1. INTRODUCTION

One of the biggest changes in the statistics curriculum in the last 25 years has been the integration of technology. Technology has the potential to allow students to work independently and explore difficult concepts. However, research has shown that students, left on their own, often miss the important concepts or subtleties that instructors deem to be important (Lane and Peres 2006). To address this issue, a project was implemented to develop and evaluate interactive modules that guide students in discovery of probability and statistics concepts. As with other simulations, their effectiveness depends on being grounded in solid pedagogical practice.

There has been a push from statistics education researchers to change the traditional lecture-only format to a more activity based format and to change from using technology as a calculating device to using it in addition as a tool that helps students understand the concepts. These changes are summarized by the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Project (Aliaga, et al. 2005), which sets forth six recommendations for teaching introductory college statistics courses. It has also been reported that students learn by constructing meaning for themselves, not just receiving knowledge. Some suggest (Marasinghe, Meeker, Cook, and Shin 1996; Mills 2003; Healy 2006) that one way to implement this theoretical framework in the statistics curriculum is via computer simulations, which can help students develop a better understanding of the more abstract concepts. Erickson (2006) even calls this approach a "constructivist's dream" (p. 3).

This paper will describe the implementation and evaluation of the Independent Interactive Inquiry-based (I^3) Learning Modules, which were funded by a National Science Foundation grant. The project originally had two primary goals: (a) to develop discovery-based interactive audio-visual instructional modules to improve student learning of post-calculus probability and statistics concepts, and (b) to evaluate the effectiveness of this learning approach and these learning objects. The project was later extended to include algebra-based statistics courses. The first objective was achieved by using existing open-source Java applets from the *Probability and Statistics Object Library* (Siegrist 2001) and the *Rossman/Chance Applet Collection* (2004), and by combining audio-visual instruction with interactive inquiry-based lessons. Each lesson is a series of questions that guide students to discover the lesson objective. The questions are side-by-side with the java applet, which allows students to explore particular aspects of the concept being addressed. These materials were designed for students to use at home to promote independent learning. Topics for some of the I^3 Learning Modules include the binomial distribution, sampling distributions, the Central Limit Theorem, confidence intervals, significance testing, and randomization. To address the second objective, the effectiveness of the modules was evaluated using pretests and posttests, demonstrating an improvement in students' overall conceptual understanding in three of the four modules discussed in this paper. These modules form a tool to use technology outside the classroom to complement other teaching techniques in communicating statistical ideas effectively.

These learning modules were administered to students at a regional four-year comprehensive university. Each module was used by students in statistics classes that have between 20-35 students at the introductory, intermediate, and advanced levels. Every student took a pretest just before using the module and a posttest just after using the module.

2. A REVIEW OF RELATED LITERATURE

2.1 Emerging Technology Tools

Technology has had an indelible impact on how statistics courses at many levels are taught and learned. The following Guidelines for Assessment and Instruction in Statistics Education (GAISE) were given by the American Statistical Association in 2005:

1. Emphasize statistical literacy and develop statistical thinking;
2. Use real data;
3. Stress conceptual understanding rather than mere knowledge of procedures;
4. Foster active learning in the classroom;
5. Use technology for developing conceptual understanding and analyzing data;
6. Use assessments to improve and evaluate student learning (Aliaga et al. 2005).

The GAISE report says, "Technology tools should also be used to help students visualize concepts and develop an understanding of abstract ideas by simulations" (Aliaga et al. 2005, p. 12). The emergence of statistical software packages (e.g. Minitab, SPSS, R) and hand-held calculators (e.g. TI-84) has allowed instructors (and, therefore, students) to focus on conceptual understanding (Blejec 2003), rather than having students perform tedious calculations by hand. It has also been shown (delMas, Garfield, and Chance 1999) that multimedia simulation activities improve students' statistical reasoning.

Implementing technology alone, however, may not be sufficient. Collins and Mittag (2005) compared two groups of introductory statistics students that were being taught hypothesis testing: one that used a calculator and one that did not. They found no statistical differences between the two groups. Additionally, studies have found that using statistical software packages to analyze data or perform simulations does not lead to deeper learning of the concepts unless these simulations are well structured (Hawkins 1996; Lane and Peres 2006).

2.2 Guided Discovery

Many researchers in mathematics and statistics education have said that students learn by constructing meaning for themselves, not just receiving knowledge (von Glasersfeld 1987; Cobb 1994; Mills 2003). Many statistics education researchers view implementing computer simulations as a tool to foster this constructivist notion of learning some of the more abstract concepts (delMas et al. 1999; Erickson 2006; Healy 2006; Garfield and Ben-Zvi 2009).

The inquiry-based guided discovery approach described by Lane and Peres (2006), in which students are led toward conceptual ideas through specific questions and directions on the instructor's part, appears to be key to implementing simulations (as well as other types of instruction) in the classroom. This has been shown in the general sciences (Vockell and Rivers 1984; Haury 1993; Edelson, Gordin, and Pea 1999) and in mathematics (Hirsch 1977; Hiebert, et al. 1996), as well as in statistics (Aberson, Berger, Healy, and Romero 2002; Lane and Peres 2006). Moreno and Mayer (2005) found that guided discovery leads students to deeper learning when compared to pure discovery. Further, some have found unguided discovery inefficient and ineffective (Rieber and Parmley 1995; de Jong and van Joolingen 1998; Mayer 2004).

In an experiment in physics, Rieber, Tzeng and Tribble (2004) evaluated the effects of interactive simulations of Newtonian motion principles. They found that participants who were given graphical explanations of the simulation were more successful than those who were not provided with these explanations. Vockell and Rivers (1984) found that biology students who were engaged in guided discovery simulations outperformed their control group counterparts (unguided discovery simulations and no simulations). Clark and Jorde (2004), using a simulation involving thermodynamics, found that students who were given additional multimedia guidance via pictures, text, and audio had a statistically significant better understanding of thermal equilibrium than the control group, which did not have the extra guidance on the simulation.

2.3 Computer Simulations In Statistics

The use of computer applications in statistics is extensive. Statistics instructors have been able to readily incorporate data analysis software in the classroom since the advent of the graphical user interface in the mid-1980s (Hawkins 1996; Mills 2005). Mills states that these software packages do not help students understand some of the more difficult and abstract concepts; rather, the students merely mimic the workings of the particular software.

Recent advances in the development of web-based interactive applications such as Java applets (many of which are freely available online) have allowed statistics instructors to easily implement various simulation activities into the classroom (delMas et al. 1999; Hodgson and Burke 2001; Blejec 2003; Mills 2003, 2005; Erickson 2006; Lane and Peres 2006; Rossman and Chance 2006b; Chance, Ben-Zvi, Garfield, and Medina 2007; Hagtvedt, Jones, and Jones 2007; Hagtvedt, Jones, and Jones 2008). In general, these simulations allow users to make changes to an input variable while observing the resulting output. This research does appear to suggest that these simulations can help students have a deeper understanding of the abstract concepts or at least make them more concrete. Lane and Peres, however, point out that just doing in-class demonstrations of these applets and interactive simulations will not necessarily lead to a deeper understanding (or even cursory understanding) of the concepts. Research shows that student learning of statistics is improved when instructors use active learning, real data, conceptual understanding, and technology (Gordon and Gordon 1992; Moore 1992, 1997). Also, educational materials designed to utilize students' different learning styles (auditory, visual, and kinesthetic) are more likely to engage them in the learning task and to improve student learning (Searson and Dunn 2001; Uzuntiryaki, Bilgin, and Geban 2003; Clark and Mayer 2008).

3. LEARNING MODULES AND METHODOLOGY

3.1 Design of the I^3 Learning modules

The web-based I^3 Learning Modules are designed to have a consistent format and design. The current version of the modules can be found at http://mcdaniel.mtsu.edu. Each module is designed to teach one basic idea to students. For example, the Binomial Distribution Module is designed to help students understand how different values for n and p affect the shape of the distribution. It does not teach when to use the binomial distribution, how to interpret the graph, or how to approximate the binomial distribution with the normal distribution.

The design theme uses a view of a set of folder-tabs at the top labeled *home*, *review*, *tutorial*, and *lesson*. For evaluation purposes pretests and posttests were constructed for each module, and a *pretest* folder-tab is 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 express their views about the modules.

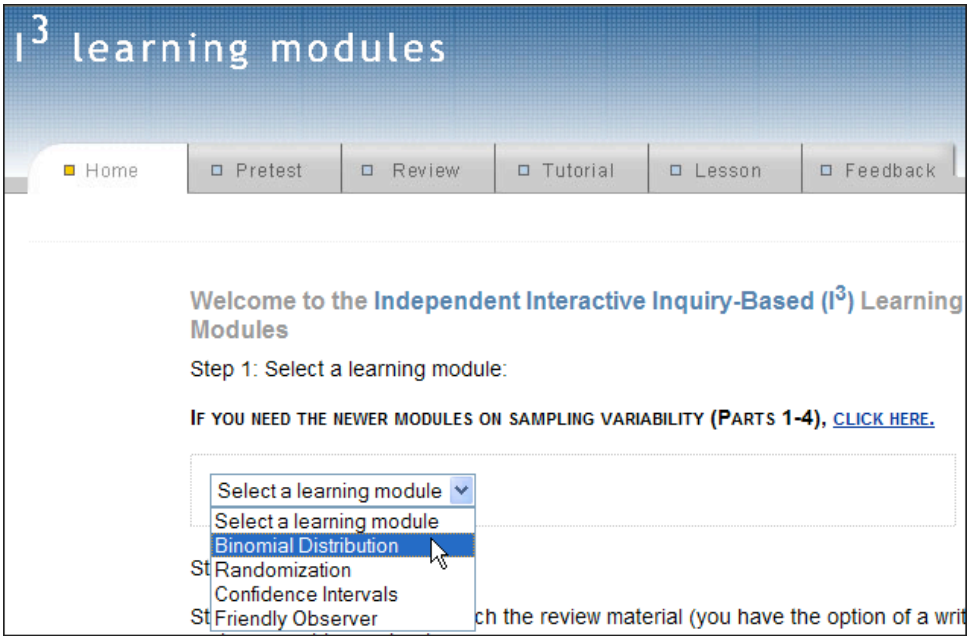

Figure 1. I^3 Learning Module Design

The home tab looks like a page the size of the screen with the learning objective for the module and very brief instructions for completing the learning module (see Figure 1). Once the module is selected, the user takes the pretest. This test consists of either open-ended or multiple choice questions about the topic at hand. After the pretest, the user continues on to the tutorial. The audio-visual tutorial tab opens the web page with a 2 to 6 minute video tutorial that carefully explains and demonstrates the interactive learning object, usually an applet, or internet-based application (see Figure 2).

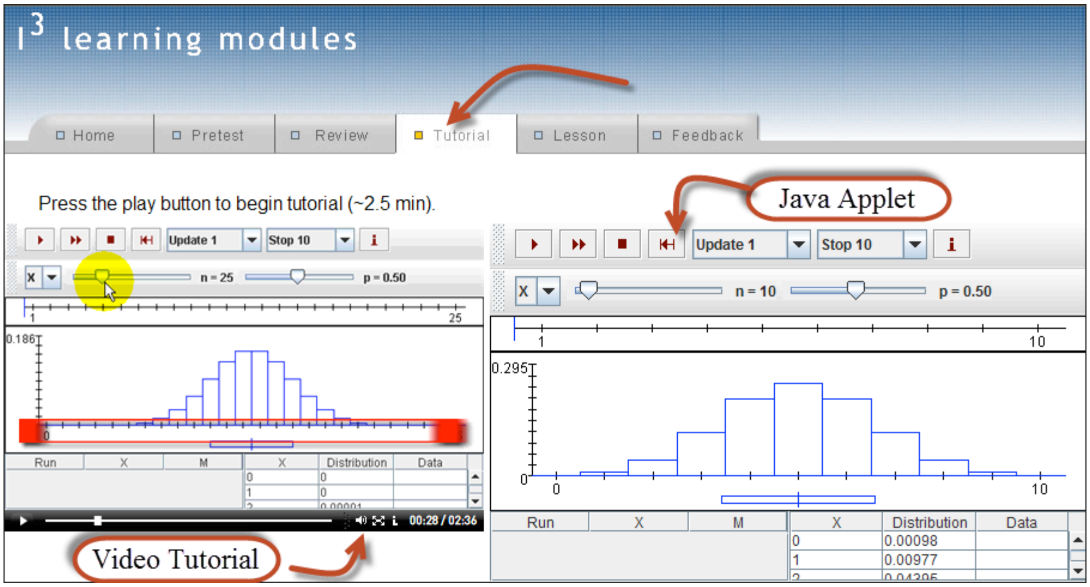

Figure 2. Tutorial for the Binomial Distribution Module

This part of the module 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 featuring a teacher's voice explaining what is happening on the screen in order to guide the student's attention. After it finishes playing, the student moves on to the lesson tab.

The lesson is a series of guided discovery questions. The questions in the Binomial Distribution Module were created by the developers of the modules. The questions in the remaining three modules are based (with permission) on *Investigating Statistical Concepts, Applications, and Methods* (Rossman and Chance 2006a). The questions are located side-by-side with the learning object, allowing students to explore the applet to help them answer the questions (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.

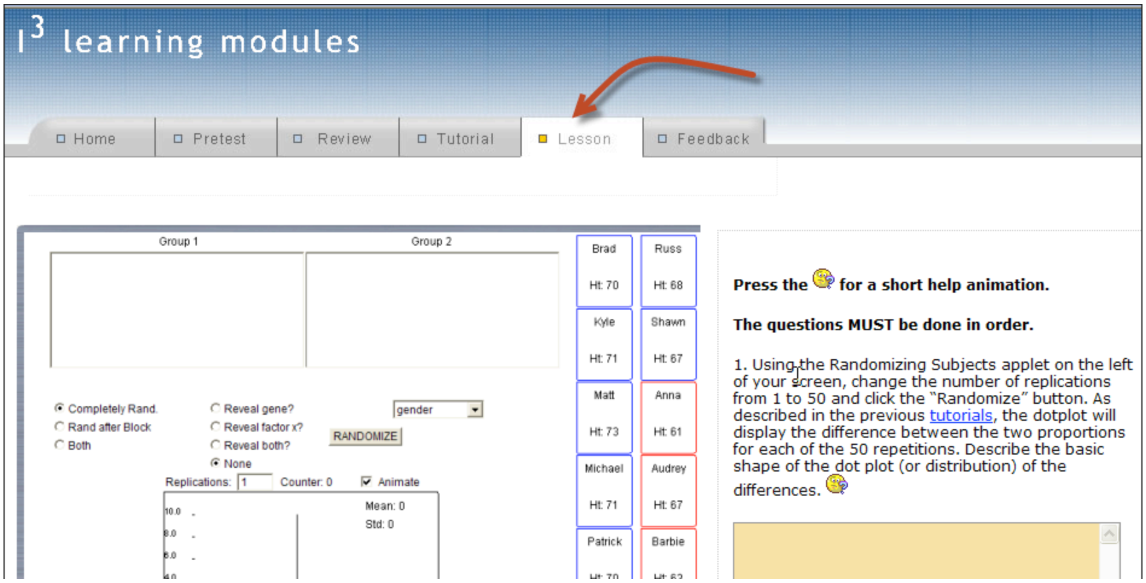

Figure 3. Lesson for the Randomization Module

There are currently eight learning modules, built around five applets. The modules are as follows: Binomial Distribution, Randomization, Confidence Intervals, Statistical Significance, and Sampling Variability (parts 1-4). For space considerations, this paper discusses four of the eight learning modules, excluding the sampling variability modules, which are similar in design to the modules we will describe.

3.2 The Binomial Distribution Module

The Binomial Distribution Module (Figure 2) is designed to teach what happens to the shape of the binomial distribution as the parameters of the distribution change. The binomial distribution describes the behavior of a random variable that counts the number of successes when a binary trial is repeated independently a fixed number of times. If the number of trials is large and the probability of success for each trial is not too big and not too small, the distribution of the number of successes shows the familiar bell-shape of the normal distribution. This concept is what this module is designed to teach.

The module is built around an applet in the *Virtual Laboratories in Probability and Statistics* developed by Siegrist (2001). The applet contains sliders to select values for n and p and shows the corresponding binomial distribution. In the lesson, which was developed by Drs. Ginger Rowell and Scott McDaniel, students are directed to select specific values and are asked questions that direct their attention to the distribution's shape, height, and width. They are asked to change p while keeping n constant, then to change n while keeping p constant. Then they are asked to describe the shapes they see.

3.3 Confidence Intervals

The Confidence Interval Module is designed to show the difference between two confidence intervals for the mean: the one based on the normal curve, and the one based on the t-distribution. A confidence interval is paired with a percentage that is interpreted as the probability of the actual value of the parameter of interest lying in the confidence interval.

This module has students use simulations to observe how closely the actual confidence levels come to the theoretical confidence level. Students also explore the widths of the two types of confidence intervals. Then they examine the robustness of t-intervals by considering different population shapes.

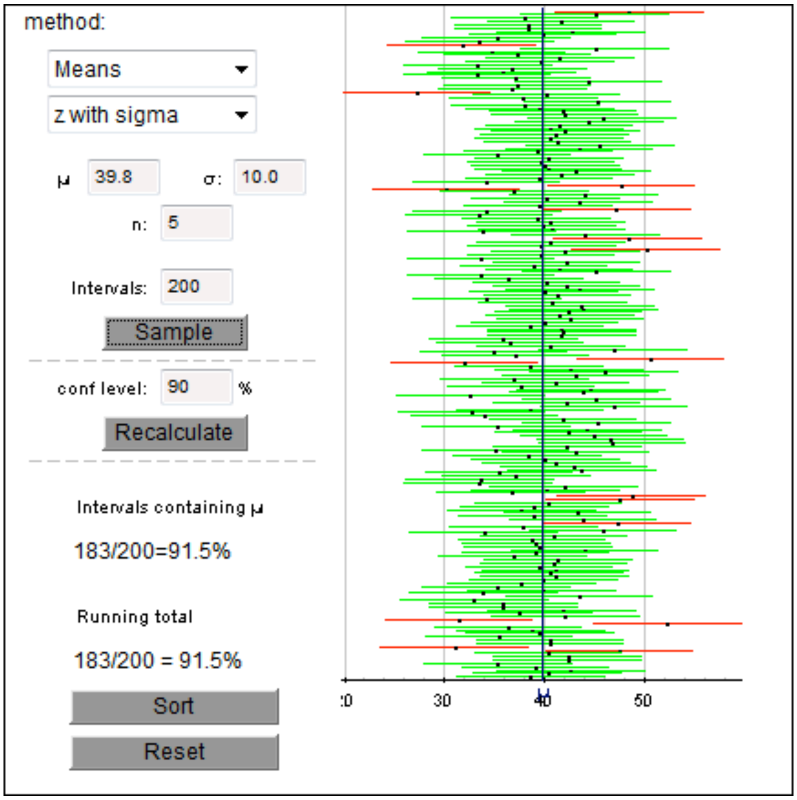

Figure 4. Applet portion of the Confidence Interval Module

The lesson is built around the Simulating Confidence Intervals applet (Figure 4) in the *Rossman/Chance Applet Collection* (Rossman and Chance 2004) and includes guided discovery questions from the accompanying text (Rossman and Chance 2006a). Students are asked to fill in some of the parameters on the confidence interval applet, such as the number of intervals to calculate, the confidence level, and the parameters of the population. The applet simulates a high number of confidence intervals, and students are asked to observe the proportion of the intervals that contain the target value.

3.4 Statistical Significance

In the Statistical Significance Module an experiment is described that compares behavior of two groups. Students are led to consider how likely the observed outcome would be if there were actually no difference between the groups.

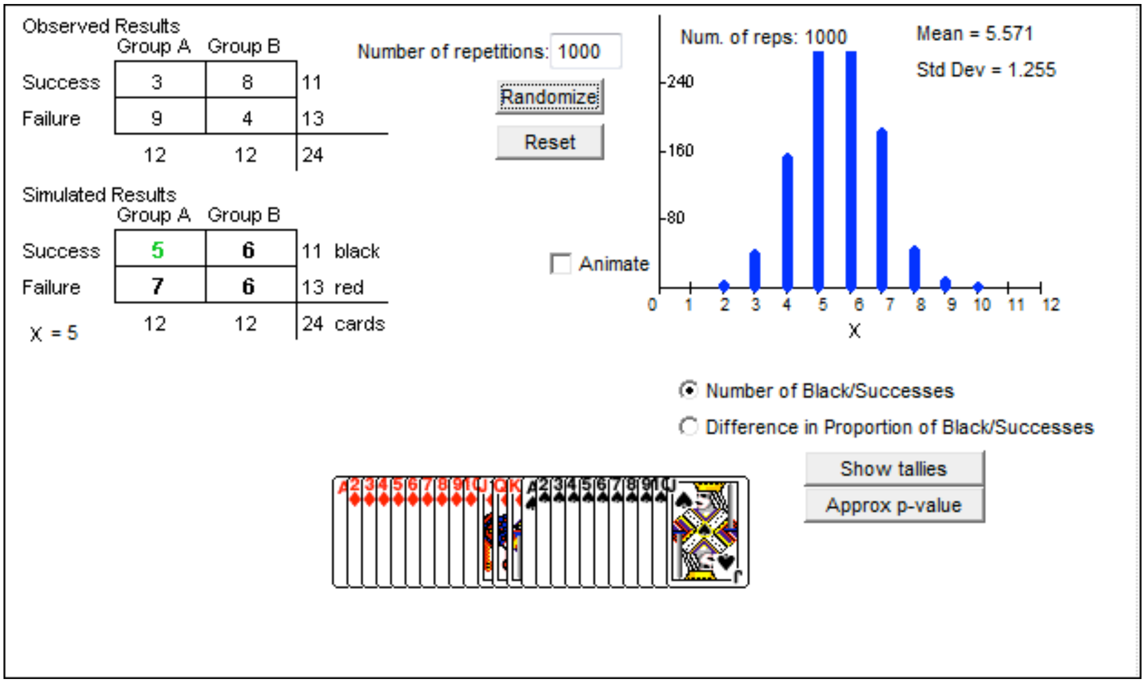

Figure 5. Friendly Observer Simulation Applet used in Statistical Significance Module

The lesson is built around the Friendly Observer Simulation applet (Figure 5) in the *Rossman/Chance Applet Collection* (Rossman and Chance 2004) and includes guided discovery questions from the accompanying text (Rossman and Chance 2006a). Students simulate an experiment with the applet. The students are able to increase the number of repetitions to investigate visually as well as numerically the concept of statistical significance and an intuitive interpretation of the p-value: how often would we get a difference this extreme (or more extreme) through the randomization process alone.

3.5 Randomization

The Randomization Module (previously shown in Figure 3) is designed to teach the students the purpose of randomization in experiments. In an experiment that requires two groups of participants, for example a control group and an experimental group, it is usually necessary that the two groups be as similar as possible. This is to ensure that any difference that is found at the end of the experiment is due to the variable of interest rather than any pre-existing difference between the groups. Dividing one large group into two smaller groups by choosing the groups randomly will result, on average, in two groups that are similar in both observable and un-observable ways.

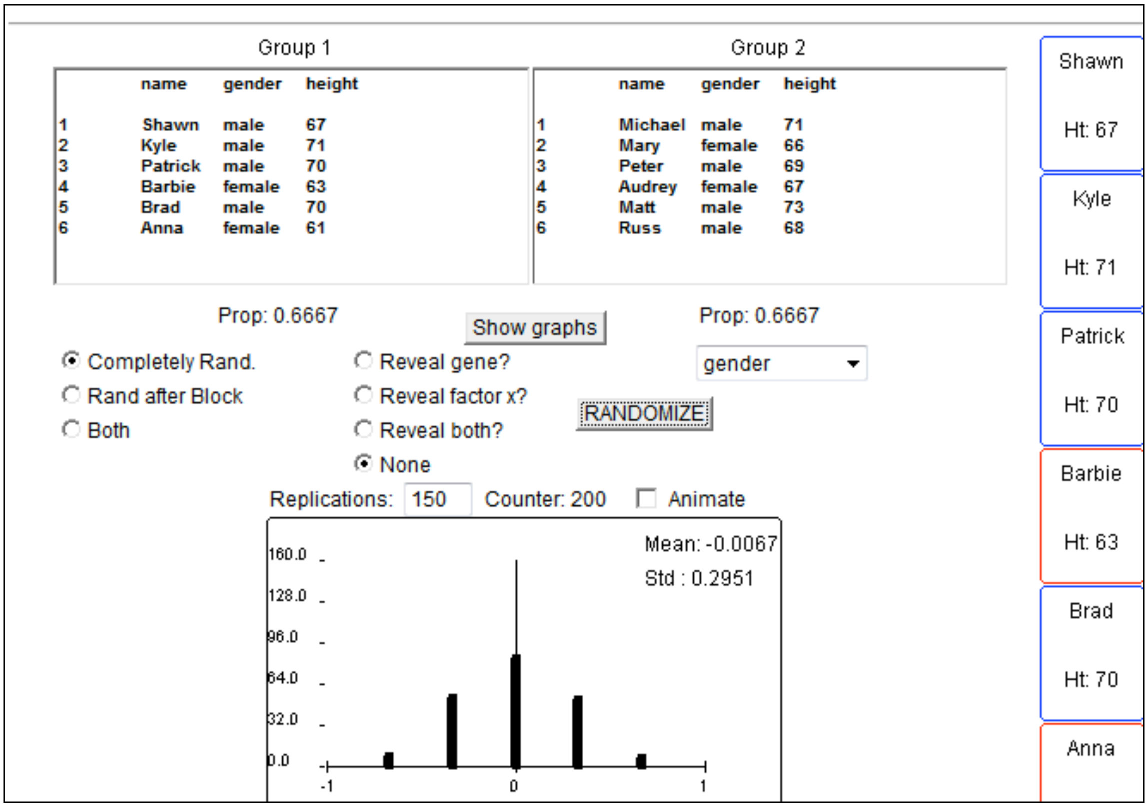

Figure 6. Randomizaton Applet

The module is built around the Randomizing Subjects applet in the *Rossman/Chance Applet Collection* (Rossman and Chance 2004) and includes guided discovery questions from the accompanying text (Rossman and Chance 2006a). In this applet (Figure 6) 12 people are randomly assigned to two groups. Then the number of males in each group is counted and the difference of those numbers is plotted in a graph. When the randomization is repeated a large number of times, this graph becomes centered at zero and is roughly symmetric. Similar graphs can be viewed for different traits such as position of a special gene and height.

3.6 Module Administration

The Binomial Distribution Module, which was generally the first one assigned, was typically given in a computer lab in order to get students familiar with the website. The other three were assigned to be done independently as homework. Other than assigning the modules, instructors did not change their teaching methods. Students took an online pretest, went through the module, then took an online posttest, except in the case of the Binomial Module for which the tests were administered by the instructor off line. No time limits were imposed on completing the module, and whether the student watched the video was not recorded. However, a de facto gateway was embedded in some of the videos making it difficult to get to the required lesson without viewing it. That is, students had to complete one part of the module before having the option to move on to subsequent parts.

The modules were administered in six different courses:

- Applied Statistics, which is a general education introductory statistics course for non-majors, and generally has 25-35 traditional-aged students ranging from freshmen to senior.
- Probability and Statistics, which is a course for science majors that has calculus as a prerequisite. Class sizes are between 20-30 traditional-aged students with a majority being sophomores.
- Probabilistic and Statistical Reasoning, which is a core requirement for the Masters of Science in Professional Science degree program that contains both statistics and non-statistics majors. Class sizes are between 15-20 graduate students in their early to mid-twenties.
- Other courses, such as Mathematical Statistics, had so few students that they were included in the overall count, but were not separated by course.

The instructors chose which modules to assign to their students, so there are differing numbers of students who tested each module. The topics covered by the modules will have also have been discussed in class. It was not recorded if students did the module before or after the class discussion. Other than assigning the modules, the instructors did not alter their teaching methods.

4. LEARNING MODULES AND RESULTS

4.1 The Binomial Distribution

In the pretest and posttest for the Binomial Module (see Figure 7), the user is presented with several graphs and several options for values of *n* and *p*. The user is asked to match values of *n* and *p* with the corresponding graphs.

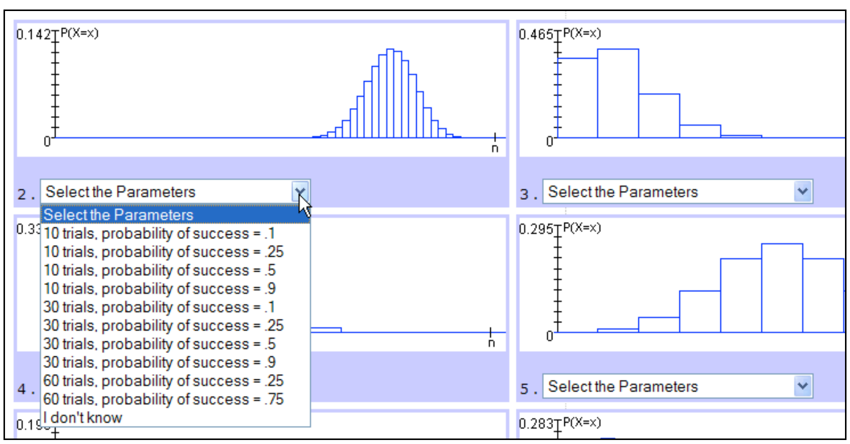

Figure 7. Pretest for Binomial Distribution Module

The posttest is given on paper to make sure that students cannot simply enter the values of *n* and *p* into the applet and answer the questions. The questions are set up as a matching problem, with ten choices of *n*-and-*p* pairs and six graphs. Therefore, responses to these questions are not independent of each other. Table 1 shows the percentage of respondents that answered the pretest and posttest questions correctly. These are the results of a total of 204 students, including students from a general-education introductory class, a calculus-based introductory class, and a mathematical statistics class.

Table 1. Binomial responses (n=204)

Question	Correct on Pretest	Correct on Posttest
n=60, p=.75	0.11	0.74
n=10, p=.1	0.30	0.70
n=10, p=.25	0.22	0.76
n=10, p=.5	0.04	0.68
n=30, p=.25	0.05	0.67
n=30, p=.1	0.10	0.63

We examined similar results broken down by class, and as may be expected, the students in the higher level classes do better than students in the general education class. For one typical question, similar proportions of students in each class missed it on the pretest, but around 83% of the students in the higher level classes responded correctly on the posttest, compared to 62% of the students in the general education class.

As mentioned above, the responses to these questions are not independent, since they are matching questions. Figure 8 shows the proportion of students with the given score on the pretest and the posttest. There were a large number of students who scored low on the pretest but scored perfect (6/6) on the posttest.

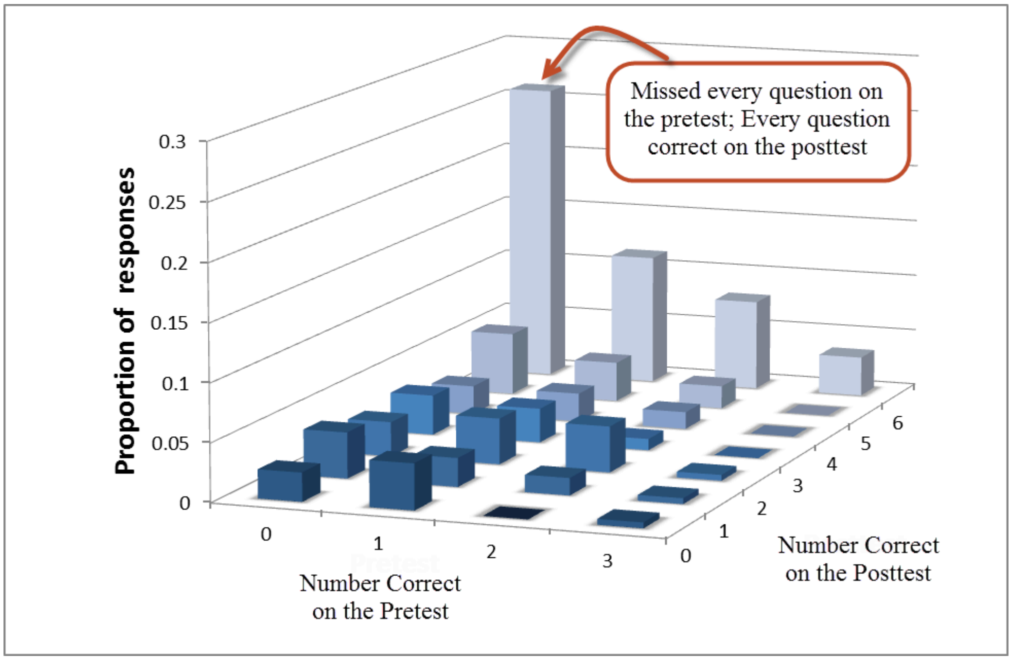

Figure 8. Binomial scores (n = 204), showing the proportion of students with the given score on the pretest and the posttest. No student scored higher than a 3 on the pretest.

4.2 Confidence Intervals

The pretest and the posttest for the confidence interval module were not multiple choice, but instead asked for descriptive answers. The pretest contains the question, “What is the difference between the z and the t one-sample confidence intervals?” The posttest contains a similar question, “Write a paragraph explaining why the t procedures are preferred to the z with s.” After we decided on a grading rubric, we graded each response separately, then met to discuss and resolve any discrepancies. The grading rubric (Figure 9) was based on a four point scale (see Lester Jr and Kroll 1991). Figure 10 gives example responses (copied verbatim) to the posttest question for this module.

4 points: Exemplary Response, All of the following characteristics must be present:

- a. The answer is correct.
- b. The explanation is clear and complete.

3 points: Good Response, *Exactly one* of the following characteristics is present:

- a. The answer is correct but there is a minor flaw in wording.
- b. The explanation lacks clarity.
- c. The explanation is incomplete.

2 points: Inadequate Response, Exactly two of the characteristics in the 3-point section are present or *one or more* of the following characteristics are present:

- a. The answer is incorrect due to a major flaw the wording but implies some understanding of the concept.
- b. Explanation lacks clarity or is incomplete but does indicate some correct and relevant reasoning.
- c. There is a partial explanation but the thought is not carried out.

1 point: Poor Response, *two* of the following characteristics must be present:

- a. The answer is incorrect.
- b. The explanation, if any, uses irrelevant arguments.
- c. The explanation just restates the problem in other words.

0 points: No Response.

- a. The student’s paper is blank or contains only wording that appears to have no relevance to the problem.

Figure 9. Grading rubric for free responses

4: The t procedures is preferred because it gives us a more accurate observed confidence level (close to the stated confidence level) than the z with s procedure; Or more of the intervals that we run using the t procedures contain the estimated population mean than that of the z with s.
3: the t intervals will be larger than z interval. that will pay for estimating the population standard deviation. t critical value adjusts the mistake.
2: The t procedures are preferred because it includes the standard deviatin of the population which leads to a more accurate finding.
1: The t procedure tends to be more accurate. It allows more information.
0: I have no idea.

Figure 10. Example responses for Confidence Intervals copied verbatim from student responses.

Table 2 and Figure 11 contain results for these questions. Comparing the number of responses given a score of 0 or 4 between the pretest and the posttest shows some improvement. The table showing the relationship between pretest and posttest shows that 45% of those who answered the first question incorrectly answered correctly on the second. The second row of the table, however, shows that of the 16 people who answered correctly for the pretest, 9 answered incorrectly on the posttest. A closer look at these 9 responses shows that the typical reason for this decrease was that the student described when to use t instead of z confidence intervals, but not why we make that decision. Since the question specifically asked about "why," the grading on this question was more strict compared to the pretest question, which only asked what the difference was. It is likely that if the pretest and posttest questions had been identical, we would have observed a higher increase in the number of people who responded correctly. It is also worth noting that fewer instructors currently cover the z with s confidence interval, skipping straight to t with s, so fewer students will have been exposed to this distinction.

Table 2. Results for Confidence Intervals Module

		Posttest Question	
		Incorrect (0 to 2)	Correct (3 or 4)
Pretest Question	Incorrect (0 to 2)	27	22
	Correct (3 or 4)	9	7

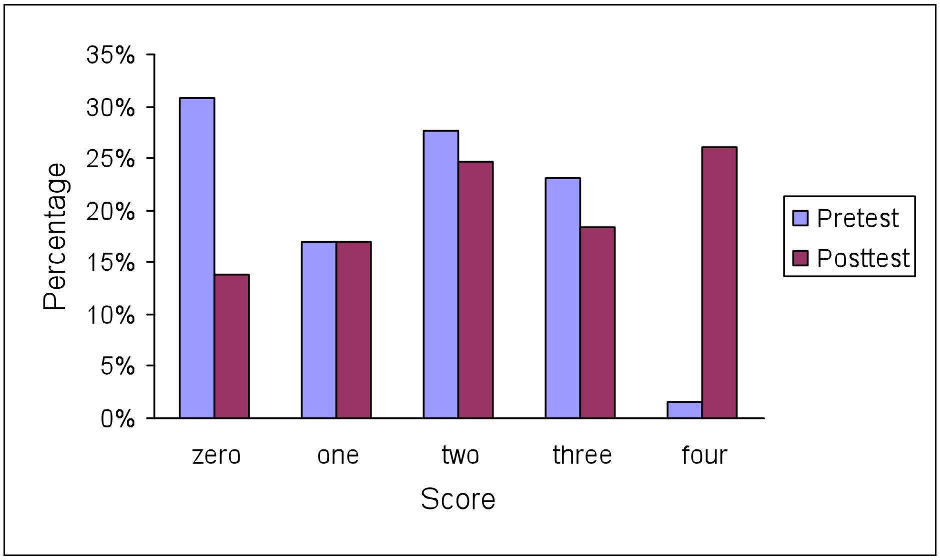

Figure 11. Results for Confidence Intervals Module. Pretest question: What is the difference between the z and the t one sample confidence intervals. Posttest question: Write a paragraph explaining why the t procedures are preferred to z with s.

4.3 Statistical Significance

The pretest for this module contains a question asking the student to define the term "p-value." The posttest contains a question that asks the student to explain their decision in a previous question which had been based on the value of the p-value. Correct answers to these questions should mention something about the probability of results occurring under the assumption that differences are due solely to chance. The grading of this module was handled the same way that the grading for the Confidence Interval Module was handled. Figure 12 contains actual sample responses to the question of defining the p-value.

4: The p-value is the probability of obtaining a result at least as extreme as the given data point. The p-value of an observed value is the probability that, given that the null hypothesis is true, it will have a value as or more unfavorable to the null hypothesis.
3. the p value provides evidence that the results do not occur by random sampling variations alone
2. P-values can indicate if there is a significant difference in values or treatments.
1. the probability that an event will occur
0. I don't know what it means.

Figure 12. Sample responses for Statistical Significance copied verbatim from student responses

The results for these questions are shown in Table 3 and Figure 13. The graphs show that quite a number of respondents increased their understanding. We also notice that there are still quite a few respondents who do not understand this term. It is not surprising that a concept such as statistical significance takes more than one lesson.

Table 3. Results for Statistical Significance Module

		Posttest Question	
		Incorrect (0 to 2)	Correct (3 or 4)
Pretest Question	Incorrect (0 to 2)	66	36
	Correct (3 or 4)	1	7

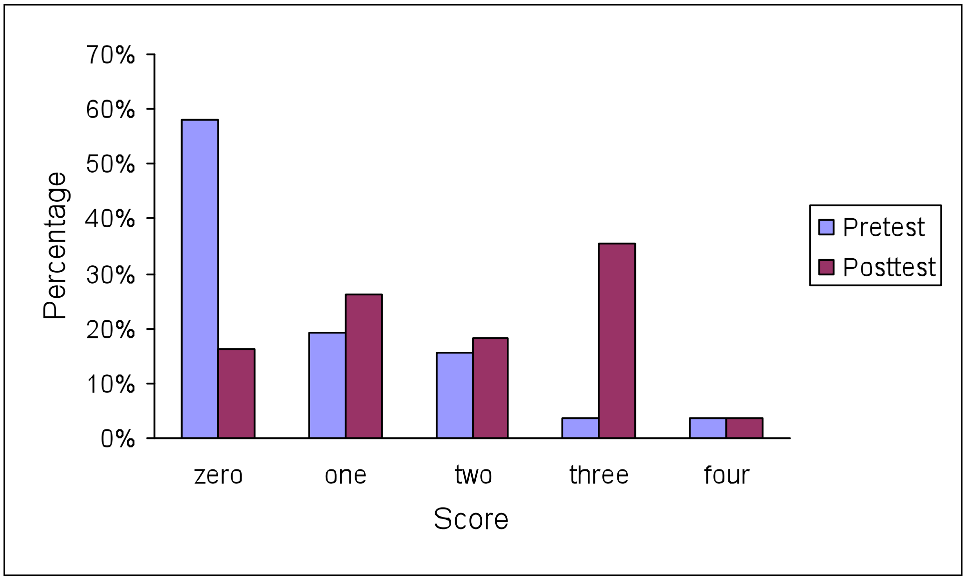

Figure 13. Results for Statistical Significance Module. Pretest question: Define p-value. Posttest question: If the p-value is around 5% or less, would you say that the difference in the two groups is likely to have occurred by some treatment other than chance? Explain your answer to [this] question.

4.4 Randomization

In the pretest an experiment describing two stumble-recovery techniques are described. Researchers wish to compare two strategies that the elderly can use to recover from tripping: elevating or lowering. Researchers divided the test subjects into two groups. One was assigned to use the lowering strategy and one was assigned to use the elevating strategy. Each group was assessed on their likelihood of falling.

The module was designed to teach why the assignment of groups is done randomly. One of the questions for this module was "Using complete sentences, describe the purpose of the randomization process used in this experiment." We paired this question with a similar one from the posttest, "Using complete sentences, describe what the process of randomization does when it is used in statistical studies." The grading of these questions was similar to the grading of the questions in the Confidence Interval Module. It was especially important for students to note that more than one trait was being affected by the randomization process. Example student responses for each point value are given in Figure 14.

4: Randomization creates groups with similar distributions and eliminates effects of potential confounding variables. It creates groups that are "balanced out" and differ only in the explanatory variable imposed. There is a tendency for the proportion of men to be "balanced out."
3: Randomization "balances out" this variable in order for a cause and effect to be determined
2: There is not a tendency for there to be a similar proportion of men between the two groups because they are radomly assorted. There are more men than there are women.
1: Most men are generally around the same height so I think that there is a tendency for them to have a similar proportion.
0: i do not know

Figure 14. Example responses for randomization question copied verbatim from student responses.

Grades for the questions on the pretest and the posttest are compared in Table 4 and Figure 15. These results show a disappointing lack of improvement. Only 21% of people who missed the pretest question answered the posttest question correctly. Similar graphs for these questions broken down by class (not shown), show similar patterns, even for the high level courses. This shows that the Randomization Module needs improvement. In examining the responses more closely, we find that students misunderstood the main terms used in the question. For example, the term “randomization” is being used to describe the process of dividing a preselected sample into two groups. But many responses referred to selecting the sample itself. The definition is included in the module, but was apparently missed by many respondents.

Based on the performance of students on this module, we would not assign this to students without a major overhaul. While we intend to structure the lesson around the stumble-recovery experiment, we feel the need to add several questions preceding the current guided-discovery questions to ensure that students understand the experiment and related vocabulary before the lesson starts.

Table 4. Results for Randomization Module

		Posttest Question	
		Incorrect (0 to 2)	Correct (3 or 4)
Pretest Question	Incorrect (0 to 2)	61	16
	Correct (3 or 4)	5	19

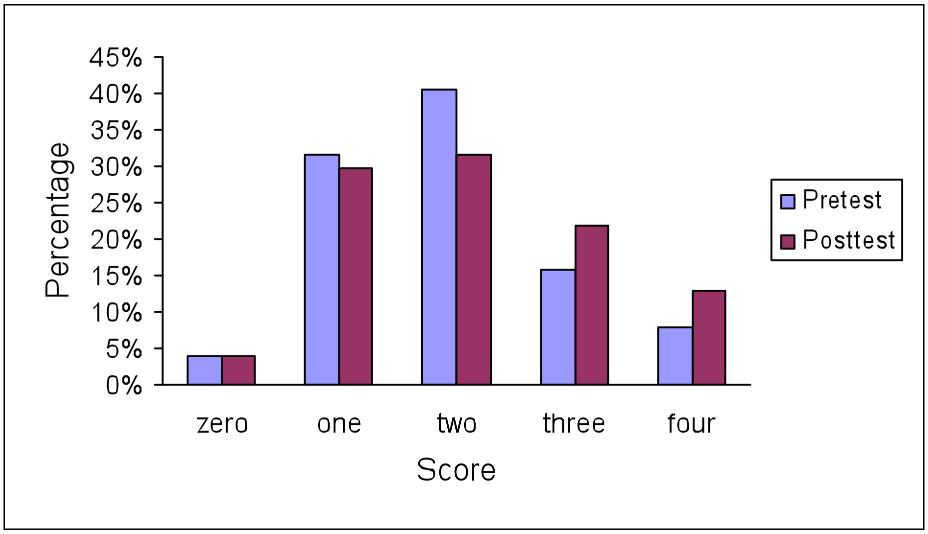

Figure 15. Results for Randomization Module. Pretest question: Using complete sentences, describe the purpose of the randomization process used in this experiment. Posttest question: Using complete sentences, describe what the process of randomization does when it is used in statistical studies.

5. FUTURE DEVELOPMENT PLANS

As mentioned above, we found that students using the Randomization Module misunderstood the meaning of the word "randomization." They tended to confuse this concept with the concept of random sampling. It is therefore necessary to redevelop the Randomization Module to be more explicit about this idea, and then to retest the new module to see if the changes improve students' learning.

While the modules were being tested, the problem of getting students' results back to the instructor of that class came up more than once. It would be useful to create a way for results to automatically be sent to the assigner of the module. We are exploring options using e-learning platforms (e.g. Moodle) to facilitate the grading task. If this does not work, we may decide to create a summary screen that students can print after completing the module.

We found during the development of these modules that, while we were designing them for students in post-calculus statistics classes, we were using them for all students, including those in introductory classes. We would like to restructure and reword some of the modules with this in mind. As a simple example, the use of the parameter name "π" for population proportion confuses the introductory students unnecessarily, so changing that symbol to a "p" seems appropriate.

Finally, we would like to develop other modules to address other statistical concepts that are standard in college level statistics, such as Type I and II error in hypothesis testing, statistical power, and the correct interpretation of confidence intervals.

6. CONCLUSION AND DISCUSSION

Research has shown that using online simulations can improve student understanding, but that students must be guided in the use of these simulations. The modules described in this paper were created to address this issue by providing guidance in an online form. They combine a simulation with audio-visual instruction in its use and with questions that guide the students to construct meaning in the topic at hand. These modules cover important ideas in statistics, such as the binomial distribution, confidence intervals, randomization, and statistical significance.

Pretests and posttests for each module show that this format can be used independently by students. We found that the number of students answering correctly on posttests was larger than that for pretests, for three of the four modules described in this paper. The lack of improvement on one module demonstrates that the format alone is not enough for student learning, that the questions and audio track should be carefully constructed using sound pedagogical techniques. It was also shown, not surprisingly, that the previous preparation of the students affected how well they learned the concepts in the modules.

REFERENCES

Aberson, C. L., Berger, D. E., Healy, M. R., and Romero, V. L. (2002), "An Interactive Tutorial for Teaching Statistical Power," *Journal of Statistics Education*, 10.

Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, B., and Utts, J. (2005), "Guidelines for Assessment and Instruction in Statistics Education (GAISE) Project College Report."

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.

Clark, D., and Jorde, D. (2004), "Helping Students Revise Disruptive Experientially Supported Ideas About Thermodynamics: Computer Visualizations and Tactile Models," *Journal of Research in Science Teaching*, 41, 1-23.

Clark, R. C., and Mayer, R. E. (2008), *E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*, San Francisco: Pfeiffer.

Cobb, P. (1994), "Where Is the Mind? Constructivist and Sociocultural Perspectives on Mathematical Development," *Educational Researcher*, 23, 13.

Collins, L., and Mittag, K. (2005), "Effect of Calculator Technology on Student Achievement in Introductory Statistics Course," *Statistics Education Research Journal*, 4, 7-15.

de Jong, T., and van Joolingen, W. (1998), "Scientific Discovery Learning with Computer Simulations of Conceptual Domains," *Review of Educational Research*, 68, 179-201.

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.

Edelson, D. C., Gordin, D. N., and Pea, R. D. (1999), "Addressing the Challenges of Inquiry-Based Learning through Technology and Curriculum Design," *Journal of the Learning Sciences*, 8, 391-450.

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. (2009), "Helping Students Develop Statistical Reasoning: Implementing a Statistical Reasoning Learning Environment," *Teaching Statistics*, 31, 72-77.

Gordon, F., and Gordon, S. (1992), *Statistics for the Twenty-First Century*, Mathematical Association of America.

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.

Haury, D. L. (1993), "Teaching Science through Inquiry (Report No. Edo-Se-93-4)," *Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.(ERIC Document Reproduction Service No. ED 359 048)*.

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.

Healy, L. (2006), "Constructing Simulations to Express Developing Statistical Knowledge," in *Proceedings of the Seventh International Conference on Teaching Statistics*, eds. A. Rossman and B. Chance, Voorburg, The Netherlands: International Statitics Institute.

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K., Human, P., Murray, H., Olivier, A., and Wearne, D. (1996), "Problem Solving as a Basis for Reform in Curriculumand Instruction: The Case of Mathematics.," *Educational Researcher*, 25, 12-21.

Hirsch, C. (1977), "The Effects of Guided Discovery and Individualized Instructional Packages on Initial Learning, Transfer, and Retention in Second-Year Algebra," *Journal for Research in Mathematics Education*, 8, 359-368.

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

Lane, D., and Peres, 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 Statitics Institute.

Lester Jr, F., and Kroll, D. (1991), "Implementing the Standards. Evaluation: A New Vision," *Mathematics Teacher*, 84, 276-284.

Marasinghe, M. G., Meeker, W. Q., Cook, D., and Shin, T. S. (1996), "Using Graphics and Simulation to Teach Statistical Concepts," *The American statistician*, 50, 342-351.

Mayer, R. E. (2004), "Should There Be a Three-Strikes Rule against Pure Discovery Learning?," *American Psychologist*, 59, 14-19.

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.

Moore, D. (1992), "Teaching Statistics as a Respectable Subject," *Statistics for the Twenty-First Century*, 14-25.

Moore, D. (1997), "New Pedagogy and New Content: The Case of Statistics," *International Statistical Review*, 65, 123-137.

Moreno, R., and Mayer, R. (2005), "Role of Guidance, Reflection, and Interactivity in an Agent-Based Multimedia Game," *Journal of Educational Psychology*, 97, 117-128.

Rieber, L. P., and Parmley, M. W. (1995), "To Teach or Not to Teach? Comparing the Use of Computer-Based Simulations in Deductive Versus Inductive Approaches to Learning with Adults in Science," *Journal of Educational Computing Research*, 14, 359-374.

Rieber, L. P., Tzeng, S.-C., and Tribble, K. (2004), "Discovery Learning, Representation, and Explanation within a Computer-Based Simulation: Finding the Right Mix," *Learning and Instruction*, 14, 307-323.

The Rossmann/Chance Applet Collection. http://www.rossmanchance.com/applets/index.html

Rossman, A., and Chance, B. (2006a), *Investigating Statistical Concepts, Applications, and Methods*, Pacific Grove, CA: Duxbury Press.

Rossman, A., and Chance, B. (2006b), "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.

Searson, R., and Dunn, R. (2001), "The Learning-Style Teaching Model," *Science and Children*, 38, 22-26.

The Probability and Statistics Object Library. http://www.math.uah.edu/stat/objects/experiments /index.xml

Uzuntiryaki, E., Bilgin, I., and Geban, O. (2003), "The Effect of Learning Styles on High School Students' Achievement and Attitudes in Chemistry," in *Annual Meeting of the National Association for Research in Science Teaching*, Philadelphia (ERIC Document Reproduction Service No. ED475483).

Vockell, E., and Rivers, R. H. (1984), "Computerized Science Simulations Stimulus to Generalized Problem Solving Capabilities," in *Annual Convention of the American Education Research Association*, New Orleans (ERIC Document Reproduction No. ED253397): ERIC Document Reproduction No. ED253397.

von Glasersfeld, E. (1987), "Learning as a Constructivist Activity," in *Problems of Representation in the Teaching and Learning of Mathematics*, ed. C. Tanner, Hillsdale, NJ: Lawerence Erlbaum Associates, pp. 3-18.