Prerequisite Testing as a Tool to Gauge Incoming Student Capability and Knowledge in an Engineering Statics Course

Roy Myose, Syed Raza, Elizabeth Rollins, Brandon Buerge, and Nicholas Smith
Aerospace Engineering Department, Wichita State University, Wichita, KS 67260-0042

Introduction

The current generation of college-aged students, referred to as "Generation Z," is said to have capabilities and needs that are quite different than previous generations of students.[1] Consequently, teaching techniques may have to be adjusted to meet their needs.[1,2] Changes are easy to make, but determining the effectiveness of those changes is more challenging to ascertain. As educators adapt their courses for the current generation, the efficacy of any alterations to the courses must be measured in terms of the impact on student performance and learner outcomes, which requires educators to know the baseline capability and knowledge level of the students entering their courses.

The first author, a 25+ year veteran engineering educator who helped develop the prerequisite test for Statics at Wichita State University (WSU),[3] has modified the examination structure of his Engineering Mechanics courses throughout his tenure to ensure that the exams accurately gauge students' abilities in his courses. One of the most recent changes was to reduce the number of homework-type working problems, which are defined here as problems involving multiple steps and calculations, on an exam. For Statics classes meeting for 50 minutes three times a week, exams were cut down from four working problems to three since a typical student needs approximately 15 minutes per problem, leaving a few minutes for them to double check their calculations. By this logic, a Statics class meeting for 75 minutes two days a week should have five working problems on their exams. However, five working problems would leave students with no extra time to review their answers. Therefore, it was decided that exams for 75-minute classes would have four working problems along with four simple concept questions acting as fillers so that the 75-minute class would have approximately the same time constraint of 15 minutes per problem with 5 surplus minutes.

Once the changes were implemented, the first author noticed that the grade point average (GPA) of the 50-minute classes was substantially lower than the GPA of the 75-minute classes. Using the GPA average of all classes taught by the first author combined together as a reference, 50-minute classes had an 8.7% lower GPA than the overall reference, while 75-minute classes had a 7.8% higher GPA. This translates to a 16.5% difference in GPA between the 50- and 75-minute classes. At first glance, this difference suggests that the concept questions were relatively easy and that students were receiving higher scores than they would have if they were taking exams based only on working problems. However, this hypothesis would be counter to what had been observed in the first author's upper-division classes, in which he routinely employed concept and short answer questions with good success. Another cause for the discrepancy in the GPAs between the 50- and 75-minute classes could arise from differences in the incoming capability and knowledge of the students, which the Statics prerequisite test is able to quantify.

To determine the reason for the difference in the class GPAs, the issue of whether concept and short answer questions can be used effectively as a measure of student learning will first be addressed through an analysis of the results this type of testing in an upper-division course. Based on the conclusion of that review, results from the final exam in Statics, which is similar in structure to the upper-division course's exam, will be examined. Following these analyses, continuing work on the use of prerequisite testing in Statics to gauge incoming student capability and knowledge by Myose *et al.*[3] will be presented. Finally, the prerequisite testing information data will be used to ascertain if the end-of-semester course GPA can be predicted using the beginning-of-semester prerequisite testing results and to determine what conclusions can be drawn from prerequisite testing about the two different types of classes.

Concept Questions and Short Answer Problems to Measure Student Performance

The junior-year Propulsion course at WSU is a typical upper-division class for Aerospace Engineering (AE). The course includes a 110-minute comprehensive final exam weighted as 30% of the semester grade. The final consists of two parts, each worth 50% of the final exam score. The first part is comprised of concept and short answer questions, with about two-thirds of the first part's points coming from concept questions. Four working problems make up the second part. Concept questions are multiple choice or simple comparators that ask the student to compare relative sizes using <, =, and > symbols. Working problems require the students to perform multiple steps of calculations to determine various propulsive performance characteristics. Since the final exam includes concept questions, short answer problems, and more complex working problems, it provides a measure of the level of difficulty that can be obtained with each type of questions.

The current dataset consisted of seven different sections from spring 2010 to spring 2018, encompassing 350 students. It should be noted that WSU uses the plus-minus grading system for final semester grades. The Pearson correlation coefficient was used to ascertain how well the scores for the first part, second part, and both parts of the final exam combined were correlated with the end-of-semester grade. The Pearson correlation coefficient ranges between +1 and –1. It is +1 when two quantities are perfectly correlated, 0 when there is no correlation at all, and –1 when an increase in one variable leads directly to a decrease in the second variable. There is less scatter in the data when the Pearson correlation coefficient approaches +/–1, while there is much more scatter when the coefficient nears zero.

The Pearson correlation coefficient between the total final exam score and the semester grade was 0.805, suggesting that the final exam is very strongly correlated with the semester grade. This is not surprising since the final exam is comprehensive, covering material from the entire semester. The Pearson correlation coefficient between the semester grade and the first part of the final was 0.691, which is still a very strong correlation with the semester grade. Between the second part of the final and the semester grade, the correlation coefficient was found to be 0.690, which showed that the first and second parts were almost identically well-correlated to the semester grade. However, the correlation coefficient for each part individually was lower than the correlation coefficient for the final exam as a whole. This difference suggests that the performance of individual students was not consistent between the first and second parts.

Figure 1 presents the results in graphical form with semester grade given in terms of grade point along the horizontal axis and final exam score in percent along the vertical axis. The vertical scale values are not included because class averages for the final exam are not disclosed to students. The average score at each grade point level is given by a blue triangle for the first part score, a green diamond for the second part score, and a pink square for the total final exam score. The solid line provides a least squares fit between the total final exam score and the semester grade. At first glance, there appears to be a large amount of scatter for grades in the D and F range. However, the data for those grades consisted of less than 20 students each, which meant that statistically significant average values were not available for the grades in the D and F range. Conversely, very little scatter exists between the total and the least squares fit line from the semester grade range of C- to A-. This lack of scatter is expected since the Pearson correlation coefficient is a very strong value of 0.805.

The blue triangles representing the average on the first part of the final exam frequently lie above the least squares fit line, while the green diamonds representing the average on the second part often lie below the line. The gap between the scores on the first and second parts of the final exam is also reflected in the Pearson correlation coefficient between the two parts which is 0.470. This again suggests that the performance of individual students varied between the two parts. The fact that the average for the first part of the exam generally was above the least squares fit line appears to indicate that the first part was slightly more difficult because a higher score was needed on the first part to

achieve a particular semester grade. This observation refutes the hypothesis raised in the introduction that concept questions may be easier than working problems, causing GPAs to be higher in the 75-minute Statics classes. Therefore, another cause for the discrepancy in the GPAs between the two sets of Static's courses must be sought.
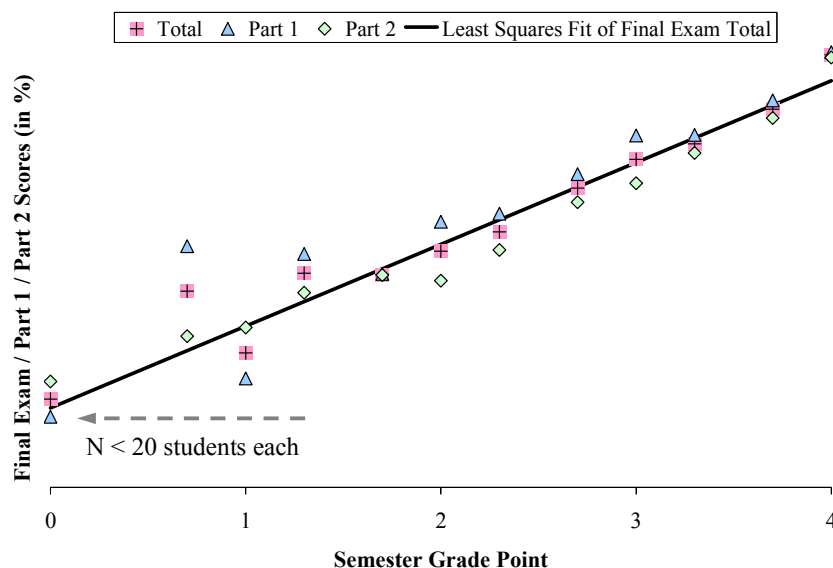

Figure 1 – Propulsion Final Exam Correlation with Semester Grade (N=350 Students).

Correlation Between the Statics Final Exam and End of Semester Grade

The sophomore year Statics course is a typical lower-division AE class. Students enroll in sections that meet either three days a week for 50 minutes or two days a week for 75 minutes during the semester. Similary to the Propulsion course, the first author administers a 110-minute comprehensive final exam when he teaches the Statics course. However, the final exam for Statics is weighted as 22% of the end-of-semester grade and consists mainly of single-step, calculation-based short answer problems with a few additional concept questions and multi-step working problems. An example of a single-step short answer question is to determine the internal force in one specific truss member. In comparison, multi-step problems might involve determining the internal force in several truss members. Students in both the 50- and 75-minute classes of Statics were given the same type of final exam under the same 110-minute time constraint.

The dataset consisted of 237 students total. Of those students, 130 were from four different 50-minute sections, and 107 were from three different 75-minute sections. The Pearson correlation coefficient between the final exam score and the semester grade for the entire dataset was 0.861, suggesting a very strong correlation. Figure 2 depicts the relationship between the final exam score and the semester grade in graphical form. The pink squares show the average final exam score for all 237 students at each grade point level, and the solid line represents a least squares fit of those final exam scores. It should be noted that the correlation level of 0.861 for Statics is slightly stronger than that the correlation level of 0.805 for the Propulsion course. The higher coefficient for Statics is expected since a student's semester grade in Statics is determined solely by their individual performance on exams, whereas Propulsion also includes team-based assignments in the calculation of the semester grades that can affect students' performance.

When the data was separated between the 50- and 75-minute classes, the Pearson correlation coefficients between the final exam score and the semester grade were 0.858 and 0.856, respectively. Therefore, there appeared to be very little difference between students from the two types of classes in terms of end-of-semester final exam performance, which essentially measures student learning

outcomes when compared to the semester grade earned. In Figure 2, there appears to be slightly more blue triangles representing the 75-minute classes lying above the overall least squares fit line compared to green diamonds representing the 50-minute classes, which suggests that students from the 75-minute classes had to perform slightly higher on the final exam than students in the 50-minute classes to earn the same grade. This trend is not expected given the fact that the GPA of the 75-minute classes was quite a bit higher. However, this observation is not a definitive conclusion because there were less than 20 students per division at almost every grade point level, making the statistical averages obtained somewhat marginal in meaningfulness.
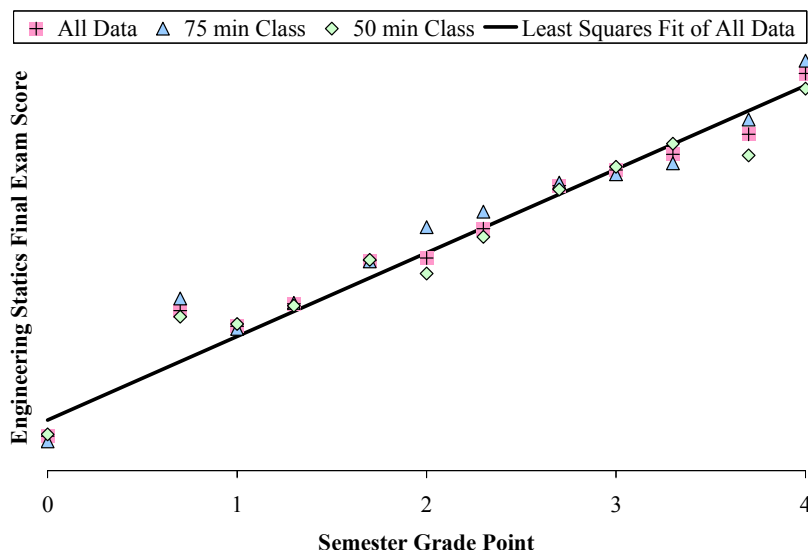

Figure 2 – Statics Final Exam Score Correlation with Semester Grade.

The results shown in Figure 2 suggest that overall, student performance and learning outcomes appear to align by the end of semester final exam irrespective of the length of the classes and the differences in the exam format throughout the semester. Since concept questions have been determined to be effective tools for measuring learning outcomes, and both types of classes are aligned in terms of their performance on the final exam, then there must be another cause for the large gap in the GPAs between the 50- and 75-minute classes. Variations in student capability and knowledge upon entering a class could affect student performance. Therefore, the incoming capability of students needs to be quantified before the teaching of new material takes place in order to observe the effect of course changes or differences. This will help accurately determine the effect of course changes independent of the capability and knowledge level of the students.

Statics Prerequisite Testing to Gauge Incoming Student Capability and Knowledge

The Statics prerequisite test has been administered at WSU each semester since 2012. The prerequisite test is used to gauge incoming student capability and knowledge. This test, referred to hereafter as the pre-test, consists of problems over six topics covered in the prerequisite Physics course. The problems, which were based on similar problems in standard Physics textbooks, have direct applications to Engineering problems in Statics. The topical areas covered are: (1) vector magnitude, (2) vector resultants, (3) friction, (4) dot product, (5) torque (i.e., moment), and (6) force equilibrium. The test structure features a combination of multiple choice and single-step calculation-based short answer problems, similar in form to the concept and short answer questions discussed in the previous sections.

The administration of the pre-test was standardized as much as possible among the sections taught by different instructors. Students took the pre-test one week after the start of the semester without any direct review of the prerequisite Physics material. The pre-test was closed book and closed notes; only

an instructor-provided equation sheet consisting of a copy of the inside front cover of the Statics textbook (Hibbeler's *Engineering Mechanics: Statics*) was given to students at the time of the test. Students were cautioned that the notation used on the equation sheet taken from the Statics textbook might not be the same as the notation in their Physics textbook. Regardless of the amount of class time available, students were given a maximum of 50 minutes to take the pre-test. It should be noted though that roughly half of the students typically finished in about 30 minutes. The only minor variation in the administration of the pre-test from instructor to instructor was the weight given to the pre-test in the overall semester grade, which ranged from 3% to 5%.

Preliminary results from this pre-test, which were presented in a previous paper,[3] were generated from a dataset of 733 students from 13 different sections taught by four different instructors. Data collection from the pre-test has continued, and the database now encompasses 1760 students from 31 different sections taught by six different instructors. Table 1 provides a comparison of the correlation of the average pre-test score to the final semester grade between the preliminary and expanded databases. Although WSU uses a plus-minus grading system, grades are presented in whole letter-grade bins in Table 1. In general, the change in the average pre-test score associated with each letter grade was negligible. The biggest differences occurred in the C grade range, which decreased from 65.2% to 62.3%, and for those who withdrew (W) from the course by the tenth week, which increased from 50.3% to 53.2%. It is not surprising to see slight adjustments in these two categories since several engineering majors at WSU require a grade of a flat C or better in Statics as a prerequisite for certain upper-division courses, which can lead to the withdrawal of students from Statics as they seek to avoid decreasing their GPAs.

Table 1. Average pre-test score (and number of students in parentheses) for each letter-grade category.

	A's	B's	C's	D's	F's	W's
Preliminary database (N=733)	78.4% (147)	67.8% (172)	65.2% (160)	56.2% (98)	53.9% (110)	50.3% (46)
Current database (N=1760)	78.1% (353)	68.0% (438)	62.3% (399)	56.8% (219)	54.9% (249)	53.2% (102)

Figure 3 presents the average pre-test score for each plus-minus letter grade category including withdrawals along with the number of students at each grade level for the current (expanded) database. Although the horizontal axis is not evenly spaced, the bar chart still suggests that a linear relationship exists between the beginning-of-semester pre-test score and the end-of-semester course grade. Figure 4 presents the results from Figure 3, but with the semester grade given in numerical grade point values along the horizontal axis. Since no grade point is associated with students who withdraw from the course, they are not a part of this figure's dataset. The pink-colored squares correspond to the average pre-test scores given in Figure 3, and the dashed line represents a least squares fit between the average pre-test score and the semester grade. The solid line depicts the least squares fit between the average pre-test score and the semester grade for the dataset of solely students taught by the first author. The solid line representing one instructor's data does not match the dashed line containing the entire dataset, indicating that some variation in the intercept and slope exists from instructor to instructor.

It is not surprising that there is a fair bit of scatter in the data since the Pearson correlation coefficient between the pre-test scores and the semester grades for the expanded database was 0.440, which suggests that there is a moderate amount of scatter in the data. This value is a slight decrease from the 0.449 correlation coefficient for the preliminary database. There are several outliers in Figure 4, occurring most prominently at the grades of C (grade point = 2.0), and A (grade point = 2.0). The outlier at a grade of C may be occurring since many engineering majors at WSU require students to achieve a grade of C or better in Statics. As a result, some students may put extra effort into passing the course and overcome their somewhat deficient prerequisite knowledge. The outlier at a grade of A

may be caused by the fact that a number of students who received an A at the end of the semester scored significantly higher than the pre-test score of approximately 78% for a grade of A predicted by the linear fit for all instructors, which is represented by the dashed line. It was not expected that information about the beginning-of-semester student capability and knowledge would precisely predict the end-of-semester performance since new content is covered during the semester. Therefore, the fact that even a moderate amount of correlation of 0.440 was found between the pre-test and the end-of-semester grade is encouraging with regard to the efficacy of the pre-test as a gauge of incoming student capability and knowledge when large numbers of students are involved.
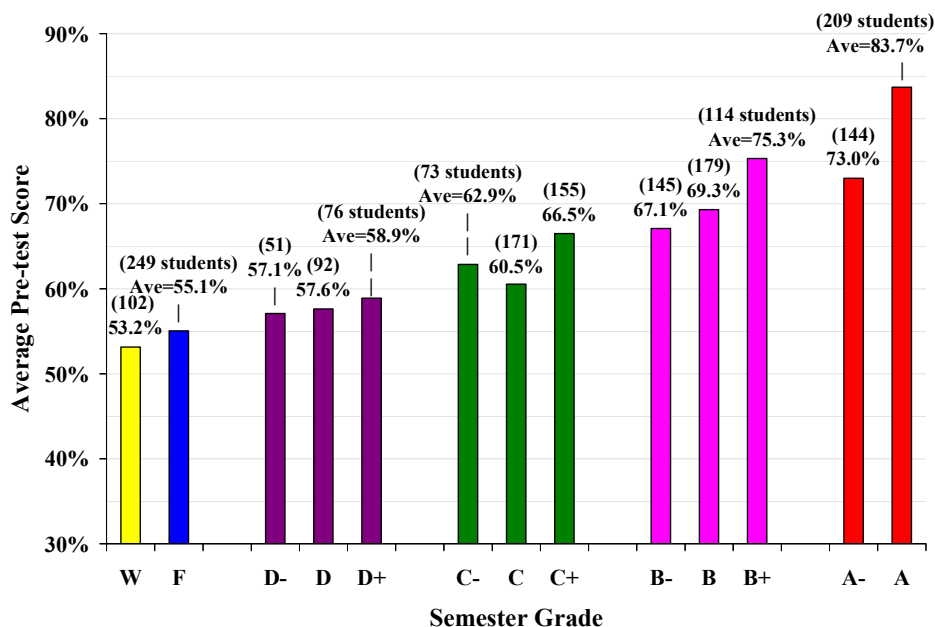

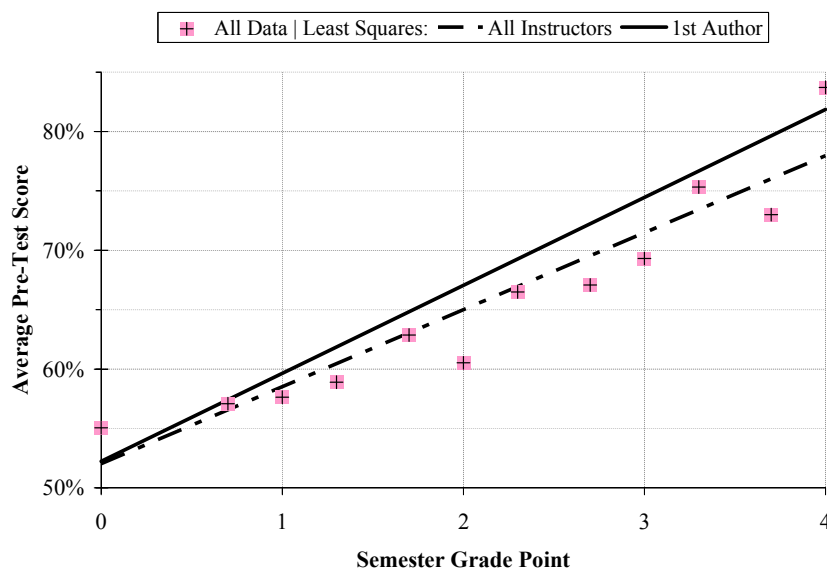Figure 3 – Average Pre-test Score as a Function of Semester Grade.

Figure 4 – Beginning-of-semester Pre-test Averages as a Function of End-of-Semester Grades.

The results of this section are consistent with the work of Stefi et al. [4], who developed a Statics concept inventory test consisting of multiple choice type questions that was administered at the beginning and end of the semester to gauge student improvement. Their concept inventory test was used at three different institutions, and correlation coefficients between the concept inventory tests and

semester-wide exams ranged from 0.332 to 0.625.[5] Thus, Stefi *et al.* found a moderate correlation between their Statics concept inventory test and semester-wide exams. The moderate correlation of 0.440 found between the WSU Statics pre-test and the end-of-semester grade is within the same range as the results of Stefi *et al.*

The discussion up to this point has focused on course-wide correlation of pre-test scores with end-of-semester grades involving large datasets of students. However, individual performance does not necessarily follow the behavior of the entire class. This is illustrated in Figure 5, which shows the distribution of pre-test scores for each whole letter grade. From the graph, it can be seen that approximately 5% of the students who received an F for their semester grade scored in the 90s on the pre-test. If the pre-test score was a precise predictor of individual performance, then these students should have earned an A in the course. Conversely, approximately 14% of the students who received an A for their semester grade scored below 60 on the pre-test. These students may have been capable students who did not take the pre-test seriously because of the low (3-5%) semester grade weight of the pre-test. Another possibility is that they had weak prerequisite knowledge but put significant effort to learn the material during the course. The varying distributions of pre-test scores for each whole letter grade shown in Figure 5 suggest that the pre-test may not necessarily be a good predictor of individual performance. However, the overall results suggest that the pre-test is a reasonably good indicator of class-wide student capability and prerequisite knowledge level, provided that the number of students is large enough (100+) to be statistically meaningful. Since the pre-test is given at the beginning of the semester to predict end-of-semester performance, the pre-test would be a useful tool for an instructor to use to measure the effect of changes in teaching format, examination structure, or active intervention for student success.
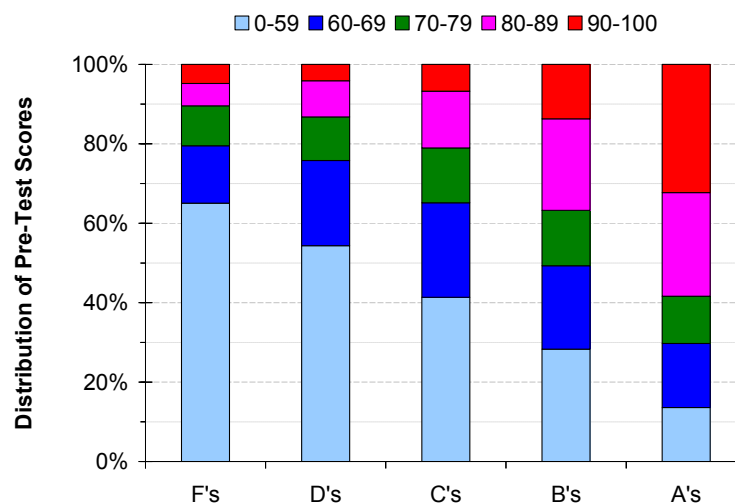

Figure 5 – Pre-test Score Distribution as a Function of Semester Grade.

Predicting Class GPA From Pre-Test Averages for the First Author's Classes

Returning to the question raised in the Introduction regarding the discrepancy in class GPAs between the 50- and 75-minute classes taught by the first author – because of the ability of the pre-test to quantify incoming student capabilities, the pre-test scores of both classes were examined to see if any conclusions could be drawn about potential differences between the two groups. Least squares fit lines between the average pre-test scores and the semester grade were generated for both the 50- and 75-minute classes of the first author to determine if there were any differences in the prerequisite knowledge between the two classes. In order to generate these models, only data from classes taught by the first author was used because of the variation in the slope and intercept of the fits for different instructors, as noted in Figure 4. The dataset included a total of 298 students: 130 students in four

sections of 50-minute classes, 107 students in three sections of 75-minute classes, and an additional 61 students from two sections of 60-minute classes taught five days a week for eight weeks in the summer. Since the structure of the final exam during the summer term is different, with the final exam being administered over two days instead of one, the summer students were omitted from the dataset during the previous discussion about the correlation between the final exam score and the course grade.

Figure 6 presents a plot of the average pre-test score versus the course grade using only data points from the first author's classes as well as least squares fits for different sets of that data. The grey squares correspond to the pre-test averages at each grade point level. The solid line in Figure 6 is identical to the one in Figure 4, representing the least squares fit for the entire dataset of 298 students taught by the first author. The Pearson correlation coefficient between the pre-test scores and the course grades for the dataset of the first author's classes is 0.457, indicating that there is a moderate level of correlation despite the scatter in the grey squares when compared to the solid least squares fit line. This value is slightly higher than the Pearson correlation coefficient of 0.440 for the dataset of all six instructors, which suggests that there is slightly less scatter in the dataset of the first author.
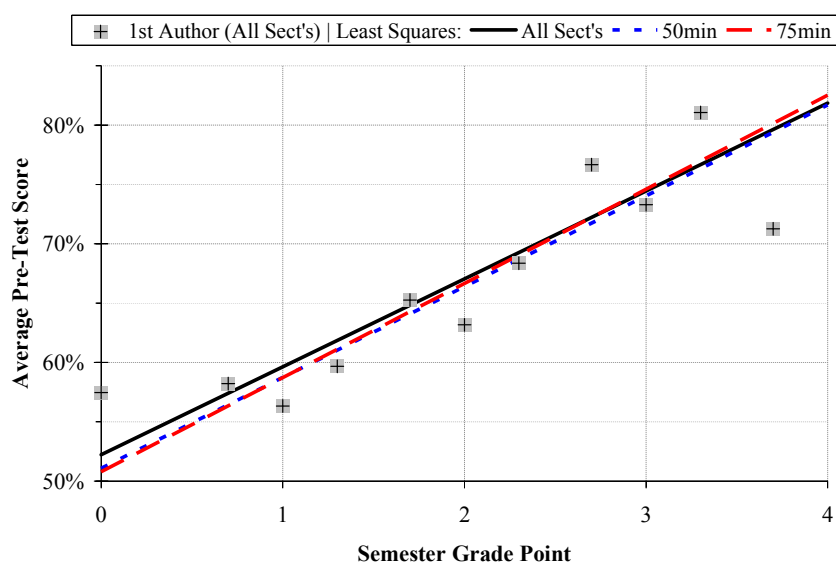

Figure 6 – Beginning-of-semester Pre-test Averages as a Function of End-of-semester Grades for the First Author's Classes.

Figure 6 also includes the least squares fit lines for data associated with the first author's 50-minute classes, shown as a blue short-dashed line, and 75-minute classes, shown as a red long-dashed line. The Pearson correlation coefficients for the 50- and 75-minute classes were 0.491 and 0.503, respectively. Compared to the correlation coefficient of 0.457 for the first author's entire dataset, these correlation coefficient values suggest that there is marginally less scatter in the data for each of the two fits. However, the levels of correlation still remain only moderate.

The general equation for the least squares fit is Pre-test Score = Slope x Grade Point + Intercept. The slope and intercept for a particular fit can be found by examining the corresponding line in Figure 6. This equation can be inverted to obtain an equation for grade point in terms of pre-test score, where Grade Point = (Pre-test Score – Intercept) / Slope. Then, this inverted equation can be used to predict the end-of-semester grade based on the pre-test score. The variability in the slope and intercept of the fits of the 50- and 75-minute classes is relatively small when compared to the fit from the first author's overall database in Figure 6. Consequently, the slope and intercept values of the least squares fit of the first author's entire database will be used in the following predictive work.

The prediction of the end-of-semester GPA for each dataset was performed two different ways. In the first method, referred to as Version A, the average pre-test score for an entire group of students was

used in the equation to predict the end-of-semester GPA for the group of students as a whole. In the second method, referred to as Version B, the pre-test scores of individual students were used in the equation to predict individual grade points, which then were averaged to obtain the GPA of the entire group of students.

Figure 7 presents the prediction results for both versions compared to the actual GPA for the first author's 50-minute classes, 75-minute classes, and all sections. The data for the 50-minute classes, 75-minute classes, and the entire dataset are shown in the left, middle, and right groups of three bars, respectively. Each group of three bars consists of the actual data on the left, version A in the middle with vertical stripes, and version B on the right with horizontal stripes. The specific vertical scale values are not included because class GPA averages are not disclosed to students. However, grid lines are drawn every 0.1 grade points to give a general scale, and a distance of 0.3 grade points represents a plus-minus letter grade, as shown on the graph. From Figure 7, it can be seen that Version A, which is based on the average pre-test score of the class, predicted the GPAs of an entire class within 0.1 grade points of the actual GPA. Version B, which obtains individual GPAs first before averaging them to find the class GPA, underestimated the GPA by 0.1 to 0.2 grade points. Thus, the pre-test can be used to predict aggregate performance of a large number (100+) of students. However, it is not reliable at predicting individual performance as discussed before.
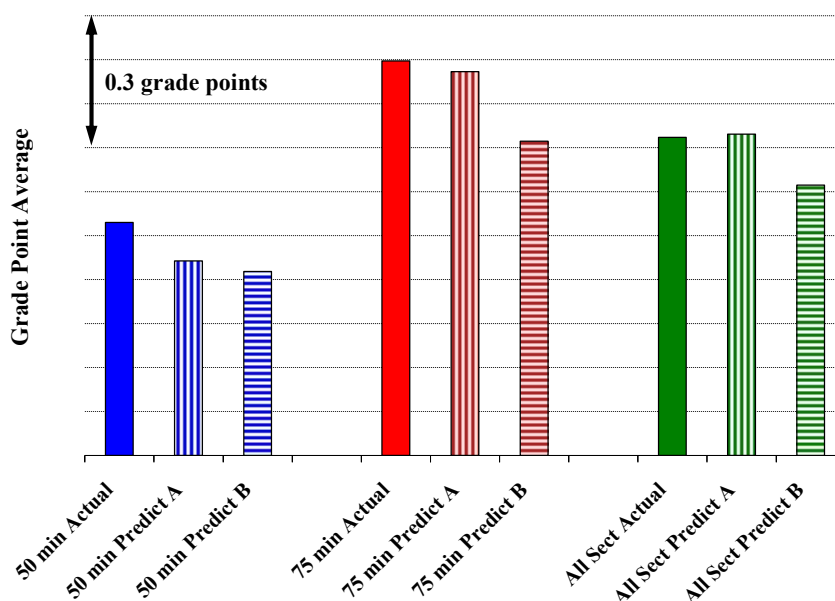

Figure 7 – Comparison of Actual versus Predicted End-of-Semester Class GPA based on Pre-test Score for the First Author's Classes.

The predictive modeling presented in Figure 7 shows that a difference of approximately 3% in the average pre-test scores between the 50- and 75-minute classes generated a difference of 16.5% in the GPAs. This result indicates that the students in the 75-minute classes were more capable and had better prerequisite knowledge than the students in the 50-minute classes, which resulted in better end-of-semester performance. Whereas individual students can overcome a 3% deficiency in preparation and knowledge level as indicated by the results in Figure 4, active intervention or changes in teaching methods will be required to change the class-wide outcome if general incoming student capability and prerequisite knowledge is lacking

Summary

A prerequisite test given at the beginning of the semester in Statics at WSU has been used since 2012 to gauge incoming student capability and prerequisite knowledge. This pre-test was shown to be

moderately well-correlated with the class end-of-semester grade even though the pre-test is given before any substantive teaching of new material occurs. Therefore, the pre-test is a good tool to use as a control metric whenever changes to the course are attempted. After showing that differences in exam structure were not the cause of a 15.5% GPA difference between the 50- and 75-minute Statics classes taught by the first author, the pre-test was used to study the incoming skill of students in the two types of classes. It was found that the 75-minute class had a 3% higher average pre-test score, indicating that those students were more capable and had stronger prerequisite knowledge, which resulted in a higher end-of-semester class GPA.

Future work under consideration includes studies on the effect of other variables such as class size or attendance on student performance while controlling for the capability and prerequisite knowledge level of the incoming students. Another avenue of future work is to examine if possible correlation between student capability and prerequisite knowledge versus course outcomes performance exists. This was not done in the present paper due to the current transition from ABET accreditation criteria a-k to a slightly different set of criteria 1-7.

Acknowledgement

Bibliography

1 K. Moore, C. Jones, and R.S. Frazier, Dec 2017, "Engineering Education for Generation Z," *American Journal of Engineering Education*, Vol. 8, No. 2, retrieved 1 Jul 2019 from https://files.eric.ed.gov/fulltext/EJ1162924.pdf.
2 M. Taylor, 2011, "Teaching Generation NeXt: Methods and Techniques for Today's Learners," *A Collection of Papers on Self-Study and Institutional Improvement – 2011 Higher Learning Commission*, retrieved 1 Jul 2019 from https://taylorprograms.com/wp-content/uploads/2018/11/Techniques_article_2011.pdf.
3 R. Myose, S. Raza, K. Hoffmann, and A. Ghoddoussi, Sep 2014, "Correlating Engineering Statics Student Performance with Scores of a Test over Pre-requisite Material Involving Problem Solving," *Proceedings of the 2014 Midwest Section Conference of the American Society for Engineering Education*, Fort Smith, AR.
4 P.S. Steif and M. Hansen, 2006, "Comparisons between Performance in a Statics Concept Inventory and Course Examinations," *International Journal of Engineering Education*, Vol. 22, retrieved 2 Jul 2019 from https://www.andrew.cmu.edu/user/steif/papers/Statics_Concept_Inventory_Steif_Hansen_2005.pdf.
5 P.S. Steif, A. Dollar, and J.A. Dantzler, Oct 2005, "Results from a Statics Concept Inventory and their Relationship to other Measures of Performance in Statics," *35th ASEE/IEEE Frontiers in Education Conference*, Indianapolis, IN, retrieved 2 July 2019 from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1611927.

Biographical Information

ROY MYOSE is a Professor of Aerospace Engineering at Wichita State University.

SYED RAZA is an Associate Engineering Educator of Aerospace Engineering at Wichita State University.

ELIZABETH ROLLINS is an Associate Teaching Professor of Aerospace Engineering at Wichita State University.

BRANDON BUERGE is an Associate Teaching Professor of Aerospace Engineering at Wichita State University.

NICHOLAS SMITH is an Assistant Professor of Aerospace Engineering at Wichita State University.