Probability and Statistics – Early Exposure in the Engineering Curriculum

Dr. Roger J Marino P.E., Drexel University

Roger Marino is an Associate Teaching Professor in the College of Engineering at Drexel University, Philadelphia Pennsylvania. His home Department is Civil Architectural and Environmental Engineering. Dr. Marino has 30+ years of field experience, and is licensed as a Professional Engineer in the State of New Jersey. His primary focus at Drexel is in the Freshman and Sophomore curriculums teaching courses across all disciplines.

Prof. Christopher M Weyant, Drexel University (Eng. & Eng. Tech.)

Dr. Weyant has been an Associate Teaching Professor in the Department of Materials Science and Engineering at Drexel University since 2011. Prior to this position, he was an Assistant Professor of Materials Science and Engineering at Stony Brook University. He earned his doctorate from Northwestern University, master's from the University of Virginia and his bachelor's from Pennsylvania State University. In addition to his experience in academia, Dr. Weyant has worked at Honeywell Aerospace, Capstone Turbine Corporation and Sandia National Laboratories.

Prof. Brandon B. Terranova, Drexel University

Dr. Terranova is an Assistant Teaching Professor in the College of Engineering at Drexel University. In his current role, he is the lead instructor for the freshman engineering program, and oversees activities in the Innovation Studio, a large-area academic makerspace. He has taught and developed courses in general engineering and mechanical engineering at Drexel. Prior to Drexel, he has taught and developed courses in physics and mathematics at SUNY Binghamton, University of Delaware, Missouri Online College, and St. Mark's High School. Dr. Terranova's research interests include plasmonics, optical tweezing, photonics, electromagnetism, and engineering education. He received his MS in Physics from SUNY Binghamton, and his PhD in Electrical Engineering with a concentration in Electrophysics from Drexel University for his work in 3D plasmonic nanostructures.

Full Paper: Probability and Statistics – Early Exposure in the Engineering Curriculum

Introduction

Probability and Statistics classes are often introduced in the second year of an Engineering Program. However, the benefits of students being exposed to these subjects during the Freshman Year have been identified by other researchers. Some of these benefits are: students' early recognition of the presence and importance of probability and statistics in addressing engineering problems; students' recognition that statistics and engineering are not in fact two distinct, unrelated entities; and the students' early exposure will benefit them in subsequent courses in their academic careers [1,2]. Major constraints in exposing students to probability and statistics in their first year are: course-space availability to accommodate an additional subject, and limited classroom time. Additionally, these constraints affect the depth at which an instructor can delve into the material [2]. Also contributing to difficulty in students understanding the material is that they may not have been exposed to the subject of statistics in high school [2].

To prepare high school students for the SAT and college, many high schools offer advanced mathematics courses such as Probability/ Statistics and Calculus. The U.S. Department of Education compiled data on mathematics courses offered at United States (US) high schools for the years 1990, 2000, 2005 and 2009 [8]. The proportion of high schools that offered a Probability/ Statistics course in 1990 was 1% compared to 10.8% in 2009. This represents an increase of 51.6% in adoption per year on average. This is the largest increase in adoption of a math course per year as compared with the other courses. The motivation for high schools to increasingly adopt a Probability/ Statistics course may be tied to the Scholastic Assessment Test (SAT), as the general SAT test includes "Center, spread, and shape of distributions", and the SAT math subject tests 1 and 2 cover "Data analysis, statistics, and probability [9]. It is noted that, although the Department of Education publication is dated 2017, no data is presented in the table for the years from 2009 to the present.

In order to introduce students to Probability and Statistics, the subjects were integrated into an existing First Year first term "Introduction to Freshman Design" course. Lecture and recitation sections were added to the existing laboratory-based course to create ENGR 111, "Introduction to Engineering Design and Data Analysis" (resulting in an increase of course credits). Three weeks of the course focused on statistical concepts. Lectures highlighted relevant statistics topics such as: central tendency, descriptive statistics, probability and distributions. Recitations were dedicated to the students working in teams performing exercises that reinforced the lecture material. Instructional assistance was provided in the recitation sections by graduate teaching assistants.

During the Fall 2018 quarter, 800 students were enrolled in the course in which there were one 50-minute lecture and one 50-minute recitation each week. Lectures contained 100-120 students and recitation sections were comprised of a maximum of 30 students. Direct assessment of the impact of lecture and recitation activities on learning of statistical concepts was accomplished through homework assignments, grading of the recitation exercises and questions on the final exam. Several effective approaches to teaching statistics have been reported and were applied to this course. They included: Multi-faceted activities [1], cognitive visualization of graphed data [2], and tactile measurement of tangible objects to understand variability of data as well as interpreting and defining outliers, averages, etc. [7]. Further insight into student perceptions of the recitation activities was garnered from comments on the course evaluations.

statistics course components

As part of the introduction to data analysis portion of this first year course, several statistical concepts were covered during three weeks of the ten-week quarter. Mean, variance, standard deviation, grouped frequency distributions, basic probability, and probability distributions were addressed. These topics were introduced through reading assignments, discussed in lecture, and applied through homework and group activities in recitation. Each week of the class consisted of a 50-minute lecture that was immediately followed by a 50-minute recitation. The low-stakes recitation activities were designed to take advantage of peer learning. Homework assignments were due one week after the lecture and recitation sessions. Further assessment of these topics was accomplished through a multiple choice final exam.

During recitation, students worked in groups of three on statistics-based activities focused in the areas of frequency distributions, uncertainty analysis and linear regression. The three activities were:

- *Analysis of Body Mass Index (BMI) Data:* Students were given a table in Excel containing the mass (in kg) and height (in m) for 40 people. They were tasked with creating frequency distributions for height and mass. Subsequently, they used the results of the frequency distributions to calculate mean and standard deviation. In addition, they had to create probability distribution curves.
- *Reaction Time and Error Analysis:* Students generated a data set by testing their reaction time grabbing a falling ruler. One student held the ruler initially positioned so that the bottom was at the top of an open hand of another student. The first student released the ruler and the second had to grab it as soon as possible. The distance along the ruler was recorded. This process was repeated to generate 15 data points. The statistical concepts highlighted in this activity included taking the standard deviation, calculating overall measurement uncertainty in drop distance and estimating a confidence interval.
- *Creating a Calibration Curve for an Ohmmeter:* Students were given a data set showing the measured and actual resistance of five resistors. From these data, they created a calibration curve by determining a linear regression model. They were not permitted to use Excel for this work and therefore needed to take averages and determine sums of squares in order to calculate the slope and intercept for their model. In addition, they

calculated the sample coefficient of determination for the model. The activity also required the calculation of the 95% confidence interval for the resistance of a population of resistors.

In an effort to encourage students' recognition of the need for statistics in engineering, the three activities above were preceded by an activity which required students to measure the diameter of a small pom-pom ball. This activity introduced students to the inherent uncertainty of measurement and highlighted the fact that measurement techniques and measurement instruments are factors which introduce variability in results. It was the hope of the designers of this exercise that students would be left wondering how to deal with such uncertainties.

assessments

In order to explore the success of integrating statistics into this first year course, both direct and indirect assessments were conducted. Direct assessment was measured through homework assignments in each week and the comprehensive final exam. Indirect assessment was conducted through pre- and post-course student surveys and a follow-up survey assessing their high school statistics preparation.

Direct assessments

Individual direct assessment of the topics addressed in recitation was accomplished through recitation activities (low-stakes), homework assignments (low-stakes) and final exam questions (high-stakes). Recitation activities were graded with simple rubrics that focused on task completion. Homework assignments were administered through the Blackboard Learn Learning Management System using multiple choice or numerical answer questions. Students answered the homework questions as individuals but were not restricted from working together. They had two attempts at each homework question. The final exam was a comprehensive multiple choice exam administered face-to-face.

Overall, students performed well on the three recitation activities listed above. As shown in Figure 1, the percentage of students scoring 90% or above was 92.8% for the BMI data analysis, 89.1% for the reaction time analysis and 76.7% for the calibration curve exercise. The calibration curve activity proved the most challenging which may be due to students being required to perform calculations using their calculators as opposed to using Excel which they did on the first two activities.

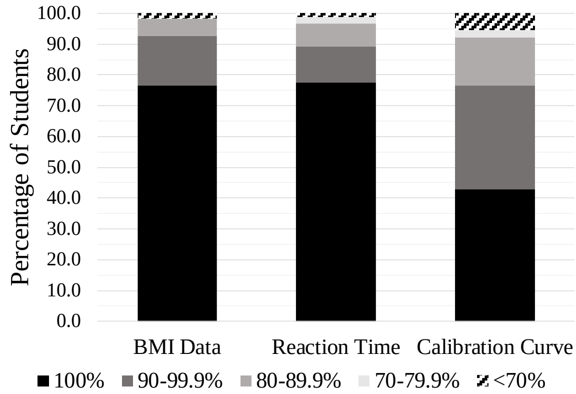

Figure 1: Distribution of grades for statistics-related recitation activities.

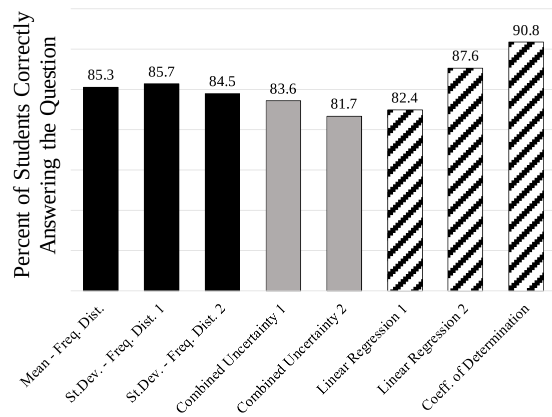

Figure 2: Percentage of students who answered select statistical homework problems correctly.

Figure 2 presents the percentage of students who correctly answered quantitative statistics-related questions on three homework assignments. The black bars are questions related to the BMI data activities. The grey bars are questions related to the reaction time exercise. The striped bars are for questions related to the calibration curve exercise. Overall, the performance on these homework questions was very good. Of particular note is the high performance on the linear regression questions which was a different trend than what was observed in the recitation activity. The main reason for this result is most likely due to students being able to use Excel for their calculations on the homework.

Although homework questions do give a sense of learning, students have all resources available to them while answering the questions, including each other. For the final exam, the only resources available were a list of equations, a calculator, and any necessary data tables needed for data analysis. Figure 3 shows performance on statistics-related questions from the final exam whose topics were covered in the recitation activities. The questions on mean and standard deviation whose results are shown in the first two columns were derived from the same data set used in the corresponding recitation activities. The second standard deviation result (third bar in Figure 3) was from a different data set. The second data set required more analysis than the first since a data range was given for each frequency instead of a single number. This higher level of interpretation was more difficult for the students. Combined uncertainty was one of the more challenging concepts that was addressed in the class and this was reflected in the final exam performance. Student performance on the linear regression analysis was good and in-line with homework. However, calculation of the coefficient of determination proved more difficult.

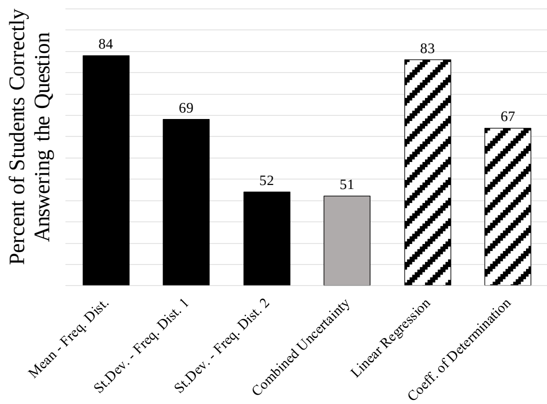

Figure 3: Percentage of students who answered select statistical final exam problems correctly.

indirect assessments

All 800 students in ENGR 111 were asked to complete a survey during the first and last weeks of the fall 2018 term. Among other course learning objective questions, students were asked to rate their perceived ability to “Analyze engineering data using basic descriptive statistics and appropriate software” based on a Likert scale (4 = Excellent, 3 = Good, 2 = Average, 1 = Poor). In the first-week survey, 706 students responded with an average score of 2.47±0.77 (between Average and Good). The reported uncertainty is the standard deviation. In the last-week survey, 642 students responded with an average score of 2.73±0.84, where the standard deviation accounts for standard error with a finite population correction. These results show that the students perceive that the Introduction to Design and Data Analysis increased their ability to analyze engineering data using basic descriptive statistics and appropriate software by 9.5%.

In a follow-up survey, a sample of 513 students were surveyed from the 800 first-year engineering students. Of these 513 students, 35.7% of them “received formal statistics training (either in a formal statistics course or statistics was covered in another course) in high school”. As referenced earlier in this paper, a national value of a 10.8% adoption of Probability/Statistics courses in U.S. high schools was reported by the Department of Education for 2009. The increased percentage of students in ENGR 111 of 35.7% which is well above the national average is likely due to the fact that the student population surveyed only consisted of those interested in engineering. Survey respondents were then asked if their high schools covered the specific topics which were the general areas of focus in ENGR 111 (basic probability, averaging, standard deviation, frequency distributions, probability distributions) and the results can be viewed in Figure 4. From their responses there is clear agreement that basic probability, averaging, and standard deviation was covered in their high school Probability/Statistics courses. Many students also had exposure to frequency and probability distributions.

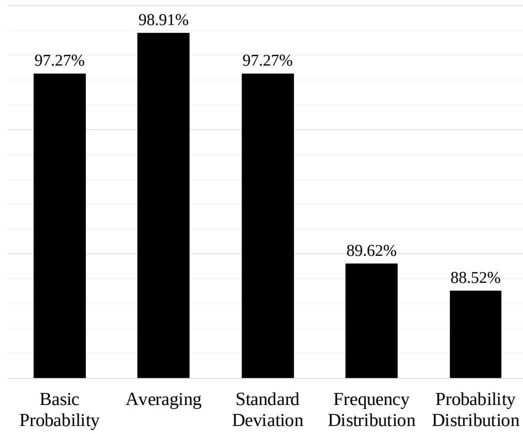

Figure 4: Perceived competence in select topics by students with high school exposure to statistics.

For those students *who had exposure to* Probability/Statistics in high school, a follow-up line of questions assessed whether ENGR 111 had “improved their understanding” of each topic using a Likert scale (1 = Strongly Agree, 2 = Somewhat Agree, 3 = Neutral, 4 = Somewhat Disagree, 5 = Strongly Disagree) and the results can be seen in Figure 5. For those students *who did not have exposure to* Probability/Statistics in high school, a follow-up line of questions assessed whether ENGR 111 had “given them an introductory understanding” of each topic using the same Likert scale and the results are displayed in Figure 6.

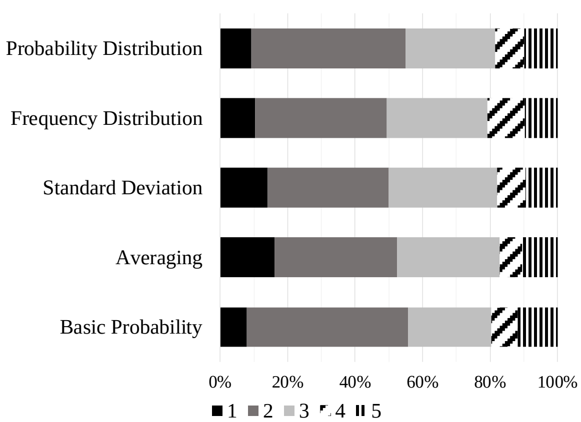

Figure 5: Perceived improved competence in select topics as a result of ENGR 111 of students who had statistics in high school

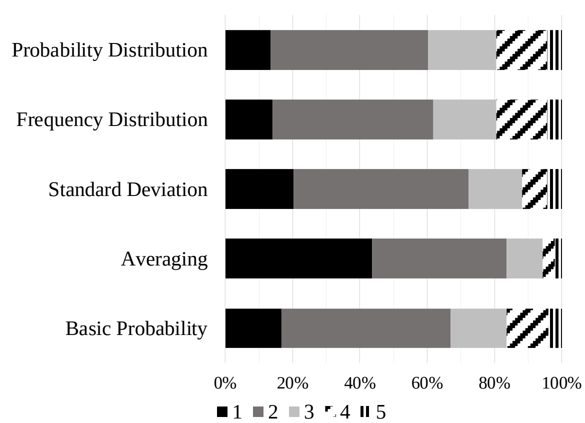

Figure 6: Perceived introductory understanding of select topics as a result of ENGR 111 of students who did not have statistics in high school

conclusions and future work

Results of the initiative to include probability and statistics into an existing Introduction to Engineering course were encouraging. Recitation activity and homework assessments were generally high – most likely due to group activity with Instructor oversight (and perhaps due to nearly 36% of the students reporting having had exposure to statistics while in high school). Exam grades were notably lower. Additionally, student surveys indicated that the cohort perceived a 9.5% increase in their own ability to analyze engineering data using basic descriptive statistics and appropriate software after taking the course. Future work may include further inclusion of relevancy examples and topical applications of the material in order to advance students' understanding and mastery of the material.

references:

[1] Jensen, D., & Kellogg, S. (2010, June), *Improving Conceptual Understanding In Probability And Statistics* Paper presented at 2010 Annual Conference & Exposition, Louisville, Kentucky https://peer.asee.org/16816

[2] Wilson, R (2002, July), *What Does This Have to Do with Us? Teaching Statistics to Engineering Students* Paper presented at ICOTS 2010 Annual Conference, Cape Town South Africa. http://iase-web.org/documents/papers/icots6/5e1_wils.pdf

[3] Reeves, K., & Blank, B., & Hernandez-Gantes, V., & Dickerson, M. (2010, June), *Using Constructivist Teaching Strategies In Probability And Statistics* Paper presented at 2010 Annual Conference & Exposition, Louisville, Kentucky. https://peer.asee.org/16660

[4] Rotante, T., & Brem, S., & Hubele, N., & Runger, G., & Kennedy, K. (2003, June), *Case Based Reasoning For Engineering Statistics* Paper presented at 2003 Annual Conference, Nashville, Tennessee. https://peer.asee.org/11787

[5] Batson, R. G. (2018, June), *How to Make Engineering Statistics More Appealing to Millennial Students* Paper presented at 2018 ASEE Annual Conference & Exposition , Salt Lake City, Utah. https://peer.asee.org/30585

[6] Jensen, D., & Kellogg, S. (2010, June), *Improving Conceptual Understanding In Probability And Statistics* Paper presented at 2010 Annual Conference & Exposition, Louisville, Kentucky. https://peer.asee.org/16816

[7] M. A. Hjalmarson, T. J. Moore and R. Delmas, "Statistical analysis when the data is an image: eliciting student thinking about sampling and variability," *Statistics Education Research Journal*, vol. 10, *(1)*, pp. 15, 2011.

[8] US Department of Education Institute of Educational Sciences, National Center for Educational Statistics, Digest of Educational Statistics, 2017, Table 225.40, 2017 https://nces.ed.gov/programs/digest/d17/tables/dr17_225.40.asp?current=yes

[9] College Board SAT Suite of Assessments, Mathematics Level 1Subject Test https://collegereadiness.collegeboard.org/sat-subject-tests/subjects/mathematics/mathematics-1