

**UCLA**

**Technology Innovations in Statistics Education**

**Title**

Innovative Activities: How Clickers can Facilitate the Use of Simulations in Large Lecture Classes

**Permalink**

<https://escholarship.org/uc/item/1jg0274b>

**Journal**

Technology Innovations in Statistics Education, 5(1)

**Author**

Kaplan, Jennifer J

**Publication Date**

2011-06-14

**DOI**

10.5070/T551000045

**Copyright Information**

Copyright 2011 by the author(s). All rights reserved unless otherwise indicated. Contact the author(s) for any necessary permissions. Learn more at <https://escholarship.org/terms>

Peer reviewed

# **Innovative Activities: How Clickers can Facilitate the Use of Simulations in Large Lecture Classes**

## **1. INTRODUCTION**

In 2006, the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Project, funded and endorsed by the American Statistical Association (ASA), produced a report about the current status of and recommended directions for introductory statistics courses at the undergraduate level. Central to the recommendations for teaching introductory statistics made by the GAISE committee were the following: foster active learning in the classroom, stress conceptual understanding rather than mere knowledge of procedures, use assessment to improve and evaluate student learning, and use real data (GAISE, 2006). At many tertiary institutions, however, the introductory statistics classes are taught in large lecture format. At Michigan State University (MSU), where the author held an academic position, such courses have enrollments of 120 – 330 students per lecture. The 2005 – 2006 total enrollment at MSU in introductory statistics courses, those that do not have a statistical prerequisite, was over 4000 students, or more than 11% of the undergraduate student body. Given these large numbers, it is unlikely that the large lecture format of these courses will change. This format makes it difficult to foster active learning and conceptual understanding and to use formative assessment efficiently to improve student learning.

Personal Response Systems (clickers) are becoming ubiquitous on campuses, particularly in large, lecture-based courses. Clickers are hand-held transmitters that allow students to respond to questions in class, with their responses recorded on the instructor's computer. These systems allow instructors to move away from didactic lecture formats towards more active learning strategies that encourage student participation and are consistent with research on active learning. Perhaps one of the best-known advocates for this approach is Eric Mazur, who uses clickers as part of his peer-instruction model in teaching conceptual physics (Mazur, 1996). Clickers have also been found to help improve student outcomes in large-enrollment engineering (Sticklen & Urban-Lurain, 2006; Urban-Lurain, Sticklen, & Buch, 2006), and biology (Zhang et al., 2005) classes. A more comprehensive review of the literature on clickers, including the limitations of previous clicker research, has been recently published in this Journal (McGowan & Gunderson, 2010) and is not repeated here.

McGowan and Gunderson's findings lead them to conclude that "as with any new technology...clickers may not be successful if they are not used in a well-planned, purposeful manner" (2010, pg 29). In response to McGowan and Gunderson (2010), this paper presents a technology case study of a purposeful implementation of clicker use in statistics. It describes the implementation of activities that marry a simulation approach to teaching statistics with clicker technology for collecting results from a large number of students. The activities were designed with the purpose of improving conceptual understanding of statistical inference and addressing the recommendations of the GAISE report in a large-lecture, introductory statistics class. The author's initial motivation for using clickers in large lecture classes was simply to foster active learning and better student-teacher interactions, two things that were lacking in her previous attempts at teaching in large lecture format. She has found, however, that the combination of

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/> or send a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

large numbers of students generating random distributions with their calculators and then reporting them using their clickers, provides not only an active learning environment, but also allows students to experience statistical concepts such as distributions, variability, the Central Limit Theorem, and the conceptual underpinnings of inference in ways that they cannot experience without these technologies. In many respects, the large class becomes a learning asset, rather than a liability, that can be leveraged to target student conceptual understanding of statistical inference.

## 2. IMPLEMENTATION

### 2.1 The course and students

Introductory Statistical Methods offered by the Department of Statistics and Probability at MSU is a 3-credit algebra-based introduction to statistics course. It is a service course for non-majors and the “catch-all” course for students since the department also offers introductory courses specifically for science majors, business majors, etc. The largest major represented is pre-nursing, but there are also a number of criminal justice, journalism, communications, and psychology majors in the course. The course fulfills the mathematics requirement at MSU. There are 120 students in each lecture. The lectures taught by the author typically consist of one-third freshman and one-quarter to one-third sophomores. The remainder of the students are upper classmen, mostly juniors with a handful of seniors.

The course textbook is *Intro Stats*, by DeVeaux, Velleman and Bock. The material covered is, in this order, data collection (surveys, studies and simulations), describing data associated with one- and two-categorical variables and one quantitative variable, probability models (discrete random variables, normal, binomial and geometric models), sampling distributions, inference for 1- and 2-proportions and 1- and 2-means, and describing bivariate data (but not inference). The semester typically ends with a brief introduction to Chi-square tests.

The professor (author) meets the students for lecture 3 times a week for 50 minutes. In addition, the students meet a TA in recitation section one day a week for 50 minutes. There are four recitation sections of about 30 students each. For assessment, there are 3 midterms and a final exam. Homework is collected weekly and graded for correctness. There are 525 total points for the semester. Homework: 100 points (19%), each Midterm: 75 points (14%), Final Exam: 150 points (28%), Clickers: 50 points (10%).

### 2.2 The clicker system and calculators

The clicker system used is the *i>clicker* first generation, which is a five button model allowing for multiple choice questions with up to five possible answers. The choice of a non-numeric input clicker by a statistics instructor may come as a surprise to the reader, but the reason for the choice along with the limitations encountered as a result are discussed in the limitations section of this paper. Students are required to purchase *i>clickers*, which cost roughly \$35 new. They may share *i>clickers* with other students (not in the same class) and use the *i>clicker* in as many classes as required without paying additional fees. While students register their *i>clickers* online,

this is only to synch the registrations with a course roster on a computer local to the university. No student assessment data is held on a remote server. At the end of the semester, students who no longer need their *i>clickers* may resell the units. Currently, used *i>clickers* available on ebay are listed at over \$20 per unit.

The clickers are used in lecture, but not in recitation. This course does not include the use of computer technology or computer labs. The activities described below are designed with the use of a TI-83 or -84 graphing calculator in mind. While the students are not required to have a graphing calculator, informal polling at the beginning of the semester in many sections has indicated that nearly all of the students already own a graphing calculator. The course is therefore taught under the assumption that all of the students have access to a graphing calculator and are bringing it with them to class.

### 2.3 Clicker Implementation Scheme

Classroom implementation of the activities has been done using both a tablet and a standard computer to run slide presentations. When using a standard computer, the slides are supplemented using either a document camera or overhead projector. The clicker questions appear on the slides. The *i>clicker* software is designed to take a screen shot of the computer running the software at the start of the polling and collects from each clicker the first and last response (recording the last response for grading purposes) as well as the time the responses were given and the number of times a student changed a response. The *i>clicker* base station displays a running tally of student responses and the instructor can display a bar chart of student responses on the computer screen during or after polling.

Clickers are used every day in class and a medium-stakes scheme is used for awarding clicker points. Clicker points, up to a possible total of 50 or 9.5% of the points available over the semester, are awarded based on the proportion of days on which the student participated in clicker questions. The distribution of the number of clicker questions per class is bimodal. “Low clicker” days tend to have 2 to 5 questions and “high clicker” days tend to have 9 to 12 questions. There tend to be a roughly equal number of “high clicker” and “low clicker” days. For a more detailed description of the author’s clicker implementation see Kaplan & Urban-Lurain (2008).

## 3. THE ACTIVITIES

### 3.1 Design of the Activities

The author has developed 12 activities for the introductory statistics class in which students perform simulations the results of which are collected via clickers. These activities are listed in Table 1. As a suite, the activities are designed to aid student development of conceptual understanding of inference. The culminating idea of most currently taught introductory statistics courses is inference: drawing conclusions about a population from sample data. There is anecdotal evidence to suggest that students can spontaneously make decisions about a population that is consistent with sample data, hence exhibiting intuitive understanding of the reasoning that

underlies a hypothesis test (personal communication, Roxy Peck, 2004). These same students, however, struggle to master the formal reasoning and enactment of a hypothesis test. In fact,

Table 1: Simulation and Clicker Activities

	Activity Name	Course Topic	Conceptual Goals for Students
Weeks 1 and 2	1. Gettysburg Address	Sampling Bias	To understand that random sampling provides unbiased estimates when compared to judgment sampling.
	2. 2008 New Hampshire Primary	Sampling Variability	To understand that values from samples vary and that sampling variability decreases as sample size increases.
	3. Cereal Toys	Simulations	To understand how simulations can be used to answer probabilistic questions
	4. Magic Lottery	Simulations	To confront the myth of the unusual event
Weeks 5 and 6	5. Tiger Woods Card (1)	Probability Models	Provide an example of the geometric model
	6. Tiger Woods Card (2)	Probability Models	Provide an example of the binomial model
	7. Rolling a Die	Probability Models	Provide an example of the uniform model
	8. Rolling 10 Dice	Sampling Distributions	Provide an introduction to the Central Limit Theorem
	9. Basketball Shooting	Sampling Distributions	Provide an introduction to the normal approximation of the binomial distribution
Weeks 7 - 9	10. Coin Flipping	Confidence Intervals	Show that level of confidence is the approximate percent of samples that produce CIs that cover the true value regardless of sample size.
	11. Cell Phone Drivers (1)	Hypothesis Testing	Provide an example of the reasoning behind hypothesis testing and the meaning of the p-value of a test.
	12. Cell Phone Drivers (2)	Hypothesis Testing	Show how Type I and II errors occur based on random sampling and how power and the probability of errors can be estimated.

Brewer claims that the area of inference is “the most misunderstood, confused, and abused of all ... statistics topics” (1985, pg. 255). In a review of the literature, Lane-Getaz (2005) identified thirteen types of misconceptions of the meaning of the p-value calculated in classical hypothesis testing. Empirically, she found these misconceptions to be held by students at all levels of study. Moreover, she found that these misconceptions were persistent; many of them were held by a sample of doctoral students taking a second graduate level statistics course. Literature in psychology has noted similar issues arising for both students and researchers. There is evidence that researchers and students have difficulty understanding both statistical significance and confidence intervals (see for example, Belia, Fidler, Williams & Cumming (2005), Haller & Krauss (2002) and Wilkerson & Olson (1997)). Noll claims “that an understanding of sampling concepts and processes is necessary for developing a robust understanding of statistical

inference” (2007, pg, 9). The set of activities outlined in Table 1 was designed with Noll’s claim in mind, beginning with activities that support understanding of sampling variability and the use of simulations, building on these foundations through probability distributions, sampling distributions toward the concepts that underlie inference, both hypothesis testing and confidence intervals.

The author begins her course with a unit on data collection so the first four activities, Gettysburg Address, 2008 New Hampshire Primary, Cereal Toys, and Magic Lottery are completed in the first two weeks of the semester. Gettysburg Address, described in more detail below, is designed to highlight the value of the use of random sampling and lay a foundation for understanding sampling variability. 2008 New Hampshire Primary specifically targets understanding sampling variability and the relationship between sample size and measures of variability of the sampling distribution. Cereal Toys and Magic Lottery are designed to help students understand the value of simulation results in learning about probabilistic situations. Completion of these activities, as well as a simulation of a randomization test done by hand in the first recitation section of the semester, are designed to prepare the students for continued use of simulation as a basis for understanding statistical inference later in the course.

In between the unit on data collection and the units on probability models and sampling distributions, the class completes a two week unit on one-variable descriptive statistics. During this time, clickers are used daily, but no simulation activities occur. Activities 5 - 7, Tiger Woods Card (1) and (2) and Rolling a Die, are done in one or two class periods after students have learned to find the expected value and standard deviation of a discrete random variable and are designed to help students understand the ideas behind probability models. This foundation is followed by the two sampling distribution activities 8: Rolling 10 Dice and 9: Basketball Shooting, which provide a foundation for understanding the theory underlying inference for proportions and The Central Limit Theorem, basically sampling distributions. These two sets of activities, probability models and sampling distributions are enacted during the fifth and/or sixth week of the 15-week semester.

Activity 10, Coin Flipping, is done after students have learned to create a confidence interval for one proportion (week 7 or 8) and is designed to show students the meaning of level of confidence and the interaction between sample size, margin of error and level of confidence. Cell Phone Drivers (1) is used in week 8 to introduce the reasoning behind hypothesis testing and the meaning of a p-value. It uses simulation, rather than calculation, to compute a p-value and is used prior to instruction on calculation of p-values from the normal approximation to the binomial distribution. Finally, Cell Phone Drivers (2) (week 9) is designed to help students understand that Type I and II errors are not mistakes per se and to connect the probability of making such errors to the notion of power and sample size in much the same way that activity 10, Coin Flipping, connects concepts about confidence intervals. The remainder of this section will describe in detail the Gettysburg Address activity and the Cell Phone Drivers (1) activity, but the author is happy to share the slides for all of the activities with interested readers.

### 3.2 Gettysburg Address

The Gettysburg Address activity used by the author is isomorphic to the Random Rectangles activity appearing in the text, Activity Based Statistics (Schaefer, Gnanadesikan, Watkins & Witmer, 1996), which has been used by Advanced Placement statistics teachers to introduce concepts associated with sampling and was developed independently of the version created by Rossman & Chance (available at <http://statweb.calpoly.edu/chance/inspire03/handouts.html>) In this version of the Gettysburg Address activity, students are given a copy of the Gettysburg Address and asked to estimate, just by looking at the Address, the average length of the words in the Address. The students are then asked to select a “reasonable” sample of 10 words and find the average length of the 10 words in their sample. Finally, the students are asked to select a random sample of 10 words, using the random integer function on their calculators, and find the average length of the 10 words in the random sample. To facilitate this part of the task, the copies of the Address have been annotated with word numbers at the end of each line. In the original Random Rectangles activity (Schaefer, et al., 1996) the population is a collection of 100 rectangles and the parameter of interest is the average area of the 100 rectangles. The slides and results for the first 3 parts of the Gettysburg Address activity are given in Figures 1a-c. When the bar graphs are compared, students are urged to notice the averages generated by the judgment samples are higher than the averages generated by the random samples.

Figure 1a

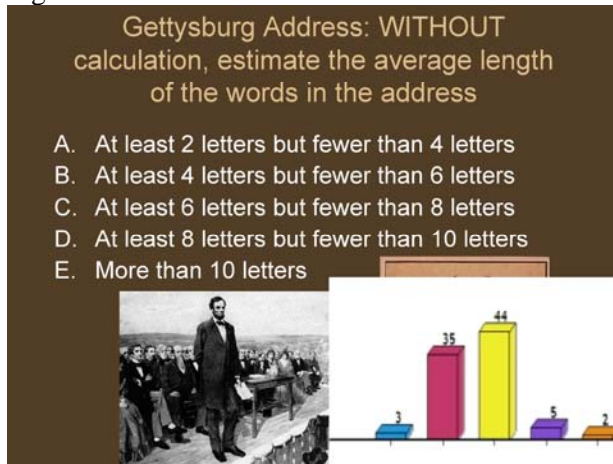


Figure 1b

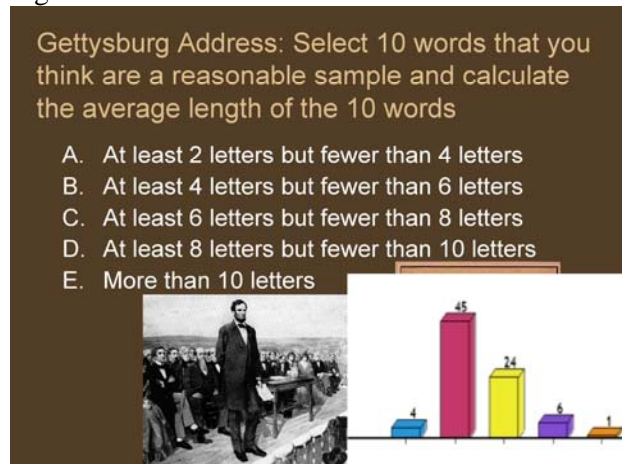


Figure 1c

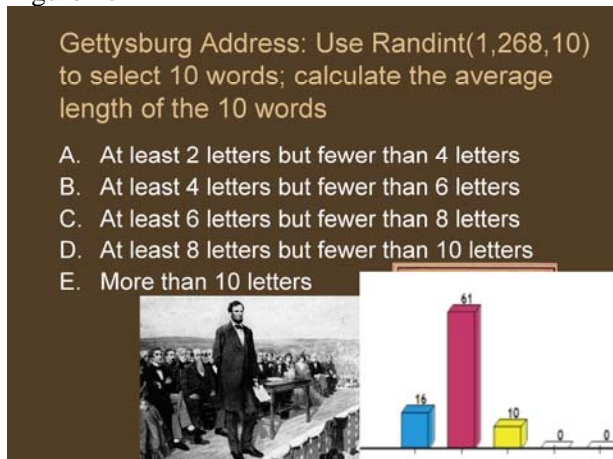
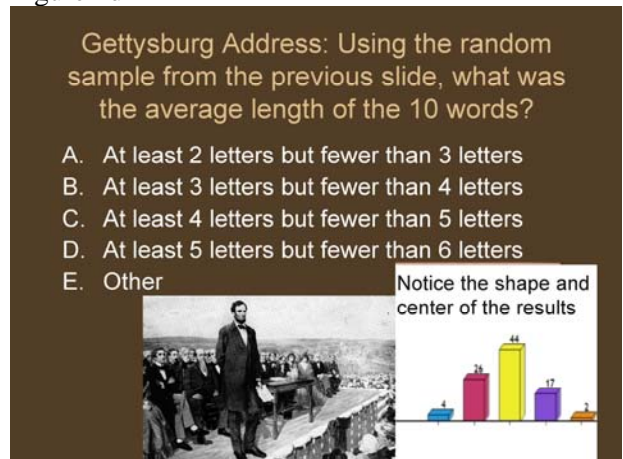


Figure 1d



The students are asked a fourth question, shown with the results in Figure 1d, with smaller bin widths than the previous three questions. This is the first experience that students have with results predicted by the Central Limit Theorem, though this is not mentioned at the time. What is discussed is that the actual (“true”) average word length of the Address is 4.3 words. Students are then asked to comment on how the results of the three methods compare. Through the wrap up discussion the instructor ensures that the following main points are made (and they are reiterated in a slide later in the presentation): (1) The true value of the parameter is estimated reasonably well by both the original guesses and the random samples, though (2) the random samples provide the estimate with less uncertainty and (3) the true value is over estimated by the judgment samples. Before the end of the discussion the instructor explains that the bias in the judgment samples occurs because our eyes are drawn to longer words and we miss the one-letter words such as “a” and “I” (and in Random Rectangles by the ease in which our eyes scan over the 1x1 rectangles in favor of the larger shapes).

This activity is an example of how activities that are already available, but designed for small classes, can be adapted using clickers for large lectures. In fact, the large lecture becomes an asset because in a short span of time a large number of samples can be drawn and the data collected and viewed by the entire class. The entire activity takes roughly 20 minutes of class time and lays the foundation for conceptual understanding of the use of random sampling, sampling variability and sampling distributions, which are specifically the concepts that Noll (2007) claims are necessary for developing a robust understanding of statistical inference.

### 3.3 Cell Phone Drivers (1)

The Cell Phone Drivers (1) activity is designed to highlight the correct meaning of a p-value and the reasoning behind conclusions made from a hypothesis test, concepts that Brewer (1985) and Lane-Getaz (2005) claim to be misunderstood, with the misunderstandings being highly resistant to change.

**Scenario for Cell Phone Driver (1) activity:** A proud legislator claims that your state’s new law against talking on a cell phone while driving has reduced cell phone use to less than 12% of all drivers. While waiting for the bus the next morning, you notice that 4 of the 10 people who drive by are using their cell phones. Does this cast doubt on the legislator’s figure of 12%? Use a simulation to estimate the likelihood of seeing at least 4 out of 10 randomly selected drivers talking on their cell phones if the actual rate of usage is 12% (DeVeaux, Velleman & Bock, 2006, pg. 262).

Students complete the above simulation as a homework assignment in week 3 of the course, 4 – 5 weeks prior to the enactment of the class activity. The activity begins with a reminder of the homework problems and the results of the simulation (Figure 2a) and then an explanation of how we would use the simulation results to make a conclusion about the congressman’s claim (Figure 2b). Students are then encouraged to think about sampling distributions of proportions, which were used the previous week to create confidence intervals for one proportion, and are asked whether the sample of 10 drivers is large enough to use the theory of sampling distributions. Once the students realize that the sample is not large enough, it is pointed out that under the theory, the sampling distribution should be unimodal and roughly symmetric and that the reason



the simulation results show right skew is precisely because the sample is not large enough for the theory to apply.

Figure 2a: Cell Phone Driver Scenario

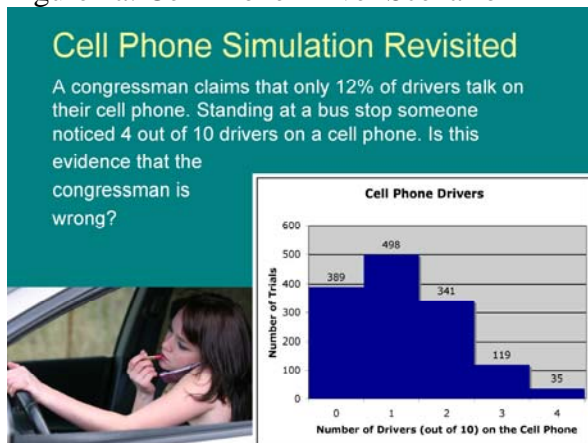
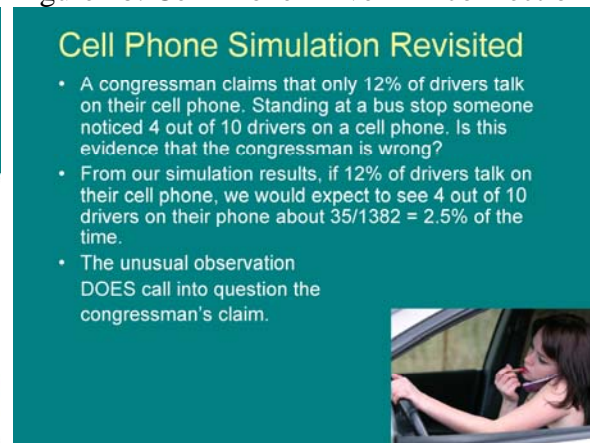


Figure 2b: Cell Phone Driver HT connection



After students agree that a sample of size 100 is large enough the students are asked for a gut reaction to the question “If the congressman is correct that only 12% of drivers talk on their cell phone, how many drivers out of 100 would have to be talking on their cell phones for you to think it was an unusually high number?” The purpose of this type of “before” question, asked prior to the enactment of the simulation, is to engage students more fully in the activity and monitor conceptions and changes to their conceptions. This is a design feature of many of the activities listed in Table 1. These questions were included based on the claim of Chance, delMas and Garfield that simulation activities that follow a “predict/test/evaluate model force students to more directly confront the misconceptions in their understanding” (2004, pg. 299) and their subsequent finding that students who enacted activities using such a model showed statistically significant improvements in posttest performance. The responses to these questions are recorded by the clicker software and are sometimes displayed, but generally there is no discussion of the results of the “before” questions until after the activity has been completed. At the end of most activities, the same question is asked again and the results from the “before” question are compared to results of the “after” question. This comparison is used as the basis of the wrap up discussion of the activity in order to maximize the chance for students to confront their misconceptions or preconceived ideas.

Some readers may have chosen to word this “before” question as, “about what percent of the drivers would have to be talking on their cell phones...” rather than “how many drivers out of 100 would have to be talking on their cell phones...” The question is worded this way, and, in fact, all of the simulation results are collected as counts rather than percentages, by specific design. There is evidence in the psychology literature on human reasoning to suggest that people can more easily estimate and solve problems when data are given as counts rather than percentages (for a summary of the literature see Kaplan & Du, 2009). While students may be trained to use representations to overcome this limitation (Kaplan & Du, 2009, Sedlmeier, 1999), the author decided to use the format that is more natural to the students. In addition, collecting results in percent format creates one more step in the process on which students can make errors. Such errors will be discussed in the Limitations section of this paper.

To simulate 100 drivers under the condition that 12% will be talking on their cell phone students are given the following directions:

- Assume that population proportion is 12%
- To simulate 100 drivers use:
  - `randint(1, 100, 100)`
- Numbers 1 - 12 are drivers on their cell phone, 13 - 100 are drivers not on their cell phone
- Count the number of drivers in your sample who are on their cell phone
- To make the counting easier, store the simulation results in a list and then sort the list

The command `randint(1, 100, 100)` returns 100 random integers between 1 and 100, inclusive, with replacement (so there will almost certainly be repeated values). By using this command, each student generates one sample of 100 drivers. Sorting the list makes the results easier to record because all of “drivers” who are “talking on their cell phones” (represented by any integers selected below 13) are at the top of the list.

Figure 3a: Cell Phone Driver Simulation Results

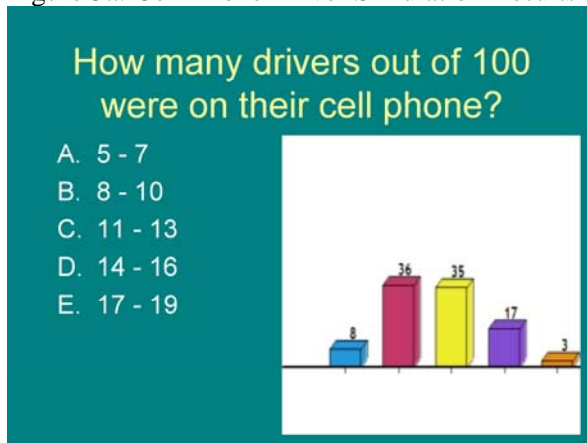
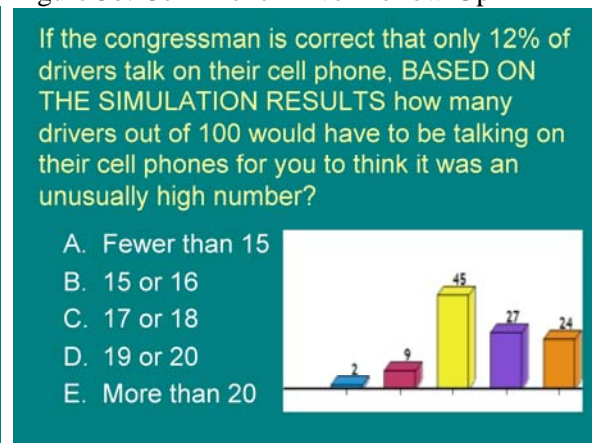


Figure 3b: Cell Phone Driver Follow-Up



The simulation results are collected via clickers using the slide in shown Figure 3a, which also shows the results of the activity. Students are asked, based on the simulation results, how many drivers out of 100 would be an unusually high number if the congressman is correct about the 12% figure (Figure 3b). Table 2 gives the p-values associated with each of the values from 14 to 21 drivers on their cell phones in the sample along with the percent of students who chose each response both before and after the completion of the activity. Notice that the category associated with 17 or 18 drivers is the smallest that would be significant at  $\alpha = 0.05$  and that only 12% of students chose numbers smaller than 17 prior to completing the activity (and this drops to 10% after the activity). While 47% of students continue to choose numbers that represent p-values less than 0.01 after the activity, this is smaller than the 55% who chose those values prior to completing the activity and more students have chosen the category that corresponds to a reasonable p-value for rejecting the null hypothesis. This lack of movement in students' selections is not considered a deficiency of the activity, rather, it is considered to be an illustration of students' definition of unlikely. While statisticians consider something that

happens less than 5% of the time to be unlikely, students may hold conceptions that things are only unlikely if they happen very rarely, less than 1% or even 0.1% of the time.

Another possible explanation for the discrepancy, and a limitation of this version of the activity, stems from the wording of the question. Students may think that 19 or 20 would be unusually high, but more than 20 would clearly be even more unusual. In future implementations the question might be reworded to ask for the smallest number of drivers observed to be talking on a cell phone that would be considered unusually high. A second limitation that may contribute to the results is the lack of refinement in the categories used to collect the simulation results. This limitation can be easily eliminated in the future with the advent of numeric input clicker systems that correctly display quantitative data using histograms. This will be discussed in more detail in the Limitations section of this paper.

Table 2: Results of Cell Phone Driver Activity

Number of Cell Phone Drivers in Sample		P-value		Students Choice (pre)	Students Choice (post)
14		0.2691		4 (4%)	2 (2%)
15	16	0.1780	0.1091	8 (8%)	9 (8%)
17	18	0.0619	0.0324	35 (33%)	45 (42%)
19	20	0.0156	0.0069	34 (32%)	27 (25%)
21		0.0028		24 (23%)	24 (22%)

## 4. LIMITATIONS

Recall that this article presents a case study of an innovative implementation of technology in the classroom rather than a research study of effects the implementation of student learning. Such a research study is certainly a logical next step and will be discussed in the next section, future directions. This section on limitations will focus on the two main limitations of or issues with enacting the activities in class: the use of a categorical clicker system and student errors in completing the simulations

Prior to implementing clickers in the classroom, the author attended meetings and presentations on campus and listened to early adopters of clicker technology before choosing a clicker system. Initially, she expected to use a numeric input clicker. There were two main reasons behind the decision not to use a numeric input system. The lesser of the two reasons was the reported ease of implementation of the *i>clicker* system coupled with the technical support resources available on campus and through the company for the system. The more important factor behind the decision was that all clicker systems available at the time, whether numeric or categorical, displayed the student responses as bar charts, rather than histograms. None of the software allowed for binning of responses and, in fact, a student who responded 14.0 would show in a different bar from a student who responded 14 to the same question. This limitation, including the one specifically discussed in the Cell Phone Drivers (1) activity, can now be addressed because the newest generation of *i>clicker* system allows the user to create a histogram of the results.

In the iteration of the activities described above, enacted prior to having histogram capabilities associated with clicker systems, the bins were created using the sampling or probability distribution for the particular situation being simulated. In some situations, the use of multiple slides cannot be avoided. For example, in the Cereal Toys activities students simulate buying boxes of cereal until 6 different toys, which are uniformly distributed in boxes, are all collected. It is not unusual to have trials in which twenty or more boxes must be “purchased” to collect all the toys. In order to collect the raw data for this activity, the first collection slide allows students to enter 6, 7, 8, 9, or 10 boxes, the second slide allows for 11 – 15 boxes and so on until all data has been collected. Meanwhile, the instructor creates a sketch of the results so the distribution can be presented to the students. For simulations in which the Central Limit Theorem applies, however, the bins are created by finding the range that contains roughly three standard deviations above and below the mean and then dividing the range into five equal intervals. Over time, the bins have been modified to be a bit less conservative. This allows for a more detailed view of distribution of the responses, but also causes the distributions to look a bit less obviously unimodal and roughly symmetric. The author does, however, look forward to the implementing the numeric entry clicker system with software that creates a histogram from numeric data so that more detailed responses can be collected, stored and analyzed in class.

The second limitation of the implementation of the activities is the errors that are generated by the students. The most difficult of these errors to address is when a student or group of students chooses not to complete the activity and, instead, clicks a button chosen haphazardly. As with any classroom management issue, the best recourse for this type of behavior is to establish a relationship of trust with the students and request that anyone who has not finished the activity to not respond (without a grade penalty) to the particular question. Because some students do not finish the simulation in the allotted time, there is no distinction in the records between students who did not finish and students who made no attempt.

Another type of student generated error occurs when a number of students complete the simulation incorrectly. It is usually clear from the responses when this happens. For example, in Gettysburg Address when students find the average of the 10 randomly generated numbers rather than the lengths of the words associated with the numbers, the averages are far larger than can be reasonably expected. Sometimes in the first enactment of Coin Flipping, in which students create a confidence interval for the percent of heads that will occur when flipping a fair coin based on a simulation of flipping a fair coin, many fewer than 95% of the students report that their 95% confidence intervals cover the true value of 50%. This “error” can be used as a teachable moment. First the instructor can explain how she knows that students have calculated their confidence intervals incorrectly (usually forgetting to multiply the standard error by 2 when creating the margin of error for a 95% confidence interval) and then the instructor can help the students correct their mistakes. This particular limitation, therefore, is actually an opportunity for learning, but instructors need to be aware that these errors may occur and be proactive in thinking about how to address such issues. Student generated issues are not limited to these activities, but these are the two issues that the author noticed as being widespread. When only a small percentage of the students were struggling to complete the simulation correctly, the instructor could help the students individually or find other students to help those who were struggling.

## 5. FUTURE DIRECTIONS

The future directions associated with this case study are dissemination and research into the effects of the implementation of such activities in the classroom. The author has implemented and modified the activities described over three semesters of teaching the introductory statistics course. She is now ready to share the activities and hopes to find instructors to implement the activities in their classes. Some research questions associated with dissemination are

1. What documentation is necessary for faithful implementation of the activities?
2. Is faithful implementation associated with teaching style, pedagogical knowledge and/or content knowledge?

Statistics instructors who are interested in implementing these activities in their classes should email the author, who will supply the slides necessary to enact the activities. Note that author has relocated to the University of Georgia and can be contacted there.

As stated in the introduction, the author began using clickers mainly to foster active learning and improve student-teacher interaction. The development of the activities was organic and occurred as part of course planning. At the time it had not occurred to the author to create a research program around the activities so baseline data on student understanding in her classes prior to the use of clickers is not available for comparison purposes. In order to enact a reasonable experimental design to study the effect of the activities on student conceptual understanding, instructors new to the activities, and less invested in their success, need to be recruited. Some research questions associated with student learning outcomes are

1. Do students develop better conceptual understandings of statistical inference from completing the activities?
2. In what way does a particular activity or set of activities interact with students' previous conception or develop students' conceptions?

It is the hope of the author that this publication will interest enough instructors to begin the type of research outlined here.

## References

Belia, S., Fidler, F., Williams, J. & Cumming, G. (2005). "Researchers misunderstand confidence intervals and standard error bars," *Psychological Methods*, 10(4), 389 – 396.

Brewer, J.K. (1985). "Behavioral statistics textbooks: Sources of myths and misconceptions?" *Journal of Educational Statistics*, 10 (3), 252 – 268.

Chance, B., delMas, R. and Garfield, J. (2004). "Reasoning about sampling distributions," *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

DeVeaux, R., Velleman, P. and Bock, D. (2006). *Intro Stats, second edition*. Boston, MA: Pearson Education, Inc.

GAISE (2006). "College Report of the Guidelines for Assessment and Instruction in Statistics Education Project," <http://www.amstat.org/education/gaise/GAISECollege.htm>

Haller, H. & Krauss, S. (2002). "Misinterpretation of significance: A problem students share with their teachers?," *Methods of Psychological Research Online*, 7(1), <http://www.mpr-online.de>

Kaplan, J.J. & Du, J. (2009). Question Format and Representations: Do Heuristics and Biases Apply to Statistics Students? *Statistics Education Research Journal*, 8 (2), 56 – 73. Available online: <http://www.stat.auckland.ac.nz/serj>

Kaplan, J.J. & Urban-Lurain, M. (2008). Personal Response Systems in Statistics: Using clickers to foster active learning and address student misconceptions. Proceedings of *the Inaugural Conference on Classroom Response Systems: Innovations and Best Practices*. Available online: <http://iclicker.com/dnn/UserCommunity/ConferencePapers/tabid/171/Default.aspx>

Lane-Getaz, S. (2005). "Summary of p-value research," Presentation at the U.S. Conference on Teaching Statistics. [www.causeweb.org/uscots/spotlight/files/R6.pdf](http://www.causeweb.org/uscots/spotlight/files/R6.pdf)

Mazur, E. (1996). *Peer instruction: A users' manual* (First ed.). Upper Saddle River, NJ: Prentice Hall.

McGowan, H. M. & Gunderson, B.K. (2010). "A Randomized Experiment Exploring How Certain Feature of Clicker Use Effect Undergraduate Students' Engagement and Learning in Statistics," *Technology Innovations in Statistics Education*, 4 (1). <http://www.escholarship.org/uc/item/2503w2np>

Noll, J.A. (2007). *Graduate teaching assistants' statistical knowledge for teaching*. Unpublished Doctoral Dissertation. <http://www.stat.auckland.ac.nz/~iase/publications/dissertations/dissertations.php>

Schaeffer, R.L., Gnanadesikan, M., Watkins, A. & Wimer, J.A. (1996). *Activity-Based Statistics*. New York, NY: Springer-Verlag.

Sedlmeier, P. (1999). *Improving Statistical Reasoning: Theoretical Models and Practical Implications*. Mahwah, N.J.: Lawrence Erlbaum Associates, Inc.

Sticklen, J., & Urban-Lurain, M. (2006). "Enhancing learning of low ability students in multi-section freshman lecture/laboratory classes," Paper presented at the ASEE Annual Meeting, Chicago, IL.

Urban-Lurain, M., Sticklen, J., & Buch, N. (2006). "High enrollment, early engineering courses and the personal response system." Paper presented at the ASEE Annual Meeting, Chicago, IL.

Wilkerson, M., & Olson, M. R. (1997). "Misconceptions about sample size, statistical significance, and treatment effect," *The Journal of Psychology*, 131(6), 627 – 631.

Zhang, B., Patterson, R., Richmond, G., Parker, J., Merrill, J., Urban-Lurain, M., et al. (2005). "Using self-response system and online learning environment in large college science classes-the technologies, instructional design, and implications," Paper presented at the International Conference on Computers in Education, Singapore.