THE SYMBIOTIC, MUTUALISTIC RELATIONSHIP BETWEEN MODELING AND SIMULATION IN DEVELOPING STUDENTS' STATISTICAL REASONING ABOUT INFERENCE AND UNCERTAINTY

Andrew Zieffler, Robert delMas, Joan Garfield, and Ethan Brown
University of Minnesota, Minneapolis, MN, USA
zief0002@umn.edu

We report preliminary results from an ongoing study of the development of tertiary students' reasoning related to statistical inference and uncertainty during a one-semester modeling and simulation-based statistics course. Comparisons of students' performance on assessments of statistical reasoning will be presented for both students enrolled in this course and those enrolled in courses that used conventional parametric methods of inference. Summaries of qualitative data from nine students who participated in problem-solving interviews will also be presented to illustrate the development of students' reasoning related to statistical inference and uncertainty. Analyses of the data indicate that students taking the modeling and simulation-based course demonstrate better understanding of the principles of study design and statistical inference and begin developing these understandings within the first few weeks of the course.

THE CATALST CURRICULUM

The CATALST curriculum, funded by the National Science Foundation, was developed as part of a three-year teaching experiment. It incorporates a different approach to introducing and building ideas regarding statistical inference. Rather than build up to inference via a sequence of common "foundational" topics, CATALST immerses students in the nuts-and-bolts of statistical inference from the first day of the curriculum using activities designed to emphasize the core logic of inference through a focus on modeling and the use of simulation.

The CATALST curriculum is composed of three primary units: (1) Modeling and simulation; (2) Comparing groups; and (3) Sampling and estimation. In the first unit, students build probability models and simulate from them to answer questions. They also begin to use the core logic of inference. In the second unit, students are introduced to the randomization test as a method to model the variation one would expect because of chance. In addition, they learn about appropriate inferences that can be made based on design facets of a study (e.g., generalizations from random sampling and causal inference under random assignment). The third unit introduces the method of bootstrapping to estimate the standard error of a statistic. It also includes an introduction to interval estimation.

The use of simulation and randomization-based methods within the CATALST curriculum to help students develop ideas of statistical inference in CATALST was inspired by George Cobb's (2005, 2007) ideas regarding randomization-based inference. Cobb points out that one advantage to teaching with a randomization-based approach to inference is that the link between statistical model, data, and analysis is much more explicit than in it is in the conventional approach (e.g., *t*-test) to inference. This, in turn, Cobb suggests, helps students focus on the core logic underlying the process of statistical inference rather than the mechanics.

The Core Logic of Statistical Inference

Activities in the CATALST curriculum were created to help students focus on "the core logic of inference" (Cobb, 2007, p. 13). When applied to randomized experiments and random samples, Cobb refers to this logic as the "three Rs": *randomize*, *repeat*, and *reject*. The CATALST project generalized this logic to encompass a broader simulation-based approach to inference as follows:

- *Model*. Specify a model that will generate data to reasonably approximate the variation in outcomes attributable to a random process.
- *Simulate*. Use the model to generate simulated data for a single trial. Specify the summary measure to be collected from each trial. Then, collect the summary measure for many trials.

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

- *Evaluate*. Use the distribution of the resulting summary measures to compare the behavior of the model to what was observed in the data. Draw inferences; make predictions, etc.

Focus on Modeling and Simulation

Paramount to understanding the core logic of inference is the more foundational aspect of models and modeling. Garfield and Ben-Zvi (2008) have pointed out that the use of models is fundamental to the practice of statistics, forming the basis for estimation and hypothesis testing. In addition, Wild and Pfannkuch (1999) suggest that developing ideas related to statistical modeling is instrumental in fostering students' statistical thinking. The CATALST curriculum highlights the importance of models and modeling through activities designed to engage students in using models to simulate data in order to evaluate claims and hypotheses. Although introduced in the first unit of the CATALST curriculum, the ideas related to modeling and simulation are revisited throughout the three-unit curriculum.

Another critical element of the CATALST curriculum is the use of simulation. Two common software tools used in the teaching of introductory statistics, especially for simulating data, are Fathom® (Finzer, 2012) and TinkerPlots™ (Konold & Miller, 2011). While Fathom® has the capability to perform the types of modeling and simulation needed for the course (e.g., Maxara & Biehler, 2006; 2007; Biehler & Prömmel, 2010), TinkerPlots™ was chosen because of its unique visual aspects that allow students to view and manipulate the models they select (e.g., sampler, spinner), and the ease in generating, collecting and representing simulated data.

Figure 1 shows a screenshot of one example of how the TinkerPlots™ software is used in the CATALST curriculum. The depicted simulation models rolling a six-sided die 10 times. Students use the simulation to explore the question, "How good are people at predicting random outcomes of common chance devices such as coins and dice?"

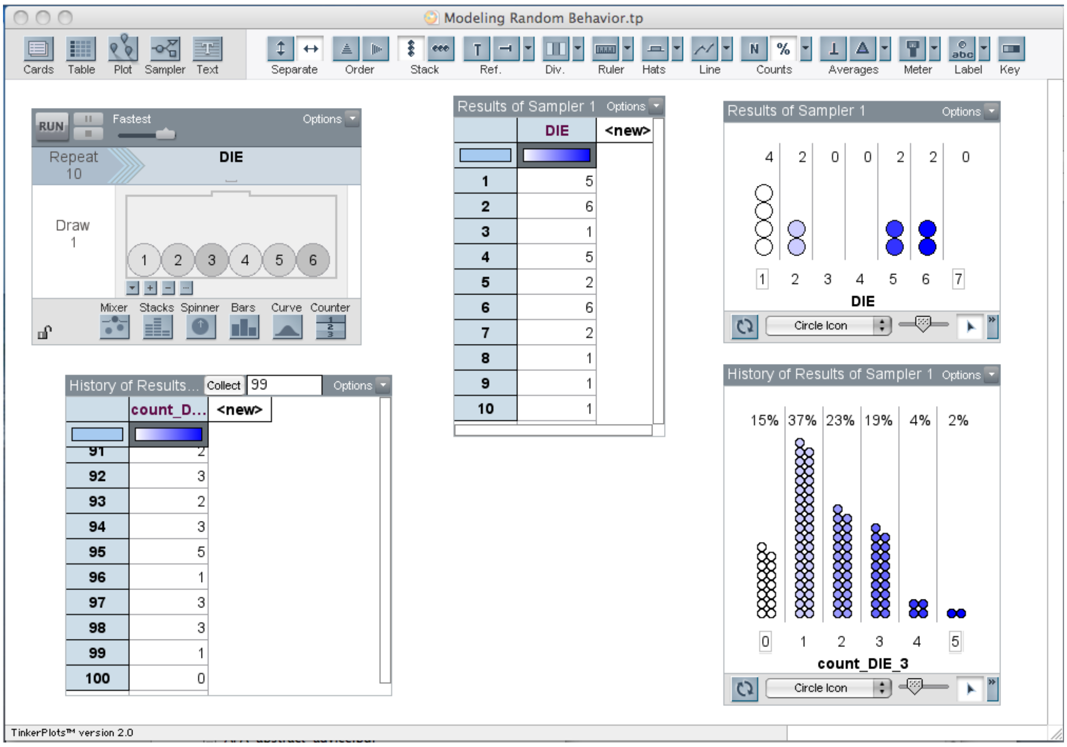

Figure 1. A screenshot of the use of TinkerPlots™ in the CATALST curriculum. The example shows a simulation of the number of threes obtained from rolling a six-sided die 10 times.

The upper-left part of Figure 1 shows the sampler, which has six equally likely values. A simulated sequence of 10 rolls is shown in the table and plot in the upper-middle and the upper-right side of Figure 1, respectively. In this simulation, students collected how many threes occurred in a 10-roll sequence (0 times in the plot on the upper-right side of Figure 1). These values are collected into the *collection table* (lower-left side of Figure 1) and can be used to create a plot of the collected summary measure over many simulated trials. The lower-right side of Figure 1 shows a plot of the summary measures (the number of threes in 10 rolls) that were collected across 100 simulated sequences of 10 rolls. Students can then use the results to answer questions they were asked prior to building the simulation (e.g., "What percentage of the time would you expect to see an outcome of five threes?").

Promoting Students' Statistical Thinking

Many current introductory statistics courses at the tertiary level present students with a wealth of material covering many topics and procedures. These courses, however, do not appear to be leading to desired students outcomes. For example, Garfield and Ben-Zvi (2008) have made note that students are often not able to remember what they have learned, and even when they do, are generally not able to transfer their knowledge to more advanced topics or new material outside the class.

A goal of the CATALST curriculum is to have students leave the course with the bigger picture of the statistical process—allowing them to solve unfamiliar problems and to articulate and apply their understanding. Using a metaphor introduced by Schoenfeld (1998), the CATALST curriculum is designed to teach students how to really "cook", rather than to only follow "recipes".

INTERVIEW STUDY

In the remainder of the paper, we provide information and results related to the ongoing evaluation of the CATALST curriculum. The primary study reported in this paper is based on an implementation of the CATALST curriculum that took place during the the fall semester of the 2011–2012 academic year. During that semester, the CATALST curriculum was taught in three sections of an introductory statistics class for liberal arts students at the University of Minnesota. Students who enroll in this course are typically not mathematics or statistics majors, and do not tend to be majoring in a STEM related discipline (e.g., physics, engineering). Each of the three sections was taught by a different graduate student (all pursuing a Ph.D. with an emphasis in statistics education). All instructors had received training in teaching the CATALST curriculum and met weekly with faculty researchers to plan, discuss, and reflect on the material and implementation.

At the end of the five-week *Modeling and Simulation* instructional unit, all students enrolled in the course were invited to participate in a one-hour interview. Nine student volunteers, one male and eight females, completed a problem-solving interview designed to investigate students' reasoning about samples and sampling variability. The second author of this paper conducted each of the interviews. The interviews, which lasted approximately an hour, were digitally recorded. In addition, the interviewees' actions on the computer were recorded using a screen-capturing tool.

The interview problems consisted of contexts where knowledge and understanding from the completed course unit could be applied, but that were novel enough to require more than a direct application of methods already encountered in the course. While the contexts were structured by an interview protocol, the questions were open-ended to allow the students to express their thinking and reasoning (Creswell, 2007). These questions were based on the informal inferential reasoning framework and assessment tasks suggested by Zieffler, Garfield, delMas and Reading (2008).

In both interview problems, students were asked to draw and support an inference about whether an observed result was unusual. The context for both problems was that of a computer game in which the player clicks on a single square from a predetermined grid size (4x4 in the first problem, and 5x5 in the second). Each square is one of four colors (blue, red, green or yellow), which is randomly filled according to a given probability model (see Figure 2).

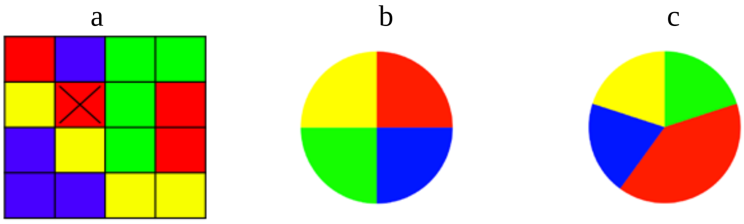

Figure 2. A screenshot of (a) the 5x5 filled grid from the second interview problem (the "x" represents a clicked square; (b) the first probability model (all four colors are equally likely); and (c) the second probability model (the color red is twice as likely as the other three colors).

RESULTS

Preliminary analysis of the interview data indicates that some students' reasoning about uncertainty at this point in the course is richer than is revealed by their initial descriptions of the methods they would use to address each problem. For example, one student described how to set up an appropriate model for the first problem, but then only described simulating a single trial of the simulation (i.e., generating a single sample). She noted that the expected number of blue for a single trial is 2.5, but she was not clear about how that information could be used to make a decision. The same student recognized the need to generate multiple simulated trials, collecting an appropriate statistic from each trial, when using TinkerPlots™. After generating the simulated distribution of the sample statistic collected from multiple trials, she used the distribution to quantify the likelihood of an outcome as or more extreme than the observed outcome, and made a statement about whether or not the observed outcome was surprising. Therefore, using TinkerPlots™, accompanied by prompts from the interviewer, revealed a deeper understanding and level of reasoning about uncertainty in the context of making a statistical inference than was revealed by the student's initial description. An implication is that while a deeper level of reasoning about uncertainty was not yet a habit of mind for some students, the reasoning is present when supported by use of the software and interviewer prompts.

After being prompted to use TinkerPlots™, most of the students interviewed used methods similar to those that they experienced during the *Modeling and Simulation* unit. The use of TinkerPlots™ appeared to facilitate their reasoning and thinking of the problems. For instance, one student initially reframed the original problem and then offered non-simulation based approaches for answering the reframed problem. However, when prompted to use TinkerPlots™, she designed an appropriate simulation for the original problem context and used the simulation results to address the original question. There were some students, however, for whom the use of TinkerPlots™ seemed not to facilitate their reasoning. For example, another student built a simulation for the purpose of convincing someone that the computer game creates random sequences of colored squares, rather than evaluating the likelihood of getting 5 blues if the game works correctly.

Several of the interviewed students appeared to be much more comfortable with frequency distributions than with sampling distributions. This is consistent with research findings from cognitive science (see Sedlmeier & Gigerenzer, 1997). As an example, many interviewees indicated that in order to decide whether the student or friend was correct, it would suffice to carry out a single trial that encompassed hundreds or thousands of "turns" and then examine the number/proportion of the individual colors to see if they were "evenly" distributed. Other students took a *sampling distribution perspective*. For example, one student suggested a simulation in which the game (composed of 10 turns) was played many times and the number of blue squares from each game would be collected in order to examine the distribution of these counts. There were also a few of the students interviewed who appeared to *blend these two perspectives* while working on the problem. One student, in fact, stated that the frequency distribution approach and the sampling distribution approach were equivalent. Another student, while looking at the simulated sampling distribution produced by TinkerPlots™, compared the percentage of games with counts of 5 blues selected out of 100 games to an expected value of 25% blue—as if she was assessing the long-run average percentage of blues for single turns of the game.

Students varied not only in their methods and thinking, but also in their attitudes and beliefs about what their simulation results told them. Some students readily updated their beliefs about the situation when the simulation results surprised them, but other students appeared to have a strong preference for *a priori* reasoning about the situation. One student even stated, after making an appropriate inference based on her simulation, that the simulation did not have much to do with her decision.

CONCLUSION

The results of this study suggest that after a five-week unit on *Modeling and Simulation*, students have an emergent ability to reason about and draw appropriate statistical inferences. It also suggests that through the use of TinkerPlots™, students were often able to identify discrepancies in their initial solutions and ways of thinking about the problem. This led many of them to update their thinking based on the results from the simulation. It is also worthy of note that many of the interviewees mimicked the TinkerPlots™ models in the vocabulary and explanations of their solutions, even before being prompted to use the software.

The TinkerPlots™ software tool seemed to promote the development of students' statistical thinking and give them a sense of what it takes to really ''cook'' rather than ''follow recipes''. This finding is consistent with previously collected evaluation data from the CATALST curriculum (e.g., Garfield, delMas, & Zieffler, 2012). Both this study and Garfield, delMas, & Zieffler (2012) found positive results regarding students' attitudes about the use of TinkerPlots™ software—a software that is rooted in how students learn as opposed to a purely analytical tool.

The preliminary results also indicate that students were at different levels of statistical understanding and thinking after five weeks of instruction. While some students produced valid models and drew reasonable conclusions, other students' understanding of statistical thinking and procedures was at a more nascent stage. For example, several students appeared to reframe the problem as testing whether or not the game was designed correctly instead of drawing an inference about whether or not an observed result was unusual. While some students show evidence of adopting a sampling distribution frame of reasoning, others were more apt to think in terms of a large sample frequency distribution, or to not make the distinction between the two perspectives. Problem-solving interviews were conducted with the nine students in the study at the end of the second and third units of the CATALST course. These interviews will be analyzed for evidence of how the students' statistical understanding and thinking developed through the remainder of the course.

The combined evidence from the CATALST evaluation data so far suggests that a course rooted in modeling and simulation-based methods is a viable option for tertiary students taking an introductory statistics course. To return to the cooking metaphor introduced earlier in the paper, although the "cooking" taught is basic—"gourmet chefs" were not prepared in a 45 hour course—students experiencing CATALST seem to develop skills that could be used in subsequent courses, as well as in daily life.

REFERENCES

Biehler, R., & Prömmel, A. (2010). Developing students' computer-supported simulation and modelling competencies by means of carefully designed working environments. In C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics* (ICOTS-8, July 2010), Ljubljana, Slovenia. Voorburg, The Netherlands: International Statistical Institute. http://www.stat.auckland.ac.nz/~iase/publications.php

Cobb, G. W. (2005). *The introductory statistics course: A saber tooth curriculum?* After dinner talk given at the United States Conference on Teaching Statistics.

Cobb, G. W. (2007). The introductory statistics course: A ptolemaic curriculum? *Technology Innovations in Statistics Education*, *1*(1). http://escholarship.org/uc/item/6hb3k0nz#page-1

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. Los Angeles, CA: Sage Publications, Inc.

Finzer, W. (2012). *Fathom® Dynamic Data*™ (v. 2L) [computer software]. Emeryville, CA: Key Curriculum Press.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York: Springer.

Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM*, *44*(7), 883-898.

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

Maxara, C., & 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.), *Proceedings of the Seventh International Conference on Teaching Statistics* (ICOTS-7, July 2006), Bahia, Brazil. Voorburg, The Netherlands: International Statistical Institute. http://www.stat.auckland.ac.nz/~iase/publications/17/7C1_MAXA.pdf

Maxara, C., & Biehler, R. (2007). Constructing stochastic simulations with a computer tool–students' competencies and difficulties. In D. Pitta–Pantazi and G. Philippou (Eds.), *Proceedings of the Fifth Congress of the European Society for Research in Mathematics Education* (pp. 762–771). Larnaca, Cyprus: Author. http://www.erme.unito.it/

Schoenfeld, A. H. (1998). Making mathematics and making pasta: From cookbook procedures to really cooking. In J. G. Greeno and S. V. Goldman (Eds.), *Thinking practices in mathematics and science learning* (pp. 299–319). Mahwah, NJ: Lawrence Erlbaum.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–265.

Zieffler, A., Garfield, J., delMas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, *7(*2), 40–58. Retrieved September 28, 2010 from http://www.stat.auckland.ac.nz/~iase/serj/SERJ7(2)_Zieffler.pdf