Civil Engineers Design High School Statistics Tasks

Marsha Shrago,[1] Laurie Garrow,[2] Marion Usselman[1]

Abstract – In fall of 2008, the Georgia Department of Education began implementing a new mathematics curriculum for high school students. This curriculum, the Georgia Performance Standards (GPS), encourages instruction that "applies mathematical concepts and skills in the context of authentic problems."[1] However, relatively few Georgia high school teachers have taught more than elementary statistical concepts, and they have limited access to authentic problems for use in teaching data analysis and probability. To address this paucity of resources, civil engineering professors and educators at Georgia Tech worked with graduate students in "Statistics in Transportation" to create tasks that would engage ninth and tenth grade students in learning and applying grade-level mathematics content using GPS. Many of the tasks are currently being introduced to high school teachers and their students. This paper discusses the assignment, highlights some of the tasks, and describes teacher and student reactions to the tasks as a learning experience.

Keywords: Georgia Performance Standards, data analysis, probability, authentic problems

INTRODUCTION

In fall of 2008, the Georgia Department of Education began implementing a new mathematics curriculum for high school students. This curriculum, the Georgia Performance Standards (GPS), encourages instruction that "applies mathematical concepts and skills in the context of authentic problems."[1] However, relatively few Georgia high school teachers have taught more than elementary statistical concepts, and they have limited access to authentic problems for use in teaching data analysis and probability. As the K-12 outreach arm at Georgia Tech, the Center for Education Integrating Math, Science, and Computing (CEISMC) recognized the need for timely and engaging resources and contacted professors in the College of Civil and Environmental Engineering to seek their assistance. Dr. Laurie Garrow agreed to assign her graduate students in "Statistics in Transportation" to create tasks that would engage ninth and tenth grade students in learning and applying grade-level mathematics content as directed in the Georgia Performance Standards.

GRADUATE CLASS ASSIGNMENT

Background and Primary Objectives

There were 23 students in Dr. Garrow's fall of 2008 graduate course "Statistics in Transportation." This is a required course for all first-year masters' and doctoral students in the Transportation Systems Program. Historically, this has been a particularly challenging course to teach, as students represent a wide range of technical backgrounds with different skills that need to be strengthened in order for them to be successful in our graduate program. On one hand, approximately 10-15% of the students come to the course with very limited exposure to and/or understanding of probability and statistics, which requires a (fast-paced) review of probability and statistics concepts typically

[1] Georgia Institute of Technology, Center of Education Integrating Science, Mathematics, and Computing (CEISMC), 760 Spring Street, Atlanta, GA 30308 marsha.shrago@ceismc.gatech.edu and marion.usselman@ceismc.gatech.edu

[2] Georgia Institute of Technology, School of Civil and Environmental Engineering, 790 Atlantic Drive, Atlanta, GA 30306 laurie.garrow@ce.gatech.edu

covered at the high school or undergraduate level. On the other hand, this is one of the first classes taken by our international graduate students that are generally very competent at solving structured math problems (i.e., they know "how" to solve problems). However, these same students generally have difficulty identifying the appropriate method that should be used to solve a problem when they are presented with unfamiliar problem context (i.e., they do not know "why" they are using a particular method to solve a problem). For these students to be successful in the graduate program, they need assignments that are less-structured, that do not necessarily have one right answer, and that force them to communicate why they feel their approach and solution is justifiable. The class project for Statistics in Transportation was primarily designed to meet the needs of these two diverse student populations.

There were four primary objectives of the class project. First, the class assignment provided the opportunity for those students with less strong mathematical backgrounds the opportunity to reinforce their understanding of fundamental probability and statistics concepts during the first month of the class. This is particularly important, as many of the advanced concepts covered later in the semester build upon the "review" material presented in the first month. Second, the assignment provided students whose primary language is not English the opportunity to develop their written communication skills, while simultaneously placing them in an unstructured problem context where it was critically important to understand "why" what they were presenting was important. Third, the class project provided the opportunity to introduce all entering transportation graduate students to the teaching resources available at Georgia Tech via CEISMC early in their studies. This is particularly beneficial, as many of our top students applying for national fellowships or scholarships need to demonstrate an awareness of these programs at their institutions and actively engage in outreach activities in order to be competitive candidates. The final objective of the class project was to develop several tasks that could be used in high school math classes across Georgia.

Graduate Assignment – Part I

The class project for Statistics in Transportation was divided into two parts. The first part of the assignment introduced students to GPS and reviewed several existing task examples. The primary goal of the first assignment was to have each of the students identify a topic or problem context they believed would be interesting to high school students and develop a rough outline of the types of questions they would ask. Importantly, students were not asked to solve their problems mathematically for the first part of the project. This was to emphasize the importance of helping students focus on explaining "why" their problem was important and to force them to express in writing why they would select different "math tools" to solve their problem. Also, this provided Dr. Garrow and Mrs. Shrago the opportunity to review the outlines, and provide feedback to students that could be incorporated prior to "finalizing" the tasks.

Excerpts from the first part of the graduate assignment are summarized below. The background context included in the assignment has been suppressed, as it is very similar to the introduction of this paper.

> As part of the class project, you will have the opportunity to develop a teaching case that incorporates probability and statistics concepts for use in high school math classes. You will have the opportunity to interact with professional curriculum development individuals from CEISMC, who will help guide your projects and provide feedback on your cases.
>
> The first part of your project is to become familiar with the GPS standards and develop an outline for your project that clearly identifies the data you will use and (some) of the math concepts you will incorporate in your teaching case. We will build on the teaching case throughout our stats class, so don't be worried (yet!) about the fact that you may not see how to apply the concept of the Central Limit Theorem to your case. Examples of the GPS standards are provided on Georgia GPS website at http://www.georgiastandards.org/math.aspx. The appendix summarizes the GPS standards related to probability, statistics, and data analysis from the website across the seven core classes (Math 1 to Math 4 and Accelerated Math 1 to Accelerated Math 3). Examples of (partially-complete) teaching cases developed for CEISMC over the summer are also included as part of this handout and have been placed on the course website for reference.
>
> 1. Identify a dataset and problem context for your project. Provide a ½-1 page description of the problem context. Make sure to explain why engineers (or planners) are interested in this topic, i.e., how your data relates to problems encountered in practice. You should try to set up this

problem context in terms of "big themes" or "interesting questions" to be investigated without mentioning specific methods that can be used to find the answer. This 'vision' will effectively enable a teacher to introduce a topic, and solicit input from the students as to how they may solve the problem. They may (already) know some of the tools they need to use to solve the problem, but will also likely identify other tools that they need to learn. This, in turn, sets the motivation for the instructor to introduce a GPS standard (like calculating means) that they can then apply to the problem. The case you develop will be "solved" over the course of 1-2 weeks of high school instruction.

2. Develop five "Essential Questions" to be investigated in your case study. The essential questions should naturally arise from your problem description (part 1).

3. Develop an outline that associates outcome measures with the methodology or questions you will use to investigate the essential questions. These outcome measures should follow a progression of either the Math Standards 1-4 OR the Accelerated Math Standards 1-3.

4. Develop the outline of problems that are designed to teach a minimum of two outcome measures from Math Standards 1-3 OR Accelerated Math Standards 1-2. You do not need to do the underlying math or solution key as part of the first assignment; the key objective is to provide me with sufficient detail for me to be able to provide suggestions or feedback that you are on the right track. I am particularly interested in having you write one sentence for each problem that provides me with an intuition behind why the problem you are asking is important. For example, we can all calculate the average number of minutes that are actually given when we put a quarter in a meter; the problem becomes much more interesting to solve is we set up an "Essential Question" that designs a dataset where there is a reason why systematic differences may exist (e.g., due to different parking meter vendors, different cities, etc.) that the students can investigate as part of the problem context. To this extent, having the "answers" in your mind, or the "essential questions" developed up front are a great asset for designing the supporting problem.

Lessons Learned from Part I of the Assignment

The division of the project in two parts proved to be quite helpful for a number of reasons. Thus, before progressing to the second part of the assignment, a brief discussion of these "lessons learned" is warranted, as these lessons influenced the design of the second part of the assignment.

One of the first lessons was that many of the students had trouble designing a single, focused task. There were two underlying – and distinct – causes for this problem. First, several of the students (and generally the top performing students who were very familiar with statistics) were very concerned that the results produced from their analysis would be of "high quality" and "very easy to detect" so as to make it easy for the high school teachers and students to understand the results. The issue, however, is that by designing tasks based on actual data, it is very difficult to obtain "perfect" results. Many students were searching 8 or more hours for an "appropriate dataset," and Dr. Garrow had to intervene to assure the students that she was not grading based on the accuracy or strength of statistical relationships in their data, but rather on the process they used to design the tasks. The second reason students had trouble designing a single, focused task emerged from students who were less competent in probability and statistics and/or communication skills. Based on a review of the assignments from the first part of the class projects by Dr. Garrow and Mrs. Shrago, a strategic decision was made to place students in teams for the second part of the assignment. Groups were formed with the primary objective of avoiding duplication in topic areas (several students proposed similar topics); however, a secondary objective was to pair students who had problems with the first part of the assignment (due to inability to find a useful dataset, etc.) with students that had a more solid base for designing the templates.

Finally, it is important to note that throughout the project, student participation, motivation, and quality of work was very high, particularly when compared against previous classes. Many of the students would consult with Dr. Garrow throughout the project, and were genuinely concerned that they were doing the "assignment correct" and that their "problems were clear" since their problems would "be used in actual high schools." Several volunteered to work with CEISMC after the project concluded, stating they were "more than willing" to make revisions. Another strategic decision that Dr. Garrow made was to have the students do the class project during the first part of the semester – prior to the midterm, and after only three (of ten) homework assignments were completed. At this point in the semester, many of the students are concerned about whether they will make an "A" or a "B" in the class, and put a lot of effort into every assignment.

Graduate Assignment – Part II

The second part of the graduate class assignment created 13 teams (10 teams had two members and 3 teams had only one member; the single-member "teams" had projects that were smaller in scope.) Mrs. Shrago attended the lecture when the second part of the project was introduced in order to help answer questions, as well as provide students with an overview of CEISMC. Excerpts from the assignment are summarized below.

> There was quite a bit of overlap in some of the project topics that came in and/or some statements that I thought could be a nightmare to put together in two weeks based on the "messiness" of the data, so for the purposes of moving forward with the second part of the project, I am recommending we combine into groups. The goals of the second part of the project are to essentially "fill in the outline," and create the problems and solution key. Please make sure to incorporate the handwritten feedback that I have given you on your original problem statement and to consider the 10 key feedback points listed below (that we will cover in class with Mrs. Shrago from CEISMC as well).
>
> 1. Make sure to define technical terms clearly. Concepts like "land use"' or "VMT" for vehicle miles of travel are second nature to us, but may not be for all high school students and teachers.
>
> 2. It is very, very important to include references to datasets or statistics that you are quoting. These teaching modules will be disseminated by CEISMC, so we must be very careful to quote facts, datasets, and any original ideas to their proper sources.
>
> 3. When you are designing your problems, it is important to make sure that you link them back to the essential questions. That is, we need to use the "hourglass" research concept[3] – we start broad with the background context, narrow in with the essential questions, provide detailed calculation suggestions, and then hit the bottom of the hourglass, where we need to broaden out and describe how the analysis addresses the essential questions and/or ask if the results can be applied to other situations.
>
> 4. When in doubt, narrow the data focus and problem. It is much better to focus in on 2-3 specific airports and/or airlines to develop questions for, and then ask the students whether they can make inferences to a population. I would recommend focusing analysis on a handful of locations, and then providing population information (versus the entire dataset) to ask the students to interpret. This will make your life much easier when developing the answer keys.
>
> 6. If you can help students "visualize" the problem by adding maps / transit maps / maps of airport locations you are examining as part of your teaching case, please do so.
>
> 7. When developing solution keys, it is going to be important to clearly describe how to use Excel functions. For complicated ideas, please provide screen shots from Excel to help the reader visualize what you are asking them to do.

[3] The "hourglass" concept was introduced by Dr. Michael D. Meyer, who gives a lecture on research methods every semester in Dr. Garrow's graduate seminar class.

8. When developing solution keys, it is going to be important to provide “templates” for the solution, as well as the detailed solution key.

9. When developing solution keys, please plan to include a detailed Excel template. This will help the teachers have an “answer key” or “fallback” solution they can use in class if something goes wrong when they are in front of the class. (I always use this trick when I give professional classes – that is, I always have the answer key ready to go in case the software crashes, I set up something up incorrectly by accident, I have to spend more time than anticipated working with students, etc.)

10. When developing solution keys and teacher notes, make sure to emphasize to the teachers the points they should emphasize, or what is “statistically meaningful” about the results.

THE PROJECTS

Dr. Garrow’s class produced 13 projects, most of which focused on an aspect of transportation planning. Each of the 13 projects contained 3 - 5 tasks, each centered on related elements of the GPS for Math 1 or 2 or Accelerated Math 1 or 2. Below are the Standards most commonly addressed:

Students will use the basic laws of probability.

a) Find the probabilities of mutually exclusive events.
b) Find the probabilities of dependent events.
c) Calculate conditional probabilities.
d) Use expected value to predict outcomes.

Students will relate samples to populations.

a) Compare summary statistics (mean, median, quartiles, and interquartile range) form one sample data distribution to another sample data distribution in describing center and variability of the data distributions.

Students will explore variability of data by determining the mean absolute deviation (the average of the absolute values of the deviations).

Students will determine an algebraic model to quantify the association between two quantitative values.

a) Gather and plot data that can be modeled with linear and quadratic functions.
b) Examine the issues of curve fitting by finding good linear fits to data using simple methods such as the median-median line and eyeballing.
c) Understand and apply the process of linear and quadratic regression for curve fitting using appropriate technology.

Each project describes a case or a concern which drives the reader (teacher or student) to investigate using statistical or probabilistic methods. Most projects lead the reader through specific directions such as “make a scatter plot of the data in your table” or “fill in column 2 of the above table and perform the calculations below.” All projects ask for interpretation of the graphs, the tables, and/or the calculations. The multiple representations encourage readers to see connections among various mathematical ideas and to appreciate a variety of ways to communicate mathematical thinking. Frequently, the readers are asked to make comparisons and draw conclusions. Technology, such as Excel spreadsheets and graphing calculators, was appropriately incorporated and there were many opportunities to build new knowledge through problem solving.

Specific Projects

While all of the projects will be considered for use with Georgia teachers and students, three appear most engaging and are being "tested" at this time. The first project, called the Honeymoon Trip, involves the planning of a bride and groom for their wedding trip. They will be flying out of Atlanta and making connections to their honeymoon destination. They want to determine the probability of their flights being late. The introductory task asks readers to calculate simple and compound probabilities as well as conditional probability. The second task requires that the reader create a linear model to examine the relationship between late arrival and length of delay. The third task guides the reader to construct a Cumulative Density Function in Excel and to use the results to answer questions about the probability of flight departures within various intervals after the scheduled departure time.

A second project, titled Health Care Statistics, looks at hospital emergency room data. The first task considers mutually exclusive events and applies basic probability rules. The second task uses conditional probability while the third task creates a scatter plot and a regression line. The fourth task requires the calculation and interpretation of the mean, the mean absolute deviation, and the standard deviation. The last task applies probability to make decisions about the purchasing of equipment for the emergency room.

A third project examines the relationship between transportation and US Greenhouse Gases. Task 1 calculates measures of spread for diesel and gasoline vehicles and compares these measures. Task 2 looks at correlation between pollution from vehicles and miles driven and task 3 develops a linear model for all sources of pollution and miles driven. The reader is asked to interpret the slope and the y-intercept and to estimate from the linear model. Tasks 4 and 5 deal with least squares regression modeling, provide an example of the Simpson paradox, and encourage the reader to consider ways to reduce pollution by changing transportation habits.

"Testing"

Most Math I classes will be initiating their units on probability and statistics in January according to the Pacing Guide provided by the Georgia Department of Education. At this time, we therefore, have no feedback from the intended audience. However, three Advanced Placement Statistics teachers have agreed to try one or more of these tasks with their classes. Feedback is expected by the beginning of March.

Many teachers are preparing for the data analysis unit in November and December. Thus, these tasks are now being read by teachers to learn the content they will be teaching and to determine if they feel confident enough in the content to use real world applications in their classrooms. (Textbook examples typically have convenient values and yield results that reinforce concepts that are currently being taught. Real world examples often have messy numbers and introduce multiple factors that may confuse or even disguise the mathematical concepts being taught.) Tasks 1 and 2 from the Honeymoon Trip were used with a group of eighth and ninth grade teachers to apply and deepen their understanding of probability. The teachers were very receptive. They wanted more practice questions to reinforce students' understanding of procedures and to increase their fluency. When it was suggested that the task could be a "Culminating Task" that allowed students to demonstrate competency rather than learn by doing, teachers showed greater appreciation of the project.

The idea of having real world problems to use in the classroom excited the teachers. They were particularly impressed that the content they teach could be placed in an engaging setting by engineers. They could tell that the questions were not written by educators, and they appreciated the effort the graduate students devoted to learning the GPS, finding reasonable data sets, and creating interesting scenarios.

Lessons Learned from the Teachers

Although teacher experience with the projects has been limited, there are a few lessons that have already been learned. Providing the teachers with electronic copies of the projects has been somewhat challenging. The files are larger than schools often allow through email. In some cases, school system email filters stop delivery of items with ".edu" addresses. We overcame these impediments by emailing to home addresses and by delivering via flash drives. For larger dissemination, we intend to post to a website.

Since the projects use real world data, the results were often messy, not like the textbook problems. For example, the Honeymoon Trip found only a 0.08 correlation between "lateness of arrival" and "length of trip delay." When teachers are teaching the concept of correlation, most prefer to use data that have a strong association so that students see the connection in the graph as well as in the calculation. In general, the projects should not be the first experiences with new concepts. (However, some projects provided step by step guidance for developing formulas such as standard deviation and using technology to produce regression lines.)

Although teachers are generally eager to receive materials they can use in their classrooms, some have such limited backgrounds in probability – especially real world cases – that they do not know what to look for in potential instructional resources. Teachers are accustomed to using spinners, drawing cards, and flipping coins to illustrate probability concepts. They are also familiar with applying algorithms, but their understanding of which rule to use when the questions do not involve the familiar manipulatives is lacking. This finding suggests that the tasks could be used to teach teachers, that is, to deepen teachers' conceptual understanding of probability. We will explore this approach as CEISMC designs new professional learning workshops.

CONCLUSIONS

Overall, the collaboration among K-12 teachers, graduate students, professors, and CEISMC seems to be very productive and beneficial. Not only will teachers receive authentic problems to use for instruction, they will strengthen their appreciation of the mathematics as CEISMC uses these quality projects in professional learning workshops. Graduate students also benefited from interaction with CEISMC; several have become interested in participating in other CEISMC activities (such as summer camp). Interaction with CEISMC on the project dramatically increased student awareness of teaching resources and educational outreach opportunities available to them at Georgia Tech. Students' lack of awareness and desire to have more formal training on curriculum development is something that Dr. Garrow had become aware of in her doctoral graduate student seminar.

REFERENCES

[1] Georgia Department of Education, "K-12 Mathematics Introduction," *Georgia Performance Standards,* August 2007, pg.1.

BIOGRAPHICAL INFORMATION

Marsha Shrago

Marsha Shrago is a Program Director for the Center for Education Integrating Math, Science, and Computing at Georgia Institute of Technology. She plans and conducts professional learning opportunities for mathematics teachers in Georgia's K-12 schools. She has also written and/or edited parent letters for grades K – 9. This letters are intended to help explain math content addressed in each unit of the new state curriculum. Prior to joining CEISMC, Marsha taught high school mathematics for 18 years, chaired her school's staff development committee, mentored new teachers to the school system, and co-taught the methods for teaching math course at Kennesaw State University for three years.

Laurie Garrow

Laurie Garrow is an Assistant Professor in the School of Civil and Environmental Engineering (Transportation Systems Group) at the Georgia Institute of Technology. She teaches undergraduate and graduate courses on probability, statistics, modeling, and urban transportation planning. Prior to joining the faculty, she worked for four years as a research analyst with United Air Lines and one year with a management consulting firm.

Marion Usselman

Dr. Marion C. Usselman is a Senior Research Scientist at the Center for Education Integrating Science, Mathematics and Computing (CEISMC) at the Georgia Institute of Technology. Marion received her B.A. in biophysics from the University of California, San Diego, and her Ph.D. in biophysics from Johns Hopkins University. She focuses on K-12 educational reform, university-K-12 partnerships, and equity issues in education.