ENHANCING STATISTICAL LITERARCY THROUGH SHORT OPEN-ENDED QUESTIONS THAT INVOLVE CONTEXT, DATA, AND UPPER LEVEL THINKING

Mahtash Esfandiari, Hai Nguyen, Yuliya Yaglovskaya and Robert Gould
Department of Statistics, University of California, Los Angeles, United States of America
esfandia@stat.ucla.edu

Communication of statistical results through writing has been recognized as one of the major objectives of any statistics course. The first step we took toward this objective was developing a bank of multiple-choice questions that involve real context and data relevant to students' lives. While the majority of students (more than 70%) are able to identify the correct answer to concepts such as confidence interval and P value, they are not able to generate correct written responses to the same concepts. We report on a National Science Foundation project to create short open-ended questions involving real context and data, three typical answers (excellent, satisfactory, and inadequate), a detailed rubric, and a computerized system to score the students' responses. Consequently, the students will enhance their statistical literacy though examination of the typical answers, generating written responses, and creating connections between major concepts covered in the lecture and data analysis labs.

INTRODUCTION

As leading statisticians and statistics educators highlighted the importance of implementing real data in the teaching of statistics (Cobb, 1991; Moore, 1998; Singer & Willet, 1990), enhancement of statistical literacy and statistical thinking was also recognized as one of the major goals of statistics education in general and introductory statistics courses in particular (Chance, 2002). This is reflected in new developments and trends in statistics education literature as well as the GAISE guidelines (Guidelines for Assessment and Instruction in Statistics Education) outlined by the American Statistical Association.

There are multiple definitions of statistical literacy and statistical thinking in the literature. In this paper we are interested in the aspect of statistical literacy and statistical thinking that relates to student's ability to interpret and communicate in writing the results of statistical calculations that are performed on real data sets. Our major objective is neither emphasis on stepwise and recipe-like calculations nor on the blind follow up of a series of instructions on how to crunch numbers by using particular software. Rather we use case studies that involve real data and context that are relevant to students' lives and model the steps that a statistician takes to answer real world questions and communicate the relevant findings to a non-statistical audience.

THE CHALLENGE

As the number of students who enroll in undergraduate statistics courses increase, we develop more concern about making it possible for the students to play an active role in constructing their own knowledge, engaging in upper level thinking, understanding statistics as a mean of answering real world questions, and learning how to communicate statistical findings in writing. By engaging in upper level thinking we imply minimizing the amount of time that the students spend on recall of information and carrying out stepwise calculations and maximizing the time that they engage in application, analysis, synthesis, and evaluation of statistical concepts and strategies that are introduced to them through case studies that involve real data and context (Esfandiari & Phares, 2004). As a solution to this dilemma, we have combined technology with regular methods of teaching to redesign Statistics 10, Introduction to Statistical Reasoning. This course was redesigned to include a "statistical thinking" and a "statistical literacy" component. The number of students who take Stats 10 is roughly 480 per term (A term is 10 weeks). Students attend lecture by the professor three times each week, in groups of about 160. Twice a week they meet in smaller groups of 40 with a graduate-student Teaching Assistant (TA); once in a discussion and answer session, and once in the computer lab.

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

TEST BANK: FIRST STEP TAKEN TOWARD SOLVING THE CHALLENGE

After redesign, the first step taken toward enticing students to communicate statistical results through writing was the development of an automated test bank of multiple-choice questions that involve data and context related to students' lives. Presently this test bank consists of approximately 2000 questions most of which are written at the upper level of thinking (Bloom, 1971) and require the students to engage in application, analysis, synthesis, and evaluation. Some of the questions in the test bank are designed to enhance the students' understanding of the major underlying concepts and create a solid foundation and scaffolding that is needed for engaging in statistical thinking and literacy (Esfandiari, 2005).

We mainly use the test bank for formative evaluation of student learning and not summative evaluation and grading. Each week the students take two on-line quizzes on the major concepts discussed that week. Moodle (the course management system we use) allows the students to see what items they missed and it also provides the instructor with summary results on the percentage of items that they missed. Consequently, the weekly quizzes entice the students to study regularly, develop an in-depth understanding of the major concepts, create a tie between statistical concepts and strategies, and play an active role in constructing their own knowledge than being passive recipients of information (Wittrock, 1974). The summary results of the on-line quiz also allows the instructor to decide what concepts need to be revisited and it also makes it possible for him/her to pinpoint the students' misconceptions by examining the incorrect options chosen by the majority of the students

Our research indicated that while the test bank helped the majority of the students to master major statistical concepts, learn what statistical method to use for answering simple real world questions, see the connection between concepts discussed in different chapters, and recognize the best way of communicating statistical findings, it was not very helpful in terms of training them to communicate the statistical results through writing. (Esfandiari et al., 2007)

OPEN-ENDED QUESTIONS: SECOND STEP TOWARD THE CHALLENGE

As the second step toward dealing with this dilemma, we have started creating a series of short open-ended questions involving real context and data, three template answers (excellent, satisfactory, and inadequate), a detailed rubric, and a computerized system to score the students' responses. The objective is to help the students enhance their statistical literacy though examination of the template answers, generate written responses, and create connections between major concepts covered in the lecture and data analysis labs. The areas covered by the open-ended questions include statistical concepts and strategies that the students are most likely to encounter in the mass media and in research literature. These areas include interpretations of plots and tables, experimental design, correlation and causation, confidence intervals, interpretation of p-values, and difference between statistical and practical significance.

The open-ended questions are posted on Moodle. The students go on-line and type their answer to the question asked. The automated essay grading software on which we have been working for the last three years scans the students' responses and classifies them into one of the three templates: excellent, satisfactory, and inadequate. This is work in progress and we aspire to improve its accuracy with time. (Esfandiari & Nguyen, 2008). However, this software is a means to an end. We are using it to entice the students to synthesize and generate their own answers. We want to communicate to them that writing is a major objective of statistics education and we also aspire to give them the training that is necessary for written communication of statistical results to a non-statistical audience.

Developing open-ended questions as well as rubrics has been a challenge. Our process is to first choose a concept, develop a question, and then have a team of instructors and teaching assistants who are involved in teaching Stat 10 examine it and offer their comments. After all the revisions are done, we develop a rubric. Then the question is administered to a pool of at least 500 students. The questions are first graded by a human. To check for inter-rater reliability a random sample of one hundred or more questions are also graded by a second human grader. Then, out of 500 essays that have been graded by a human around 90% are used for training the software and 10% for testing it. So far our overall level of accuracy is around 70%, although accuracy varies between categories.

AUTOMATED TEST BANK IS NECESSARY BUT NOT SUFFICIENT

Our findings indicated that the majority of the students were able to choose the correct answer to multiple-choice questions on a certain concept, but they were not equally successful in generating the correct written response to a similar question on the same concept. In the following an example is provided on the concept of confidence interval.

Two Sample multiple-choice questions on confidence interval are given below:

1) Andy wants to find out whether the average amount of credit card debt is different for male and female juniors who attend a four-year college with 20,000 students. For a random sample of 300 female juniors and 250 male juniors, he finds that the average debt is $600.00 more for females. He also finds the 99% confidence interval to be $500.00 to $700.00

What would be the best way to interpret this confidence interval?

a) Andy should reject the null hypothesis and conclude that the average amount of credit card debt is different for the sample of male and female juniors at this college. (2%)
b) Andy can be 99% confident that the proportion of credit card debt is $500.00 to $700.00 more for female than male juniors at this college. (9%)
c) Andy should fail to reject the null hypothesis and conclude that there is no relationship between gender and the average amount of credit card debt for male and female juniors at this college. (7%)
d) Andy can be 99% confident that the average amount of credit card debt is $500.00 to $700.000 more for female than male juniors at this college. (80%)

2) In a school district they want to estimate the average number of hours that middle school students play video games per month. They select a random sample of 250 students from the population of 5000. They find the average number of hours per month to be about 7 with a standard deviation of 2. They use the sample statistic and construct the 95% confidence interval to be from 6.75 to 7.25 hours.

What would be the best way to interpret the confidence interval?

They can be 95% confident that

a. on the average the middle school students in this district play between 6.75 to 7.25 hours of video games per month. (91%)
b. 95 out of 100 middle school students in this district play between 6.75 to 7.25 hours of video games per month. (6%)
c. if they took one hundred random samples of size 250 from this population, in 95 of them the confidence interval will be between 6.75 to 7.25 hours. (1%)
d. on the average the middle school students in this sample play between 6.75 to 7.25 hours of video game per month. (2%)

In the case of both multiple-choice items, the majority of students chose the correct response. But when a similar question was used as an open-ended query, the success rate was much lower:

Sample open-ended question on confidence interval that requires a written response:

We want to estimate the average credit card debt for students at a college with an enrollment of about 5000 students. We select a random sample of 250 students and find that their average debt is about $3000.00 with a standard deviation of $500.00. We use the sample statistic and construct the 95% confidence interval from $2938.00 to $3062.00

Here is *an inaccurate* interpretation. *Explain explicitly* why you think that the interpretation is *inaccurate*.

We are 95% confident that the credit card debt for this population is between $2938.00 to $3062.00 and if we choose one hundred random samples of size 250, in 95 out of 100 cases the confidence interval computed will be between $2938.00 to $3062.00.

This question was administered to 129 students and the percentages of the responses in the inadequate, satisfactory, and excellent categories (as determined by human graders) were 70, 23, and 7 respectively. This shows that the students can memorize the correct phrasing well enough to make good decisions on a statistical concept presented in a multiple-choice format, but not be able to write the correct response to the same concept presented in the form of an open-ended question.

CONCLUSIONS AND DISCUSSION

Based on detailed observations that were conducted on the Teaching Assistants prior to the redesign, the TAs spent most of their time solving the homework problems on the board and the coefficient of correlation between the exam scores and the homework was around 0.2. Based on in-depth evaluations that were conducted after the redesign, student attendance in section increased (from 56% to 92%), TAs focused on concepts and not homework solutions, and the TAs changed their role from lecturer to facilitator. Also, the coefficient of correlation between exam scores with homework and quiz scores increased to 0.60; implying that after redesign there is a transfer of learning from what happens in section to what the students learn in the course. The homework questions prior to and after redesign were open-ended questions from the textbook or made by the instructor. The quiz component was added after the redesign and all of the questions were multiple-choice. The final exam was the same prior to and after redesign and it consisted of a combination of multiple-choice and open-ended questions. Thus, we have been able to take some major steps toward redesigning Stat 10 and aligning it with the major objectives of GAISE (Guidelines for Assessment and Instruction in Statistics Education) (Garfield et al., 2007). However, in order to overcome the challenge of enhancing writing in large classes of 160 or more we need to continue our work toward developing short open-ended questions, typical answers, and rubrics that prepare the students for the challenge of communicating statistical findings through writing. We also aspire to fine-tune the automated essay grading system to make the job of grading thousands of essays doable (Esfandiari et al., 2007, Esfandiari & Nguyen, 2008).

REFERENCES

Bloom, B., Hastings, J., & Maduas, M. (1971). Handbook of Formative and Summative Evaluation of Student Learning. Mc Graw Hill.

Chance, B. L. (2002). Components of Statistical Thinking and Implications for Instruction and Assessment. *Journal of Statistics Education*, *10*(3).

Cobb, G.W. (1992). Teaching statistics: more data, less lecturing in *Heeding the Call for Change*, Lynn Steen, editor, MAA Notes, Number 22, 3-43.

Esfandiari, M., & Phares, S. (2004). “Testing for Upper Level Thinking in Lower Division and Upper Division Statistics Classes and the Role of Technology”. UCLA Department of Statistics Lecture Series.

Esfandiari, M., Barr, C., & Sugano, A.P. (2007). Using Technology to Enhance Teaching, Learning, and Assessment in Statistics Education. University of California, Riverside. Invited Paper.

Esfandiari, M., & Nguyen, H. (2008). Development of an Automated Essay Grading Software for Statistics (AEGSS): A Prototype. UCLA Department of Statistics Lecture Series.

Garfield, J., Aliaga, M., Cobb, G., Cuff, C., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, R., Utts, J., Velleman, P., & Witmer, J., (2005). “GAISE College Report”, American Statistical Association. Online: http://www.amstat.org/education/GAISE/ GAISECollege.htm.

Moore, D. S. (1998). Should mathematicians teach statistics? *The College Mathematica Journal*, *19*, 3-35.

Singer, J. S., & Willet, J. (1990). Improving the teaching of applied statistics: Putting the data back into data analysis. *The American Statistician*, *44*(3), 223-230.

Wittrock, M. C. (1974). A generative model of mathematics learning. *Journal for Research in Mathematics Education*, *5*(4), 181-197.