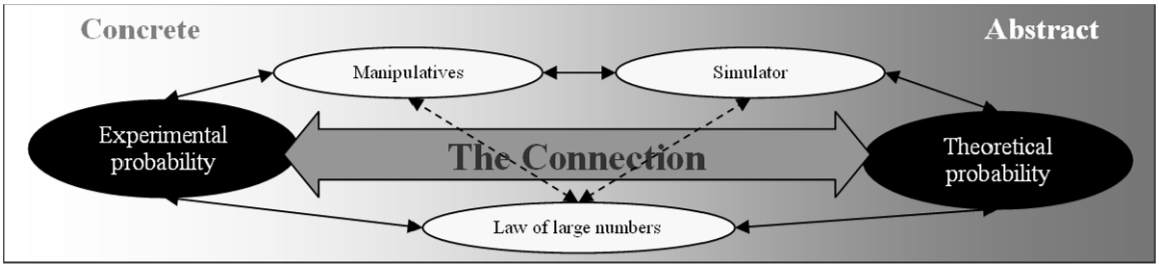


Figure 2. Concrete to Abstract conceptual framework.

2. METHODOLOGY

Subjects

The study took place during the first term of 2007 in a grade 5/6 classroom (ages 10–12) in a government primary school in Hobart, Tasmania, Australia.

- The class consisted of 27 students, 11 grade fives (5 males and 6 females) and 16 grade sixes (7 male and 9 female), with varying degrees of mathematical competence.
- This class was chosen based on their age in relation to Piaget's theory of cognitive development (1955), which places the students' understanding on the cusp of the transition from the concrete operations stage to the formal (abstract) operations stage.
- Prior to the research intervention the students had experienced lessons in probability with their usual teacher as suggested by the Tasmanian mathematics curriculum (Department of Education, 2008).

- Activities included working with dice, coins and basic probability, determining if certain games were fair, and ranking the likelihood of events in their lives.

Whole-class procedure

A test was administered to the class prior to the lessons on probability. The section relevant to this paper included questions on basic probability chosen from the test items used by Watson & Callingham (2003), an assortment of open and closed questions, often including a request for explanation or justification. These questions are found in Appendix A. The responses were used to judge the level at which to begin the lessons and later to assist in selecting students to be interviewed.

The two extended lessons were taught by the first author in conjunction with the classroom teacher as part of the normal curriculum. The first lesson began with a structured discussion of Probability including

(i) its meaning in terms of events like the sun setting,

(ii) the ordering of chance phrases, and

(iii) meanings of specific terms like "variation" and "random."

Then a coin was introduced and students were asked to predict outcomes as it was tossed and outcomes were recorded. After 10 trials and discussion, the idea of "theoretical probability" was introduced, including favorable and total possible outcomes, the fraction 1/2, decimal 0.5, and percent 50%. These ideas were then linked back to the earlier trials and the term "experimental probability" introduced. The relationship between theoretical and experimental was mooted and the question asked about how many trials would be required for the experimental probability to confirm the theoretical probability.

The lesson then turned to the coin tossing activity with students being given recording sheets, asked to record their estimates about how many heads they would get in 10 tosses, and allowed to carry out the 10 tosses and record the fraction of outcomes that were heads. Sticky labels were used by students to record their outcomes on a class chart; these were compared to the theoretical value of 5/10 to illustrate the variation involved. Students then combined their results with neighbors to create samples of size 20 and then 40. These were also recorded on the class chart (see Figure 3). The discussion at the end of the first lesson focused on the closeness of the results to the theoretical probability and the effect of sample size.

The following lesson continued with the same format and involved the use of the *TinkerPlots* Sampler to investigate the connection between the theoretical probability and the observed experimental outcomes in larger trials of 100, 1,000, and 10,000 (see Figure 4). In the whole class lesson, the Mixer was used to simulate a coin. Figure 5 is a screen dump of part of the actual activity. The students were able to see: the Mixer, a running tally of both heads and tails in the form of a bar graph, a dot graph tracking the variance between heads and tails, and a stacked dot plot graph tracking the number of heads in a row.

Work samples were collected from the students in the form of a recording sheet and a reflective summary at the end of the second lesson. The reflective summary asked students to answer three questions: What is experimental probability? What is theoretical probability? How are they connected? Responses to these questions helped to confirm the choice of students for interview.

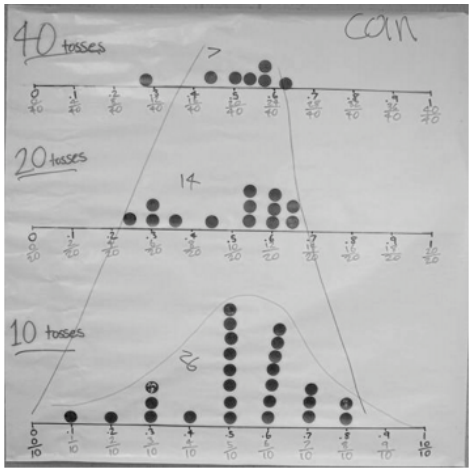


Figure 3. Tracking the tossing of a coin.

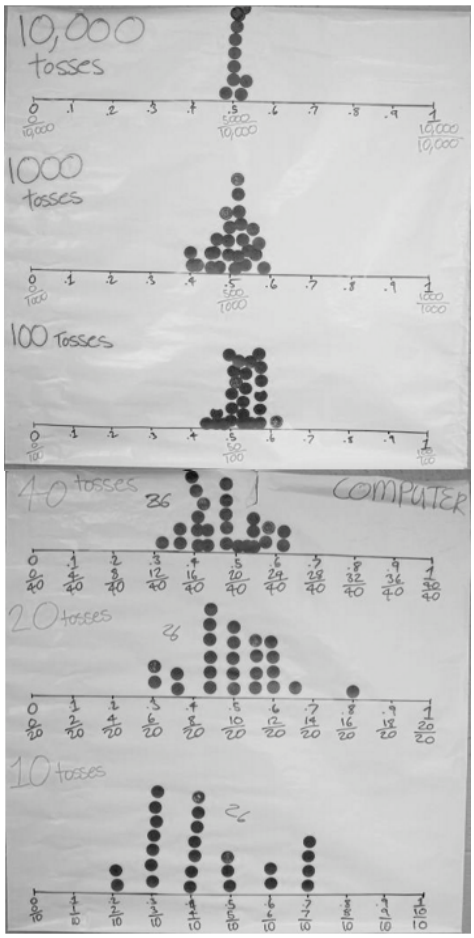


Figure 4. Tracking the simulated tossing of a coin.

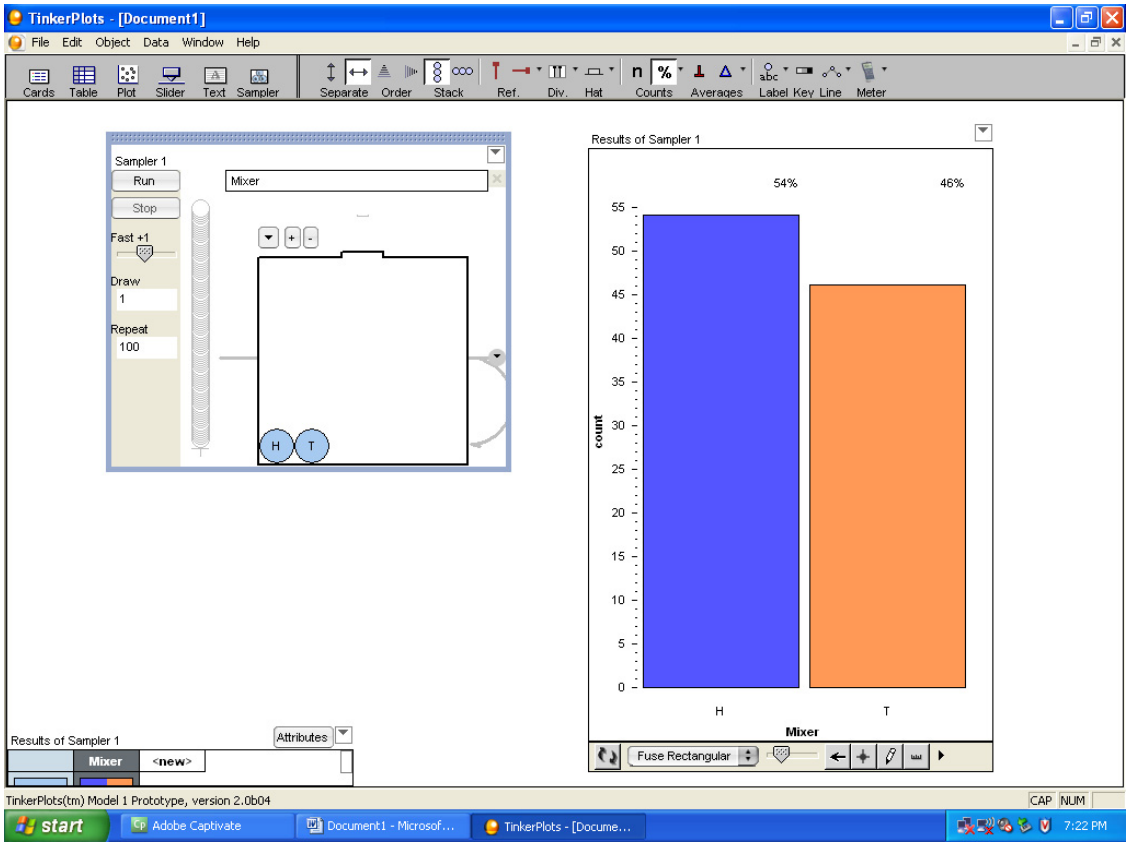


Figure 5. Screen dump of the Sampler whole class coin activity including three graphs (click to see a video).

Interview procedure

The eight students chosen for interview, in consultation with the classroom teacher, were those who performed well in the pre-test and the reflection sheets and who were considered by the teacher to be articulate and comfortable talking to a researcher one-on-one. The interview protocol used a die and the *TinkerPlots* Sampler. The introduction of a die in the interview protocol, instead of coins, provided an opportunity for students to demonstrate their understanding in a new context (Blythe, 1998). During the interview, students were asked to explain, generalize, find evidence, and apply their current probabilistic understanding of dice.

As part of the interview protocol, students were exposed to three scenarios involving the simulator and a die. The Spinner was chosen to simulate the die as there was more flexibility in the loading of the outcomes. Three scenarios were developed: one without loading, one with a large loading, and one with a slight loading. These loadings were hidden from the student. Figure 6 is a screen dump of one of the actual activities. The students could see the Spinner but not its contents, the table of results, and a stacked dot plot with the decimals recording the proportion of the experimental outcomes above each column.

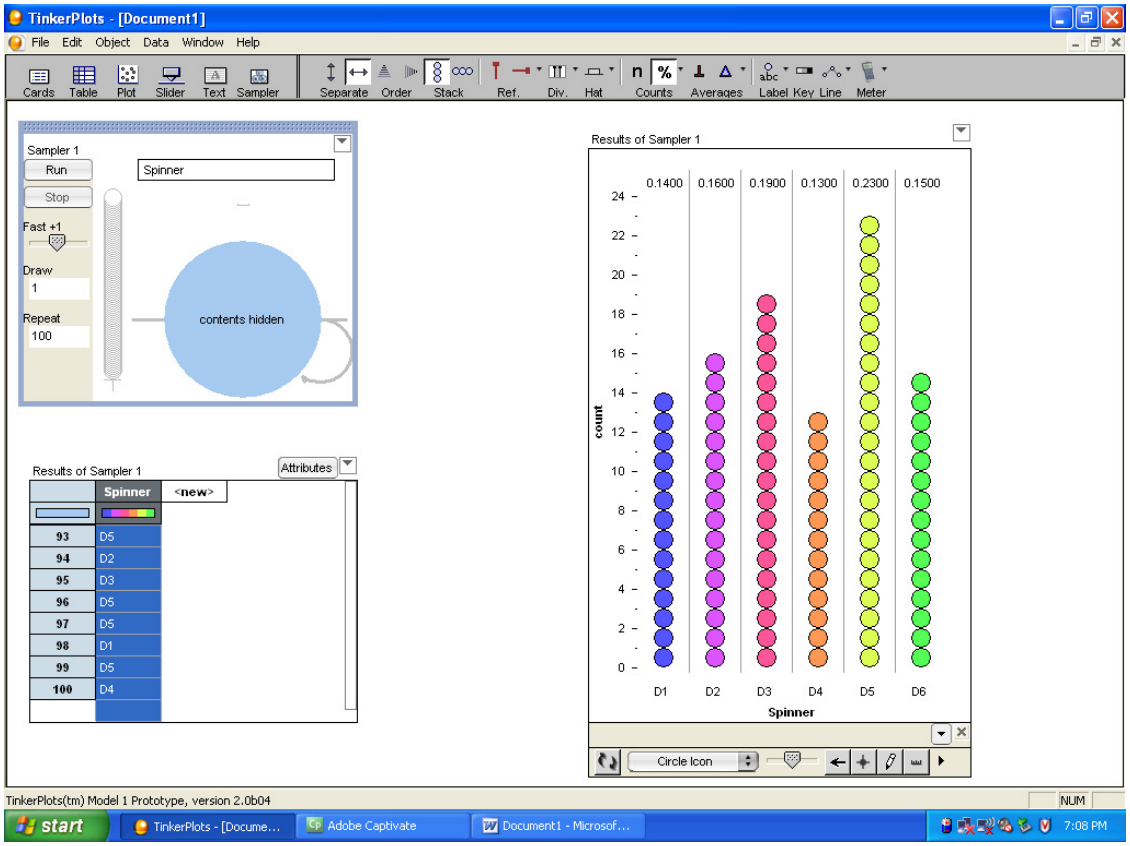


Figure 6. Screen dump of the Sampler personal interview die activity including a table and graph – click to see a video.

The specific questions in the protocol were designed to cover the concept framework in Figure 2. The interview began with a discussion of a tangible die and the associated theoretical probability. Then the student was asked to draw graphs estimating the outcomes for tossing the die 10, 100, and 1,000 times and to discuss how different they would be. Students were also asked what the outcomes from a loaded die would look like. Then the Sampler was introduced with discussion of its relationship to the real die on the table and three cases with loaded or unloaded dice were considered. The interview ended with a discussion of the relationship between experimental and theoretical probability as experienced during the interview. The main interview structure is provided in Appendix B.

Analysis

The determination of the level of student response was based on mappings of the ideas expected to be included in the students' responses in relation to the five elements in Figure 2. These are shown in Figure 7. The connections among the five elements themselves are shown in Figure 8. The way the components of the figures were combined, for both Figures 7 and 8, was assessed using the SOLO model of Biggs & Collis (1982, 1991) (cf. Watson & Moritz, 2003). Four levels of the model were employed in this study:

- Unistructural responses (U) were characterized by single aspects of the element and a lack of recognition of contradictions when they occurred.
- Multistructural responses (M) contained a series of aspects of the element with contradictions likely to be recognized but unresolved.
- Relational responses (R) demonstrated a linking together of aspects of the element, resolving to a large extent any conflict that arose.
- If responses did not appear to employ any of the ingredients contributing to the elements, they were assigned to the Prestructural level (P).

Hence for each of the five elements in Figure 2, responses of the students (including classroom observations and interviews) were assigned a level (P, U, M, or R) in relation to the related sub-elements displayed for each in Figure 7. Each of the elements was considered independently at this stage in relation to its expected sub-elements. The relative difficulty of combining sub-elements for the different elements was not an issue at this point; hence it is acknowledged that some elements were more difficult for the students to construct from sub-elements than others.

Finally, the links established in the responses among all five elements to display understanding of the relationship between concrete experimental probability and abstract theoretical probability were assessed to assign a level of response for the overall meta-task. This process is similar to that of other SOLO analyses where more than one cycle of cognitive functioning is observed; the first cycle documents the building of a concept, such as average, whereas the second employs the concept in a higher order, more complex problem-solving environment (see e.g., Watson, Collis, & Moritz, 1997).

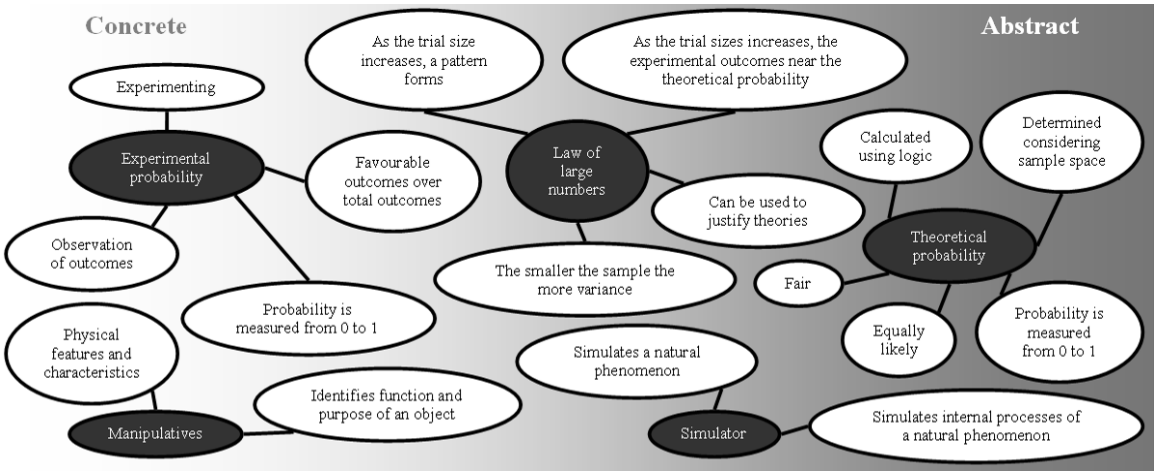


Figure 7. Mapping in relation to each of the five elements in Figure 2.

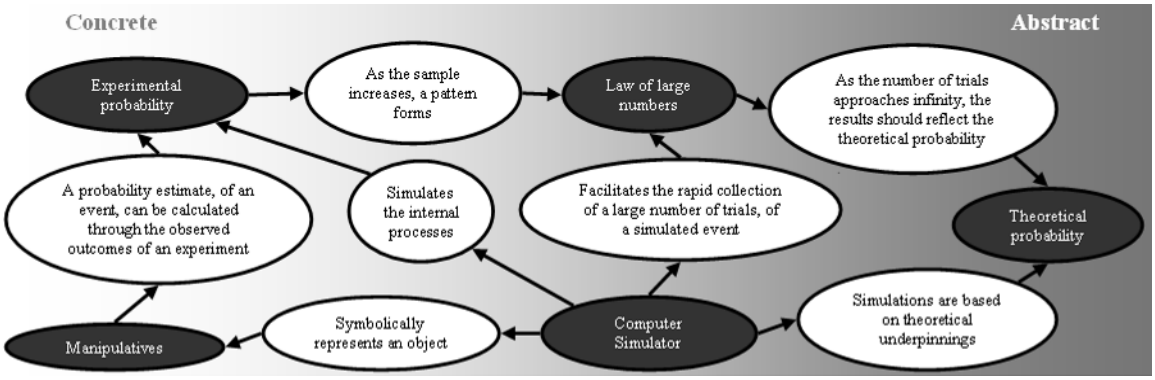


Figure 8. Connections among five elements in Figure 2.

3. RESULTS

The test administered to 26 students before the lessons was scored out of a possible total score of 30. Of the students, 10 achieved a score greater than or equal to 20, 13 achieved between 10 and 19; and 3 received a score of 9 or less. The results (for details click here) indicate that a third of the students had a good basic appreciation of calculating probabilities and some intuitions on the variation involved in an experimental setting.

During the classroom lessons, it was observed that some students had difficulty using decimals to the hundredths and thousandths places and converting between fractions and decimals. Students' sketches on work sheets however indicated that 70% appreciated the leveling out of the experimental outcomes as the sample size increased as seen in Figures 3 and 4.

Construction of a scale

The responses to the three reflection questions asked at the end of the second lesson were coded at four levels, with 0 being idiosyncratic, and 1 to 3 representing a greater degree of understanding of the topic in the question. Summaries of the results for the 23 students who completed the reflection are shown in Table 1. Finding the language to answer the questions in Table 1 was not easy for students. Most had a general idea of what was involved but tended to miss the essential elements.

Code	What is experimental probability?		What is theoretical probability?		How are they connected?	
0	Idiosyncratic	2	Idiosyncratic	7	Idiosyncratic	2
1	Experimenting with chance, probability	8	50/50 chance, contextual	7	They are similar or same, 50/50 chance	10
2	Experimenting with coins, etc, try something, do it	10	Using logic, thinking about it	5	"Prove" each other	4
3	Looking for patterns to confirm ideas	3	Calculation based on sample space	4	Eventually come together & join. In the long run they converge	7

Table 1. Number of Students Achieving Each Level of Reflection for Three Questions about Probability.

The eight students who were interviewed had scores ranging from 20 to 23 on the initial class test. Of a possible total score of 9 for the three reflection questions summarized in Table 1, seven of the interviewed students had scores between 6 and 9, and one had a score of 2. In some cases these two scores and classroom contributions were relevant to determining a level of response.

Because the interview questions were designed to explore the five elements of the model in Figure 2, it was possible to assign SOLO levels to the students' observed understanding of each, in terms of the sub-elements in Figure 7. Considering the entire framework together related to the connections among the elements in Figure 8, it was also possible to assign a SOLO level to the observed understanding of the connection between experimental and theoretical probability.

The levels assigned for the eight students are given in Table 2. It can be seen that all eight students interviewed displayed an integrated, relational understanding of the Manipulatives in the context they encountered. This was demonstrated by the students' explanations of what dice are,

students' intended use of dice during the interviews, and observations of students' use of coins during the classroom activities. All understood the purpose of dice, for example in playing games, and for creating data to explore chance in the classroom.

	Manipul-atives	Simulat-or	Experimental Probability	Theoretical Probability	Law of large Numbers	Overall Connection
S_3	R	R	R	R	R^-	R
S_8	R	R	M	M^+	R	R
S_1	R	R	R	M	R	M
S_4	R	R	R	M	M	M
S_2	R	R	R	M^+	U	U
S_5	R	R	M	M	U	U
S_6	R	R	M	M	U	U
S_7	R	R	U	U	P	P

Table 2. SOLO Levels for the Five Elements and Connections for Five Interviewed Students – P = Prestructural, U = Unistructural, M = Multistructural, R = Relational

Understanding of the simulator

In terms of building the understanding associated with the Simulator, all eight students' responses were judged to have reached the relational level of appreciating the reproduction of a process within the computer that produced random outcomes.

- Although students could not explain how the random process might operate, they saw for example the equal size of icons in the software as representing the equal chance of sides of the die. Some students reached this level spontaneously when confronted with the two versions of the Sampler (Mixer and Spinner) during the interview, whereas others required prompting.
- In discussing the Simulator as a reflection of a real die, some students preferred a die because of their sense of control or because of the belief that a die was "more random," whereas others preferred the Sampler as being faster and more interesting.
- Establishing that the eight students had a firm understanding of the Manipulatives they had used and the Simulator they observed meant that making judgments on the other three elements in Figure 2 could have meaningful starting points.

Experimental probability

- In relation to 'experimental probability', S_7, during the interview, struggled to move past his Reflection statement written at the end of the lessons, that experimental probability is doing experiments to find probability; the response was classified as Unistructural. Three responses were judged to be Multistructural, in that separate ideas were expressed but not linked to an appreciation of the purpose of Experimental Probability.

- S_5, for example said on the Reflection sheet,

  "I think it is when a coin is tossed, you're not sure what side it will land at."

  Short term variation was seen as the outcome of experimenting by S_5 during the interview. On the other hand, S_6 and S_8 indicated that the experimenting was needed to work out if the theory were really true.

- The other four students demonstrated Relational understanding with respect to Experimental Probability. S_2 for example was able to estimate, measure, and record outcomes from the class experiments, using fractions and decimals, as well as graphs, demonstrating the purpose of experimenting. She also highlighted the purpose of experimenting by suggesting that a test for fairness was to throw a die a number of times.

Theoretical probability

For 'theoretical probability' it was more difficult for students' responses to reach the Relational level.

- S_3 was the most successful in being able to refer without prompting to the elements of a sample space, explain equally likely outcomes, and use proportional reasoning (logic) to solve one of the problems on the test. Most of the other responses covered some of the ideas in Figure 7 but did not link them together in a coherent fashion, being assessed as Multistructural.

- S_6 and S_8 for example could discuss "equal likelihood" and "fair" in terms of "even outcomes." They also had the idea of a 1 in 6 chance of coming up in theory without demonstrating other proportional reasoning skills.

- S_7's Unistructural response was based on an "anything can happen" view of probability with very little ability to quantify chance or possible outcomes without specific prompting.

Law of Large Numbers

The Law of Large Numbers was the most difficult of the elements in Figure 2 for students to grasp. As can be seen in Figure 7, the sub-elements related to expected outcomes from larger or smaller samples create conflict that makes theoretical predictions difficult: pattern and variation are both observed. This is also seen in the students' initial sketches of distributions of outcomes for tossing a die an increasing number of times, which are shown in Figure 9. The conflict between sketches and verbal explanations for some students reflects the observations of Borovcnik & Bentz (1991) in their analysis of test items relating sample space and symmetry and relating frequency and random distribution.

- Three students, S_1, S_3, and S_8, were assessed as providing responses demonstrating Relational understanding. S_1 struggled at times with what was the appropriate theoretical distribution, for example for a single die, but demonstrated well the value of large numbers of trials in making decisions during the interview about whether the simulated die was loaded or not. This is shown in the following extract.

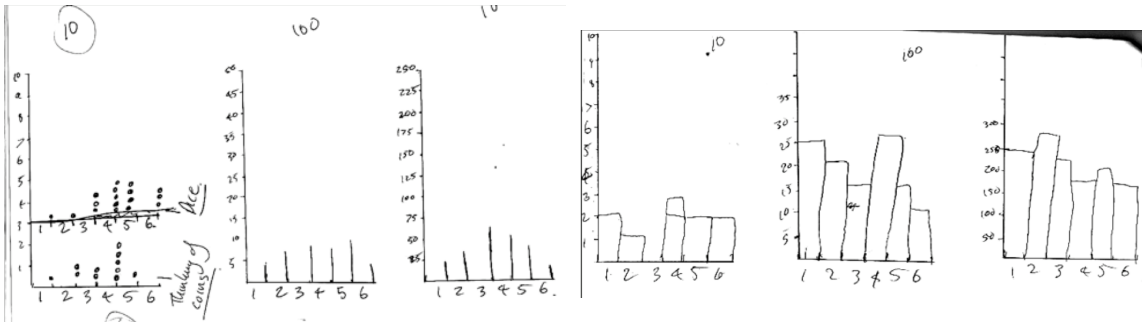
  > I think the most worthwhile ones [samples] are 1,000, 1,0000, and 15,000, so let's start with the 1,000 [rejecting smaller sample sizes] … [Requests 1,000 again.] We got 5 up again … Five is up a fair way. Let's change to 10,000 … Side 5 is winning. This seems too much of a coincidence [Requests 10,000 again.] Side 5 is winning. I think it is an uneven dice. [Later.] I found that with the more tosses we did it became more clear.

  Although S_3 had difficulty in reconciling his drawing of graphs for increasing sample sizes with the outcomes later observed from the simulations, he could articulate well the limiting value of 0.167 in decimal form for the fair die. With the loaded die he started with small samples sizes, developed a theory about how they might be loaded and used larger sample sizes (e.g., 10,000) to test his theory. S_8 drew an even distribution for 1,000 tosses of a die before experiencing the simulations. For the fair simulated die, he increased the sample size from 100 to 500, noting

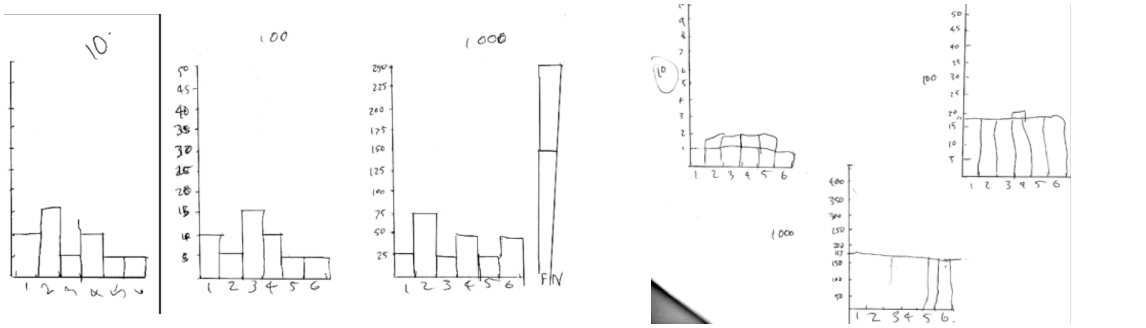
  > "it is getting more equal."

  After two samples of 1,000, he said,

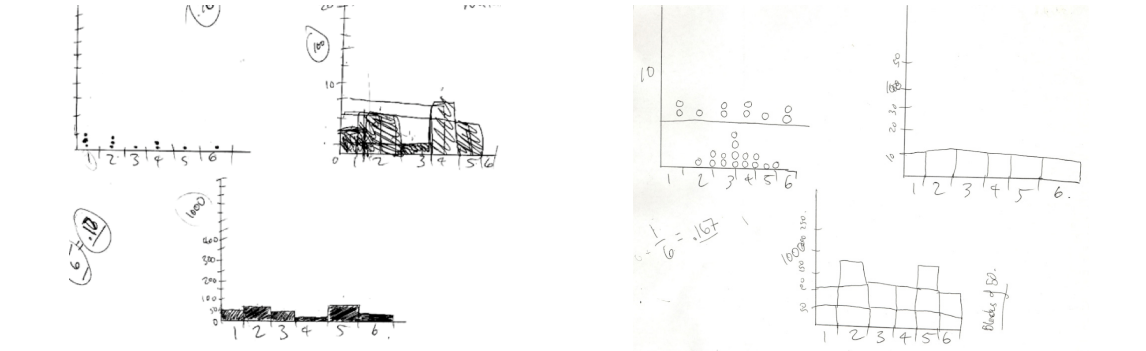
  > "That's kinda like the dice, because the dice is never going to be the same but the more you do, the more the same they are going to be."



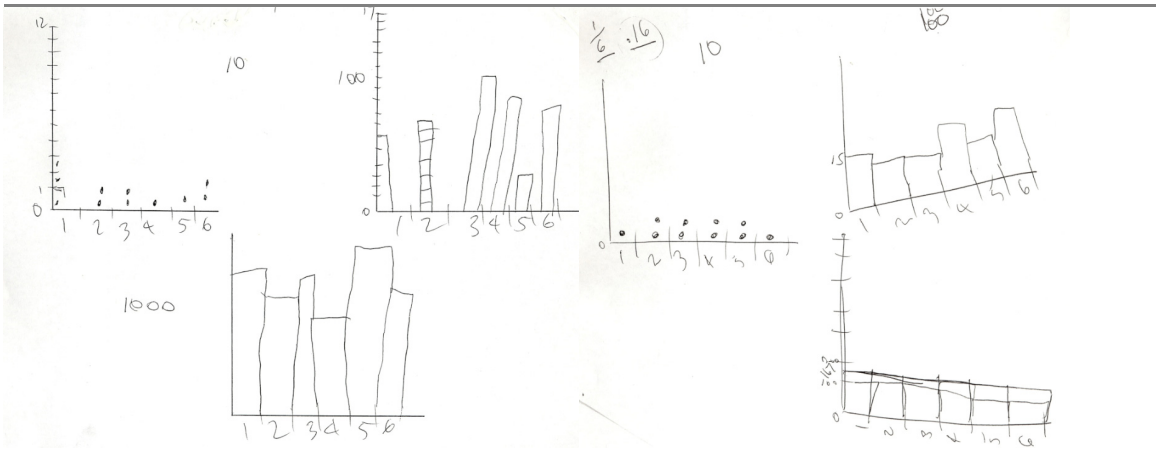
Student 1 Student 2



Student 3 Student 4



Student 5 Student 6



Student 7 Student 8

Figure 9. Students' drawings of estimated roles of a die, 10, 100, and 1,000 times.

- S_4 was the only student to be judged to display Multistructural understanding in relation to the Law of Large Numbers. S_4 displayed virtually no variation in the graphs produced to show outcomes from a fair die (cf. Figure 9) and could calculate expected numbers of outcomes for different outcomes (e.g., 167 for each outcome from 1,000 tosses). He was, however, satisfied with simulations of 100 or less, focusing on sizes that were multiples of 6, apparently with the expectation of obtaining even numbers of outcomes. He relied more on comparing outcomes across samples than on increasing the sample size, recognizing the conflict when outcomes differed but not being able to resolve it. He was therefore deemed to display multi-structural understanding.

- Three responses were considered to be Unistructural in relation to the Law of Large Numbers. As an example of the difficulties experienced by these students, S_2 produced large variation in her drawings of all sample sizes for a fair die and during the simulation trials jumped around from trials of sizes 10, 100, 1,000, and back to 100, stating,

  "I'm actually testing each one a few times and then comparing them."

  Her expectation that large samples should show the same variation as small ones was displayed when she reduced the sample size for a loaded die from 1,000 to 100 stating,

  "it is a bit weird and reminds me of a joke,"

  because the smaller size had more variation,

  "so that is more fair."

  She could not resolve the conflict that the large number of trials highlighted the loaded number whereas the smaller number of trials produced more variation and disguised the loaded number. The other two students displayed similar conflict when exposed to different sizes of trials during the simulation part of the interview, after some initial comments about expecting more information from more trials.

- The response of S_7 with respect to the Law of Large Numbers was considered Prestructural because he could not label graphs for his predicted outcomes and there was large variation drawn for 1,000 trials. For the simulation he repeated small samples of size 10 and agreed to 50 when prompted. Even with help with decimals he could not appreciate the higher frequencies associated with the loaded number (5) and concluded that the die was fair because the other numbers had heights that were nearly the same as each other.

Overall connection

In relation to the overall connection between experimental and theoretical probability reflecting the concrete to abstract continuum, two students, S_3 and S_8, were able to articulate and demonstrate this understanding clearly in relation to the elements and connections in Figure 8.

- Although S_3's initial graphs (Figure 9) did not reflect his explanation that followed, he discussed a leveling out of the results in the larger trials and used the larger trials to determine fairness, and was therefore assessed to have a Relational understanding. S_3 also showed a more sophisticated understanding of the relation between the theoretical and experimental probabilities by using decimals. It was common among the other students to rely heavily on the visual representation of the leveling out of the data provided by the software.

  S_3, however, tracked the experimental outcomes of *TinkerPlots* Sampler trials on a piece of paper. After using a trial of 100,000 he explicitly demonstrated a connection between his experimental outcomes and the theoretical probability; he stated,

  > "they are all .16 [referring to the experimental outcomes], just as we found last time, the theoretical probability was .167, here they are all [.167 roughly]."

  The use of the decimals was considered to show a sophisticated proportional understanding for this age group.

- Although S_8, when talking about the connection of experimental and theoretical probability, suggested

  > "you need to experiment maybe to work out if the theory is true,"

  at one time he saw the simulator as being experimental because the outcomes were not all the same but later when asked if the simulator were closer to the real die on the table or to the theory of the die, he said,

  > "closer to theory … because it has been programmed to work around the theory with slight changes because it's random."

  As noted previously, because S_8's initial sketches of outcomes for different sizes of samples reflected appropriate understanding of the changing expectation involved and he was able to articulate an understanding of the connection between Theoretical and Experimental Probability, he was deemed to display a Relational understanding.

Two students, S_1 and S_4, were considered to demonstrate in their responses a Multistructural appreciation of the connection between experimental and theoretical probability:

- Although S_1 spoke of a smoothing out of the results and the images reflected this in relation to her conceived model (Figure 9), her initial intuitions regarding the distribution of numbers were incorrect, believing that she hardly ever rolled a 1 or a 6. This posed conflict, which she recognized, but it is not clear that she finally resolved it.

- S_4, during the hypothetical estimated trials of 10, 100, and 1,000 die rolls, demonstrated an understanding that experimental outcomes should reflect the theoretical probability, dividing each hypothetical trial size into roughly equal portions of .167. This was reflected in his initial graphs (Figure 9).

  Concern was raised, however, when he estimated the leveling out of data with small trial sizes of less than 100. This concern was reinforced later when he expressed surprise at the variation in outcomes in a trial size of 24, expecting a leveling out of the data. Even though S_4 was able to identify correctly the loaded die in each hypothetical scenario, he was not judged to demonstrate a Relational understanding. As S_4 used multiple samples of less than 100 to determine fairness, and did not recognize the significance of "large" trials, he was unable to use the Law of Large Numbers to determine fairness and demonstrate the link between the theoretical and experimental probabilities.

Three other students, although able to articulate a basic understanding, were unable to apply this understanding in the hypothetical and simulated scenarios.

Two students, S_2 and S_5, at times articulated an understanding of a leveling out of the data in the larger number of trials, however a belief that fairness was demonstrated by short-term variance in the outcomes hindered the development of an understanding of the connection between experimental and theoretical probability. It appears that these students saw the connection as: theoretically, the outcomes are equally likely and as the results are random, short term variance in experimental outcomes demonstrates fairness.

As there was a variety of contradictory unresolved issues, seen for example in their initial sketches (Figure 9), a Unistructural understanding was demonstrated.

The third student, S_6 could not articulate the links through the Law of Large Numbers, being reluctant to suggest large sample sizes and struggling to reconcile simulations with a belief that outcomes are

> "theoretical when it's all even."

One student in the interview had difficulty expressing the links among the elements.

- Especially for links between the elements related to the Law of Large Numbers, a lack of appreciation of fairness, and inability to interpret the numbers that might demonstrate fairness for the "complete" die rather than for "some numbers and not others." His response (S_7) overall in the interview was considered to be Prestructural, even though his written reflection response at the end of the lessons was scored as 6 out of 9.

4. DISCUSSION

All of the students interviewed were able to articulate, at some stage during the research, a basic level of understanding of the relationship between experimental and theoretical probability, mentioning a leveling out of results or "proving" of one another. S_3 and S_8 provided the richest examples of students making the connections between experimental and theoretical probability. They were able to connect the experimental outcomes on the simulator with the appropriate theoretical probability, as well as demonstrating an understanding of the law of large numbers, proportional reasoning, and the underpinning theoretical concepts of fairness and equally likely outcomes.

These findings support the conclusions of Stohl, Rider, & Tarr (2004) that students who used larger sample sizes made appropriate inferences regarding fairness and of Abrahamson & Cendak (2006) that students recognized the greater proportion of some binomial combinations with the use of larger samples in a simulation setting. Recent research by Prediger (2008), although not carried out in an environment with simulation software, reached a similar conclusion working with students in a gaming context.

The conflict experienced by students in differentiating pattern and variation as significant for different sample sizes reflects the analysis of Borovcnik & Bentz (1991). In addition to these outcomes, it is apparent that a linked network of concepts, proportional reasoning, and an understanding of the implications of the underpinning theoretical concepts of fairness and equally likely outcomes are key to developing students' understanding of the connection between the experimental outcomes and the associated theoretical probability. Although not explicitly associated with the demonstration of the Law of Large Numbers, the outcomes reported by Abrahamson, Janusz & Wilensky (2006) support these other aspects of building understanding.

In relation to the other interviewed students, the lack of understanding of the implications of the underpinning theoretical concept of fairness was a barrier to connecting theoretical probability to the experimental outcomes. The determination of fairness and the testing of the theoretical probability appeared conceptually different to the students' minds. The students understood fairness to mean short-term randomness, so to test this required a small trial. When testing the theoretical probability some students appeared more comfortable with larger samples. S_3 demonstrated this responding "6 rolls" to a question about determining fairness but then proceeding to use trials of 10,000 to test the theoretical probability. S_1 demonstrated the clearest connection, referring back to the whole class lesson and the large number of flips it required to test the fairness of the coin.

Linking these two concepts will help students understand better the connection between the theoretical probability and the experimental outcomes. When considering the students' current exposure to sample sizes, such as in common board games that involve 10 to 20 rolls of a die, it is understandable that they consider fairness in terms of short term variation. It is therefore imperative to expose students to broader real world contexts that incorporate larger samples and discussions of randomness and fairness in consideration of these sample sizes. *TinkerPlots* provides further opportunity for this type of discussion by means of a table that records the number of times a specific outcome is achieved in succession in a large number of trials. Many opportunities hence exist for providing cognitive conflict for preconceived ideas of fairness and sample size.

When engaging in the simulated trials the students demonstrated a reliance on the visual representations to determine fairness. In the continuum from concrete to abstract the reliance on the graphs from simulations is conceptually in the middle, the link to understanding more abstract concepts. An over reliance on the graphs however raised concern in this study, as scale plays a large role in providing the perceived leveling out of the data. The Sampler automatically adjusts the scale of the graph from the smaller number of trials to the larger number of trials, keeping the overall graph the same size.

Appropriate interpretation of these graphs relies on a high level understanding of proportional reasoning. Van de Walle (2004) identified proportional reasoning as an important curriculum content connection to the understanding of probability. If the Sampler did not alter the scale, then the graph would not visually show a leveling out of the results. This highlights the importance of focusing the students' attention on the decimal summary of outcomes, as it gives results in proportion to the total and is easily comparable to the theoretical probability.

The use of a probability simulator such as the *TinkerPlots* Sampler can provide a meaningful link to concrete manipulatives and can make large samples, and potentially the Law of Large Numbers, easily accessible. Students in the classroom and in the interviews were able to follow the generation of outcomes in relation to the corresponding tangible experimental outcomes, which in turn assisted in developing the overall connection between experimental and theoretical probability.

5. CONCLUSION

As Fischbein (1982) noted, creating new correct probabilistic intuitions requires guessing outcomes, performing experiments and evaluating outcomes, not just practicing probability calculations. This study provides some helpful directions on how this regime might be implemented. In documenting students' understandings of the connections between theoretical and experimental probability many students demonstrated the building blocks necessary to bridge understandings of the two concepts (see Figures 7 and 8).

It is evident from this study that it is insufficient for educators to assume that students understand theoretical probability if they can calculate it. Instead, there is need to investigate and develop students' understanding of the underpinning concepts of fairness, equally likely, and random in large and small numbers of trials. It is also apparent from this study that it is insufficient for educators to focus on the calculation of the theoretical probability and the observation of experimental outcomes to develop students' understanding of the connection between these two concepts; this connection needs to be taught and experienced explicitly.

The authors agree with Konold & Kazak (2008) that it is important to begin these experiences when students are young to give time for the exposure over many years. The results of this investigation indicate that there is hope for understanding the link between the concrete and the abstract.

ACKNOWLEDGEMENT

The authors thank Cliff Konold for allowing the new development version of *TinkerPlots* Sampler to be used in this study. The original research associated with this paper was part of the first author's honors dissertation in the Faculty of Education at the University of Tasmania.

REFERENCES

Abrahamson, D. & Cendak, R. M. (2006). The odds of understanding the law of large numbers: A design for grounding intuitive probability in combinatorial analysis. In J. Novotná, H. Moraová, M. Krátká, & N. Stehlíková (Eds.), *Proceedings of the Thirtieth Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 1–8). Prague: Charles University. Retrieved June 5, 2009, from http://gse.berkeley.edu/faculty/dabrahamson/publications/Abrahamson-Cendak_PME30.pdf.

Abrahamson, D., Janusz, R., & Wilensky, U. (2006). There once was a 9-Block ... – A middle-school design for probability and statistics. *Journal of Statistics Education*, 14(1). Retrieved June 5, 2009, from http://www.amstat.org/publications/jse/v14n1/abrahamson.html.

Abrahamson, D., & Wilensky, U. (n. d.). *Connected Mathematics: ProbLab*. The Center for Connected Learning and Computer-Based Modeling. Retrieved June 5, 2009, from http://ccl.northwestern.edu/ProbLab/ .

Abrahamson, D., & Wilensky, U. (2007). Learning axes and bridging tools in a technology-based design for statistics. *International Journal of Computers for Mathematics Learning*, 12(1), 23–55. Retrieved June 5, 2009, from http://ccl.northwestern.edu/papers/2007/04-learningaxes.pdf.

Basson, A., Krantz, S., & Thornton, B. (2006). A new kind of instructional mathematics laboratory. *Primus,* 16(4), 332–348.

Ben-Zvi, D., Gil, E., & Apel, N. (2007). What is hidden beyond the data? Helping young students to reason and argue about some wider universe. In D. Pratt & J. Ainley (Eds.), *Reasoning about Informal Inferential Statistical Reasoning: A collection of current research studies*. Proceedings SRTL-5. Retrieved June 5,, 2009, from http://srtl.stat.auckland.ac.nz/srtl5/view_presentation/presentation_id=49.

Biggs, J. B. & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (structure of the observed learning outcome)*. New York: Academic Press.

Biggs, J. B. & Collis, K. F. (1991). Multimodal learning and the quality of intelligent behaviour. In H. A. H. Rowe (Ed.), *Intelligence: Reconceptualisation and measurement* (pp. 57–76). Hillsdale, NJ: Lawrence Erlbaum Associates.

Blythe, T. (1998). *The teaching for understanding guide.* San Francisco: Jossey-Bass.

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

Clements, D. (1999). Concrete manipulatives, concrete ideas. *Contemporary Issues in Early Childhood*, 1(1), 45–60.

Department of Education. (2008). *The Tasmanian Curriculum: Mathematics-numeracy. K–10 syllabus and support materials.* Hobart: Author. Retrieved June 5, 2009, from http://www.education.tas.gov.au/curriculum/standards/maths.

Erickson, T. (2006). Using simulation to learn about inference. In A. Rossman & B. Chance (Eds.), *Proc. Seventh Intern. Conf. Teaching Statistics: Working Cooperatively in Statistics Education, Salvador, Brazil.* Voorburg: International Statistical Institute. [CDROM].

Fischbein, E. (1982). Intuition and proof. *For the Learning of Mathematics*, 3(2), 9–19.

Ireland, S. (2007). *Making connections between concrete objects and abstract concepts: Developing students' understanding of the connection between observed experimental probability outcomes and the associated theoretical probability, aided by computer simulation software.* Unpublished Honours dissertation. Hobart: Education Faculty, University of Tasmania.

James, G. & James, R. (1959). *Mathematics dictionary.* Princeton, NJ: D. Van Nostrand Company Inc.

Jones, G. (2005). Reflections. In G. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 367–372). New York: Springer.

Konold, C. (2006). Handling complexity in the design of educational software tools. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics: Working Cooperatively in Statistics Education, Salvador, Brazil.* [CDROM] Voorburg: International Association for Statistical Education and the International Statistical Institute.

Konold, C., Harradine, A., & Kazak, S. (2007). Understanding distributions by modelling them. *International Journal of Computers for Mathematical Learning*, 12, 217–230.

Konold, C. & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education*, *2*, Article 1. Retrieved June 5, 2009, from http://repositories.cdlib.org/uclastat/cts/tise/vol2/iss1/art1.

Konold, C. & Lehrer, R. (2008). Technology and mathematics education: An essay in honor of Jim Kaput. In L.D. English (Ed.), *Handbook of international research in mathematics education* (2nd ed.) (pp. 49–71). New York: Routledge.

Konold, C. & Miller, C. D. (1993). *ProbSim*. [Computer software] Amherst, MA: University of Massachusetts.

Konold, C. & Miller, C. D. (2005). *TinkerPlots: Dynamic data exploration.* [Computer software] Emeryville, CA: Key Curriculum Press.

Malhotra, N., Hall, J., Shaw, M., & Crisp, M. (1996). *Marketing research: An applied orientation*. Sydney: Prentice Hall Australia.

Mills, J. (2002). Using computer simulation methods to teach statistics: A review of the literature. *Journal of Statistics Education*, *10*(1). Retrieved June 5, 2009, from http://www.amstat.org/publications/jse/v10n1/mills.html.

Paparistodemou, E. & Meletiou-Mavrotheris, M. (2008). Developing young students' informal inference skills in data analysis. *Statistics Education Research Journal*, 7(2), 83–106. Retrieved June 5, 2009, from http://www.stat.auckland.ac.nz/~iase/serj/SERJ7(2)_Paparistodemou.pdf.

Piaget, J. (1955). *The construction of reality in the child*. London: Routledge & Kegan Paul.

Pratt, D. & Noss, R. (2002). The microevolution of mathematical knowledge: The case of randomness. *Journal of Learning Science*, 11, 453–488.

Prediger, S. (2008). Do you want me to do it with probability or with my normal thinking? Horizontal and vertical views on the formation of stochastic conceptions. *International Electronic Journal of Mathematics Education*, 3(3), 126–154. Retrieved June 5, ,2009, from http://www.iejme.com/032008/ab1.htm.

Rider, R. & Stohl Lee, H. (2006). Differences in students' use of computer simulation tools and reasoning about empirical data and theoretical distributions. In A. Rossman & B. Chance (Eds.), *Proc. Seventh Intern. Conf. Teaching Statistics: Working Cooperatively in Statistics Education, Salvador, Brazil.* Voorburg: International Statistical Institute. [CDROM].

Sarama, J. & Clements, D. (1998). Using computers for algebraic thinking. *Teaching Children Mathematics*, 5(3), 186–190.

Steen, L. (2007). Every teacher is a teacher of mathematics. *Principal Leadership*, 7(5), 16–20.

Stein, M. & Bovalino, J. (2001). Manipulatives: One piece of the puzzle. *Mathematics Teaching in the Middle School*, 6(6), 356–359.

Stohl, H., Rider, R., & Tarr, J. (2004). *Making connections between empirical and theoretical probability: Students' generation and analysis of data in a technological environment.* Retrieved June 5, 2009, from http://www.probexplorer.com/Articles/LeeRiderTarrConnectE&T.pdf

Stohl, H. & Tarr, J. (2002). Developing notions of inference using probability simulation tools. *Journal of Mathematical Behaviour*, 21, 319–337.

Van de Walle, J. (2004). *Elementary and middle school mathematics: Teaching developmentally* (5th ed.). Boston: Pearson Education, Inc.

Watson, J. M. & Callingham, R. A. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, 2(2), 3–46. Retrieved June 5, 2009, from http://www.stat.auckland.ac.nz/~iase/serj/SERJ2(2)_Watson_Callingham.pdf.

Watson, J. M., Collis, K. F., & Moritz, J. B. (1997). The development of chance measurement. *Mathematics Education Research Journal*, 9, 60–82.

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

Wilensky, U. (1991). Abstract meditations on the concrete and concrete implications for mathematics education. In I. Harel & S. Papert (Eds.), *Constructionism* (pp. 193–204). Norwood, NJ: Ablex Publishing Corporation. Retrieved June 5, 2009, from http://ccl.northwestern.edu/papers/concrete/.

Wordnet. (n. d.). Computer simulation. *Wordnet: A lexical database for the English language.* Retrieved June 1, 2007, from http://wordnetweb.princeton.edu/perl/webwn?s=computer%20simulation.

ALL APPENDICES

Appendix A and B is also contained in the paper. For all appendices to this paper, follow this link.

Authors : Seth Ireland and Jane Watson
E-mails : seth.ireland@education.tas.gov.au – Jane.Watson@utas.edu.au
Address (for both) : Faculty of Education,
University of Tasmania, Hobart, Tasmania, Australia

APPENDIX A

Section 1 of Class Test: Probabilistic intuitions

Methodology section

1. Complete the sentence

(a) One thing that **will certainly** happen today is _____

_____

_____

(b) One thing that is **impossible** today is _____

_____

_____

(c) One thing that **might possibly** happen today is _____

_____

_____

2. What things happen in a "random" way?

3. One day, Claire won Tattslotto with the numbers

1; 7; 13; 21; 22; 36.

So she said she would always play the same group of numbers, because they were lucky.

What do you think about this?

4. Consider rolling one six-sided die. Is it easier to throw?

- ❒ (1) a one; or
- ❒ (6) a six; or
- ❒ (=) are both a one and a six equally easy to throw?

Please explain your answer.

5. Imagine you threw the die 60 times. Fill in the table below to show how many times each number might come up.

Number on Dice	How many times it might come up
1	
2	
3	
4	
5	
6	
TOTAL	60

Why do you think these numbers are reasonable?

6. A mathematics class has 13 boys and 16 girls in it. Each pupil's name is written on a piece of paper. All the names are put in a hat. The teacher picks out one name without looking.

Is it more likely that

❒ (b) the name is a boy

or

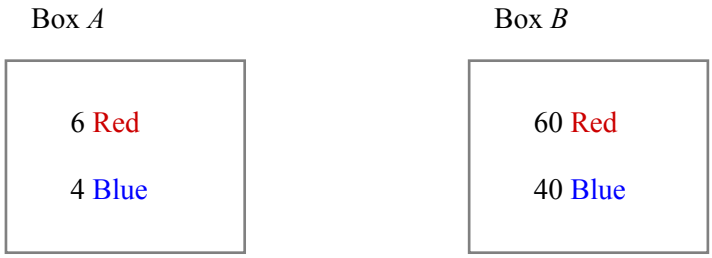
❒ (g) the name is a girl

or

❒ (=) are both a girl and a boy equally likely?

Please explain your answer.

7. Box *A* and Box *B* are filled with red and blue marbles as follows:



Each box is shaken. You want to get a blue marble, but you are only allowed to pick out one marble without looking.

Which box should you choose? (Circle your answer)

❒ (*A*) Box *A* (with 6 red and 4 blue).

❒ (*B*) Box *B* (with 60 red and 40 blue).

❒ (=) It doesn't matter.

Please explain your answer.

8. If you flipped a coin what is the chance you will get a heads?

Explain your answer.

APPENDIX B

Personal Interview Protocol Interview procedure in methodology section

Issue addressed	Questions
Introduction	
Relaxing	• Show Die
Participants	• How did you learn from the lessons? What was easy what was hard?
Manipulatives	
General	• Describe to me the object in your hands? • What other objects are this shape? • Where do we use this object? And how?
Theoretical	• What happens if we toss the die? • What are the possible outcomes? • What do you think the probabilities of these outcomes might be? • Does this mean that all outcomes are "Equally likely"?
Experimental	• What would it mean if they were not "Equally likely"?
Experimental Concrete	• I was playing a game with someone and they said that they never get a 6, they believed the die was loaded. What ways can I find out if it is loaded? Be prepared for responses like "feel if it is heavy on one side," "check the numbers," or "all dice are fair, you don't have to check." • In probability we make the assumption that the dice is fair. What is meant by fair?
Experimental & Theoretical	• How many rolls do you think you would need to make to be certain it was fair? Why? • GET PEN AND PAPER. DRAW GRAPHS. • If it was a fair die, what would you expect the results to look like after 10 rolls? Why? If I did 10 rolls another 3 times? Would it be different? • 100 rolls? Why? If I did it another 3 times? • 1,000? If I did it another 3 times? • Would the results be different if it was an unfair dice? How? • How fair (equally likely) do you consider a normal dice to be? Why?

Computers	
Computers & manipulations	• Explain to me what is happening on the screen when I run a trial? • How is that similar to what we did in class with the coin? • What does the spinner do? • In your own words how does the simulator represent the die? (Show spinner) • What are the possible outcomes? • What do you think the probabilities of these outcomes might be? • Does this mean that all outcomes as "Equally likely" as the die? Why? (Show mixer)
Computers manipulations & Theory	• What are the possible outcomes? (Why same as before?) • What do you think the probabilities of these outcomes might be? • Does this mean that all outcomes as "Equally likely" as the die? Why? • Of the spinner and mixer, which do you think is a better representation of the die? Why? • How do you think the computer randomly chooses the numbers from 1–6?
Computer to manipulations	• The sampler has been set up the same as this die. It has the same theoretical probability. Let's run some samples, you tell me when you are satisfied that the die is fair or not. ○ Start with small samples of 10, 20, 50. Ask for student input on sample size. ○ What do you notice about the results? ○ What do you think will happen after we added these additional samples? ○ Are you satisfied yet? ○ How did you come to that conclusion?
Experimental & theoretical	• Explore simulations with loaded dice. Have students explore how many tosses they need to infer what the theoretical probability might be and how many times they need to test it. • Show correct loadings of the loaded dice and ask where does this (the spinner loadings) fit into the range from theoretical to experimental probability? Why? • Did you prefer working with the dice or computer better? Why? • How fair do you consider the computer simulation to be? Why? • Is the computer fairer then the dice? Why/why not? • Is the simulator "closer" to the real die we have here on the table or "closer" to the concept of a fair die we talked about with equal probability for each number? Why? • In your own words explain how theoretical and experimental probability of a die are related? Can you give me an analogy?