Student Approaches to Constructing Statistical Models using TinkerPlots™

1. INTRODUCTION

We have increasing access to new powerful technologies that allow us to capture all types of data. Modern technology not only allows us to perform quick computations and manipulations of data, it also allows us new ways of organizing, representing, summarizing and modeling data (and thus, new ways of perceiving the world we live in) (see Gould, 2010). Data from technology (smartphones, web, apps, etc.) are more available to the non-expert to use, manipulate, and interpret. With increased access to data come an increased responsibility to prepare every citizen to develop their ability to organize, represent, summarize and model data, as well as use the evidence gathered from data to make inferences. These are skills that more businesses and organizations require of their employees, and how these employees make inferences from data may in turn influence important business and policy decisions.

Nolan and Lang (2010) argue that computational literacy is now "fundamental to statistical practice… vital to all facets of a statistician's work… yet it occupies an astonishingly small proportion of the statistics curricula" (p.98). Cobb (2007) argues that technology offers statistics educators an opportunity to place more emphasis on the key concepts of inference (i.e., chance models and determining statistical unusualness) and less emphasis on procedures (i.e., formulaic hypothesis tests like *z* and *t* –tests). He states, "we may be living in the early twenty-first century, but our curriculum is still preparing students for applied work typical of the first half of the twentieth century" (p. 7). Statistics educators argue that many of the components of our introductory statistics courses (e.g., using a *z*-score to calculate a 95% confidence interval) are relics dating back to the 1900's historical roots of statistics and need to be reconceived in light of a data-driven, technologically based world (Cobb, 2007; Nolan & Lang, 2010).

Based on these recommendations, educators are developing new approaches to the teaching of statistics that align with recent trends in the way statisticians work with data. New curricula have emerged that focus on modeling, randomization techniques, and computer simulations (e.g., Garfield, delMas, & Zieffler, 2012; Lock et al., 2013; Morgan, 2011; Tintle, VanderStoep, Holmes, Quisenberry, & Swanson, 2011). These statistics educators now advocate teaching inference from an empirical perspective through modeling and simulation, which, they argue, helps students better understand how statistical decisions are made (Chance, delMas & Garfield, 2004; Cobb, 2007; Garfield & Ben-Zvi, 2008; Tintle, Topliff, VanderStoep, Holmes, & Swanson, 2012). Garfield and Ben-Zvi (2008) describe statisticians using models in two different ways: (1) "select or design and use appropriate *models* to simulate data to answer a research question", and (2) "[f]it a statistical model to existing data" (p. 145, italics in original). In their description a statistical model can be a random generator used to answer a statistical problem. A strong understanding of models is key to understanding the power of statistics, yet Garfield and Ben-Zvi (2008) note that most introductory statistics courses do not explicitly focus on modeling. They also indicate that there has been little attention to how students learn and use statistical models. Shaughnessy (2007) asserts that there is also little research related to how technology might support students' ability to create and understand statistical models. In order to change the "culture" of the statistics classroom, aligning it more closely to the way statisticians work with data, as recommended by Cobb (2007) and Nolan and Lang (2010), there needs to be a better understanding of *how* technology implemented with particular curricula material impacts students' development of statistical modeling.

The work presented here makes a contribution to statistics education research through classroom observations involving students in two statistics classrooms that experienced the Change Agents for

Teaching and Learning Statistics (CATALST) curriculum (see Garfield et al., 2012). The CATALST curriculum focuses on modeling and simulation integrated through the innovative use of TinkerPlots™ technology (Konold & Miller, 2015). In particular, we studied the ways students in these two classrooms created statistical models using TinkerPlots™ technology to answer a statistical inference problem. The research questions that guided our investigation are:

1. How do students who receive the CATALST curriculum and use TinkerPlots™ technology software conceive of setting up a statistical model? What aspects of the statistical problem do they relate to the model they construct?
2. What are the affordances of the CATALST curriculum coupled with TinkerPlots™ technology?
3. What are the challenges of the CATALST curriculum coupled with TinkerPlots™ technology?

2. BACKGROUND

2.1 EDUCATIONAL TECHNOLOGIES

In 1985, Pea described the impact technologies could have on education and learning. He provided two metaphors, "amplifier" and "reorganizer", from which to frame the role of technology in education. When technology is used to perform calculations more quickly, then the technology is considered an amplifier. For example, using technology (e.g., a computer or calculator) to calculate standard deviation is an amplifier use of technology because the task remains the same, computing a measure of variability, but the technology "amplifies" or "extends" our capabilities in completing the task. *Using*, as opposed to *constructing*, a ready made applet or computer program to compute many trials of a simulation thereby rapidly producing an empirical sampling distribution, rather than carrying out a simulation by hand (e.g., rolling a die 2000 times) is another example of an amplifier use of technology. Researchers (e.g., Ben-Zvi, 2000; Pea, 1985) argue that when technology is used as an amplifier, students are not necessarily forced to change the way they think about a problem rather, the technology allows them to perform tedious calculations quickly. In contrast, a reorganizer use of technology has the power to create cognitive shifts in students' thinking, helping them "transcend the limitations of the mind" (Pea, 1985, p. 91). Pea's philosophical view is that "our productive activities change the world, thereby changing the ways in which the world can change us" (p. 169). He argues that technology has the potential to fundamentally change the way humans think and learn when utilized appropriately in the classroom. Pea encouraged the education community to develop curricula that would require both amplifier and reorganizer uses of technology.

TinkerPlots™ was designed to be an educational technology. Konold and Lehrer (2008) state, "the objects that students build with this tool, and the inscriptions they create to organize and explore the output, are dynamic forms of mathematical expression which give rise to and facilitate their thinking about the domain, and that these ideas would not be readily available to them if they were restricted to purely written symbolic forms of mathematics" (p. 65). The description quoted here suggests that TinkerPlots™ may support students in a reorganizer way. The software requires students to determine how to organize data, make representations, create data summaries, and construct models, whereas other technologies (excel, statistical applets, etc.) contain ready-made representations, which subsequently, do not give students an opportunity to construct their own representations and thus their own meaning. Ben-Zvi (2000) researched middle school students statistical reasoning using TinkerPlots™ technology through the amplifier and reorganizer metaphors. He notes several important "reorganizing" effects of using technology in statistics education. In his work with middle school students on graphical representations of data, he argues that dynamic technologies allowed different cognitive levels of activity such as creating plots of data as well as manipulating and transforming data representations. Ben-Zvi's work suggests that the technology expanded the possible choices of representations for students to work with. This expansion provided them a different set of objects to study.

2.2 THE ROLE OF MODELING IN EDUCATION

Lesh and Doerr's (2003) Modeling and Modeling Perspective (MMP) is a philosophical stance that cognition is really about constructing models that help us to describe the world we live in. Hestenes (1992) states that,

> The great game of science is modeling the real world, and each scientific theory lays down a system of rules for playing the game. The object of the game is to construct valid models of real objects and processes. Such models comprise the content core of scientific knowledge. To understand science is to know how scientific models are constructed and validated. The main objective of science instruction should therefore be to teach the modeling game. (p. 732)

Hestenes' (1992, 2010) perspective on the use of modeling in math and science has major pedagogical implications. In particular, he argues that curriculum needs to be organized around models because models form the primary building blocks of knowledge. He suggests that classroom instruction should focus on students' construction of models as well as verifying and comparing their models with their peers.

Research in mathematics education has shown mathematical modeling to be an effective instructional method, preparing students to solve "real world problems" and think about unfamiliar situations flexibly and creatively (English, 2006; Lesh & Doerr, 2003). Mousoulides and his colleagues (2010) suggest that modeling activities can cultivate critical thinking in a way that extends beyond the recommendations of the National Council of Teachers of Mathematics (NCTM, 2000). In particular, they state, "modeling activities set within authentic contexts, allow for students' multiple interpretations and approaches, promoting intrinsic motivation and self regulation" (Mousoulides, Pittalis, Christou, & Sriraman, 2010, p. 120). When carefully designed, modeling problems can offer students rich learning opportunities because the problems presented are not already orchestrated for students.

2.3 TECHNOLOGY AND MODELING IN STATISTICS EDUCATION

In the field of statistics, Cobb (2007) suggests that modeling should play an important role in statistics instruction. He states, "[e]very statistical method is derived from a model and one of our goals in teaching statistical thinking should be to encourage students to think explicitly about the fit between model and reality" (p. 2). Despite the development of new curriculum and the recommendations of statistics educators to bring modeling to the forefront of our courses there is a paucity of research studying curricular materials focused on statistical modeling, and more specifically, focused on the ways students create statistical models. Maxara and Biehler (2006, 2007) and Garfield, delMas, and Zieffler (2012) are a few notable exceptions.

Maxara and Biehler (2006, 2007) studied how students constructed models and simulations of stochastic phenomena using Fathom. They hypothesized that there are three steps for students as they work on simulations – setting up a model, writing the plan of simulation, and putting the plan into action in Fathom (2007, p. 764). They noted that problems arise in transforming the model into a correct simulation in Fathom because students may pick the wrong simulation, select an incorrect number of cases (trials of the experiment), or use an incorrect formula for running the simulation. In addition, they noted that students might have problems in the naming of objects, which can lead to problems interpreting results later on. Maxara and Biehler (2006) also explored a six step modeling process that may be useful for students learning modeling activities: "1. Intuitive theories and expectations, 2. Building a stochastic model, 3a. Generating a simulation plan, 3b. Theoretical analysis of the problem, 4. Comparing simulation and theoretical analysis, 5. Comparing the intuitive theory with theoretical analysis and/or simulation; debugging of misconceptions, 6. Exploring further questions: Varying assumptions, contexts and questions" (p. 6).

Garfield, delMas and Zieffler's (2012) work did not focus on the specifics of student-constructed models, rather they studied general impacts of a curriculum focused on modeling and simulation approaches using technology. They developed the CATALST curriculum[1]. They compared students in the CATALST courses and students throughout the country in traditional introductory statistics courses with respect to their responses on the Goals and Outcomes Associated with Learning Statistics (GOALS) assessment, which is a comprehensive content exam with 20 forced-choice and three open-ended items. On many of the assessment items they observed little difference between the scores of CATALST and non-CATALST students. However, they observed that CATALST students performed very well on GOALS items that were particularly focused on interpreting visual displays of data. They plan to conduct additional implementations of CATALST and continue to study differences between CATALST and non-CATALST students. It is worth noting that student attitudes toward the use of TinkerPlots™ software as an important component of learning statistics were overwhelmingly positive. Though this research did not focus on the details of students' modeling activities, it does show that students using modeling curriculum with TinkerPlots™ can perform at least on par with students receiving a more traditional introductory statistics curriculum.

If we are to fully realize the recommendations of Cobb (2007) and develop radically different curriculum for the introductory statistics course that is more in line with the practice of doing statistics, then more needs to be done in the statistics education research community to continue the development of CATALST's non-traditional approach (or that of other new curricula focused on modeling and simulation) to teaching statistics. In particular, we need to (1) better understand the affordances and challenges of curriculum focused on statistical modeling using technology, (2) know how we can continue to improve upon new curriculum materials, and (3) study what students' modeling activities look like in a classroom focused on modeling and simulation using technology. A rich modeling approach requires students to solve statistical problems in ways that are meaningful to them, and such an approach encourages students to develop generalizable solutions and express problems using a variety of representations (for examples in mathematics see English, 2003; Mousoulides, 2007). We hypothesize that using TinkerPlots™ technology to *construct a model* in order to answer a particular statistical question offers a rich approach to developing students' understanding of introductory statistics. We hypothesize that the modeling approach described here provides an opportunity for students to use the technology as a reorganizer because it provides the students with the freedom to construct their own model. This freedom has the potential to create a fundamental change in how students come to model a statistical problem and the types of inferences they might make based on their choices of model representations and modeling approaches. The work presented here provides an initial glimpse into studying students' model construction activities using TinkerPlots™ software.

2.4 CATALST CURRICULUM

The CATALST curriculum was derived as part of a National Science Foundation funded project to develop and implement innovative instructional materials for an introductory college statistics course (Garfield et al., 2012). The curriculum consists of three units that together form a complete course for a one-semester introductory statistics class. The first unit focuses on chance models and simulation. The

[1] Further details about this curriculum will be discussed in the next section. In this section, though, we discuss CATALST in relation to research that has focused on modeling in statistics education.

second unit focuses on models for comparing two groups. The third unit focuses on estimating models using data and transitions students to formal tests of inference such as *t* or *z*-tests. The CATALST curriculum is based on a theory of learning that emphasizes conceptual understanding and critical thinking, and places the learner in an active role – constructing knowledge through engagement in classroom activities. According to Garfield and her colleagues (2012), four principals that guided the development of the CATALST curriculum were, "(1) model-eliciting activities (e.g., Lesh & Doerr, 2003; Lesh, Hoover, Hole, Kelly, & Post, 2000); (2) inventing to learn and the role of prior knowledge (e.g., Schwartz, Sears, & Chang, 2007); (3) instructional design principles (e.g., Cobb & McClain, 2004); and (4) the modeling work (Konold et al., 2011) and software developed by Konold and Miller (2011)" (p. 884). These four guiding principals are discussed in the paragraphs that follow in order to give the reader a better sense of the CATALST curriculum and its development.

The CATALST materials use model-eliciting activities (MEAs), based on the work of Lesh and his colleagues (2000), to introduce each new unit. According to Garfield and her colleagues (2012), the MEAs were "designed to encourage students to build mathematical models in order to solve complex problems, as well as provide a means for educators to better understand students' thinking" (p. 884). MEAs are based on six key principles (see Lesh et al., 2000 for further details): model construction, reality, self-assessment, construct documentation, shareability and reusability, and effective prototype. While engaging with MEAs students focus on constructing models with little instruction and have a chance to share their models with their peers, revise their models within their small groups, test their models using new data sets, and again revise their models based on the results of their tests. For example, the first unit in CATALST begins with a MEA investigating the random feature of the iPod shuffle. The activity begins by having students try to explain what it means to be random and how they might tell if their iPod shuffle was truly random. Students are given a play list for a hypothetical person's iPod and information about what songs were played when the iPod was set to random. Students are asked to devise a rule or set of rules (e.g., the iPod should not play the same artist for more than two songs in a row) for determining if the iPod Shuffle was playing songs at random. After they develop these rules they are given additional playlists generated by the hypothetical person's iPod so they can test their rules and revise as they think necessary. After small groups of students have a chance to create, test, and revise a set of rules, a whole class discussion ensues in which groups share their rules and the class has an opportunity to question and critique them.

Following each MEA, there are several activities in each CATALST unit that guide students through key ideas raised in the MEA (e.g., randomness, chance/null model, informal inference based on a single population, p-value) that relate to the overall unit goal. The activities that follow the MEAs continue to stress active learning in small groups, developing and testing conjectures, modeling statistical problems with TinkerPlots™, and using real data or simulating data through the software. Each of these foci in turn supports students' statistical thinking. In addition, the units that follow each MEA also relate to the second and third guiding principals used during the design of the CATALST curriculum: inventing to learn and building on prior knowledge, as well as the instructional design principles of Cobb and McClain (2004). In particular, the materials for each of the three units encourage "students to make and test conjectures, work in groups while using technology-tools, and engage in whole class and small group discussions" (Garfield et al., 2012, p. 885). These activities provide a more general approach to modeling in a way that supports students' abilities to find an appropriate statistical model and generate data to answer their question as opposed to applying a traditional statistical procedure such as a *t* or *z*-test. This modeling approach is also well aligned with GAISE recommendations for teaching statistics such as developing statistical literacy and thinking, using real data, stressing conceptual understanding, fostering active learning, and using technology (American Statistical Association, 2005).

The fourth guiding principal that Garfield and her colleagues (2012) used when developing the CATALST curriculum was that of technology. The curriculum was designed to be implemented using

TinkerPlots™ technology. According to Garfield and her colleagues, TinkerPlots™ was chosen "because of the unique visual capabilities it has, allowing students to see the devices they select (e.g., sampler, spinner) and to easily use these models to simulate and collect data, which allows students to examine and evaluate distributions of statistics in order to draw statistical inferences" (Garfield et al., 2012, p. 886). The unique modeling features of TinkerPlots™ correspond naturally to the designers' guiding principal of modeling and in addition supports active student learning as students work together to create TinkerPlots™ models, run simulations, and answer statistical inference questions based on their work. As such, TinkerPlots™ is a key feature of the course and used in order to achieve their pedagogical goals of having students develop models and conduct simulations. The majority of activities in the curriculum materials use TinkerPlots™.

3. METHODS

3.1 DESCRIPTION OF PARTICIPANTS AND TASK

The participants in this study were students enrolled in one of two small introductory statistics classrooms at a large urban university in the United States. The students enrolled in these two classrooms utilized the CATALST curriculum (Garfield et al., 2012) and TinkerPlots™ technology (Konold & Miller, 2015) during their 10-week term. The introductory course where we implemented the CATALST curriculum is normally intended for students who need a one term introductory statistics course for their majors. Many of the students in this course major in criminal justice, sociology, women's studies, psychology or other majors in the social sciences. Traditionally this course covers descriptive statistics, graphing, and some basic probability so that students are either prepared for the standard two term introductory statistics sequence or so that they can satisfy the required basic mathematics course needed for graduation in their major.

In the first classroom (Classroom 1) the first author was the classroom instructor and in the second classroom (Classroom 2) a graduate student, who had previously mentored with the first author, was the classroom instructor. In Classroom 1 a total of 17 students enrolled in the course and all students consented to be participants in the study. One student dropped the course after the midterm and is not counted in the data analysis here. In Classroom 2 a total of 21 students enrolled in the course and all consented to participate in the study. However, one student did not complete the final exam and is not included in the data analysis presented here. Thus, the total number of students in the sample is 36. All but one student (a student from Classroom 2) identified themselves as poor math students and expressed low confidence in their mathematical abilities.

Data collection for these two classrooms consisted of all student work on in-class activities, task-based semi-structured interviews (with a subset of the students from each class), and work associated with student assessment items. For the purpose of this study, data were drawn from students' written work on the final exam. Each classroom took the Models of Statistical Thinking (MOST) assessment (see Garfield et al., 2012) at the end of the course as part of their final exam. Students had access to TinkerPlots™ for the assessment. We selected the Facebook task (see Figure 1) from the MOST assessment to analyze student responses for this paper specifically because it is a relatively straightforward one-proportion test. We wanted to begin our study around how students conceive of a typical statistical inference problem that contains little in the way of complications. Given that most of the students' experiences during the 10-week course[2] were with activities from the first unit of the CATALST curriculum, which is devoted to

[2] The original CATALST materials were designed for a one-semester course so this 10-week course only implemented the first two units of the curriculum.

one population type statistical problems, we wanted to get a sense of what emerged in the way of students' statistical modeling ideas after a substantial amount of time had been spent on these types of problems.

> Facebook is a social networking Web site. One piece of data that members of Facebook often report is their relationship status: single, in a relationship, married, it's complicated, etc. With the help of Lee Byron of Facebook, David McCandless–a London-based author, writer, and designer–examined changes in peoples' relationship status, in particular, breakups. A plot of the results showed that there were repeated peaks on Mondays, a day that seems to be of higher risk for breakups.
>
> Consider a random sample of 50 breakups reported on Facebook within the last year. Of these 50, 20% occurred on Monday. **Explain how you could determine whether this result would be surprising if there really is no difference in the chance for relationship break-ups among the seven days**. *(Be sure to give enough detail that someone else could easily follow your explanation.)*

Figure 1. Facebook task from CATALST's MOST assessment

3.2 DEVELOPING A RUBRIC FOR ASSESSING STUDENT-GENERATED MODELS

One of the unique aspects of the CATALST curriculum is that students have the freedom to construct, test, assess, and revise (when needed) their own statistical models using TinkerPlots™. As such, when using TinkerPlots™ students need to attend to many aspects of the statistical problem in order to construct a model that will appropriately model the "real-world" phenomenon presented in the problem. Given that technology and modeling curriculum is relatively new to statistics education, a qualitative, descriptive approach was used to explore student work. In particular, during our analysis we attended to two components of the students' written work for the Facebook task: 1) the TinkerPlots™ sampler they constructed, and 2) their written justification for the model setup.

The rubric for assessing the student-generated models developed in part by reflecting on classroom instruction and assessment and in part by reviewing student responses to the Facebook task. In particular, we focused on classroom instruction and the assessment of student knowledge throughout the course in order to construct a generalization of the statistical modeling process with TinkerPlots™. In turn, this generalization of the statistical modeling process with TinkerPlots™ became a measure for analyzing student models in our research. Figures 2 and 3 discuss the modeling process and an example of a TinkerPlots™ model we hoped our students would create to solve the Facebook task, respectively. Table 1 discusses how we coded student-generated models. First we outline what the classroom sessions looked like, in terms of the modeling process, and then share how our rubric for analysis developed out of that.

A typical classroom approach for tasks like the Facebook task (i.e., one population inference problems) was to encourage students to create a sampler device. For the Facebook task, one appropriate model is a mixer with seven balls set to sample with replacement (which is indicated by the closed bucket in the TinkerPlots™ mixer, see Figure 3), "draw" set to 1 and "repeat" set to 50. Class discussion included asking students for justification for their choice of sampler and sampler setup. For example, in the Facebook task, if students selected a mixer like that in Figure 3 then we would ask why the mixer contained seven balls. A student might reply one ball for every day of the week a couple could break-up and that each day is just as likely to be a day for a break up as any other day (null hypothesis). We would

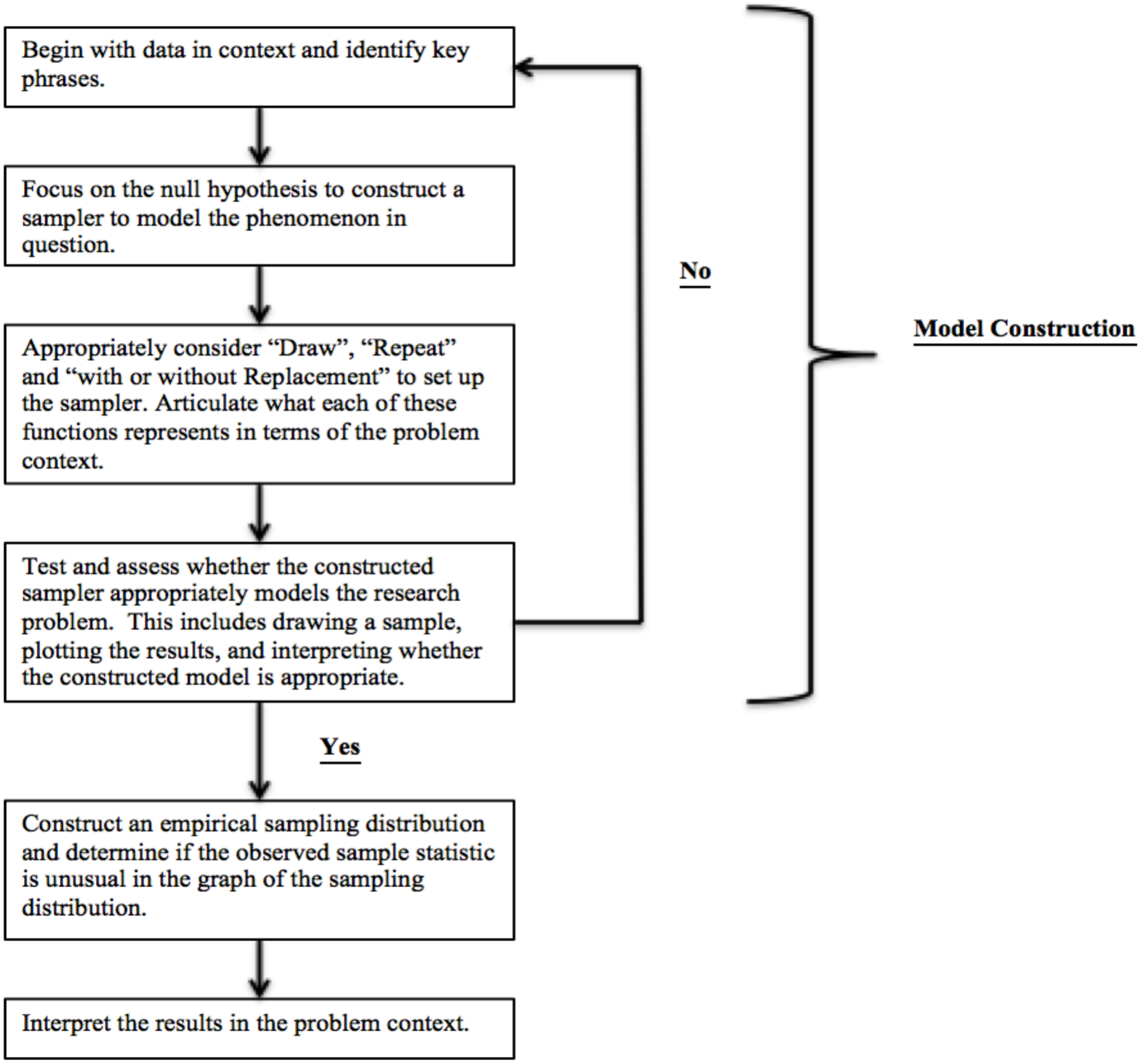

Figure 2. Generalization of modeling statistical problems with TinkerPlots™ in CATALST activities

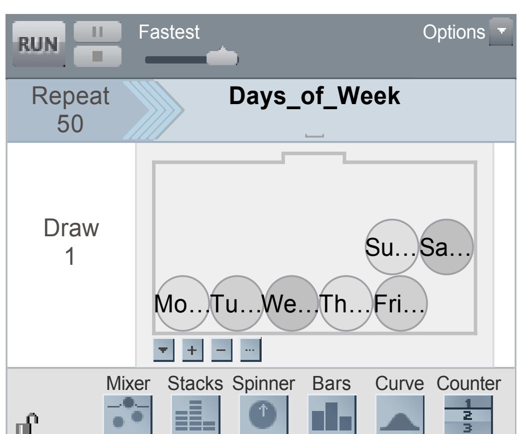

Figure 3. A TinkerPlots™ device used to model the Facebook task

also ask why a device would be set to "with replacement" or "without replacement" and what "draw" and "repeat" are set to and why. Again, in the Facebook task the mixer is set to model sampling without replacement since each couple is in theory independent of the next couple in terms of which day of the week they might break up. "Draw" is set to 1 and "repeat" is set to 50 because we can draw 1 couple at a time and the problem described sampling 50 couples' break-up status. The detail around the model choice and model set up (i.e., "draw", "repeat" and "with" or "without replacement") presented in this Facebook example is representative of the ways similar types of modeling problems were discussed in class. These classroom discussions set the standard expected of student work throughout the term and supported our development of a rubric for assessing student-generated models.

During the analysis of student responses to the Facebook task both authors initially reviewed student responses independently, taking note of themes in the data or interesting student responses. We identified the ways students worked through the modeling process as outlined in Figure 2 and whether or not their model in TinkerPlots™ was similar to or different from that of the model shown in Figure 3. We then discussed our observations together looking for areas of agreement and disagreement. Upon reaching an initial consensus, the first author partitioned student responses into several types of categories based on the type of model they constructed and the description they used to support their TinkerPlots™ model. The second author then reviewed the codes of the student responses and another discussion ensued until consensus was reached again. The result of our coding process is described in the following paragraphs.

During our analysis, we noticed that two main categories of TinkerPlots™ models emerged: single device models (SDMs) and linked device models (LDMs). A SDM is one random generating device such as a spinner or a mixer, where as a LDM contains two random generating devices. In the LDM an element is selected at random from the first device and then paired with a randomly selected element from the second device[3].

A detailed review of the students' justification for the construction of their models led to the generation of further codes. We present a detailed description of those codes here. Within each of the two main categories of TinkerPlots™ models, students gave (1) robust TinkerPlots™ models/model descriptions, (2) nearly robust TinkerPlots™ model/model description but neglected to mention the issue of replacement, (3) contradiction between TinkerPlots™ model and model description, or (4) highly problematic models. Student descriptions of their modeling approach were considered robust if the TinkerPlots™ model constructed had the potential to answer the Facebook task *and* the student provided justification for the model setup within the problem context (much like the example provided and a model like that shown in Figure 3). A highly problematic model was typically one in which there was significant inconsistency between the TinkerPlots™ model and model description or the student model did not adequately represent the null hypothesis. These problematic models will be detailed in this section and provide the reader a sense of the challenges in this particular modeling approach. Table 1 gives the reader a detailed description of the four codes as well as some general clarifying examples and discussion of codes[4].

---

[3] Linked and single device models will be discussed in detail throughout the Results section and will give the reader a very clear picture of what those types of models look like.

[4] The coding scheme and discussion of student work is focused on student generated models and justification for their models. In this paper we do not discuss whether or not students went on to successfully answer the statistical question posed to them. In general students who did not have robust models were not able to successfully answer the statistical question and some students who did have robust models still could not successfully answer the statistical question posed. However, this information is not relevant to our research questions, which focus exclusively on students' construction of models and justification for the models they construct.

Codes	Description of Code	Clarifying Example(s)/Discussion
Robust TinkerPlots™ Model/Model Descriptions	TinkerPlots™ model and written description could adequately answer the Facebook task. Provided strong justification for model set up and explained model within the context of problem.	Model device like that in Figure 4 or 13. Clear statement of null hypothesis. Complete description of the meaning of draw, repeat and replacement within context of problem.
Nearly Robust TinkerPlots™ Model/Model Description	TinkerPlots™ model and written description could adequately answer the Facebook task. Provided strong justification for model set up and explained model within the context of problem, but neglected to discuss replacement	Model device like that in Figure 4 or 13. Clear statement of null hypothesis. Complete description of the meaning of draw and repeat within context of problem. However, forgot to discuss replacement or gave no justification for using "with" or "without replacement".
Contradiction between TinkerPlots™ Model & Model Description	TinkerPlots™ model and written description contradicted each other primarily with respect to issue of replacement. Otherwise the model and description were robust or nearly robust. Note: One student gave contradiction with description of draw function.	1. Student uses a spinner model but suggests replacement is not needed or model should be sampling without replacement. 2. Creating a model set to "without replacement" but suggesting the model should be "with replacement". 3. Suggesting draw should be set to 7 because of representing each day of the week, but then using 50 for draw.
Highly Problematic TinkerPlots™ Models	Models that cannot answer the Facebook research question and contain many possible incorrect conceptions of null hypothesis or incorrect conceptions of running statistical simulation.	1. Setting up a model with observation data. 2. Inputting 50/50 model into the problem. 3. Not relating or incorrectly relating model to context.

Table 1. Coding scheme for student-generated TinkerPlots™ models/model descriptions

4. RESULTS

Painting a picture of students' TinkerPlots™ models for the Facebook task is much more complicated than simply providing counts of the number of correct models versus incorrect models. Since students are given the freedom to create their own model and are responsible for justifying the model that they construct, special attention was given to the types of models that they constructed and the reasoning they used to support their model construction. As such, this study attempts to delve deeper into the learning challenges and opportunities students experienced as a result of the curriculum and technology through a more detailed account of student descriptions of their modeling approach using TinkerPlots™.

The students overwhelmingly provided detailed descriptions of how they would model the Facebook task using TinkerPlots™ and in most cases they also provided pictures of their TinkerPlots™ sampler device. Only three students (two from Classroom 1 and one from Classroom 2) gave responses that did not describe a simulation through the lens of a TinkerPlots™ sampler device. In fact, these three students did not provide a description of a model at all, rather they provided subjective arguments for why the observed 20% break-ups on Mondays was or was not surprising. The other 33 students provided a description (and in most cases a picture) of a TinkerPlots™ sampler device for their model. Table 2 shows the breakdown of students' model choices.

Classroom	No Model and No Description of a Model	Single Device Model (SDM)	Linked Device Model (LDM)
	Type of TinkerPlots™ Model		
Classroom 1	2	5	9
Classroom 2	1	13	6
Both Classrooms	3	18	15

Table 2. Summary counts of student TinkerPlots™ models for the Facebook task

The remainder of the results section is divided into two subsections. The first subsection shares the SDMs that the students constructed and the second subsection shares the LDMs that the students constructed. In each subsection we provide illustrative examples of student responses to highlight the successes and challenges students faced when constructing and justifying their TinkerPlots™ models.

4.1 SINGLE DEVICE MODELS

Overall eighteen students constructed SDMs in response to the Facebook task. Six students (3 from Classroom 1 and 3 from Classroom 2) provided a robust SDM. Eight students (2 from Classroom 1 and 6 from Classroom 2) provided a nearly robust SDM. These eight students provided a TinkerPlots™ model using a device set to “with replacement” (and, thus, a model that could adequately address the statistical problem). That is to say these students constructed devices that modeled sampling with replacement, yet they neglected to discuss why they constructed a device that modeled sampling with replacement or even acknowledge that the device they constructed modeled sampling with replacement. Three students’ (from Classroom 2) TinkerPlots™ models contradicted their model description. Finally, one student from Classroom 2 experienced a significant challenge in adequately modeling the null hypothesis.

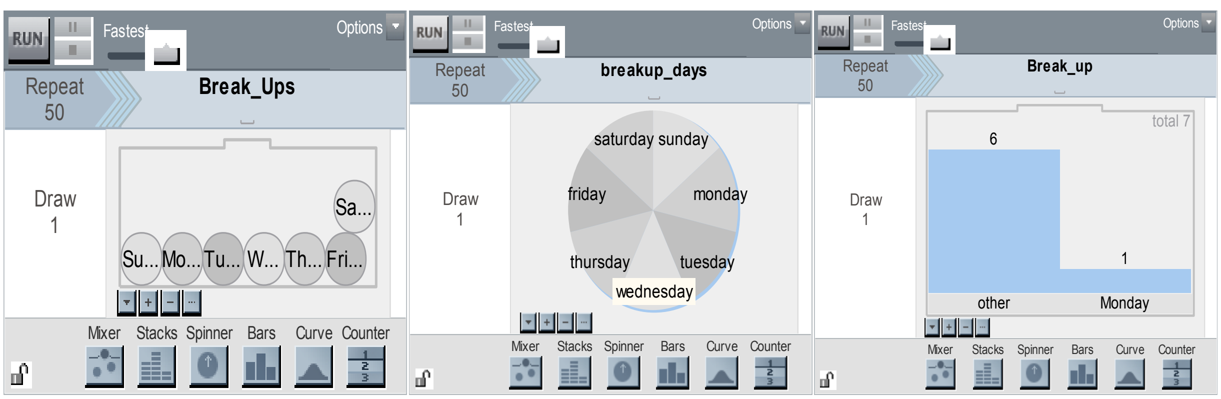

Figure 4. Three student-constructed models for the Facebook task (from left to right: mixer, spinner, and stacks)

Figure 4 shows the three types of SDM devices (mixer, spinner, and stacks, respectively) students used in setting up their models. In each of these student-generated models, “draw" is set to 1 and "repeat" is set to 50 because we can draw 1 couple at a time and the problem described sampling 50 couples’ break-up status. In both the mixer and spinner (left and middle of Figure 4, respectively) each day of the week is explicitly represented, highlighting that each day is just as likely to be a day for a break-up as any other day. In the device using stacks (right of Figure 4), the student has partitioned the days of the week into “Monday” and “other”. Each of the individual bars in the “other” stack implicitly represents a day of the week other than Monday (i.e., Tuesday – Sunday) and since each individual bar has the same “weight” the student is still modeling that each day is just as likely to be a day for a break up as any other day. All three of these devices are set to model sampling with replacement. More specifically, the spinner, by default, always models sampling with replacement, while the closed bins on both the mixer and stacks (left and right of Figure 4, respectively) signify that the device is set to "with replacement". All three of these student-constructed devices have the potential to adequately model the null hypothesis of the Facebook task.

As mentioned above six students provided a robust model and model description when responding to the Facebook task. Here we provide two examples of student responses and discuss why these were categorized as robust. Student A (Classroom 1) set up a spinner SDM, like the middle device in Figure 4. She provided a robust written explanation of her model (Figure 5).

> **Student A:** I set up a spinner with each piece representing one day of the week. Drawing 1 represents one break-up. Repeat representing all 50 break-ups. I set it to with replacement because people could break up any day of the week and more than one couple could break up on each day of the week.

Figure 5. Student A's written description of the model set up

In her description she explicitly relates "draw", "repeat" and "with replacement" from her TinkerPlots™ model back to the problem context. She draws 1 to look at one couple at a time and repeats 50 for the 50 couples in the problem. She explains why her model is set to "with replacement" – since more than one couple can break up on a given day. Student B (Classroom 2) also explicitly discussed "draw", "repeat" and "replacement" and related them to the problem context. His work is shown in Figure 6.

> **Student B:** To do this you would have to make a sampler with seven outcomes of equal probability with replacement (the 7 days of the week with replacement because you can never run out of days of the week unless your event encompasses a specific length of time). I used a spinner because it is easiest for me to use when the trial must be with replacement, but a mixer or stacks could be used as well. The draw should be 1 because you are drawing one break up at a time and you should draw 50 times because the study deals with 50 couples (50 breakups, which also constitutes one trial).

Figure 6. Student B's written description of the model set up

There were 8 other students that also selected one of the devices shown in Figure 4 and also related the "draw" and "repeat" functions back to the problem context. The only reason these were not identified as fully robust models is because these students either did not mention the idea of "replacement" or stated the model would be "with replacement" but did not state why. For example, Student C (Classroom 2) used a mixer like the first device shown in Figure 4 and he described "draw", "repeat" and the issue of replacement, but did not explain why he set the model to "with replacement" (see Figure 7).

> **Student C:** I set up my sampler as a mixer with 7 different balls that represented the 7 days of the week that people could break up. I set the draw value to 1 and the repeat value to 50 so that there would be one day drawn at a time, but there would be 50 trials. This was set to "with replacement".

Figure 7. Student C's written description of the model set up

It is possible that the students who either neglected to mention issues related to sampling with or without replacement or simply stated their models were set to "with replacement" understood why they were setting up their models this way, but simply did not take the time to explain why. The question did not specifically ask students to do this, however both classes spent a significant amount of time discussing the issue of replacement in problems like this. In addition, it is impossible for us to analyze how these students imagine the role of sampling with or without replacement in the context of this problem without any written data about their thinking.

There were three students whose models and descriptions of their models showed evidence of inconsistency, though again each of these students did provide a TinkerPlots™ model that would adequately model the statistical problem. For example, Student D (Classroom 2) provided an appropriate TinkerPlots™ model. He used a spinner divided into 7 parts for each day of the week (middle device in Figure 4). His description adequately described setting up the repeat value and the draw value. However, his description of why he used the spinner, that "replacement isn't needed" appears contradictory (see Figure 8).

Student D: For my understanding, I used Tinkerplots and made a sampler with a repeat value of 50 and a draw value of one. The repeat value is for the number of couples that broke up during the week; draw value is one because we only want to know about one couple at a time. I used a spinner because replacement isn't needed and that way there is an equal chance that the break up could happen on any day of the week.

Figure 8. Student D's written description of the model set up

The spinner is, by default, set to sample with replacement, but Student D's comments could indicate that he thought the model should sample without replacement since he said "replacement isn't needed". It is also possible that he knows he is using a device that is by default sampling with replacement so perhaps he is saying he does not have to worry about issues of replacement. Yet, his explanation is unclear, leaving us to question how he actually conceived of sampling with or without replacement.

Another student also appeared confused about using "with" or "without replacement" in her description of the model set up. Student E (Classroom 2) set up a TinkerPlots™ spinner sampler (middle device in Figure 4) and her description is shown in Figure 9. She suggested setting up a sampler "without replacement", which contradicts her TinkerPlots™ spinner model, which is automatically sampling with replacement. In addition, her justification: "you can never run out of days to break up" seems to suggest sampling with replacement rather than her statement of "without replacement". Again, it is unclear if there is simply a typo in what she wrote or a deeper confusion about when to model sampling with or without replacement.

Student E: First, create a mixer with seven options, one for every day of the week. Set it to without replacement, as you can never run out of days to break up with someone, no matter how many other people broke up that day as well. Set the Draw value to 1, because you can only break up with someone on one day (or so we'll say, to keep it simple). Set the Repeat value to 50, because that is the size of the random sample we are simulating.

Figure 9. Student E's written description of the model set up

Student F (Classroom 2) appeared to experience some struggle with respect to interpreting "draw" in TinkerPlots™. The written description of his TinkerPlots™ model is shown in Figure 10.

Student F: I came up with making a simple Sampler of 7 elements. Each would represent an individual day and the sampler would be put at 50 repeats. I could of made it where the sampler had 50 elements and made the draw value to 7 and repeat to 50. The repeat would represent 50 trials and the draw value would represent 7 days, but the problem with this is that you wouldn't know which day is which, so that is why I chose to make the sampler 7 elements with a repeat value of 50 and draw value of 1. This would represent 50 different breakups and the draw value would represent 1 day being chosen for the breakup.

Figure 10. Student F's written description of the model set up

Student F gave two different explanations of how the sampler could be configured. His actual TinkerPlots™ sampler is a correct model (mixer, see first device in Figure 4), but he also described another model where his sampler would contain 50 elements, his draw value would be set to 7 to represent each day of the week and repeat would be set to 50 for the 50 trials. However, he did not explain what the 50 elements in the sampler would then represent. Based on his explanation and work it is plausible that he envisioned the sampler with 50 elements as couples and by setting a draw value of 7 for each day of the week, he imagined pulling a couple out of the sampler and assigning that couple a day to break up through the draw value. This idea has merit, but cannot produce a viable simulation because the "draw" function on TinkerPlots™ is not capable of randomly assigning a day of the week to each couple for breaking up. The "draw" function randomly selects the designated number of elements from the sampler (e.g., 1 element or 7 elements), but does not label them with a value such as day of the week. If this was the way he interpreted this alternative set up it is also unclear what the 50 "repeats" now represent since he still suggested having 50 trials. In addition, this student did not discuss the issue of sampling with or without replacement for his model. He used a sampler set to "with replacement", which is indicated by the closed bin in the TinkerPlots™ mixer model shown in Figure 4. However, because TinkerPlots™ default is set to model sampling with replacement we have no idea of whether or not he considered the issue of replacement as he constructed his model.

One student, who used a SDM approach, appeared to have significant challenges in modeling the Facebook task. Student G (Classroom 2) constructed a SDM using the spinner device. However, she used the percentages from the observed data in the problem. Her spinner contained two sections, one labeled "Monday" at 20% and the one section labeled "Other Days" at 80% (see Figure 11).

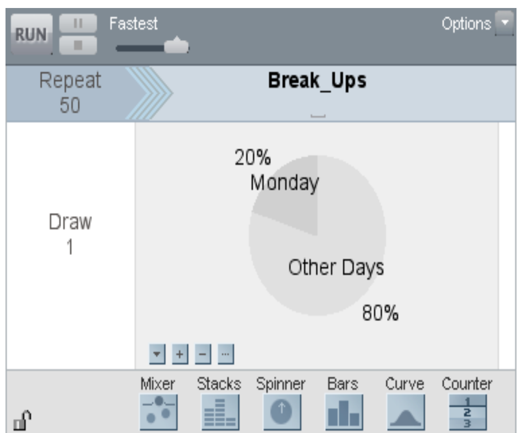

Figure 11. Student G's TinkerPlots™ model

> **Student G:** When setting up the sampler, I used the spinner as the representation, due to the example giving a percentage when determining how often breakups occur on a particular day. Since the occurrence of break ups weren't given for the other days, I labeled it as such and indicated the 20% chance of break ups occurring on Monday. The draw is set to one since you have only 2 possible outcomes; being split up on Monday or not. The repeat value is the sample space, which is given in the problem, 50 break ups. Also as far as replacement goes the sampler is set to spinner, it's automatically set to "with replacement".

Figure 12. Student G's written description of model set up

Her description of the modeling process (see Figure 12) suggests she does not have a strong sense of modeling the null hypothesis. This model randomly assigns each of the 50 couples in the problem to breakup on either Monday (with a probability of 20%) or another day of the week (with a probability of 80%). That is, rather than assuming each day of the week is equally likely to result in a day that a couple breaks up, modeling this equally likely assumption, and then observing where 20% breakups on Monday falls in a sampling distribution based on the assumption that each day is equally likely to result in a break up, the student instead constructed a statistical model that assigns a probability of breakups on Mondays equal to the observed proportion of breakups on Mondays (20%). Technically, if one were to use this model to construct an empirical sampling distribution of the percentage of breakups on Monday, they would obtain a symmetric sampling distribution centered at 20% (reflecting the assumption that the person is making in the model construction phase). As such, the model constructed does not accurately model the Facebook task. In addition, this student did not discuss why she wants a device that models sampling with replacement. Of the eighteen SDM models, this last model poses a real challenge for teaching using the modeling approach because the underlying assumptions of this model are distinctly different than those of the other models that have at their basis the assumption that each day of the week is just as likely to result in a break up as any other day of the week. This last model assumes the 20% chance of break-ups on Mondays observed in the sample rather than each day of the week is just as likely as any other day to break up. The difference between these two assumptions gets at the heart of the underlying conceptual idea surrounding the null hypothesis.

4.2 LINKED DEVICE MODELS

Overall 15 students constructed LDMs in response to the Facebook task. Four of these 15 responses (2 from Classroom 1 and 2 from Classroom 2) gave appropriate models for answering the Facebook task and provided robust descriptions of their model. Two of these 15 students (1 from each classroom) provided TinkerPlots™ LDMs that appropriately answered the statistical problem and gave robust descriptions except they did not mention the issue of replacement or relate it back to the context of the problem. Four of these 15 student responses (3 from the Classroom 1 and 1 from Classroom 2) contained a contradiction between the model and the description with respect to the issue of replacement. Two of these 15 students (1 from each classroom) provided a robust TinkerPlots™ model but no explanation of their model[5]. Three of these 15 students had significant challenges in their TinkerPlots™ models. Two (one from each classroom) inputted a 50% chance of breaking up or staying together on any given day and one student (from the Classroom 1) used the observed data to construct his null model in addition to having challenges with "draw" and "repeat". The different student responses are elaborated on in this section.

Four of the 15 students used what we termed a robust, but *redundant* model (i.e., LDMs) and gave detailed descriptions of their model and how it related to the problem context. Students H and I (Classroom 1 and Classroom 2, respectively) are examples of robust, but redundant, LDMs (see Figures 13, 14, 15, 16 for student work).

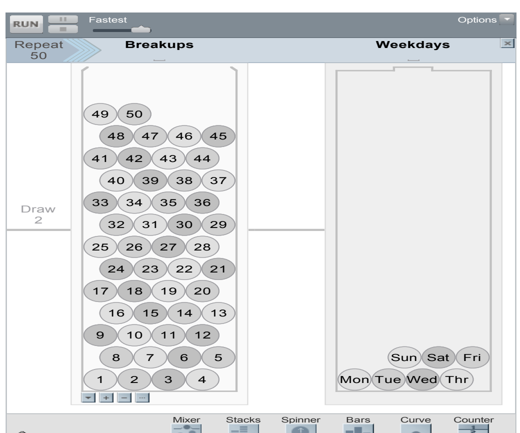

Figure 13. Student H's TinkerPlots™ model for the Facebook task

---

[5] The work from these two students is not shared here. They provided an adequate model, such as the one shown in Figure 13, but did not provide additional work for us to analyze how they might have thought about their models.

Student H: I would match two devices on the sampler: the first one with a mixer with numbers from 1-50 (for breakups), and a second one with the seven days of the week. The first one would be without replacement whereas the second one would be with replacement as we have to match 50 breakups with seven days of the week and see how many times Monday is chosen. Repeat would be at 50 so the 50 numbers are assigned to one day of the week.

Figure 14. Student H's written explanation for model of the Facebook task

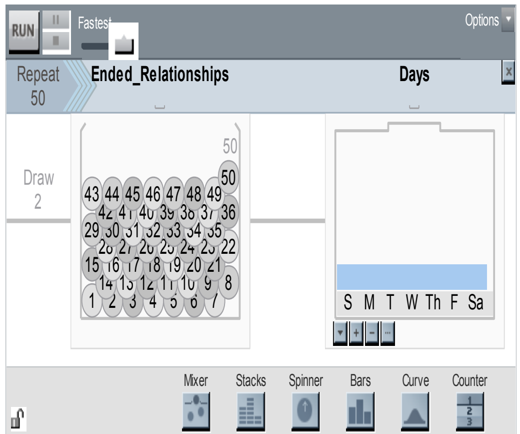

Figure 15. Student I's TinkerPlots$^{\text{TM}}$ model for the Facebook task

Student I: The sampler above uses two devices but note that our results only need to show the choices from one. The first device labeled "Ended Relationships" is set up such that the computer will not repeat a drawing of the same number. In other words, once a relationship has been corresponded to a day of the week, it is 'taken out' of the sampler so that it will not be chosen again, ensuring that we have only percentages of 50 relationships just as Lee Byron's experiment had been concerned with. Of course, the "Days" device *does* replace the 'days,' so that a day can be chosen more than once.

Figure 16. Student I's written explanation for model of the Facebook task

Both students H and I described assigning a couple from the first mixer to a day of the week to breakup. They also noted that once a couple is selected from the first bin that the couple cannot be selected again, but that the day of the week can be selected more than once. They both set the repeat to 50, Student H described assigning 50 numbers to each day of the week, though neither spent much time describing what "repeat" represented in the context of the problem. Had they attended more to the repeat function they might have realized that setting repeat to 50 modeled 50 couples each being assigned a day to break-up. Though redundant, this type of LDM provides a more concrete conceptualization of the TinkerPlots$^{\text{TM}}$ model to the actual problem by providing a visualization of both the 50 couples and the days of the week and how they get matched together. It is also interesting that Student I seemed to acknowledge that his model contained repetition when he noted that he really only needed one device.

Two of the fifteen students using LDMs (were coded as nearly robust TinkerPlots$^{\text{TM}}$ models because their descriptions that did not relate replacement back to the context. For example, student J (Classroom 2) provided a robust LDM (see Figure 17) and his description discussed aspects of how he set up his sampler, but we do not know specifically how he related the idea of replacement to the problem context (see Figure 18).

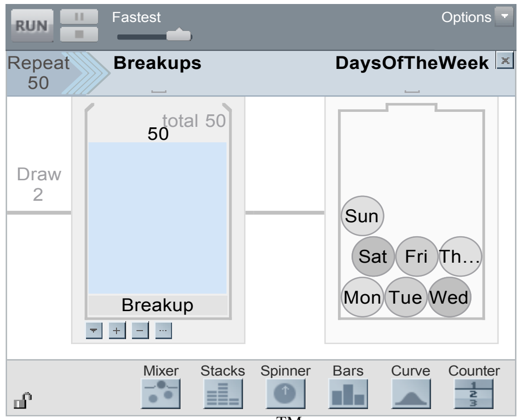

Figure 17. Student J's TinkerPlots™ model for the Facebook task

> **Student J:** I set up a sampler with a stacked and mixer sampler. My stacked sampler was set to without replacement and labeled breakups with 50 stacks, and the mixer was set up with days of the week with replacement. I then set the sampler to repeat 50 times.

Figure 18. Student J's written explanation for model of the Facebook task

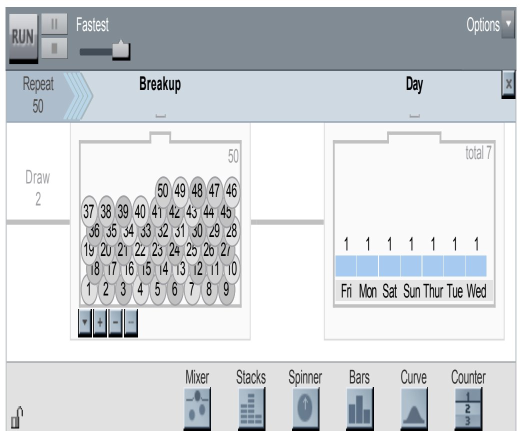

Figure 19. Student K's TinkerPlots™ model for the Facebook task

> **Student K:** To determine if the observation from the random sample is surprising, I will use TinkerPlots… to determine the probability that the result of 20% of breakups happen on Mondays if there really is no difference in the chance for relationship break-ups among the seven days. The null model assumes that there is an equal probability for a breakup to occur on each of the seven days of the week. First I create a sampler with two linked devices. The device on the left is a mixer with 50 balls in it – one to represent each of the 50 breakups in the sample. The device on the right of the sampler is a stack device with seven equally probable stacks, one for each day of the week.

Figure 20. Student K's written description of model set up for the Facebook task

Student K (first Classroom 1) did not mention the issue of replacement in his explanation (see Figure 20), but the picture of his model has both devices set to sample with replacement (closed bins in TinkerPlots™, see Figure 19). Technically, if the 50 elements in the first mixer device represent break-ups then he should have this device set to "without replacement" otherwise he runs the risk of pulling the same couple out multiple times. However, because the mixer with 50 elements is redundant in the model it makes no difference in the results whether or not he has "with" or "without" replacement selected. Had the problem been set up in such a way that sampling with or without replacement would have made a difference, one wonders if, after inspecting the results of his simulation, he would have noticed this.

Four of the 15 students provided models that contained contradictory language related to the issue of replacement. For example, student L (Classroom 1) did not provide a picture of her sampler, but her description suggests she would set up her model "without replacement" (see Figure 21). However, she does not specify whether both devices or only one of the devices would be set to "without replacement" leaving us unable to fully characterize her response as robust. If her days of the week stacks device was actually set to "without replacement" she would not be able to run the simulation she described. Since she only provided a description of her model and the process we do not know if she tried to run her simulation nor if she considered sampling with or without replacement in both devices.

Student L: I would place 50 balls in a mixer and link it to 7 stacks, which are representing each day of the week Monday through Sunday. I would run the sampler without replacement…

Figure 21. Student L's written description of TinkerPlots™ model

Finally, three students constructed LDMs that revealed significant challenges in learning how to model statistical problems. Two of these students used a 50/50 spinner in their TinkerPlots™ model. The third student used the observed data as the ratio in his TinkerPlots™ spinner model, but also had other significant challenges.

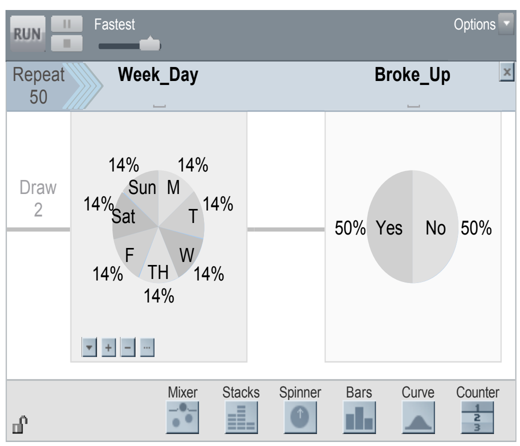

Figure 22. Student M's TinkerPlots™ model of the Facebook task

Student M: Two spinners will be used to replicate this experiment. One spinner will represent the days of the week. The second spinner will represent if the couple did break up (yes) or did not break up (no). The percentage for each spinner will be set so there is equal chance. In other words 14% x 7 = 100% and a 50/50% chance for yes and no. The spinners are used so there is replacement. The draw value will be set to 2 because we want a day picked and also if the couple broke up or not. The repeat value will be set to 50 representing couples.

Figure 23. Student M's written explanation for TinkerPlots™ model

Two of these three students created a spinner for breaking up or staying together. They divided the spinner into two equal parts and assigned a 50% chance of breaking up and a 50% chance of staying together. Then each of these students created a linked device for selecting one of the seven days of the week. We dubbed this strategy the *50/50* model. Student M's work (Classroom 2) is shown in Figure's 22 and 23.

This model first selects a day of the week at random and then randomly assigns it a “Yes” or “No” for breaking up or staying together. This essentially assigns each couple not just a day of the week for breaking up but also assigns them a probability of breaking up. The problem with this model is that these students inserted an extra probability into the model. They seem to assume that part of the null model is the equally likely chance of breaking up or staying together. It is as if they read into the problem that the couples may or may not break up. However, the focus should be on 50 couples that did break up, assigning them a day of the week (at random) for their break up.

Finally, Student N (Classroom 1) created a TinkerPlots™ model using the observed data. That is, he set up his spinner with 20% of the area assigned to Mondays and 80% assigned to all other days (see Figure 24). This strategy bears some similarity to two of the other student modeling strategies, but at the same time it is distinctly different. It is similar to student G’s SDM model where she used the 20% from the observed sample as the null hypothesis (the probability a couple breaks up on Monday). However, it also melds into part of Student M’s model in that it is also adding in an extra probability. Student M assumed couples had a 50/50 chance for breaking up where as this student’s model inserts a 20% chance for couples to break up. That is, this student used the observed data from the study (20% of couples broke up on a Monday), to assign the probability of whether a couple breaks up or not. Then he assigned that couple a day of the week for which they either break up or stay together.

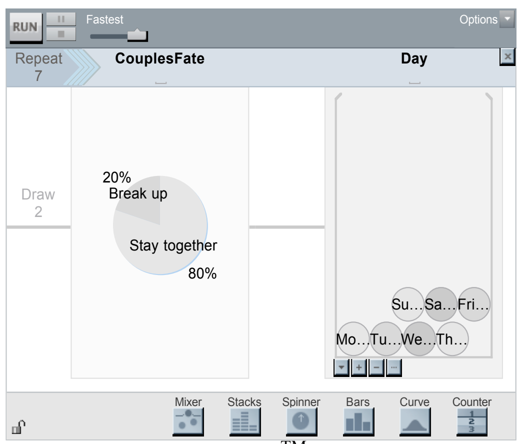

Figure 24. Student N’s TinkerPlots™ model for the Facebook task

> **Student N:** I would use this specific set-up because it assumes the null hypothesis that each day of the week has an equal chance that a couple will break up on that day. As the percentage we were given was 20% for Monday, that is the percentage we're using in this experiment. This gives each day a 20% chance to entertain a break u without effecting the other days of the week. ...After assembling the model I would run a single trial of 7 days (as we have 7 days in a week without replacement) and plot the results.

Figure 25. Student N’s written work for the Facebook task

In addition to adding in this extra probability of a couple breaking up or staying together based on the observed data, there are a number of other challenges with this student’s model. He does not model the 50 couples described in the problem. Rather he sets his “repeat” to 7 so that he only draws 7 couples and his mixer with the 7 balls (one for each day of the week) is set to “without replacement”. Thus, his model seems to select a couple and assign that couple a 20 chance of breaking up and then selects a day of the week for which that couple has either broken up or stayed together. Once he selects 7 couples he is out of days of the week in his mixer. Thus, this student has created a rather unique model with respect to the underlying assumptions of his null hypothesis, his “repeat” function, and his choice of “without replacement” for his days of the week mixer device.

5. DISCUSSION

For these CATALST students, the modeling approach using TinkerPlots™ software appeared to support their ability to think about the model within the context of the problem. This result is not surprising, as the focus of the CATALST course is on modeling. They used TinkerPlots™ to visualize the problem or perhaps based on their work with the TinkerPlots™ sampler they visualized statistical problems through that medium. Lesh and Doerr (2003) suggest that student-generated models "are not simply processes, that students use on the way to producing 'the answer'…They are the most important component of the response that is needed. So the process is the product" (p. 3). Lesh and Doerr's comment seems appropriate in the context of the given research. We observed that the student-generated models and how the students made sense of the models they created were key processes in their way to producing a solution to whatever statistical problem they were attempting to solve. The TinkerPlots™ sampler devices were products of their work and a key process in their development of solving statistical inference problems through modeling and simulation. Their choice of device, how they interpreted "draw", "repeat", and "replacement" were important aspects of their process and became part of the product of their work. The student responses shared in this paper highlight some of the affordances and challenges of the curriculum and technology on student reasoning. Given that all but three students used TinkerPlots™ as a means for solving the Facebook problem, it seems clear that the sampler device in the software framed the way students made sense of statistical problems. In addition, most students experienced success in modeling the Facebook task. While we observed that many of the students did not mention sampling with or without replacement or had small issues with their TinkerPlots™ models, most still managed to make considerable progress in their modeling activities. This suggests that TinkerPlots™ and the classroom activities afforded some positive opportunities for impacting students' statistical development.

We also found evidence of challenges in student models. There was some evidence in the work presented here that some students struggled with "draw", "repeat", and "replacement" as they setup their TinkerPlots™ models. These struggles were observed throughout the course and most often around issues of replacement as they set up and described their TinkerPlots™ models. Many of the student models presented here contained some amount of ambiguity, particularly around the idea of sampling with or without replacement. Future work needs to be done to better understand how classroom discussion around the issue of replacement as well as tasks that might help highlight differences between problems where replacement is warranted and ones where it is not.

Student responses to the Facebook task also showed some challenges around students making connections between the null hypothesis and their TinkerPlots™ model. In particular, some students were challenged by (1) teasing apart the null model from the observed data, or (2) teasing apart a null model of no difference from an equally likely (50%) chance of breaking up or not breaking up. The 50/50 model, that is inputting a 50/50 probability somewhere into their model, appears to be a pervasive issue in other areas of statistics education as well. For example, Noll and Shaughnessy (2012) observed middle school students gravitating toward 50/50 when making an estimate for the proportion of red candies in a jar when given a set of empirical sampling distributions of the counts of red candies in handfuls of ten, despite the fact that the sampling distributions were clearly not centered at 50%. Konold (1989) observed what he called the outcome approach when students attempted to "predict the outcome of a single next trial" (p. 61). In this case students tended to assign probabilities to uncertain events as yes or no with a chance of 50% to each outcome. This tendency for students to gravitate toward 50/50 models may be a result of instruction, beginning in primary school and continuing through college, primarily focused on binary outcomes. When students are not given an explicit probability for an event, they may be inclined to assign a probability of 50% to random events. In other words, students may view a situation where they do not know the likelihood as a safe place to assume that any particular outcome must have about a 50% chance of happening.

The use of observed data in setting up the null model in TinkerPlots^{TM} was seen in student work throughout the CATALST implementation and needs to be addressed in future work. It makes sense that without a greater understanding of probability and the conditional probability statement inherent in the null hypothesis (how likely is a result of 20% given an assumption that any day is just as likely to result in a break-up) that one could confuse these ideas and put the 20% in the model set up. Instruction must face the issue of how to help students tease apart the underlying null hypothesis assumption and the observed data from the sample. It may be useful to give an extension around tasks like the Facebook task where an activity like this begins with students constructing their own models and then the class moves on to comparing, contrasting, and evaluating their peers' models. If some students create models using the observed data in place of the null assumption than classroom discussion may be able to address this issue. It may also be useful to change the underlying null assumption away from always assuming equally likely probabilities (in this case any day is as likely to be a day for a break-up). Perhaps having extension problems where students are told that data from the previous 10 years suggests break-ups on Mondays are around 5%, but that a sample taken this past year resulted in 10%. Given the new result is there reason to suspect that the percent of break-ups on Mondays has changed. Varying the problem like this may help students see the underlying null hypothesis more clearly. That curriculum materials typically contain an underlying, implicit assumption of equally likely for the null hypothesis may make it difficult for student to see the null as explicit in the problem.

6. CONCLUSIONS

The CATALST curriculum coupled with the TinkerPlots^{TM} technology seemed to impact the way these students thought about statistics. The software appeared to create cognitive shifts for students where they visualized statistical problems through the TinkerPlots^{TM} sampler. That is to say, the TinkerPlots^{TM} sampler appeared to serve as a re-organizer tool for these students, framing the way they thought about statistical problems. All but three of the students either described a TinkerPlots^{TM} sampler device or actually created the sampler device as a way to frame the Facebook problem. Though we did not discuss it in this paper, these students also used the models they constructed and the data from their simulations to make their inferences. We hypothesize that the dynamic aspects of the TinkerPlots^{TM} sampler tool concurrently with the activities that focused on modeling and simulation (re)framed students' ways of viewing statistical inference. There is some evidence from Garfield, delMas and Zieffler's (2012) study that TinkerPlots^{TM} also significantly impacted the ways their students viewed statistics because they found that 85% of the students in their study felt that learning to create TinkerPlots^{TM} models was important for learning to think statistically.

There are limitations to this study. In particular, the sample sizes for both classes were small. Yet, this study does shed light on the ways students using CATALST and TinkerPlots^{TM} approach statistical modeling. The work points out affordances of the CATALST curriculum coupled with TinkerPlots^{TM}, as well as challenges. Future studies could scale up with multiple classrooms or classrooms at multiple institutions to see if similar challenges (e.g., ideas of replacement, 50/50 models, or using observed data models) exist as well as ways instruction might better support those challenges when they arise. In addition, this study reports on students' modeling activities using a new curriculum and TinkerPlots^{TM} software, but it raises the question as to which aspect (the curriculum or the software) of the course created more impact in students' modeling behaviors. Would another software have created the same kinds of modeling behaviors for students? In particular, we wonder what the impact would be if different software, one with fewer visualization affordances (such as the sampler in TinkerPlots^{TM} offers), was used? Additionally, what would be the impact on students' modeling behaviors using TinkerPlots^{TM} with another curriculum? We suspect that because the curriculum was developed with a modeling perspective in mind that it worked in positive ways with TinkerPlots^{TM} technology (a technology that affords opportunities for constructing dynamic and visual statistical models), yet more research is needed to tease apart the impact of the curriculum and the technology. There also needs to be further investigation into the ways in which using TinkerPlots™ for statistical modeling can reframe students' ways of thinking about statistics. For example, how might students model other kinds of statistical problems (e.g., two populations)? Research investigating students' statistical modeling behaviors is in its infancy. There is much work to be done in this area if we are to move our introductory courses in the direction of modeling and simulation approaches.

REFERENCES

American Statistical Association. (2005), "Guidelines for Assessment and Instruction in Statistics Education (GAISE) Project: College Report," Washington, DC. Available at http://www.amstat.org/education/gaise/GaiseCollege_full.pdf.

Ben-Zvi, D. (2000), "Toward Understanding the Role of Technological Tools in Statistical Learning," *Mathematical Thinking and Learning*, *2*(1 & 2), 127–155.

Chance, B., delMas, R., and Garfield, J. (2004), "Reasoning about Sampling Distributions," In D. Ben-Zvi and J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking* (pp. 295–323). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Cobb, G. W. (2007), "The Introductory Statistics Course: A Ptolemaic Curriculum," *Technology Innovations in Statistics Education*, Vol. 1: No. 1, Article 1. Available at http://escholarship.org/uc/item/6hb3k0nz.

Cobb, P., and McClain, K. (2004), "Principles of Instructional Design for Supporting the Development of Students' Statistical Reasoning," In D. Ben-Zvi and J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking* (pp. 375–395). Dordrecht, The Netherlands: Kluwer Academic Publishers.

English, L. D. (2003), "Reconciling Theory, Research, and Practice: A Models and Modeling Perspective," *Educational Studies in Mathematics*, *54*(2 & 3), 225–248.

English, L. D. (2006), "Mathematical Modeling in the Primary School: Children's Construction of a Consumer Guide," *Educational Studies in Mathematics*, *63*(3), 303–323.

Garfield, J., and Ben-Zvi, D. (2008), *Developing Students' Statistical Reasoning: Connecting Research and Teaching Practice*, US: Springer.

Garfield, J., delMas, R., and Zieffler, A. (2012), "Developing Statistical Modelers and Thinkers in an Introductory, Tertiary-level Statistics Course," *ZDM – The International Journal on Mathematics Education*, *44*(7), 883–898.

Gould, R. (2010), "Statistics and the Modern Student," *International Statistical Review*, *78*(2), 297–315.

Hestenes, D. (1992), "Modeling Games in the Newtonian World," *American Journal of Physics*, *60*(8), 732–748.

Hestenes, D. (2010), "Modeling Theory for Math and Science Education," In R. Lesh, P. L. Galbraith, C.

R. Haines, and A. Hurford (Eds.), *Modeling Students' Mathematical Modeling Competencies* (pp. 13–41). New York: Springer.

Konold, C. (1989). "Informal Conceptions of Probability," *Cognition and Instruction, 6*(1), 59-98.

Konold, C., and 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*, *Second Edition* (pp. 49-72). New York: Routledge.

Konold, C., Madden, S., Pollatsek, A., Pfannkuch, M., Wild, C., Ziedins, I., Finzer, W., Horton, N. J., and Kazak, S. (2011), "Conceptual Challenges in Coordinating Theoretical and Data-centered Estimates of Probability," *Mathematical Thinking and Learning*, *13*(1 & 2), 68–86.

Konold, C., and Miller, C. (2011), TinkerPlots™ Version 2 [computer software]. Emeryville, CA: Key Curriculum Press.

Konold, C., and Miller, C. (2015), TinkerPlots™ Version 2.2 [computer software], Learn Troop. Available at http://www.tinkerplots.com/.

Lesh, R., and Doerr, H. M. (2003), "Foundations of a Models and Modeling Perspective on Mathematics Teaching, Learning, and Problem Solving," In R. Lesh and H. M. Doerr (Eds.), *Beyond Constructivism: Models and Modeling Perspectives on Mathematics Problem Solving, Learning, and Teaching* (pp. 3–34). Mahwah, NJ: Lawrence Erlbaum Associates.

Lesh, R., Hoover, M., Hole, B., Kelly, A., and Post, T. (2000), "Principles for Developing Thought-Revealing Activities for Students and Teachers," In A. Kelly and R. Lesh (Eds.), *Handbook of Research Design in Mathematics and Science Education* (pp. 591–646). Mahwah, NJ: Lawrence Erlbaum Associates.

Lock, R.H., Lock, P.F., Lock Morgan, K., Lock, E.F., and Lock, D.F. (2013), *Statistics: Unlocking the Power of Data* (1st ed.), Hoboken, NJ: John Wiley & Sons.

Maxara, C., and Biehler, R. (2006), "Students' Probabilistic Simulation and Modeling Competence after a Computer Intensive Elementary Course in Statistics and Probability," In A. Rossman and B. Chance (Eds.), *Working Cooperatively in Statistics Education. Proceedings of the Seventh International Conference on Teaching Statistics* (ICOTS7, July 2006). Salvador da Bahia, Brazil. Available at www.stat.auckland.ac.nz/~iase/publications/17/7C1_MAXA.pdf.

Maxara, C., and Biehler, R. (2007), "Constructing Stochastic Simulations with a Computer Tool – Students' Competencies and Difficulties," In *Proceedings of the Fifth Conference of the European Society for Research in Mathematics Education* Lamaca, Cyprus.

Morgan, K. L. (2011), "Using Simulation Methods to Introduce Inference," Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) webinar [online]. Available at http://www.causeweb.org/webinar/teaching/.

Mousoulides, N., Pittalis, M., Christou, C., and Sriraman, B. (2010), "Tracing Students' Modeling Processes in School," In R. Lesh, P. L. Galbraith, C. R. Haines, and A. Hurford (Eds.), *Modeling Students' Mathematical Modeling Competencies* (pp. 119–129). New York: Springer.

Mousoulides, N. (2007), "*The Modeling Perspective in the Teaching and Learning of Mathematical*

Problem Solving," unpublished Doctoral dissertation, University of Cyprus, Nicosia.

National Council of Teachers of Mathematics (NCTM) (2000), *Principles and Standards for School Mathematics*, Reston, VA: NCTM

Nolan, D., and Lang, D. T. (2010), "Computing in the Statistics Curricula," *The American Statistician*, *64*(2), 97–107.

Noll, J., and Shaughnessy, J. M. (2012), "Aspects of Students' Reasoning about Variation in Empirical Sampling Distributions," *Journal for Research in Mathematics Education, 43(5)*, 509-556.

Pea, R. D. (1985), "Beyond Amplification: Using the Computer to Reorganize Mental Functioning," *Educational Psychologist*, *20*(4), 167–182.

Shaughnessy, J. M. (2007), "Research on Statistics Learning and Reasoning," In F. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 957–1009). Greenwich, CT: Information Age Publishing, Inc. and NCTM.

Schwartz, D. L., Sears, D., and Chang, J. (2007), "Reconsidering Prior Knowledge," In M. C. Lovett and P. Shah (Eds.), *Thinking with Data* (pp. 319–344). Mahwah, NJ: Erlbaum.

Tintle, N., Topliff, K., VanderStoep, J., Holmes, V., and Swanson, T. (2012), "Retention of Statistical Concepts in a Preliminary Randomization-Based Introductory Statistics Curriculum," *Statistics Education Research Journal*, 11(1), 21–40. Available at http://iase-web.org/documents/SERJ/SERJ11%281%29_Tintle.pdf.

Tintle, N., VanderStoep, J., Holmes, V., Quisenberry, B., and Swanson, T. (2011), "Development and Assessment of a Preliminary Randomization-Based Introductory Statistics Curriculum," *Journal of Statistics Education*, Vol. 19: No. 1, Article 4. Available at http://www.amstat.org/publications/jse/v19n1/tintle.pdf.