Using Applets and Video Instruction to Foster Students' Understanding of Sampling Variability

ABSTRACT

Online instructional modules that combine an applet, audio-visual tutorials, and guided discovery questions were created to teach the concept of sampling variability. The modules did contribute to an increase in understanding. However, they are a supplement to, not a replacement for, traditional instruction. The researchers found, using pretests and posttests, that student understanding of sampling distributions increased. There is room for futher improvement, which could be accomplished in two ways. A focus on designing for the introductory, rather than advanced, statistics student could be helpful. Also, giving students more feedback could help their performance in later modules.

Keywords: *Statistics education research, Sampling variability, Central Limit Theorem, simulation, guided learning, discovery learning, constructivism, Java applets, online learning*

1. INTRODUCTION

Research has found that using applets helps students develop understanding of sampling variability. However, research also says that students learn best when guidance is specifically provided. Online learning modules were created that involve use of an existing java applet paired with a guided discovery lessons and with audio visual tutorials. In this paper, we describe these modules and the results of pretests and posttests. We have found that the modules increase understanding but can be improved in two ways. First, the results show that students who are in upper-level courses showed more improvement. Therefore, the modules might be redesigned with students' prerequisite knowledge in mind. Second, there is dependence between the four modules; that is, students who did well on the first modules (1 and 2) were more likely to do well on the later ones (3 and 4). This indicates that providing more feedback on the early modules might improve performance on the later modules.

2. A REVIEW OF RELATED LITERATURE

One of the more challenging topics in an introductory statistics course is the concept of sampling distributions (Tversky and Kahneman 1971, Hagtvedt, Jones and Jones 2007, Garfield and Ben-Zvi 2008). Cobb (2007) states that students have a difficulty making a process into a mathematical object. Garfield and Ben-Zvi claim that understanding the nature of sampling variability is critical for understanding statistical inference, on which Cobb states the statistics curriculum should be centered. Garfield and Ben-Zvi report that many statistics educators are relying on technology, such as computer simulations, to foster sound statistical reason in students.

There is an emerging body of literature that encourages the use of statistical simulation for helping students develop a better understanding of some of the more abstract concepts in statistics (delMas, Garfield and Chance 1999, Hodgson and Burke 2001, Blejec 2003, Mills 2003, 2005, Erickson 2006, Lane and Peres 2006, Rossman and Chance 2006, Chance, Ben-Zvi, Garfield and Medina 2007, Hagtvedt, Jones and Jones. 2007, Hagtvedt, Jones and Jones 2008). Many of these simulation activities, such as java applets, are freely available on the Internet and allow students to investigate concepts interactively on the computer. Students are able to manipulate certain parameters (such as sample size or population proportion) and explore the resulting outcome. Collectively, these studies stress that using simulations as a pedagogical tool enables students to make connections from the abstract to the concrete more readily, the sampling distribution being one of these concepts.

There have been several studies that have investigated the effects of using technology to buttress statistical reasoning about sampling distributions. delMas, Garfield, and Chance (1999) investigated the effects of a sampling distribution program and found a statistically significant difference between pretest and posttest scores. Notably, they found that in order to have the greatest impact, students' attention must be focused on the relevant aspects of the simulation. Building on this research, similar results were found in a study that was replicated by Lunsford, Rowell, and Goodson-Espy (2006). Lane and Tang (2000) investigated the effectiveness of how well students transfer knowledge gained from using a simulation on sampling distributions to real world applications. They compared students who used a textbook approach to those who used a computer simulation. Those in the simulation group outperformed their textbook group counterparts.

Mills (2005) investigated the use of simulation pertaining to the Central Limit Theorem in an introductory statistics class. Students were divided into two groups: a simulation group and a traditional group. Mills found that the group that experienced the simulated activities outperformed their traditional counterpart group on pretests and posttests. Another experiment involving a simulation tool used to investigate sampling distributions and the Central Limit Theorem, Hagtvedt, Jones, and Jones (2007) found understanding of sampling distributions improved after students experimented with the simulation activity. From qualitative feedback, they also found that students preferred the simulation software compared with the traditional lecture. Moreover, as other researchers have posited (Hawkins 1996, Lane and Peres 2006, Garfield and Ben-Zvi 2008), activities of this ilk work best when students are supervised and/or guided by the instructor.

Not all studies involving sampling distribution simulations have yielded positive results. A study that evaluated the effectiveness of an Internet tutorial on sampling distributions compared two groups: one used an interactive tutorial on sampling distributions while the other group attended a lecture and a demonstration. After administering pretest and posttest, it was found that there were no significant differences between the two groups (Aberson, Berger, Healy, Kyle and Romero 2000).

3. LEARNING MODULES AND RESULTS

3.1 Design of the I^3 Learning Modules

The web-based Interactive, Independent, Inquiry-Based (I^3)learning modules for sampling variability combine an online applet with an audio-visual tutorial and guided discovery questions. The four Sampling Variability modules were created to have a consistent format and design. The design theme uses a set of links at the top labeled *Sampling Variability Home*, *Pretest*, *Warm Up*, *Tutorial*, and *Lesson* (see Figure 1). For evaluation purposes pretests and posttests were constructed for each module. The pretest was used to collect the content-knowledge baseline student information in an electronic format to facilitate analysis. The posttests are accessed by a separate link to keep students from viewing the posttest until the appropriate time. A *Feedback* tab also allows students to make their views known about the modules.

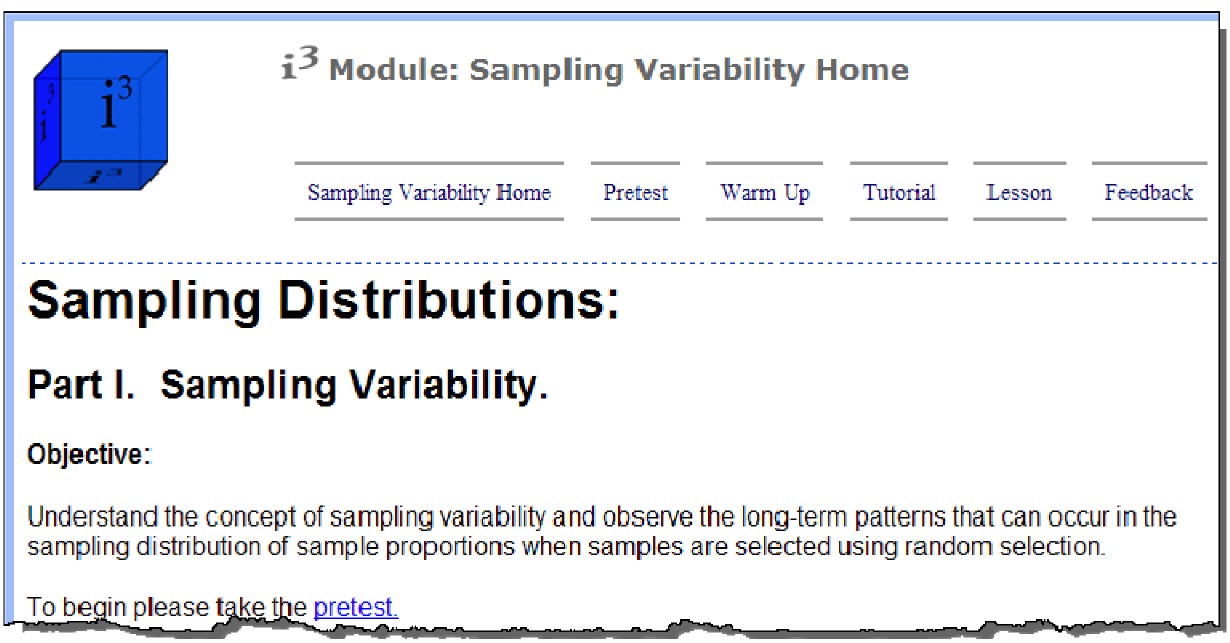


Figure 1: I^3 Learning Module Design

The home page includes the learning objective for the module and a link to the pretest (see Figure 1). Once the module is selected, the user takes the pretest. This test consists of multiple choice questions about the topic at hand. After the pretest, the user is directed to the "warm up," which typically consists of terminology the students should be familiar with prior to completing the module. This terminology would have either been discussed in class (e.g. parameter vs statistic) or addressed in an earlier module (e.g. confidence intervals). After the warm up, the user continues on to the tutorial. The audio-visual tutorial tab opens the web page with a video tutorial that carefully explains and demonstrates the java applet (see Figure 2). This tutorial was designed to focus students' attention on the part of the applet that is relevant to the particular module.

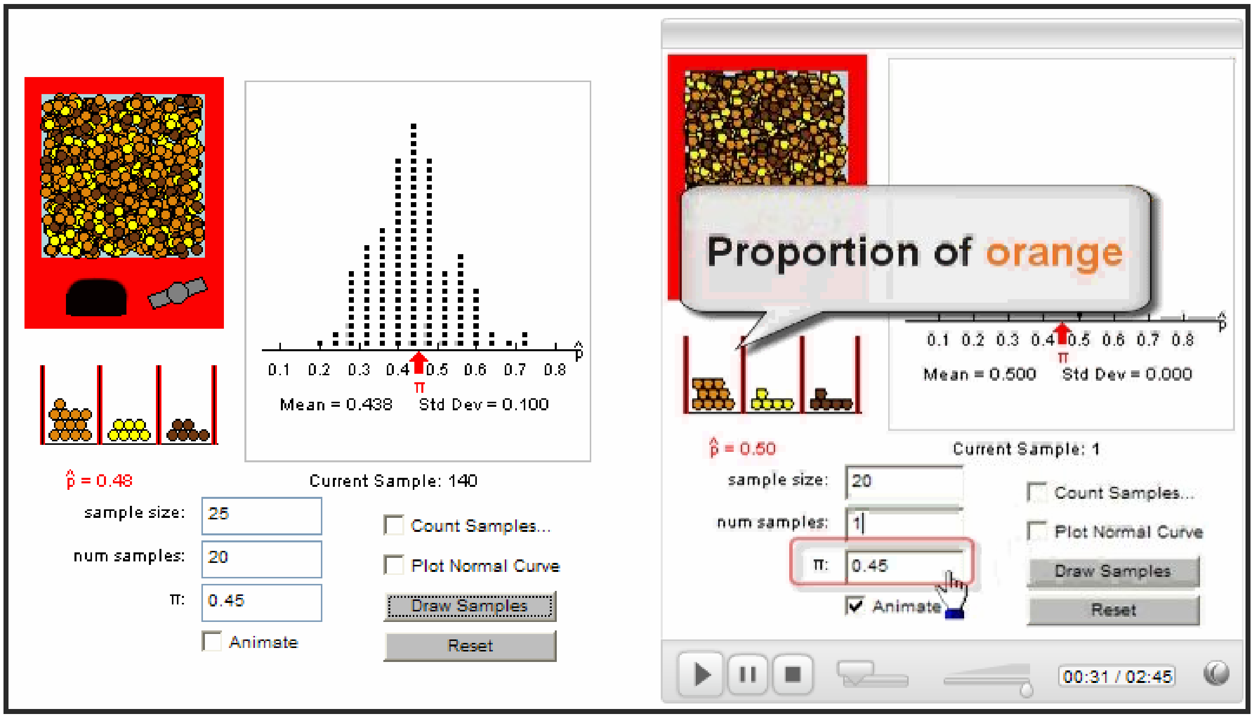


Figure 2: Tutorial for the Sampling Distribution module

The audio visual tutorial is designed to feel as if the teacher is demonstrating the material to the student. The visual component is an animation of the applet, highlighting important and sometimes complex features, showing the results of selecting certain components. During the animation, an audio component plays, which is a teacher's voice explaining what is happening on the screen and guiding the student's attention. The tutorial is short, typically 2-4 minutes long. After it finishes playing, the student is directed to the lesson.

The lesson is a series of well-planned questions based on Investigating Statistical Concepts, Applications, and Methods (Rossman and Chance 2005). The questions guide the student to discover the lesson objective. The questions are located side-by-side with the learning object, allowing students to explore the applet to help them answer the questions (see Figure 3). Once the lesson is complete, there is a wrap-up section, which summarizes the learning objective. Then the user takes a posttest and is directed to the feedback page.

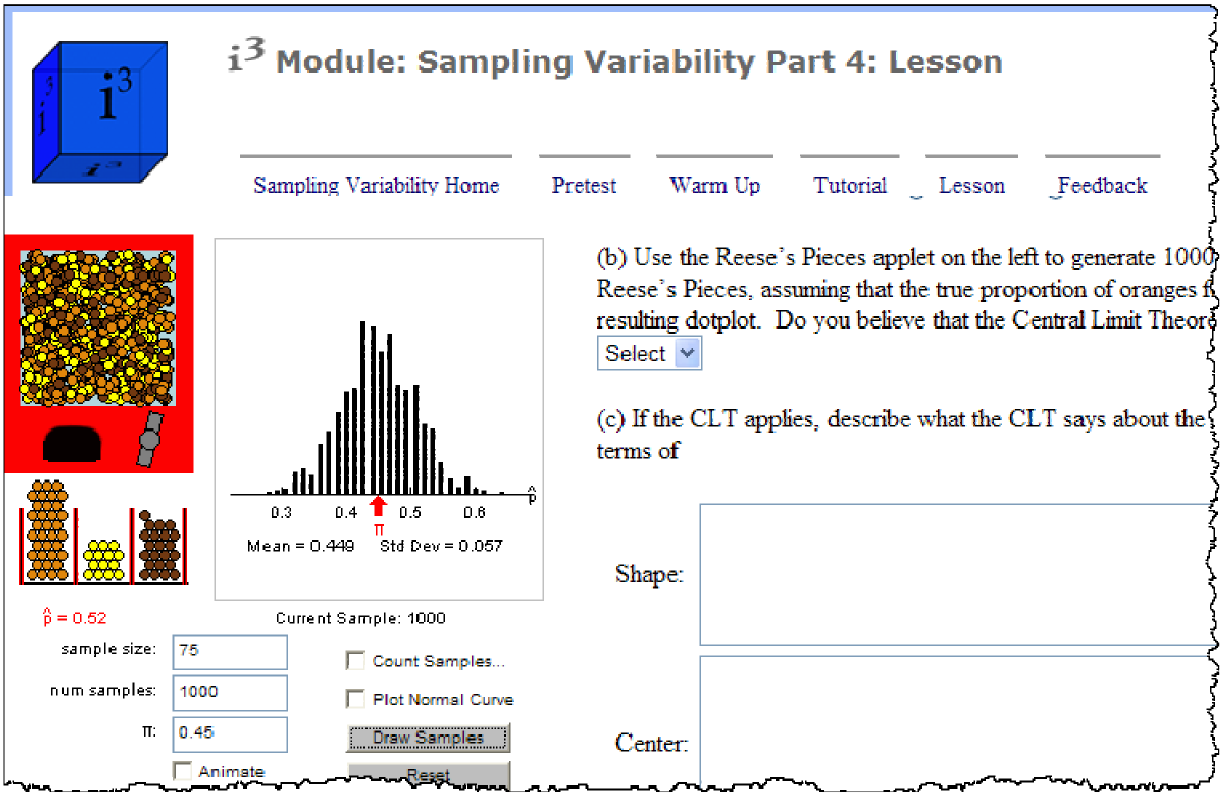


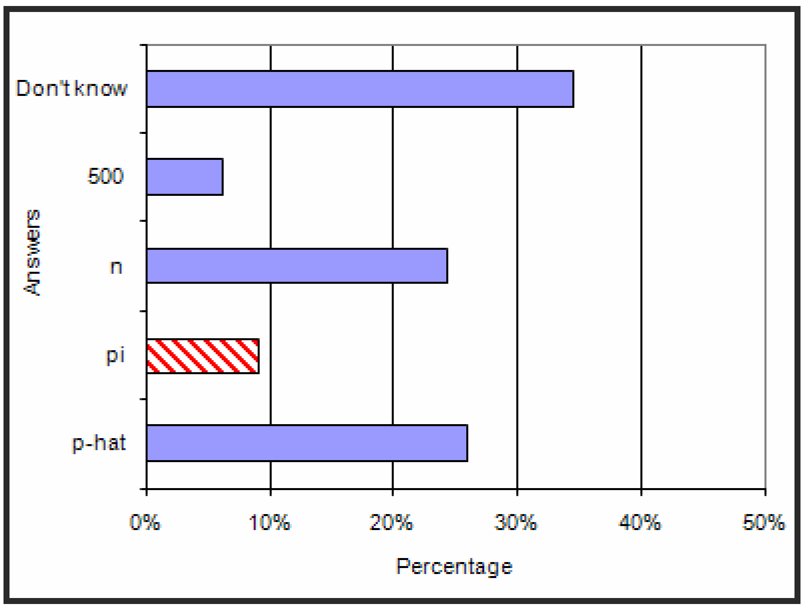
Figure 3: Lesson for the Sampling Variability (Part 4) module

Four modules were designed around the idea of sampling variability. Sampling variability encompasses the idea that a statistic is a random variable, and therefore has its own distribution and its own mean and standard deviation. Because each random sample taken from a population is different, the statistics calculated on them will differ. Understanding the distribution of the statistic allows one to decide if an observed value is unusual. Sampling distributions are famously difficult for students to understand. Yet, research (Garfield and Ben-Zvi 2008) is showing that understanding the concepts behind sampling distributions is critical for understanding the core of many statistics courses: statistical inference. Students often arrive in a statistics class with an intuitive understanding of sampling, and even the correct view that different samples will produce different sample statistics. It is much more challenging, however, for them to view the sampling distribution as a theoretical construct of all of the possible samples chosen from the population.

These modules are built around the Reese's Pieces Samples applet in the *Rossman/Chance Applet Collection* (Rossman and Chance 2004). The applet simulates the selecting of different samples of Reese's Pieces candies. Students input variables such as sample size and number of samples and then investigate the results. Each uses the same applet, but has a slightly different learning goal, each learning objective building on the knowledge from the previous module.

3.2 Sampling Variability Part 1

The first module, Sampling Variability Part 1, is designed to have students observe the long-term patterns that occur in the sampling distribution of sample proportions. The guided discovery questions start by pointing out the statistic of interest (the proportion of orange candies in the sample) and the fact that it is an estimate of an unknown parameter. Then the questions ask about the shape, center, and spread of the distribution of this statistic after 500 samples have been drawn. The pretest and posttest were multiple choice. A representative question on the pretest was, "The center of the sampling distribution of proportion with the sample size of 25 and with 500 replications will be the same as the value of: (pi)." (The correct answer is shown in parentheses.) It was paired with a question from the posttest, "The mean of the sample proportion values will approach (pi) when we have 500:" Responses for these questions are compared in Figure 4.



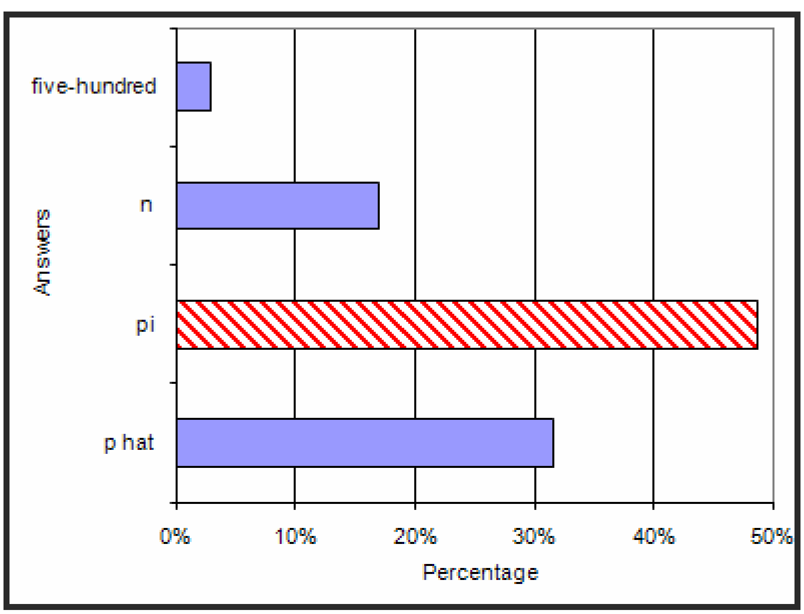
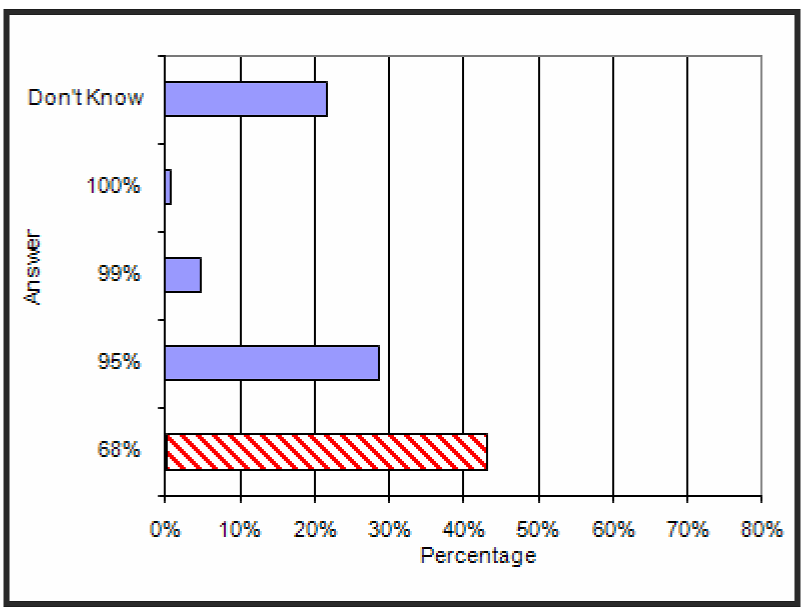


Figure 4: Results for Sampling Variability Part 1 (n = 177). Pretest question: *The center of the sampling distribution of proportion with the sample size of 25 with 500 replications will be the same value of: ( ).* Posttest question: *The mean of the sample proportion values will approach ( ) when we have 500:*

Although these results are not ideal, they do show improvement. The percentage of students who answer correctly increases from 9% on the pretest to 49% on the posttest. Further examination of the responses shows that students in higher-level courses have a greater percentage of correct answers than those for the lowest level course. In a graduate course, 35% of students answered correctly on the pretest, and 81% answered correctly on the posttest. We conjecture that the lower performance of the introductory students is caused by confusion with the use of the Greek letter "π" to mean something other than the ratio of the circumference of a circle to its diameter.

3.3 Sampling Variability Part 2

The second module has students investigate the number of sample percentages that are within one, two, and three standard deviations from the mean. The pretest contains the question: "The percentage of the sampling distribution of proportion which will be included by the interval around the mean with one SD will be about: (68%)." This is paired with the following question from the posttest: "The percentage of the sampling distribution of proportion which will be included by the interval around the mean with two SD will be about: (95%)." Responses to these questions are summarized in Figure 5.



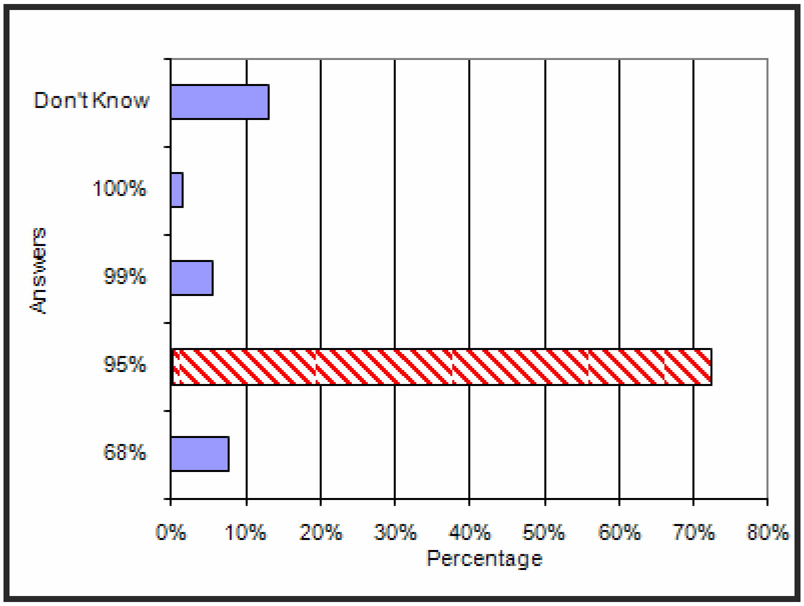
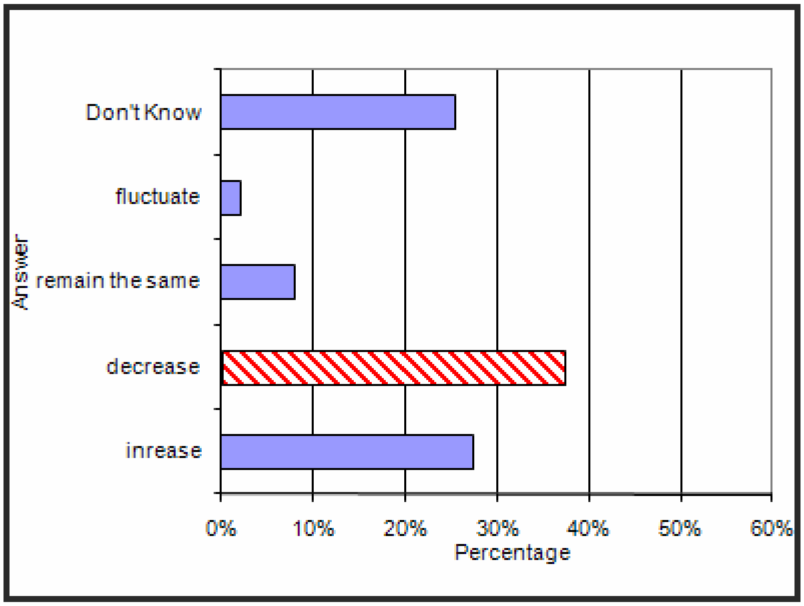


Figure 5: Results for Sampling Variability Part 2 (n = 130). Pretest question: *The percentage of sampling distribution of proportion which will be included by the interval around the mean with one SD will be about:* Posttest question: *The percentage of sampling distribution of proportion which will be included by the interval around the mean with two SD will be about:*

As observed in part 1, there is improvement between the pretest and the posttest, with 43% answering correctly on the pretest, and 72% answering correctly on the posttest. Also as before, we see a difference in performance by class. The graduate students increased from 70% correct on the pretest to 90% correct on the posttest, while the introductory class increased from 37% on the pretest to 68% on the posttest.

3.4 Sampling Variability Part 3

The third module explores how the sample size affects the behavior of the sampling distribution and introduces the Central Limit Theorem for proportions. The pretest contains the question: “If we increase sample size from 25 to 45 for the sampling distribution of proportion in building an interval around the sample proportion, the standard deviation of the sampling distribution of proportion will: (decrease)." This question is paired with a posttest question addressing the same idea: "If we reduce sample size from 45 to 25 for the sampling distribution of proportion in building an interval around the sample proportion, the standard deviation of the sampling distribution of proportion will: (increase)." Results for these questions are shown in Figure 6.



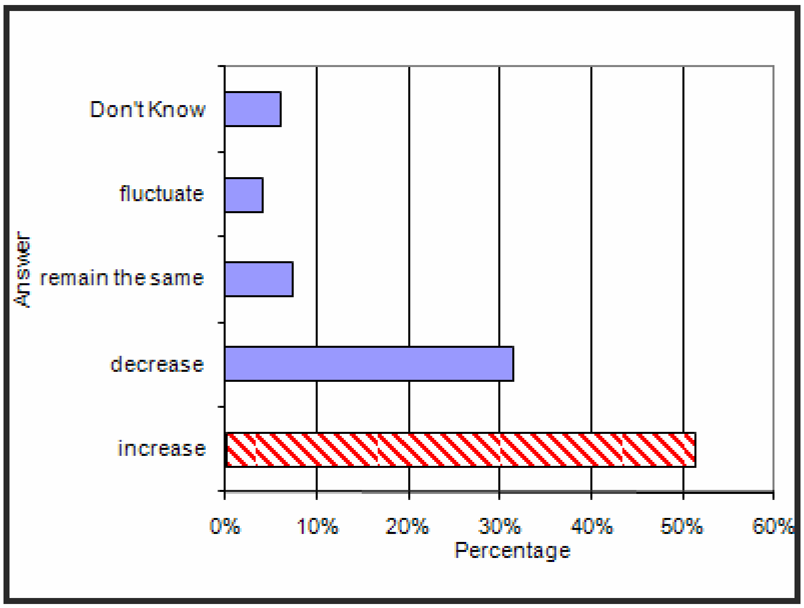


Figure 6: Results for Sampling Variability Part 3 (n = 150). Pretest question: *If we increase sample size from 25 to 45 for the sampling distribution of proportion in building an interval around the sample proportion, the standard deviation of the sampling distribution of proportion will*: Posttest question: *If we reduce sample size from 45 to 25 for the sampling distribution of proportion in building an interval around the sample proportion, the standard deviation of the sampling distribution of proportion will:*

We observe some improvement here, 38% answered correctly in the pretest, and 50% answered correctly in the posttest, while the percentage of "Don't Know" answers dropped from 27% to 6%. However, this seems to consist primarily of students who started with "Don't Know" as an answer moving to the correct answer. In examining the responses divided by class, we find that the students in the general education introductory course show responses that lead one to believe that they were guessing between the two plausible answers. On the pretest, 33% correctly answered "decrease" and 31% incorrectly answered "increase," while on the posttest, 43% correctly answered "increase" while 36% incorrectly answered "decrease." Students in the upper-level courses showed greater improvement. On the pretest, 48% correctly answered "decrease" and 28% incorrectly answered "increase," while on the posttest, 76% correctly answered "increase" while 17% incorrectly answered "decrease." The module was designed for the post-calculus course and, although it does not contain any calculus, it does seem to be more accessible to the more advanced students.

3.5 Sampling Variability Part 4

Finally, the fourth module encourages students to use their understanding of the sampling distribution to decide whether a particular outcome can be considered surprising. The pretest and posttest for this module contained a question that was designed to ask about the interpretation of a p-value, but was worded in a way that was confusing to the students. ("The true proportion value is .45, we obtained the sample proportion, the z-value, & the z-score. If the probability of the z-score is .20, we can say that it is (impossible/unlikely/very likely/100% likely/Don't know) that the sample proportion did not come from the population.") We found no change at all in students' understanding with this question, in fact graduate students with extensive background in statistics answered this question incorrectly. We speculate that this is due to the double negative phrasing in "*unlikely* that it did *not*."

However, another question that addresses p-values, but not in as conceptual a way, was phrased on the pretest as "If the computed z-value is small, the tail probability of the z-distribution cut off by the z-value will be: (large).". On the posttest, it was phrased as, "If the computed z-value is large, the tail probability of the z-distribution cut off by the z-value will be (small)." The responses to these questions are summarized in Figure 7.

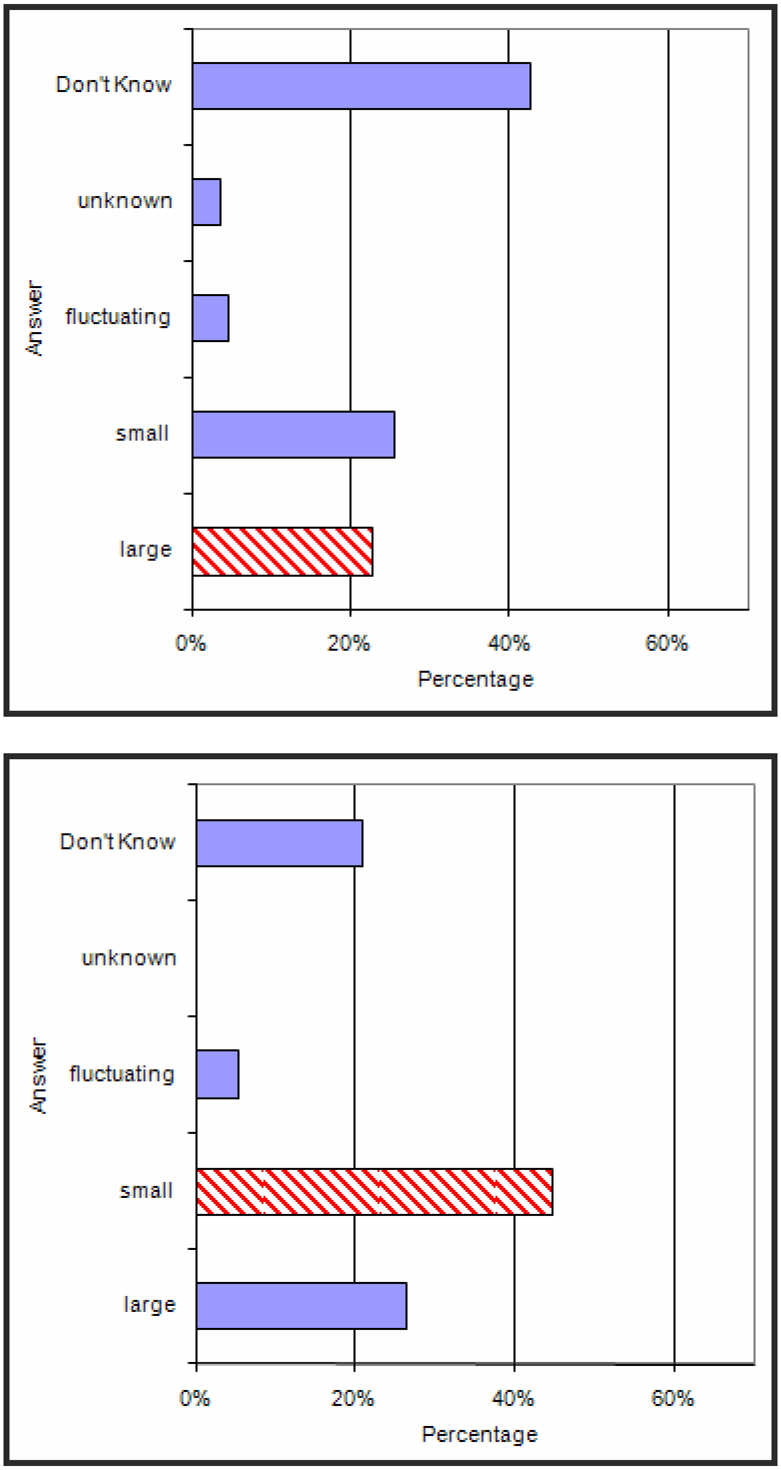


Figure 7: Results for Sampling Variability Part 4 (n = 110). Pretest question: *If the computed z-value is small, the tail probability of the z-distribution cut off by the z-value will be:* Posttest question: *If the computed z-value is large, the tail probability of the z-distribution cut off by the z-value will be*

As in the previous modules, we observe an increase in the percentage of correct answers and found that 21% answered correctly in the pretest, while 44% answered correctly in the posttest. Comparing the different level classes, the introductory course had 16% correct on the pretest and 40% on the posttest, while the advanced class had 50% correct on the pretest and 60% correct on the posttest.

3.6 Comparing Performance Across Modules

We also noticed a trend in the posttest scores across modules. In the question described above for the posttest in module 1, only 49% answered correctly, but this is actually a very large improvement from the pretest (9%). After module 1, we see 72% answering correctly in the posttest of module 2, 51% in module 3, and only 45% in module 4. We believe that this is due to the dependence on each module on using the information from previous modules as prerequisite. Evidence for this is found in comparing similar questions from the posttest of module 1 and module 4, and from the posttest of module 2 and module 3. For example, question 7 on the posttest of module 1 asks, "The mean of the sample proportion values will approach (pi) when we have 500." A similar question is question 2 on the posttest of module 4, which asks, "CLT states that the center of the sampling distribution of sample proportion will be the same as: (the population proportion value)." We were able to compare the responses on these questions since the name of the student was recorded. There were a total of 102 students who answered both questions. Of the 49 who answered incorrectly on module 1, only 14 (29%) answered correctly on module 4. Of the 53 who answered correctly on module 1, 32 (60%) answered correctly on module 4. This implies that performance on the previous module is predictive of performance on the later module.

Similarly, question 2 on the posttest of module 2 asks, "If the width of an interval around the mean score is getting smaller, the resulting interval will include (less) portion of the distribution." A similar question is question 1 on the posttest of module 3, which asks, "The narrower the interval, the (less) confidence we have that a randomly chosen proportion will be included in the interval." There were a total of 104 students who answered both questions. Of the 33 who answered incorrectly on module 2, only 18 (55%) answered correctly on module 3. Of the 62 who answered correctly on module 2, 42 (68%) answered correctly on module 3. (These totals do not sum to 104 because "don't know" was a choice.) Again we observe that performance on the previous module is predictive of performance on the later module.

In summary, the sampling variability modules increased understanding of this topic. However, the modules by themselves are clearly not enough to teach such a difficult topic. The percentages of correct answers on the posttests are still not very large. The differences we observe between the performance of students in the different level courses suggests that the modules are aimed at a higher-level student. A redesign may make them more accessible to the lower-level students. For example, vocabulary could be defined several times throughout the modules rather than once at the beginning. Also, the evidence of a dependence on later modules of a correct understanding of earlier modules implies that additional feedback between modules may be helpful. The modules were designed with the assumption that instructors would discuss the modules between uses, but in practice we find that instructors tend to assign all four modules at the same time, and students attempt to do all of them within a short time period. Therefore, the modules themselves could be redesigned to give more feedback to students.

4. FURTHER RESEARCH

The authors have created the Independent Interactive Inquiry-based (I^3) Learning Modules, which form a tool to use technology outside the classroom to complement other teaching techniques in communicating statistical ideas effectively. Due to the overall success of these modules, based on pretest and posttest data, the authors would like to extend their prior research. There are three main objectives to help further the research in this arena.

The first objective is to modify the existing modules to make them ready for dissemination and modify them for a dual audience (introductory statistics and post calculus-based statistics). For example, vocabulary should be defined in multiple places within the modules. Also, the module data can only be access by the researchers. We have applied for funding to allow for teachers to access student results automatically within a module management system. While the modules are not ready for general release, they can be accessed here: http://mcdaniel.mtsu.edu/newIcubeLessons/Home2.htm.

We would also like to be able to measure other variables that may affect how well students progress through a module or through multiple modules that are strongly connected (e.g Sampling Variability parts 1-4). For instance, does receiving feedback at various stages of the modules, (thus reinforcing the correct answer or correcting a wrong answer) help the student? Students could receive feedback from two sources: the lesson module itself or from the instructor. We would also want to examine benefits of spacing out the modules to determine if students who have time between the modules do better or worse than those who do them all at once.

Finally, the third objective is to add additional topics to the I^3 learning modules. Currently we have the sampling variability modules described here and 4 other topics: the binomial distribution, confidence intervals, randomization, and statistical significance. Other topics could be added, such as the interpretation of confidence intervals, correlation and regression, variation, the law of large numbers, randomization tests, approximation with the normal distribution, sampling distributions, errors in hypothesis testing, power, and ANOVA.

5. CONCLUSION

Online modules that combine guided discovery with audio-visual tutorials contribute to an increase in understanding. However, they are a supplement to, not a replacement for, traditional instruction. The researchers found that student understanding of sampling distributions increased although there is room for futher improvement. A focus on designing for the introductory statistics student could be helpful. Also, giving students more feedback could help their performance in later modules.

6. ACKNOWLEDGEMENTS

We appreciate the support of a grant from the National Science Foundation (NSF CCLI – 0443088). We also are indebted to Dr. Ginger Rowell, whose participation made this project possible. We would also like to thank the following graduate students: Marisella Castro, Kai Guo, Megan Hall and Kacie Kleja. We would like to thank the following MTSU instructors and professors for in-class testing: Carla Adamson, Dr. Curtis Church, Christine Snearly, and Dr. Dennis Walsh.

7. REFERENCES

Aberson, C., Berger, D., Healy, M., Kyle, D., and Romero, V. (2000), "Evaluation of an Interactive Tutorial for Teaching the Central Limit Theorem," *Teaching of Psychology*, 27, 289-291.

Blejec, A. (2003), "Teaching Statistics by Using Simulations on the Internet," in *IASE Satellite Conference on Statistics*, Berlin, Germany.

Chance, B., Ben-Zvi, D., Garfield, J., and Medina, E. (2007), "The Role of Technology in Improving Student Learning of Statistics," *Technology Innovations in Statistics Education Journal*, 1.

Cobb, G. (2007), "The Introductory Statistics Course: A Ptolemaic Curriculum?," *Technology Innovations in Statistics Education*, 1.

delMas, R., Garfield, J., and Chance, B. (1999), "A Model of Classroom Research in Action: Developing Simulation Activities to Improve Students' Statistical Reasoning," *Journal of Statistics Education*, 7, 1-12.

Erickson, T. (2006), "Using Simulation to Learn About Inference," in *Proceedings of the Seventh International Conference on Teaching Statistics*, eds. A. Rossman and B. Chance, Voorburg, The Netherlands: International Statitics Institute.

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

Hagtvedt, R., Jones, G., and Jones, K. (2008), "Teaching Confidence Intervals Using Simulation," *Teaching Statistics*, 30, 53-56.

Hagtvedt, R., Jones, G. T., and Jones, K. (2007), "Pedagogical Simulation of Sampling Distributions and the Central Limit Theorem," *Teaching Statistics*, 29, 94-97.

Hawkins, A. (1996), "Myth-Conceptions!," in *Research on the Role of Technology in Teaching and Learning Statistics*, eds. J. Garfield and G. Burrill, Voorburg, The Netherlands: International Statistical Institute, pp. 1-14.

Hodgson, T., and Burke, M. (2001), "On Simulation and the Teaching of Statistics," *Teaching Statistics*, 22, 91-96.

Lane, D., and Tang, Z. (2000), "Effectiveness of Simulation Training on Transfer of Statistical Concepts," *Journal of Educational Computing Research*, 22, 383-396.

Lane, D. M., and Peres, S. C. (2006), "Interactive Simulations in the Teaching of Statistics: Promise and Pitfalls," in *Proceedings of the Seventh International Conference on Teaching Statistics*, eds. A. Rossman and B. Chance, Voorburg, The Netherlands: International Statistics Institute.

Lunsford, M., Rowell, G., and Goodson-Espy, T. (2006), "Classroom Research: Assessment of Student Understanding of Sampling Distributions of Means and the Central Limit Theorem in Post-Calculus Probability and Statistics Classes," *Journal of Statistics Education*, 14.

Mills, J. (2003), "A Theoretical Framework for Teaching Statistics," *Teaching Statistics*, 25, 56-58.

Mills, J. (2005), "Learning Abstract Statistics Concepts Using Simulation.," *Educational Research Quarterly*, 28, 18-33.

Rossman, A., and Chance, B. (2004), "The Rossmann/Chance Applet Collection. ," Available at http://www.rossmanchance.com/applets/index.html.

Rossman, A., and Chance, B. (2005), *Investigating Statistical Concepts, Applications, and Methods*, Pacific Grove, CA: Duxbury Press.

Rossman, A., and Chance, B. (2006), "Using Simulations to Teach and Learn Statistics," in *Proceedings of the Seventh International Conference on Teaching Statistics* (Vol. 7), eds. A. Rossman and B. Chance, Voorburg, The Netherlands: International Statitics Institute.

Tversky, A., and Kahneman, D. (1971), "Belief in the Law of Small Numbers," *Psychological Bulletin*, 76, 105-110.