TEACHING INTRODUCTORY STATISTICS USING STUDENT GENERATED DATA IN A LARGE CLASS

Aklilu Zeleke[1] and Carl Lee[2]
[1]Michigan State University, United States of America
[2]Central Michigan University, United States of America
zeleke@msu.edu

During the past few years, we used hands-on activities generated by students in a large introductory statistics course for science majors. In this paper, we will share some strategies of incorporating in-class and online real time activities and discuss issues related to group projects, group formation and learning outcome assessment. Our experience suggests that student attitudes towards statistics and motivation of learning statistics have increased, and the learning outcomes related to conceptual reasoning have also increased. We will also address approaches we took to overcome some major challenges we encountered.

INTRODUCTION

Literature has shown that students tend to perform better when what is presented uses problems and applications related to their major. In such courses they become active participants in the learning process and enhance their understanding (Zetterqvist, 1997). It is less common to witness this level of enthusiasm and interest in students who take introductory level statistics courses that are required but not directly related to specific majors. Consequently it is very important to create environments that will engage students in the process of learning so as to improve their performance and attitudes towards statistics. To achieve this goal, the Guidelines for Assessment and Instructions in Statistical Education (GAISE) (http://www.amstat.org/education/gaise/), recommends, among others, to emphasize statistical thinking, foster active learning in the classroom, incorporate technology, implement cooperative learning, use real and class generated data to formulate statistical questions and stress conceptual understanding rather than mere knowledge of procedures. Various approaches using the GAISE guidelines have been successfully implemented for small classes (e.g., Rossman and Chance, 2001; Watkins, et al, 2004). Examples of using GAISE guidelines in a large class setting can be found in Zachropoulou (2006) and some challenges of implementing these guidelines in large classes are discussed in Magel (1998). One of the main reasons for instructors to stick to the lecture based teaching approach in large classes is because it is easier and it is safer (Benjamin, 1991, Magel, 1996). Research shows, however, that–use of a variety of methods in our teaching has multiple advantages: students' performances improve, their level of interest in the subject increases and their attitude towards statistics in general changes. As a result teaching statistics in a way that promotes an overall increase of statistical literacy and statistical thinking as defined in Chance (2002) will be in the best interest of our students and all who care about statistics education. Some examples of studies that show the use of cooperative learning in large statistics classes are (Garfield, 1993; Giraud, 1997) In these works the authors discuss issues related to "Why Cooperative Learning?", "How to Form Groups?" and in-class and out of class activities where data are provided to students in worksheet format. In addition to discussing some of these issues, the main focus of this article is on the use of "real data" generated by students in the classroom to enhance understanding of statistical concepts. We will demonstrate how we implemented several recommendations made by the GAISE report in our class. The article is organized as follows: Background, Activities Assessment and Challenges.

BACKGROUND

At Michigan State University (MSU) we experimented on using student generated data to teach STT 231 (Statistics for Scientists), a 3-credit calculus based introductory statistics course. All students taking this course were science majors. Topics covered include univariate and bivariate data analysis, linear and non-linear regression models, discrete and continuous probability distributions (binomial, hypergeometric, exponential, normal), sampling distributions and estimation of parameters (using confidence intervals and hypothesis testing). Student enrollment was approximately 150 per semester. Traditionally instructors have lectured in this course using

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

textbook type problems. Classical examples such as “Flipping a coin”, “Rolling a die” or “Color distributions in a bag of M&Ms” were used over and over again. Student evaluations at the end of the semester showed very low enthusiasm and interest in this course. It was very common to read comments such as “*Worst class ever*”, “*Boring*”, “*Is statistics all about coins, dies and M&Ms?* “. In 2006, Zeleke decided to teach this course using data that students themselves generated in the classroom using the real-time online hands-on activities developed by Lee & Famoye (2006). On the first day of class we did a survey to find out what type of learning/teaching approaches will be most appealing to students. Their responses were somewhat mixed. While 78% of the students rated “Listening to a Lecture” as their top learning tool, 85% also put “Using Hands-on Activities” in the top group as seen in Table1.

Table 1. Summary of responses for five categories

Lecture	Follow Textbook	Use Computer	Projects	Hands-on Exercises
117 (78%)	88 (59%)	77 (51%)	93 (62%)	127 (85%)

To some extent, this response validated what we wanted to do in this class. During the first week, students formed their own groups for the in-class activities. The instructor helped some (about 20 out of 150) to find/form a group. Each group had four or five members. The total number of groups was 35. Each group selected a captain for a particular activity. Captains were responsible for making sure that all group members participate in all aspects of an activity (data generation, analysis and project write up). One report per group was submitted for grading. All group members were asked to put their signature to validate their participation in the activity. Members who have not signed off the report received 0 points. Group activity scores counted for 20% of their total grade. We implemented three group activities. Each activity had four components: Data Generation [DG], descriptive Data Analysis [DA], relating information from data to statistical Model [M], making Conclusion [C]. Each of these four components falls under the GAISE recommendations listed above. Project reports ranged from 5 to 10 pages.

ACTIVITIES USED IN THE CLASS

Students conducted three hands-on, in class activities (listed below) and collected their own data. The statistics instructor’s role during these activities was mainly to facilitate the discussions (e. g. posing lead questions to keep the discussions going) and, in rare cases, giving advice to help some groups stay on track.

1. *Is hand size a good predictor of height?*

Students were asked to record hand size and height for their group members [DG]. Standard ruler was used for measuring hand size. Initially no information was given on how to measure hand size. Some measured vertically from the tip of the middle finger to the wrist. Others measured from tip of thumb to tip of pinky finger. Back and forth discussion within the group and across the group erupted immediately. Almost everyone was involved in intense discussion as to why we should do it one way or the other, or whether it matters to have one standard way of measurement. Finally the class settled to measure hand size from the tip of the thumb to the tip of the pinky finger, with all five fingers stretched to the maximum. Once these data were generated for all students and shared to the class, students were asked to analyze the data they just collected [DA]. Again no information was given as to what method they could use. Many realized quickly that this was a bivariate data analysis and worked on concepts like scatterplots, correlation, linear regression and residual plots. They used the model they created for prediction and discussed issues like measurement errors and extrapolation [DA/M]. After they worked on their own data, the instructor provided data from the real-time online database by Lee & Famoye (2006) to allow students compare their results with other students’ data from across the nation [C].

2. *How many raisins in a 0.5 oz. raisin box?*

0.5 oz boxes of raisins were distributed to the entire class. First students estimated the number of raisins in the box. Then they were asked to count and find the exact number [DG]. After this they made comparisons between guess count number and the actual count number. They displayed the information in histograms [DA] and were asked to guess the average number of raisins in the box by constructing confidence intervals and by performing a hypothesis test [M]. After comparing similar results with data from other institutions collected from the real-time online database by Lee and Famoye (2006), they made conclusions about their findings [C].

3. *How many macaroni shells in a 16 oz box?*

16 oz boxes of macaroni shells and permanent color markers were distributed. Students were asked to take a handful of macaronis and mark these with the marker. Then they mixed the marked macaronis with the others and vigorously shook the box for 30 seconds. After this, they selected a random sample of macaronis and counted the number of colored macaronis in the second sample [DG]. They displayed the data they generated in a two way table [DA] and used this information to estimate the number of macaronis in the 16 oz box. This activity is simple enough to implement in the classroom and simulates the capture-recapture model. Students make connections to estimation of population size, hyper geometric distribution, binomial coefficients and proportions. Discussions immediately started when they encountered events such as no marked macaroni in the second sample, too many colored macaronis in the second sample, size of the initial marked sample, etc [M]. After they have done this activity for about 40 minutes, they documented their finding in their report [C].

ASSESSMENT

The assessment tool we used for measuring the success of the active learning environment was the pre/post survey that we administered at the beginning and end of the semester along with comments. The following graph shows the number of students who responded about the use of tools that help them learn the best. From the graph below, we witness a slight digression in these learning tools away from a lecture based format to an active learning environment.

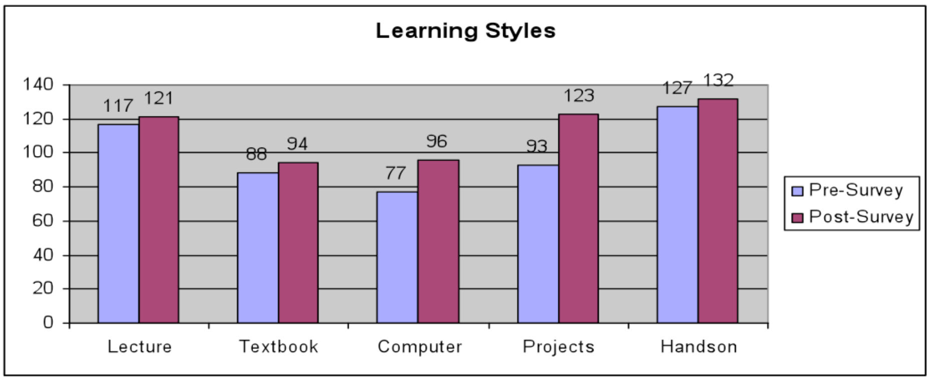

Figure 1. Student Performance Based on Different Learning Tools

By far the majority of the comments were in favor of more hands on activities and using student generated data, hence showing an increased level of interest in statistics. Here are some sample comments: "*The in class activities and projects helped me most*" , "*Finally I see real examples used in statistics class*" , "*The raisin count activity was the best, not to mention we got to eat the raisins afterwards*", "*This is the first large class I took where we actually DID something in the lecture room. I wish my other large classes follow the same approach. I learned a lot about statistics from the in class activities*". These kinds of responses are encouraging and tell us that students want us to engage them in the learning/teaching process. They motivate us to create more activities that students use to generate data in the classroom.

CHALLENGES AND DISCUSSIONS

Our experiment in a large class seems to show a similar advantage as small classes in engaging students in the learning process and motivating their interest in statistics. However, we also encountered three major challenges.

1. Group formation and group dynamics

The student population at MSU is very heterogeneous (commuters versus non-commuters, traditional versus non-traditional, international students, diverse academic, ethnic and cultural backgrounds, etc). In order to form the groups in a timely fashion, we first ask students to from their own group, then, placing randomly the students who are not able to form groups to the formed groups. By following this approach in a class of 150 students, only 20 students were not able to find a group by the end of first week. Among the 20 who did not find a group, only 4 were not able to form a group by the end of the second week. These were placed randomly in other groups and they were able to fully participate in the activities.

2. Conflicts among group members

Different kinds of conflicts surfaced among group members at different times (slacking, disagreements on how to complete assignments, some members dominating others, missing submission deadlines, etc). For this type of challenge we suggest making project grades common and a significant part of their course grade. Students respond well when their work is significantly tied to their grade. Reward good team work by offering few extra credit points and penalize inefficient groups to make sure they get the message. Discourage individual work (some want to complete the work on their own and submit early) and make every member accountable for work done/not done. Ask group members to report problems to captains and captains in return report to instructor about free riders and good/bad work ethics. This way the instructor deals with one member at a time rather than the entire group. Rotate captains and make captains sign off the work completed. Send regular reminders by e-mail and post announcements on course websites about approaching deadlines. Have groups submit progress report every two or three days and intervene immediately if group dynamics slows down. For example if a report is delayed by a day first send e-mail reminders to captains and inform them that points will be deducted if progress reports are not sent. Implement deadlines strictly. Do not give extensions for various excuses (sickness, family emergency, etc).

3. Lack of sufficient time to complete the activities

It is very important that all instructions about the activities be sent to students at least a day before the activity. Students must be well aware of what they are doing before they come to class. Post detailed step-by-step instructions on course websites and send them to students by e-mail. If available, have graduate TAs or group captains help during the activity sessions. If possible use online resources to collect data and generate surveys.

REFERENCES

Benjamin, L. T., Jr. (1991). Personalization of Active Learning in the Large Introductory Psychology Class. *Teaching of Psychology*, *18*(2), 68-74.

Chance, B. L. (2002). Components of Statistical Thinking and Implications for Instruction and Assessment. *Journal of Statistics Education* 10(3). Online: www.amstat.org/publications/jse/v10n3/chance.html.

GAISE Report (2005). www.amstat.org/education/gaise/.

Garfield, J. (1993). Teaching Statistics Using Small-Group Cooperative Learning. *Journal of Statistics Education*, *1*(1). Online: www.amstat.org/publications/jse/v1n1/garfield.html.

Giraud, G. (1997), Cooperative Learning and Statistics Instruction. *Journal of Statistics Education, 5*(3). Online: www.amstat.org/publications/jse/v5n3/giraud.html.

Lee, C., & Famoye, F. (2006). Teaching statistics using a real time online database created by students. *Proceedings, International Conference on Teaching Statistics*, Brazil.

Magel, R. (1998). Using Cooperative Learning in a Large Introductory Statistics Class. *Journal of Statistics Education*, *6*(3). Online: www.amstat.org/publications/jse/v6n3/magal.html.

Rhonda, M. (1998). Using Cooperative Learning in Large Introductory Statistics Class. *Journal of Statistics Education*, *6*(3). Online: www.amstat.org/publications/jse/v6n3/raonda.html.

Rossman, A., & Chance, B. (2001). *Workshop Statistics*. Key College Pub.

Watkins, A., Schaeffer, R., & Cobb, G. (2004). *Statistics in Action.* Key Curriculum Press.

Zachropoulou, H. (2006). Two learning activities for a large introductory statistics class. *Journal of Statistics Education, 14*(1). Online: www.amstat.org/publications/jse/v14n1/zacharopoulou.html.

Zetterqvist, L. (1997). Statistics for Chemistry Students: How to Make a Statistics Course Useful By Focusing on Applications. *Journal of Statistics Education, 5*(1). Online: www.amstat.org/publications/jse/v5n1/zetterqvist.html.