

THE ROLE OF PROBABILITY IN DEVELOPING LEARNERS' MODELS OF SIMULATION APPROACHES TO INFERENCE

HOLLYLYNNE S. LEE

*NC State University
hollylynne@ncsu.edu*

HELEN M. DOERR

*Syracuse University
hmdoerr@syr.edu*

DUNG TRAN

*Victoria University, Australia
Dung.Tran1@vu.edu.au*

JENNIFER N. LOVETT

*NC State University
jnickel@ncsu.edu*

ABSTRACT

Repeated sampling approaches to inference that rely on simulations have recently gained prominence in statistics education, and probabilistic concepts are at the core of this approach. In this approach, learners need to develop a mapping among the problem situation, a physical enactment, computer representations, and the underlying randomization and sampling processes. We explicate the role of probability in this approach and draw upon a models and modeling perspective to support the development of teachers' models for using a repeated sampling approach for inference. We explicate the model development task sequence and examine the teachers' representations of their conceptualizations of a repeated sampling approach for inference. We propose key conceptualizations that can guide instruction when using simulations and repeated sampling for drawing inferences.

Keywords: *Statistics education research; Simulation models; Teachers*

1. INTRODUCTION

Making inferences from data is central for doing statistics. Repeated sampling approaches to inference have recently gained prominence in statistics education and probabilistic concepts are at the core of this approach. A simulation that uses repeated sampling can be an important tool to help students develop a deep understanding of the abstract statistical concepts involved in inferential reasoning (Burrill, 2002; Maxara & Biehler, 2006). Indeed,

Cobb (2007) suggests that educators can help students develop an understanding of inference through the “three R’s: randomize data, repeat by simulation, and reject any model that puts your data in its tail” (p.12). In this paper, we aim to unpack this “three R’s” approach to illuminate the role of probability in such a simulation approach to inference.

Simulations of repeated sampling have been used in several curricula efforts in the United States at the collegiate level, and researchers have reported modest results in improvement of students’ understandings of inference through this approach (e.g., Garfield, delMas, & Zieffler, 2012; Tintle, Topliff, Vanderstoep, Holmes, & Swanson, 2012). In addition, new curriculum standards in some countries, such as the United States and New Zealand, suggest such an approach for high school students (Council of Chief State School Officers, 2010; New Zealand Ministry of Education, 2006).

Many curriculum developers recommend that learners experience repeated sampling methods in a physical way before using computing power (e.g., Cobb, 2007; Rossman, 2008). The physical enactments are intended to assist learners in conceiving of the process of sampling as a repeatable action (e.g., Watson & Chance, 2012). Thus, physical enactments serve as a way to model the underlying probability assumptions in a problem (e.g., is it equally likely for two events to occur? does each person have an equal chance to be assigned to a treatment group?). However, in many curricula, the repeated sampling process is often created by the instructor or the curriculum materials, and students are asked to use a prescribed process with physical objects or are told exactly what to input in a computer simulation (e.g., Cumming, Miller, & Pfannkuch, 2014; Roy et al., 2014). A simulation approach (using physical and computer tools) seems to be an appropriate way to help students develop statistical inference conceptually. However, while the “three R’s” process may seem simple, understanding all parts of a simulation is conceptually complicated. In fact, learners who know *how* to conduct a simulation may not have a robust understanding of *why* they are conducting a simulation, *what* is being simulated, and *how* to make appropriate conclusions based on a simulation. Thus, our research aims to consider *how learners and teachers could build a (conceptual) model for a repeated sampling approach to inference that highlights the role of probability models, uses both the power of physical models and the capabilities of computer simulations, and can be generally applied to many inference contexts.*

2. BACKGROUND AND THEORETICAL PERSPECTIVES

There are two predominate ways of using models in statistical thinking: 1) “select or design and use appropriate models to simulate data to answer a research question”, and 2) “fit a statistical model to existing data or data you have collected through a survey or experiment in order to explain and describe the variability” (Garfield & Ben-Zvi, 2008, p. 147). In our work, we are focused on the first type of model. Simulations are in and of themselves models of a real world process. When designing and using a simulation, one is trying to model some process so as potentially to understand better the inputs, inner workings, and outputs of the process. At the core of a simulation approach to inference is randomness—the first of the “three R’s”. However, at every step of a simulation based repeated sampling approach to inference there is much deeper understanding of other issues concerning probability models that we aim to help unpack in our teaching and research. What follows are descriptions of the theoretical perspectives we use in our work related to probability models, models and

modeling perspective on learning, and repeated sampling representations and processes described by other researchers.

2.1. ROLE OF PROBABILITY MODELS

Probability is an abstract concept that cannot be directly measured. There are many objects or real world events whose outcome cannot be completely determined unless you know all information about actions on the object or event in world (e.g., we can not determine the outcome of a coin toss unless we know all the physical forces acting upon the coin while it is tossed). In the absence of such knowledge, we can build models to express a probability of an event occurring. A simple example is to consider that in a given toss of a coin, without knowing all the forces acting upon the coin, we can not determine the outcome of which side will land facing up; thus, we build a probability model to estimate the likelihood of each side occurring.

Three of the most common ways to approach developing a probability model are the classical (or Laplacean), frequentist, and subjectivist. In each of these approaches, there is an aspect of defining and structuring the problem at hand along with looking forward and looking backwards (Eichler & Vogel, 2014). When looking forward, one is building a model with predictive power of what is expected to happen in the future. When looking backwards, one is building a model based on data that has already occurred or on new data based on a simulation using a model for a probability of an event. In every instance, the models have underlying assumptions. Depending on the approach taken, looking forward and backwards may occur at different times in solving a problem and serve particular purposes.

Chaput, Girard, and Henry (2011) describe three parts of a probability modeling process that includes embodying observations and assumptions into a pseudo-concrete working model, mathematizing the model into a hypothesis driven probability model that can be enacted, validating a model through examining the fit to data, and interpreting the model in the context of the problem. Many have advocated that we want students and teachers to understand the bi-directional relationship between probability models and data, and between empirically-developed models and theoretically-developed models (e.g., Eichler & Vogel, 2014; Konold & Kazak, 2008; Lee & Lee, 2009; Pfannkuch & Ziedins, 2014; Pratt, 2011; Stohl & Tarr, 2002; Wild, 2006). We claim that in using a repeated sampling approach to inference, the assumptions and model-building process in a simulation should be made more explicit. Pfannkuch and Ziedins' (2014) description of probability models and their purpose provides a useful perspective on the strong role that probability models have in a repeated sampling approach to inference:

A probability model will often be associated with the idea of a system evolving dynamically over time... a model is usually built to answer a particular question or questions about a system, sometimes just to understand its behavior better, but often in order to optimize some measure of its performance, or alternatively, to predict performance under some alternative scenario. ...although they [models] are only approximations to what happens in the real world, these approximations can help us better understand the behavior in the real world. (p. 103)

In a repeated sampling approach to inference, students and teachers should be conceiving of the observed outcome (from an observational study or an experimental design) as resulting from a process that is repeatable, and that repeating the process may result in a different

outcome. Thus, the question becomes, how unusual is what happened in the particular instance that we know about already? In other words, what is the likelihood of a particular outcome occurring if a process is repeated many times? That is the end goal. But to begin with, we need to make sense of the problem we are trying to solve and ask what are some of the underlying assumptions we need to consider? What is the process in the problem that is being repeated and what is the role of randomness and probability in that process? All of these questions require a conceptual understanding (or model) of the simulation process and its underlying assumptions.

2.2. MODELS AND MODELING PERSPECTIVE

Given our focus on using probability as a model, it made sense to situate our work in a models and modeling perspective on teaching and learning mathematics as articulated by Lesh and Doerr (2003). In this perspective, the goal for the learner is to build a model that can be generalized to other problem situations and productively re-used in a range of contexts. Thus, we are particularly interested in how teachers can develop a robust model of using repeated sampling for making inferences for problem situations. Such a model includes understanding the relationships among the problem situation, physical enactments of sampling, representations of those enactments, computer representations, the underlying randomization (i.e., the probability models discussed above), the distribution of the statistics of interest, and how to interpret and use such a distribution to make a decision. To achieve this goal, we designed and implemented a sequence of model development activities (Årlebäck, Doerr, & O’Neil, 2013; Doerr & English, 2003; Hjalmarson, Diefes-Dux, Bowman, & Zawojewski, 2006; Hjalmarson, Diefes-Dux, & Moore, 2008).

Model development sequences are structurally related tasks that begin with a *model eliciting activity* (MEA) and are followed by *model exploration activities* (MXA) and *model application activities* (MAA), as shown in Figure 1. These sequences provide a way of organizing instruction on a central concept such as proportional reasoning, rates of change, or variation and distribution. The MEA elicits learners’ initial models of a realistic problem situation. These models are systems of elements, relationships, operations, and rules that can be used to describe, explain or predict the behavior of the realistic situation. The model exploration activities then engage learners in thinking *about* the initial models that were elicited. The model exploration activities focus on the underlying structure of the model and on the strengths of various representations and ways of using them productively. Model application activities engage learners in thinking *with* their models by applying them to new contexts. This results in learners making adaptations to their model, extending previously explored representations, and refining language for describing, explaining or predicting the behavior of some realistic phenomena. Each component of a model development sequence engages learners in multiple cycles of descriptions, interpretations, conjectures and explanations that are tested, revised and refined while working with other learners. In this way, a central concept such as a repeated sampling approach to inference is not understood all at once, but rather learners deepen their understandings as their model (or conceptual system) is developed and revised over time through a sequence of modeling tasks.

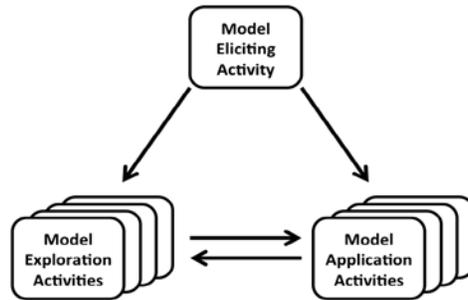


Figure 1. The general structure of a model development sequence (Ärlebäck et al., 2013, p. 317).

2.3. VISUAL REPRESENTATIONS OF REPEATED SAMPLING

Much research and curriculum development in recent years has focused on understanding inference and simulation approaches. For the purposes of our paper, we will highlight the work of several researchers whose visual representations and descriptions of repeated sampling informed the design of the model development sequence, its implementation and our research. In 2002, Saldanha and Thompson reported that when students can visualize a simulation process through a three-tier scheme, they develop a deeper understanding of the process and logic of inference. This scheme is centered on “the images of repeatedly sampling from a population, recording a statistic, and tracking the accumulation of statistics as they distribute themselves along a range of possibilities” (p. 261). The diagram in Figure 2 is meant to draw attention explicitly to the multiplicative relationship among a population, sample(s), and a distribution of sample statistics. Their work also had students experience and attend to three levels in the sampling process: 1) randomly draw items to form a sample of a given size and record a statistic of interest, 2) repeat Level 1 process a large number of times and accumulate a collection of statistics, and 3) partition the collection of statistics to determine what proportion lies beyond a given value.

Several researchers have built from Saldanha and Thompson’s (2002) multi-tier scheme and the models and modeling work of Lesh and Doerr (2003) for creating other visual representations that can assist students when using simulation approaches, or repeated sampling techniques, for inference (Garfield et al., 2012; Lane-Getaz, 2006). Lane-Getaz offered the Simulation Process Model (SPM) including three tiers: population parameters, random samples, and distribution of sample statistics. The SPM diagram resembles Saldanha and Thompson’s representation and verbal description of three levels, but uses more explicit language in the diagram itself. The first tier is to describe the population distribution as the beginning of the simulation process, then, random samples are drawn from the population, and a sample statistic is selected related to the simulation process for Tier 2. In the last and third tier, the distribution of the sample statistics is formulated, and used to evaluate the likelihood of the event happened in the original problem (Figure 3). Lane-Getaz described how she used the SPM diagram as an organizer to help students understand the general process of inference. She then adapted the SPM to specific examples used in her course so that students can see how the diagram frames the simulation process used in different contexts.

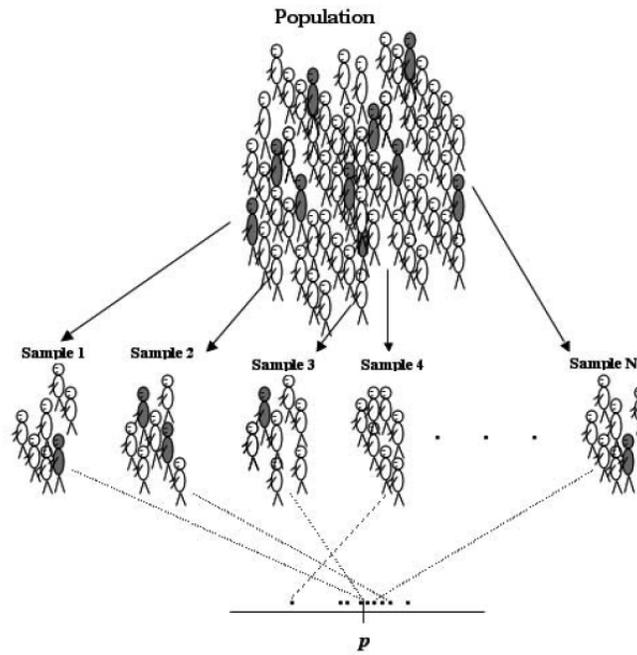


Figure 2. Diagram for sampling conception (Saldanha & Thompson, 2002, p. 267).

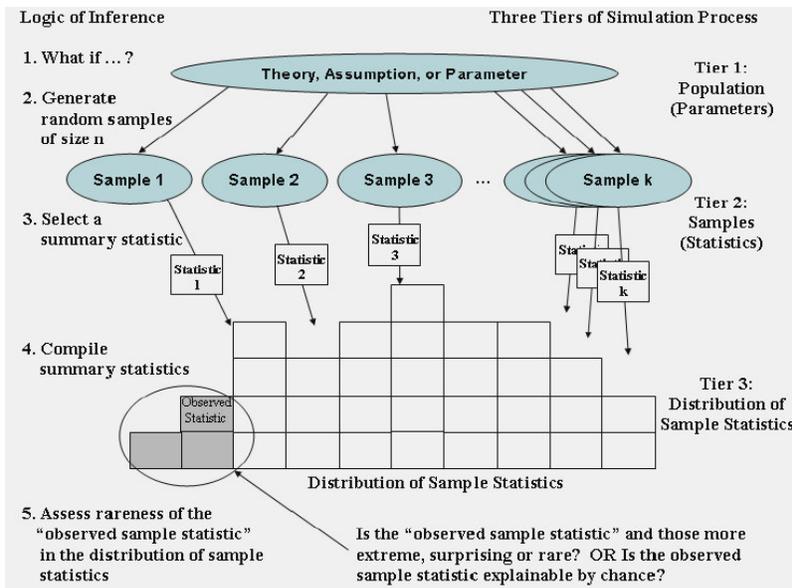


Figure 3. Representation of Simulation Process Model (Lane-Getaz, 2006, p. 280).

In line with Lane-Getz's suggestion, Garfield et al. (2012) used a models and modeling approach in the design and research of the CATALYST curriculum (Catalyst for Change, 2012). They describe a generalized structure to the logic of a simulation approach to inference that includes specifying a model to be used to generate data, using the model to generate simulated data for a single trial and then multiple trials, each time collecting a statistic of interest, and finally using the distribution of the collected summary measures to compare the observed data with the behavior of the model (Garfield et al., 2012, see p. 887 for more details). Garfield and colleagues advocated using a general structured diagram with students to help organize their thinking about the general simulation process and for specific problems. Figure 4 shows their three-level diagram applied to a particular problem (Cereal Box prizes) that includes specifying a model, randomizing and repeating samples and collecting numerical summary measures, and evaluating a distribution of the numerical summary measures.

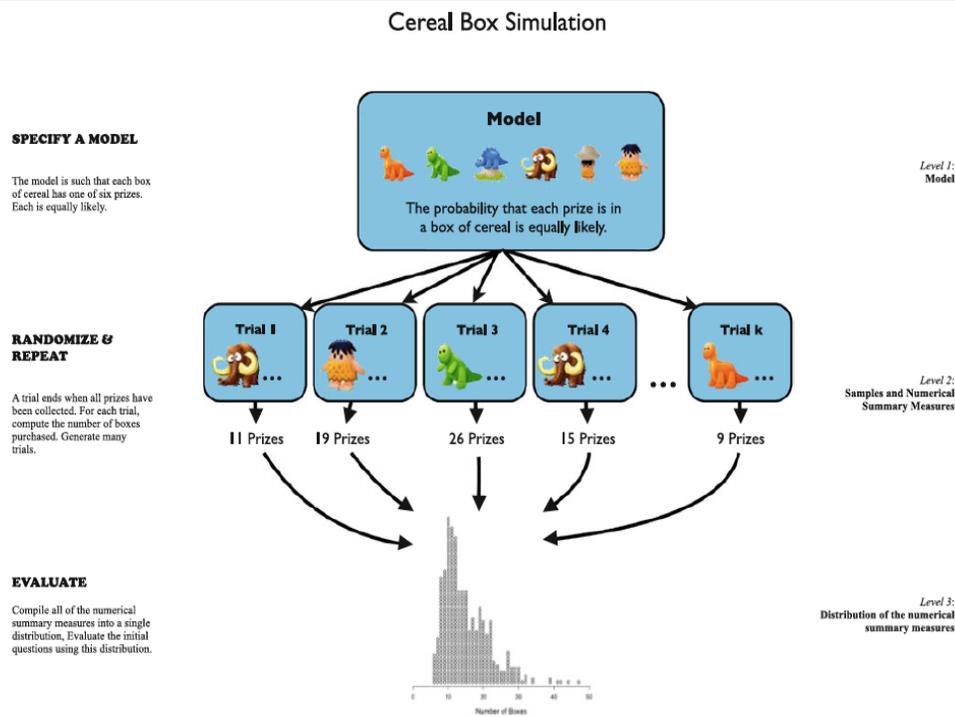


Figure 4. Diagram of CATALYST modeling and simulation process (Garfield et al., 2012, p. 890).

Our understanding of the literature on probability models and repeated sampling approaches to inference, and the representations used by others, informed our design of a sequence of model development tasks to use with participants in a graduate course on teaching and learning statistics. What follows is a description of the course, participants, and a set of tasks that took a repeated sampling approach to inference using both physical and computer-based simulations. The set of tasks served as a model development sequence (Lesh, Cramer, Doerr, Post, & Zawojewski, 2003) that enabled us to support the development of

participants' understanding of a simulation approach to inference while also revealing and eliciting their thinking.

3. THE STUDY

3.1. SETTING AND PARTICIPANTS

A team of four instructors from two institutions met weekly via videoconference for an academic year to design a 15-week graduate-level course offered at each institution, and to discuss issues and alter plans as the course was taught. The course consisted of opportunities for participants, mostly secondary and post-secondary teachers, to engage in statistical investigations with real data and tasks designed to develop understandings of distribution, samples and sampling distributions, and inferential statistics, especially using repeated sampling approaches. Across the two institutions, the course served a variety of graduate students ($n = 27$). Participants consisted of one undergraduate pre-service teacher, six pre-service teachers in a masters program, 11 in-service teachers enrolled in a masters program, one full-time masters student in mathematics education, eight doctoral students in mathematics or mathematics education. Twenty-one participants were female and six were male. Six participants indicated that English was their second language. Most participants had completed the equivalent of an undergraduate degree in mathematics, and all but two had at least one prior course in statistics.

3.2. THE MODEL DEVELOPMENT SEQUENCE

Our instructional goal for our participants was to support them in developing a general model that they could use to approach inference situations using a repeated sampling approach, and for them to be able to assist other learners in using a simulation approach. This model includes understanding the relationships among the problem situation, physical enactments of sampling, representations of those enactments, computer representations, and the underlying randomization (i.e., the probability models discussed above), the distribution of the statistics of interest and how to interpret and use such a distribution to make a decision. The model development sequence we designed and implemented is about having learners (in this case participants who are interested in teaching statistics) develop a model for repeated sampling as an approach for making inferences. In order for participants to develop that model (and the entailments needed for teaching that model), they have to be able to make connections to and use the underlying probability model of repeatable actions with unpredictable outcomes. What follows is a detailed description of the model development sequence of tasks, and how we thought each task would contribute to the development of participants' general models for using a repeated sampling approach to inference. Further details about our planned and enacted trajectory for our learners are in Arnold, Confrey, Jones, Lee, and Pfannkuch (forthcoming).

Model eliciting activity 1 As we noted above, at the core of a simulation approach is a random process. By this we mean a repeatable action whose outcome cannot be determined ahead of time. While we were certain that our participants would be familiar with coin tosses, spinners and rolls of a die, we did not assume that they fully appreciated the role of

such repeatable actions in simulations. Thus, as with all model development sequences, the first model eliciting activity (MEA1) was designed to bring forth the concept of using probability models to describe results (expected or observed) from a repeatable action. We engaged students in using coins, spinners, marbles, plastic cups and both fair and unfair dice to identify the repeatable action, describe the possible outcomes of those actions, identify an event of interest, and then repeatedly perform some action with the hands-on material they were given, record the event of interest, and then collect the results. Finally, they needed to estimate a probability distribution for the event of interest. This experience made it clear to our participants that assumptions about randomness and equiprobability needed to be grappled with and were not always evident. It also drew participants' attention to the need to explicitly describe what action was being repeated and what event was actually of interest. They seemed to walk away with a new founded appreciation that using physical devices in probability and statistics lessons required much more careful attention and consideration than they had typically experienced as a learner.

Model eliciting activity 2 A few weeks later, after engaging in informal inference tasks using probability simulations of unfair dice and examining students' reasoning (Schoolopoly task, see Lee, Angotti, & Tarr, 2010), we focused more explicitly on using a simulation approach to inference by engaging in a second model eliciting activity (MEA2), adapted from the *Paul the Octopus* task (Lock, Lock, Morgan, Lock, & Lock, 2013). In this task, the famous octopus, Paul, had achieved eight out of eight correct predictions by swimming in a tank to choose one of two boxes marked with the country flags to eat from. His choice of boxes was deemed as his prediction and then compared with the actual winner of the game to see if he was correct. This task was designed to elicit learners' initial models of repeated sampling, using coins in a cup as a repeatable random process, to determine the likelihood of a single proportion, namely correctly guessing the winners of eight out of eight games. Given a set of eight coins and a cup, the learners were asked to generate a way to simulate the actions in the real scenario that were being repeated. They were easily able to see how they could use one coin toss to simulate the octopus choosing a box to eat from. However, a rich discussion ensued concerning whether to simulate another coin toss for the actual winner and compare it to the prediction coin toss. In addition, when it was suggested to consider the coin toss as representing whether the prediction was correct or not, a further discussion ensued about whether they needed to do eight tosses with one coin, or if they could toss eight coins simultaneously. This discussion made it obvious to us that the mapping of the real world scenario to the simulation was not at all obvious. Eventually, they agreed to work in pairs with eight coins in a cup, spill the coins out, record the number of heads representing correct predictions, and repeat this four times. The resulting collection of number of correct predictions were marked on a dotplot displayed on the board. This was followed by a discussion of the shape of the distribution, what the distribution was actually of (which also caused lots of rich discussion and was not obvious to many participants), and which results constituted a likely or unlikely event (such as predicting all eight games correctly).

Model exploration activity 3 Following MEA2, the learners engaged with a model exploration activity (MXA3). The goal of the MXA3 was to explore the representations used in the *Paul the Octopus* task and to examine the underlying structure of the learners' models for a repeated sampling approach. This was done in three computational environments:

Statkey (<http://lock5stat.com/statkey/>), an applet available as part of the Rossman/Chance Applet Collection (<http://www.rossmanchance.com/applets/OneProp/OneProp.htm>), and in *TinkerPlots* (Konold & Miller, 2011). Each of these environments allowed the learners to explore various parameters of their simulation models (e.g., number of samples, using probability other than 50/50), and how these affected the distribution of sample proportions (or frequencies), which would not be possible with their by-hand simulations with coins.

Model application activity 4 Model application activities apply and adapt models that were initially elicited in an MEA. In this sequence, we wanted to apply the repeated sampling model for understanding the likelihood of a single proportion (MEA2 and MXA3 with *Paul the Octopus* task) to using repeated sampling to compare two proportions from an experimental design study (MAA4). To do this, we modified the *Dolphin Therapy* task (Catalysts for Change, 2012) to ask our participants to create a by-hand simulation using index cards that would answer the question posed in the task: can swimming with dolphins be therapeutic for patients suffering from depression? In the experiment, in the dolphin swimming group (treatment), 10/15 improved their depression, while 3/15 improved in the control group. They were given 30 index cards marked with the therapeutic results from the study of the 30 participants in two groups of 15, control and treatment; thus, 13 cards were marked with “YES” for benefiting with swimming with dolphins, and 17 cards marked with “NO” for not benefiting from the swimming. We anticipated that it would not be at all obvious how to create such a simulation. The simulation requires making an assumption that patients would improve (or not) their levels of depression, regardless of which group they are assigned. The assumption that each person has the same chance of being assigned to either group and thus that under random assignment we would not expect a difference in the proportion of those improved in either group. Essentially, the learners could shuffle the cards representing the 30 patient improvements, and deal the cards into two groups of 15 (a repeatable action of randomizing them and assigning into two groups). Hence by repeating this action and computing the difference in proportion of YES’s in each group, the learners should be able to examine a distribution of the difference in proportions and consider how likely it is that the benefits of therapy happened by chance alone. A variety of methods were created by the learners and, after discussion, the groups came to consensus on a method so that class level results could be aggregated in a dotplot, discussed and interpreted.

Model exploration activities 5 and 6 The *Dolphin Therapy* hands-on experience was followed by another model exploration activity (MXA5) where the representations and structure of the simulation and decision making using the sampling distribution was explored again in *Statkey* and *TinkerPlots*. Teachers were also able to further explore the structure of their developing models by engaging with two articles (Lane-Getaz, 2006; Lee, Starling, & Gonzalez, 2014) to read on their own in which authors used diagrams (Figure 3, and another similar to Figure 2) to illustrate the simulation approach used in the tasks discussed in the articles (MXA6).

Model exploration activity 7 After the two simulation tasks (*Paul the Octopus* and *Dolphin Therapy*), the instructor summarized the models that were designed and the resulting simulation process used. The intent of this summary was to help make connections across the two simulation tasks and generalize the processes used to develop an overall conception of

the modeling and simulation process. After this summary, and before the assigned readings had been discussed, the instructor used the following task to allow teachers an opportunity to express their developing conceptions of the simulation process in terms of how they would help their students to understand the process. Teachers formed nine groups of 2-3 to create diagrams on a large poster. The teachers were given the following task (MXA7):

Suppose you were going to use a randomization approach with your students to help them use a simulation (with physical objects or computer models) to investigate if an observed statistic is likely to occur, or not unlikely to occur. Draw a diagram you could use to help students understand the general process used for applying randomization techniques for solving these types of tasks.

We consider this last task a particular kind of model exploration activity in that it focuses on representations of the structure of models of repeated sampling for drawing inferences that would serve a pedagogical purpose. That is, the intended audience for this representation would be the future students of our participants and this representation hence served a perceived purpose of explaining the structure of models of repeated sampling to other learners.

Model application activity 8 Following this MXA, we noticed the wide variety of representations expressed in participants' diagram, and that many participants seem to be "hand-waving" about the "magic" that happens in the randomize and repeat phases in a simulation approach. Thus, we designed an additional MAA task (*The Jump Rope* task) that was structurally similar to the *Dolphin Therapy* task but required an adaptation of their previous model since it involved comparing means for two unequally-sized groups (MAA8). In addition, we deliberately changed the form of the manipulatives (using unmarked craft sticks) to push the learners further in understanding the role of randomization in their model of repeated sampling. The participants had varied approaches to recognizing what the repeatable action was in this scenario. Many used the craft sticks in some way (with slight variations from each other, to indicate jump rope scores and repeatedly re-assigning those scores into two different unequal-sized groups. Some of our participants really struggled with this task and did not create very viable ways of representing the scores or reassignment. Their attempts at applying their model for a repeated sampling approach to inference to create this simulation in such a different context really illuminated for us how fragile their models may be.

4. DEVELOPING MODELS OF A REPEATED SAMPLING APPROACH TO INFERENCE

In what ways could our participants express their general model for a repeated sampling through a simulation approach to inference? In this section, we carefully examine our participants' diagrams generated in the MXA7 task that they thought would be useful for helping students learning to use a repeated sampling approach. We chose five diagrams created by our participants that are representative of the collection of nine diagrams, and were specifically selected to illustrate our findings. These diagrams are in the Appendix and labeled Diagram A, B, C, D, and E. Though we began our open coding of the diagrams informed by literature and the ways others had represented repeated sampling and simulation approaches (Figures 2, 3, 4), we will explicate how our analysis of the diagrams created by

our participants led us to identify aspects that seemed to be more or less salient in the participants' models of using a repeated sampling approach to inference. What follows is a description of the major themes we identified in the representations and descriptions in the diagrams that indicated to us how well developed participants' models were. It is these themes that are shaping our vision for ways to be more explicit about the underlying role of probability in simulations and about the design of tasks that support the development of learners' models of repeated sampling in simulation approaches to inference.

4.1. REPRESENTATIONS OF THE PROBLEM AND/OR MODEL

Lane-Getaz (2006) presented the process of using simulation to develop the logic of inference starting with a question in mind, "what if", to investigate a problem (see Figure 3). In this step, students need to specify a "theory, assumption, or parameter" for further sampling. The "Model" level of Garfield et al. (2012), shown in Figure 4, explicitly unpacks the real world cereal box into statistical terms (six equally likely prizes). We consider this step as crucial in creating a probability model of the real world problem. The purpose of this step is to express the problem of interest in statistical terms that include a set of assumptions (e.g., likelihood of an event occurring). We found that the diagrams varied in the degree to which they made explicit the role of probability with its assumptions. For example, the top row of "steps" in Diagrams A and E rudimentarily addresses the importance of creating a probability model of the real world problem. Diagram C shows that the group of participants decided to use a coin flip as a model of a mother's reaction (yes or no) to whether children can have a party. Implicitly, this coin flip model makes the assumption that probability of a head/tail (likely assumed to be 0.5) is congruent to the probability of the mother's response of yes/no. In Diagram D, participants elaborated steps needed to model a real world problem by stating, "determine parameter of interest, determine assumptions for proportion(s)", and "simulation model (based on assumptions)".

Emerging from our analysis of these diagrams, we found that three groups of our participants attended to identifying the event of interest and the statistical result in the original problem that would be later used (collected) in the simulation process and for inferential decision-making. For example, the Diagrams A, B, and E explicitly marked or indicated a quantity of interest from the original problem and referred to it later in the simulation process as the statistics to collect and that one needs to locate that statistic in the distribution of sample statistics for decision making. Thus, three of these five diagrams emphasized the process of creating a model for the real world problem and attended to the importance of the original statistics of interest in the problem. We note that this level of detail attended to by our participants *does not* appear explicitly in the models in the research literature that we have shown in Figures 2, 3, and 4.

After the probability model for the real world problem is formulated, one needs to identify the process that is being repeated in the simulation. In this step, one decides which types of random-generating devices (physical or technological tools) can be used to accurately represent the repeated action in the problem, whether it is a random selection from a population, a randomly generated outcome, or a random assignment to a group. When designing and using the simulation, one must consider the assumptions behind the use of the tools and their alignment with the assumptions made in the question of interest. For example, when using a coin to model the chance of success for a mother saying yes to a party

(Diagram C), an implicit assumption being made is that the coin is fair and that the probabilities of a head and a tail are equal and equal to 0.5. Similarly, if one decides to use a simulation tool, one needs to specify the probability distribution and how it matches the assumptions in the problem. In Diagram D, the group of participants referred to these assumptions (“Simulation Model: Based on Assumptions”) but did not explicitly specify the need to examine the match between the probability distribution and the problem situation.

4.2. REPRESENTATIONS OF THE RANDOMIZE AND REPEAT PROCESS

The participants had very different ways of representing the process of generating a random sample of size n , observing an event of interest and computing a statistic, repeating this process k times, and collecting and displaying statistics from all k samples. To begin with, the notion of a repeated random sample (or randomization process) was not always explicit. While Diagrams B and E indicated random sampling, Diagram C may have implied a random process by noting the “coin flip”. The notion of a single sample of size n , repeated k times was also not always well represented. For the diagrams using a specific example (B and C), the n that was noted in the diagrams matched the problem context ($n=15$ in each group, and $n= 5$ responses from the mother). We also see in Diagram B that participants represented the repeatable process of combining all participants together and reassigning to groups for the *Dolphin Therapy* example. However, the participants who drew or described a general process did not explicitly state the need for k samples of size n , though phrases such as “many samples” and “high number”, and the pictorial diagrams in Diagram C and D implied a repeated process. As deliberately pointed out in the Lane-Getaz’s (2006) diagram (Figure 3), “the samples of size n ” are important to distinguish from k samples (often many) in repeated sampling. Both n and k are critical parameters in designing and running a simulation, especially since these parameters are often inputs required in software such as *TinkerPlots*.

All teachers explicitly or implicitly indicated that the simulation process includes recording, collecting, and graphing a statistic of interest from each sample. However, the level of detail in their representations of this process varied greatly. Because these diagrams were meant for participants to express a representation they could use to help students understand this process, we were left wondering whether they really understood the randomize, repeat, and collect process.

4.3. REPRESENTATIONS OF AN EMPIRICAL SAMPLING DISTRIBUTION

A key representation needed for using a simulation approach to inference is the resulting empirical sampling distribution of the statistics collected in each repeated sample about the event of interest (e.g., the proportion of correct predictions from *Paul the Octopus*, or the difference in proportions of participants with improved depression in *Dolphin Therapy*). All but one diagram shown in the Appendix included an image of an empirical sampling distribution, with the exception shown in Diagram D. Three of the five diagrams explicitly indicated that the original statistic from the problem context should be located in the distribution (A, B, and E) and used to assess likelihood that the original statistics would occur under the assumed model of random selection or assignment. These diagrams explicitly indicated where to look in the empirical sampling distribution and how to estimate a

probability (proportion or p -value) of the actual observed event by examining the tail(s) of the distribution. These participants seemed to have a more fully developed notion of how the p -value is an indicator of how unusual the event of interest is when considering the repeatable actions from the original problem under the null hypothesis (no change/no difference or equally likely).

5. DISCUSSION

The model development sequence we designed and implemented and the successes and struggles our course participants had in developing robust conceptual models for repeated sampling approaches to inference led us to deep reflection about how attending to the role of probability in a simulation approach to inference may be useful. Especially when considering our participants' diagrams as external representations of their conceptual models, we saw the need to propose aspects of a repeated sampling approach that should be made much more explicit for learners and teachers.

All the diagrams drawn by prior researchers (Saldanha & Thompson, 2002; Lane-Getaz, 2006; Garfield et al., 2012) include three tiers or levels for a simulation process (Figures 2, 3, 4), including problem and models, repeated samples, and a sampling distribution, even though these terms are either implicitly or explicitly used. More recent work by Podworny and Biehler (2014) explicitly draws learners' attention to each of these aspects of a simulation process as they use a particular structure in their tasks for students that have them record the work and interpretations they are making in each part of a simulation with *TinkerPlots*.

In designing our model development sequence, we certainly were attending to creating experiences for our learners to incorporate these aspects into their conceptual model. The prior work of the first author focused on probability simulations also highly influenced the nature of the first MEA (MEA1 with physical simulations with common random number generating devices), and how probability language was used throughout the task sequence. During the task sequence there were instances when, as with all real classroom research, conversations had to be cut short, and ways that instructors expressed conceptualizations or posed questions to our learners were not well phrased. Thus, we view our enacted model development sequence as a critical, but not final, step in understanding how instructors can best design experiences for their learners that lead to a robust and general conceptual model on how repeated sampling approaches to inference can be used in a variety of contexts.

To assist us in pushing forward, we used both our empirical data from our learners' diagrams and the approaches, descriptions, and diagrams used by others (Saldanha & Thompson, 2002; Lane-Getaz, 2006; Garfield et al., 2012). We outline several key conceptualizations that we believe would be powerful components of learners' conceptual models for a repeated sampling approach to inference that emphasizes the role of probability.

First and foremost, we suggest that more attention needs to be given to the modeling process, the explicit role of probability in inference, and use of probability language. We feel that there is a two-part modeling process that should be made explicit. The first is to create a local specific *model of* the real world context in statistical terms. The second is creating a simulation process that models the repeatable actions in the original problem and can be used to generate random samples. Most previous work has combined these two aspects into a single "model" or "population" level. We also suggest being more explicit concerning

building a distribution of sample statistics, viewing the distribution as a probability distribution, using the distribution to reason about the observed statistic, and making a claim about the chance of that observed statistic occurring. In Table 1, we explicate the key conceptualizations we believe are critical in a repeated sampling approach to inference, and what such conceptualizations may afford in learners' capabilities.

Table 1. Key conceptions and capabilities for learners' general models for using repeated sampling for inference.

Conceptualization	Capabilities this conception affords
Conceive of events in the real world problem as a result from a repeatable action	<ul style="list-style-type: none"> • Identify the underlying probability model of the event of interest (what is repeatable?) • Consider what results would be considered unusual, or what would be considered usual or "to be expected". • Express a usual expectation as a null hypothesis. • Specify the observed statistic and the statistic of interest that should be observed when each action is repeated.
Conceive of and create a method for simulating the repeated sampling process	<ul style="list-style-type: none"> • Identify the repeatable action that needs to be enacted. • Choose tool (physical or computer) and map the action in the real word to a simple repeatable process using the tool.
Conceive of repeated sampling as a way to generate simulated statistics	<ul style="list-style-type: none"> • Recognize the need to enact the process for selecting a random sample of same size n and record the statistic for event of interest. • Repeat the random sampling process k times (large number) and collect the statistic from each sample for event of interest.
Conceive of how collected statistics from repeated samples vary with respect to likelihood	<ul style="list-style-type: none"> • Build a distribution of the recorded statistics. • Notice what seems to be usual (typical, or more likely to occur), and what is unusual (or unlikely to occur). • Locate the original observed statistic in the distribution and consider whether it was in a range of "likely to happen" or "unlikely to happen".
Conceive of the inferential decision as involving deciding if the observed statistic and those more extreme are explainable by chance.	<ul style="list-style-type: none"> • Use proportional reasoning to evaluate the likelihood that the observed event, and those more extreme, happened under the random process used to generate repeated actions and simulated statistics.

We claim the conceptualizations noted in Table 1 are important for learners to develop in order to have a robust way of conceiving how a repeated sampling approach using simulations can be used to engage in inference. Because our target learners are those interested in teaching statistics, our focus is on assisting them to understand how probability is used in this approach and to develop a generalized model as a connected conceptual system that they can draw upon themselves as they assist their own learners in using a repeated sampling approach to inference. It is important to recall that all of our learners had previous exposure and experience with learning traditional inference techniques, and some had experiences in teaching such techniques. In our group of learners, we only had two who had

previous experience in using a repeated sampling approach in their own curriculum materials with their students. Thus, our sequence of tasks was designed with all these learners in mind. However, we believe the key conceptualizations we have outlined would be useful for all learners to develop if we want them to develop a general model of one can use a repeated sampling approach to inference. We also suggest that using a models and modeling perspective to task development that places particular emphasis on learners developing their own models (not prescribed by teachers or curriculum materials) would be useful to guide further classroom-based research.

ACKNOWLEDGEMENTS

The authors wish to thank Tina Starling and Bridgette Jacob for their contributions to the design and implementation of the model development sequence. Their role as co-instructors with Hollylynne Lee and Helen Doerr were an invaluable part of our teaching and research.

REFERENCES

- Ärlebäck, J. B., Doerr, H. M., & O'Neil, A. H. (2013). A modeling perspective on interpreting rates of change in context. *Mathematical Thinking and Learning, 15*, 314–336.
- Arnold, P., Confrey, J., Jones, S., Lee, H. S., & Pfannuch, M. (forthcoming). Learning trajectories in statistics education. In D. Ben-Zvi, J. Garfield, & K. Makar (Eds.), *International Handbook of Research in Statistics Education*. New York: Springer.
- Burrill, G. (2002). Simulation as a tool to develop statistical understanding. In B. Phillips (Ed.), *Developing a statistically literate society* (Proceedings of the Sixth International Conference on Teaching Statistics, Cape Town, South Africa). Voorburg, The Netherlands: International Association for Statistical Education and the International Statistical Institute. Retrieved from http://iase-web.org/documents/papers/icots6/7d1_burr.pdf
- Catalysts for Change. (2012). *Statistical thinking: A simulation approach to modeling uncertainty*. Minnesota, MN: Catalyst Press.
- Chaput, B., Girard, J.-C., & Henry, M. (2011). Frequentist approach: Modelling and simulation in statistics and probability teaching. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics – Challenges for teaching and teacher education* (pp. 85-95). The Netherlands: Springer.
- Cobb, P. (2007). The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education, 1*(1).
- Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. Washington, DC: Author.
- Cumming, J., Miller, C., & Pfannkuch, M. (2014). Using bootstrap dynamic visualizations in teaching. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education* (Proceedings of the Ninth International Conference on Teaching Statistics, Flagstaff, USA). Voorburg, The Netherlands: International Association for Statistical Education and the International Statistical Institute. Retrieved from http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_3D1_CUMMING.pdf
- Doerr, H. M., & English, L. D. (2003). A modeling perspective on students' mathematical reasoning about data. *Journal for Research in Mathematics Education, 34*(2), 110–136.

- Eichler, A., & Vogel, M. (2014). Three approaches for modelling situations with randomness. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 75-99). Dordrecht, The Netherlands: Springer.
- Garfield, J., & Ben-Zvi, D. (2008). Chapter 7: Learning to reason about statistical models and modeling. In J. Garfield & D. Ben-Zvi, *Developing students statistical reasoning* (pp. 143-163). Dordrecht, The Netherlands: Springer.
- Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM*, *44*, 883-898.
- Hjalmarson, M. A., Diefes-Dux, H. A., & Moore, T. J. (2008). Designing model development sequences for engineering. In J. S. Zawojewski, H. A. Diefes-Dux, & K. J. Bowman (Eds.), *Models and modeling in engineering education: Designing experiences for all students* (pp. 37-54). Rotterdam: Sense Publishers.
- Hjalmarson, M. A., Diefes-Dux, H. A., Bowman, K. J., & Zawojewski, J. S. (2006). Quantifying aluminum crystal size part 2: The model development sequence. *Journal of STEM Education: Innovations and Research*, *7*, 64-73.
- Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education*, *2*(1), 1-37.
- Konold, C., & Miller, C. D. (2011). *TinkerPlots: Dynamic data exploration* [computer software, Version 2.0]. Emeryville, CA: Key Curriculum Press.
- Lane-Getaz, S. J. (2006). What is statistical thinking and how is it developed? In G. Burrill (Ed.), *Thinking and reasoning with data and chance: Sixty-eighth yearbook* (pp. 272-289). Reston, VA: National Council of Teachers of Mathematics.
- Lee, H. S., & Lee, J. T. (2009). Reasoning about probabilistic phenomena: Lessons learned and applied in software design. *Technology Innovations in Statistics Education*, *3*(2).
- Lee, H. S., Angotti, R. L., & Tarr, J. E. (2010). Making comparisons between observed data and expected outcomes: Students' informal hypothesis testing with probability simulation tools. *Statistics Education Research Journal*, *9*(1), 68-96.
- Lee, H. S., Starling, T. T., & Gonzalez, M. D. (2014). Using data to motivate the use of empirical sampling distributions. *Mathematics Teacher*, *107*(6), 465-469.
- Lesh, R. A., & Doerr, H. M. (Eds.). (2003). *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Lesh, R. A., Cramer, K., Doerr, H. M., Post, T., & Zawojewski, J. S. (2003). Model development sequences. In R. A. Lesh & H. M. Doerr (Eds.), *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching* (pp. 35-58). Mahwah, NJ: Lawrence Erlbaum Assoc.
- Lock, R. H., Lock, R.H., Morgan, K. L., Lock, E. F., & Lock, D. F. (2013). *Statistics: Unlocking the power of data*. Hoboken, NJ: Wiley.
- Maxara, C., & Biehler, R. (2006). Students' probabilistic simulation and modeling competence after a computer intensive elementary course in statistics and probability. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education* (Proceedings of the Seventh International Conference on Teaching Statistics, Salvador, Brazil). Voorburg, The Netherlands: International Association for Statistical Education and the International Statistical Institute. Retrieved from http://iase-web.org/documents/papers/icots7/7C1_MAXA.pdf

- New Zealand Ministry of Education. (2006). *The New Zealand curriculum: Mathematics and statistics*. Retrieved from: Nzcurriculum.tki.org.nz
- Pfannkuch, M., & Ziedins, I. (2014). A modelling perspective on probability. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 101-116). Dordrecht, The Netherlands: Springer.
- Podworny, S., & Biehler, R. (2014). A learning trajectory on hypothesis testing with TinkerPlots - Design and exploratory evaluation. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education* (Proceedings of the Ninth International Conference on Teaching Statistics, Flagstaff, USA). Voorburg, The Netherlands: International Association for Statistical Education and the International Statistical Institute. Retrieved from http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_9A2_PODWORNYPDF.pdf
- Pratt, D. (2011). Re-connecting probability and reasoning about data in secondary school teaching, In *Proceedings of the 58th World Statistics Conference, Dublin* (pp. 890-899). Voorburg: The Netherlands: International Statistical Institute. Retrieved from <http://2011.isiproceedings.org/papers/450478.pdf>.
- Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal*, 7(2), 5-19.
- Roy, S., Rossman, A., Chance, B., Cobb, G., Vanderstoep, J., Tintle, N., & Swanson, T. (2014). Using simulation/randomization to introduce p -value in week 1. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education* (Proceedings of the Ninth International Conference on Teaching Statistics, Flagstaff, USA). Voorburg, The Netherlands: International Association for Statistical Education and the International Statistical Institute. Retrieved from http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_4A2_ROYPDF.pdf
- Saldanha, L., & Thompson, P. W. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics*, 51, 257-270.
- Stohl, H., & Tarr, J. E. (2002). Developing the notions on inference using probability simulation tools. *Journal of Mathematical Behavior*, 21, 319-337.
- Tintle, N., Topliff, K., Vanderstoep, J., Holmes, V.-L., & Swanson, T. (2012). Retention of statistical concepts in a preliminary randomization-based introductory statistics curriculum. *Statistics Education Research Journal*, 11(1), 21-40.
- Watson, J., & Chance, B. (2012). Building intuitions about statistical inference based on resampling. *Australian Senior Mathematics Journal*, 26(1), 6-16.
- Wild, C. (2006). The concept of distribution. *Statistics Education Research Journal*, 5(2), 10-26.

HOLLYLYNNE S. LEE
 Campus Box 7801
 502C Poe Hall
 Raleigh, NC 27695

APPENDIX: FIVE DIAGRAMS CREATED BY PARTICIPANTS

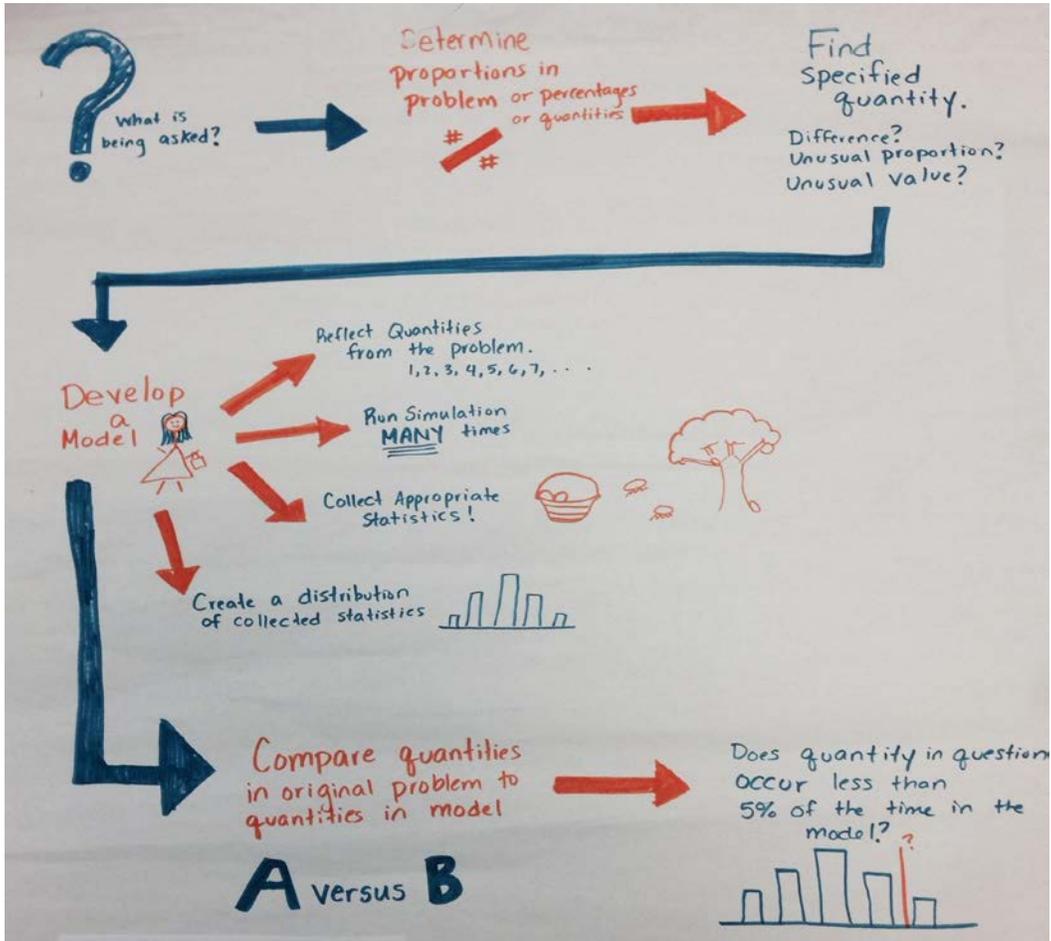


Diagram A

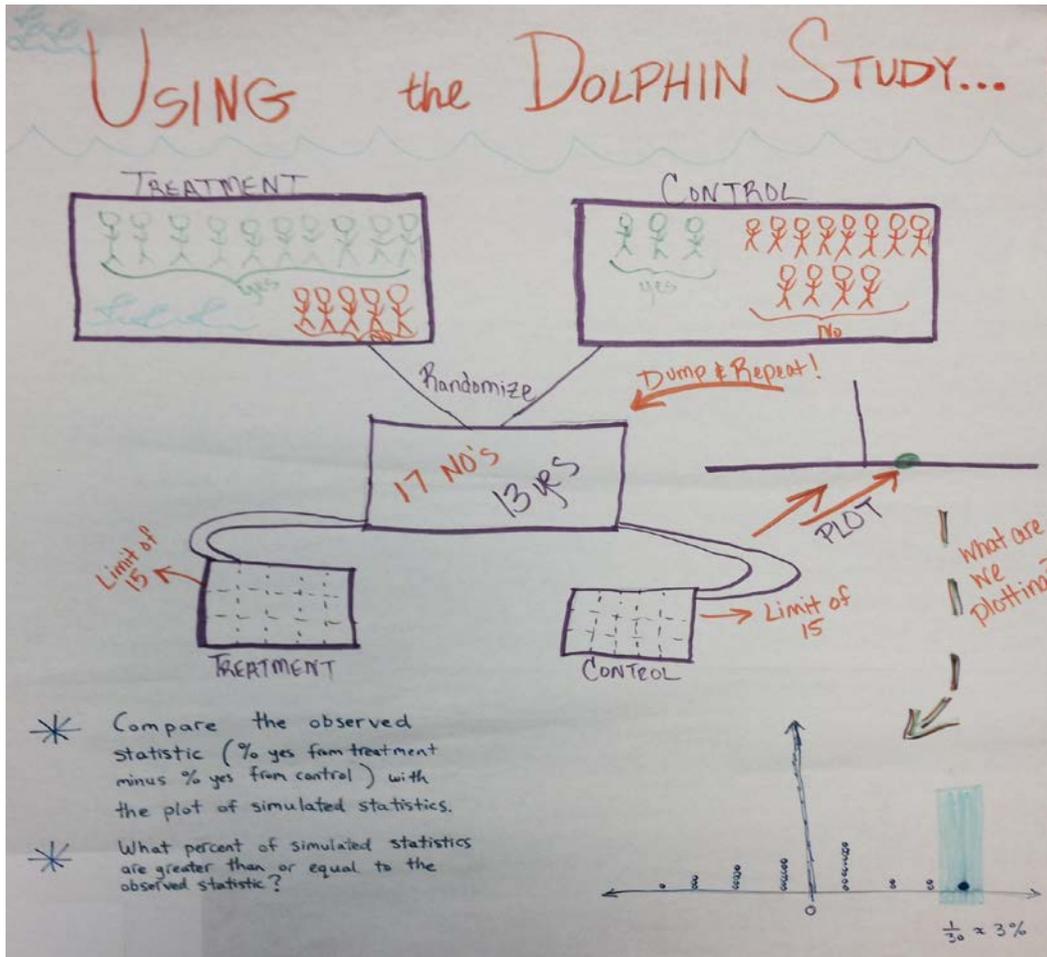
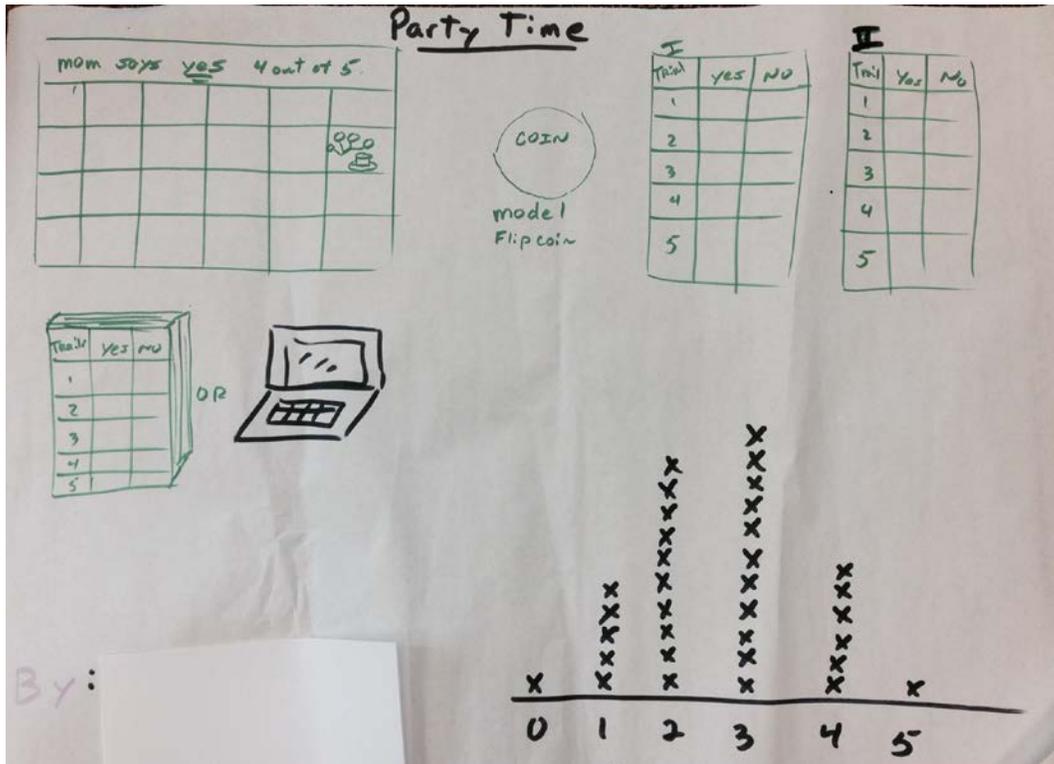


Diagram B



By:

Diagram C

- DETERMINE PARAMETER OF INTEREST
(BASED ON RESEARCH QUESTION)
- DETERMINE ASSUMPTIONS FOR PROPORTIONS
- SIMULATION MODEL
(BASED ON ASSUMPTIONS)
- DATA COLLECTION & GRAPHICALLY DISPLAY
(USING RELATIVELY HIGH NUMBER OF SAMPLES FROM SIMULATION)
- LOCATE INITIAL VALUE OF STATISTIC OF INTEREST
- FROM GRAPHIC REPRESENTATION, DETERMINE
PROBABILITY OF OBTAINING INITIAL VALUE
IN REGARDS TO RESEARCH QUESTION
(EQUAL TO, GREATER THAN, LESS THAN)
- DISCUSS / DETERMINE IF THAT PROBABILITY
IS STATISTICALLY SIGNIFICANT

Diagram D

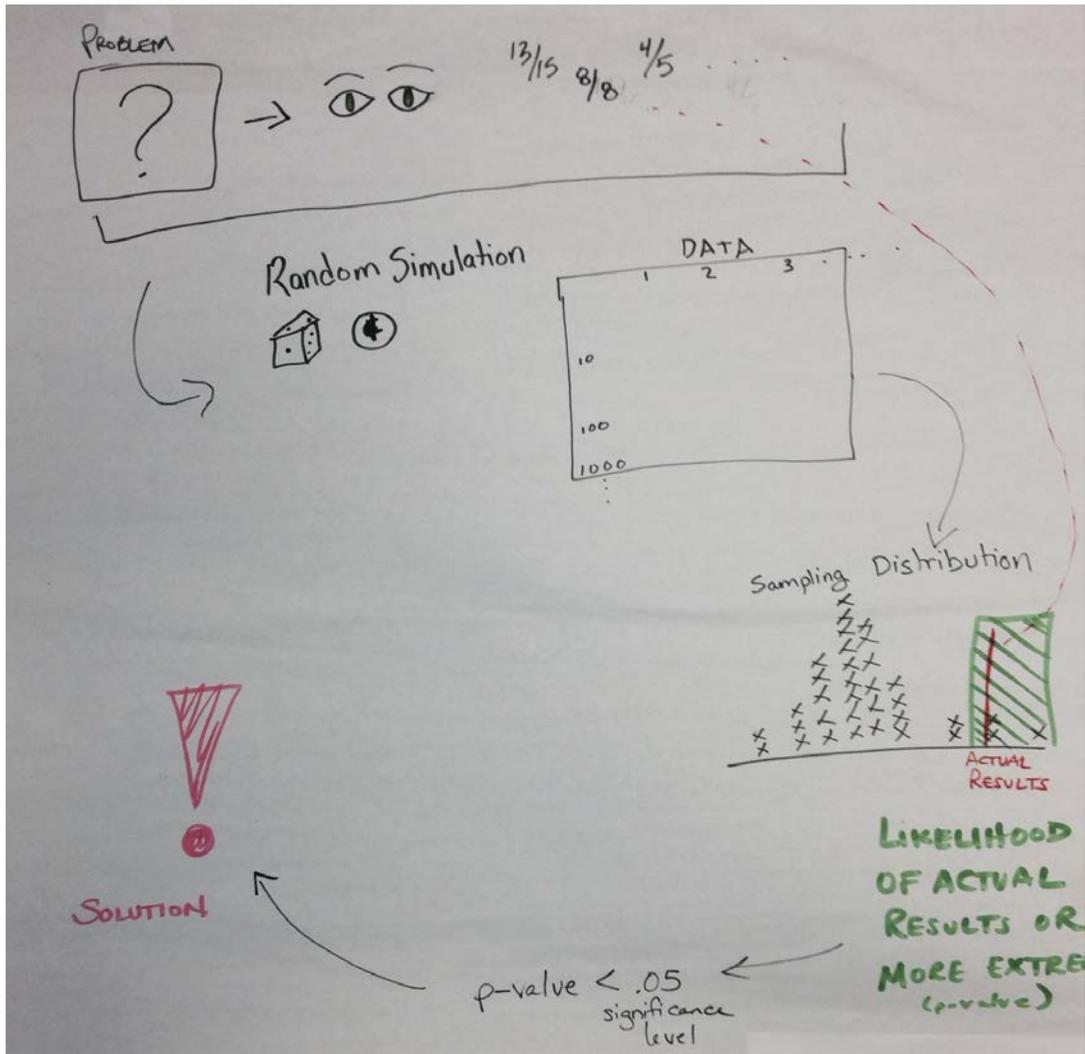


Diagram E