

CLICKERS, SIMULATIONS, AND CONCEPTUAL UNDERSTANDING OF STATISTICAL INFERENCE

Jennifer J. Kaplan
University of Georgia, USA
jkaplan@uga.edu

This paper is a technology case study that addresses the theme of using clicker technology in a large lecture format undergraduate introduction to statistics class to develop student conceptual understanding of inference. The paper will present one of a suite of activities designed to help students develop conceptual understanding inference. The activity, targeting understanding of the process of hypothesis testing and the meaning of the p-value, has been implemented in a large lecture introductory statistics course in which the students use calculators to perform a trial of a simulation and clickers to report the results of the trial. Design features of the activity along with slides and student responses will be shared. An extension of the activity designed to improve student understanding of Type I and II errors and power as well as classroom implementation issues and future directions for research will be discussed.

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 (PRS or “clickers”) are a technology that allows 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. McGowan and Gunderson (2010) provide a comprehensive review of the literature on clickers, including the limitations of previous clicker research; their 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” (p. 29). In response to McGowan and Gunderson, this paper presents a technology case study of a purposeful implementation of clicker use in statistics. It describes the implementation of one of a suite 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 combination of 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.

SETTING

Introductory Statistical Methods, in which the activities were implemented, 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. There are 120 students in each lecture. The students meet the professor for 3 lectures per week for 50 minutes. In addition, the students meet a TA in recitation section one day a week for 50 minutes. Clickers are used in lecture, but not in recitation sections.

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 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. 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. This course does not include the use of computer technology or computer labs. The activities are designed with the use of a TI-83 or -84 graphing calculator in mind.

ACTIVITY DESIGN

The author developed 12 activities for the introductory statistics class in which students perform simulations the results of which are collected via clickers. As a suite, the activities were 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, Brewer claims that the area of inference is “the most misunderstood, confused, and abused of all ... statistics topics” (1985, p. 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; Wilkerson & Olson, 1997). Noll claims “that an understanding of sampling concepts and processes is necessary for developing a robust understanding of statistical inference” (2007, p. 9). The suite of activities 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.

Each activity has an *activation question*. The purpose of the activation 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. 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, p. 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 activation 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 activation question are compared to results of the same question asked after the activity has been completed. 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.

SAMPLE ACTIVITY: CELL PHONE DRIVERS

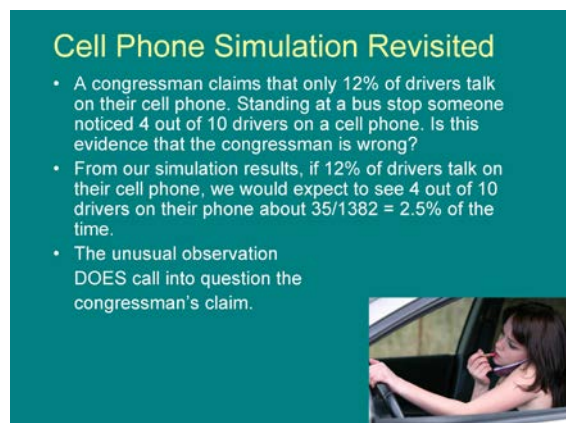
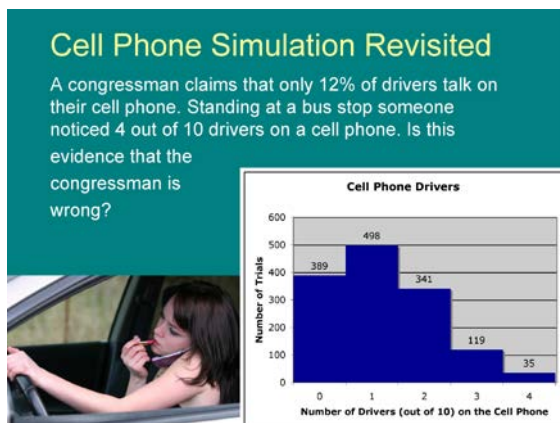
The Cell Phone Drivers (1) activity is designed to highlight the correct meaning of a p-value and the reasoning behind conclusions made from a hypothesis test.

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, p. 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 1a) and then an explanation of how we would use the simulation results to make a conclusion about the congressman’s claim (Figure 1b). 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 1a: Cell Phone Driver Scenario

Figure 1b: 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 activation 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?”

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)` on a TI-84 calculator 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.

The simulation results are collected via clickers using the slide in shown Figure 2a, 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 1 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.

Figure 2a: Cell Phone Driver Simulation Results Figure 2b: Cell Phone Driver Follow-Up

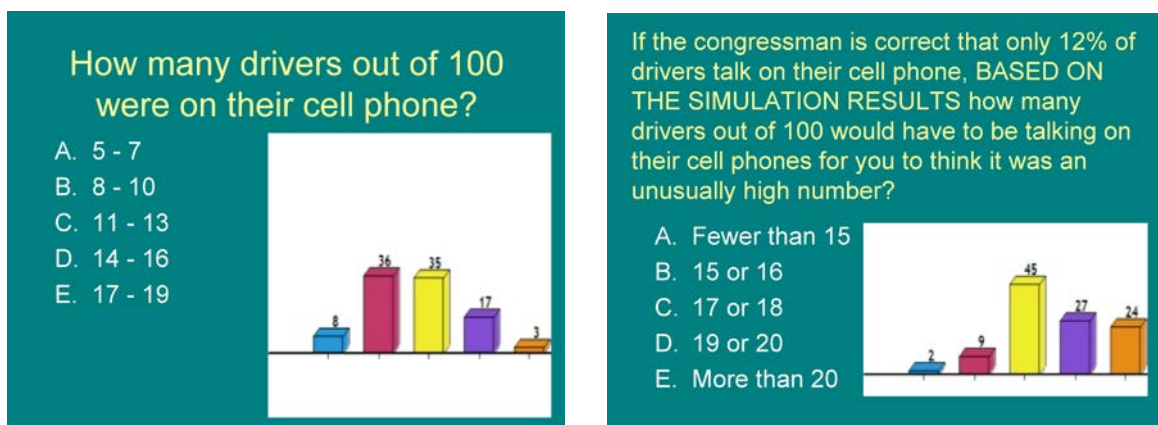


Table 1: Results of Cell Phone Driver Activity

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%)

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.

The activity can be extended to the teaching of errors of Type I and II and power. Using the same context, or any context associated with one proportion, students can simulate a sample proportion as described above for which the null hypothesis is true. The Type I error rate is the proportion of the simulated samples for which we would reject the null hypothesis (and the students who generated those samples would have made a Type I error). To demonstrate Type II errors and power, students can simulate samples proportions based on alternate parameter values (i.e. simulating the situation in which we should reject the null hypothesis). The Type II error rate is the proportion of the simulated samples for which we would fail to reject the null hypothesis (and the students who generated those samples would have made a Type II error) and the power is the proportion of simulated samples for which we find evidence to reject the null hypothesis. Factors such as the alternative parameter value and sample size could be varied to demonstrate the relationship between these factors and the power and error rates.

LIMITATIONS

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 activity 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. For simulations the activity described, in which the Central Limit Theorem applies, 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 and the difference between the expected outcome of the simulation and the reported outcomes can be used to diagnose student misunderstandings. 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.

CONCLUSION

This paper provides an overview of the implementation clickers that marries a simulation approach to teaching statistics with clicker technology for collecting results from a large number of students and details of one of a suite of activities designed to build student conceptual understanding of inference. More detailed information on the implementation of clickers by the author and the suite of activities can be found in Kaplan (2008; 2011).

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., & 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., & 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. (2011). Innovative activities: How clickers can facilitate the use of simulations in large lecture classes. *Technology Innovations in Statistics Education, 5*(1). <http://escholarship.org/uc/item/1jg0274b>
- Kaplan, J. J. (2008). *i>clicker pedagogy case study*. Invited paper for the iclicker.com website. Available online: http://iclicker.com/dnn/Portals/0/Kaplan_Case_Study.pdf
- Lane-Getaz, S. (2005). *Summary of p-value research*. Presentation at the U.S. Conference on Teaching Statistics. www.causeweb.org/usscots/spotlight/files/R6.pdf
- 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. (2011). Graduate teaching assistants' statistical content knowledge of sampling. *Statistics Education Research Journal, 10*(2), 48 – 74. http://iase-web.org/documents/SERJ/SERJ10%282%29_Noll.pdf
- Wilkerson, M., & Olson, M. R. (1997). Misconceptions about sample size, statistical significance, and treatment effect. *The Journal of Psychology, 131*(6), 627 – 631.