The Future of Engineering Education
2024 Annual Conference & Exposition
Oregon Convention Center
Portland, OR . June 23 - 26, 2024
Paper ID #42451

Boosting Achieved-Learning Outcomes with Maritime-Specific Projects in a Machine Learning Course

Dr. Paul Marty Kump, Kansas State University
Ian August

BOOSTING ACHIEVED LEARNING OUTCOMES WITH MARITIME-SPECIFIC PROJECTS IN A MACHINE LEARNING COURSE

0: Abstract

In 2022, we developed a maritime-specific course in machine learning (ML) for undergraduate maritime engineering and naval architecture students in an effort to boost low levels of achieved student outcomes as articulated by the Accreditation Board for Engineering and Technology (ABET). As a major component to the course, we designed a set of mini projects that all utilized the same maritime-related dataset—hypothesizing that domain-specific projects would increase student performance—and we made the projects and solutions publicly available for students and like-minded instructors. This work was met with high praise from colleagues and students, with several positive comments and solicitations for downloads and solutions. But just how important is domain-specific material to an ML course? In this paper, we report on the effects our course had on student learning. The results and lessons learned from our study are valuable information for course developers, instructional designers, and educators looking to boost student performance and craft a domain-specific ML course, be it maritime or other.

We measure the effects of our course on student learning by applying the difference-in-differences (DiD) statistical technique to course data before and after the 2022 course redesign. Included in this analysis are the results of course surveys completed by students, achieved level of ABET learning outcomes, and students' final grades in the course. We find significant increase for ABET learning outcomes; a result that pleased ABET during our institution's most-recent review cycle. We also find a significant increase of +0.50 grade point average (out of 4.00) to students' final grades. With regard to student attitude and perception via course evaluations, a positive change is observed, but we are unable to conclude that the change is statistically significant.

1: Introduction

Incorporating career-related examples in college courses benefits students in a variety of ways. These range from providing students with authentic learning experiences aligned with specific skills and activities one would perform on the job, to a deeper learning and growing of students' interest in the study area. Examples can take the form of course content, real-world case studies, authentic assessments [1], and service-learning opportunities. In today's world, instructors are competing for students' attention with a multitude of engaging, if not distracting, activities like perusing social media on smartphones. The more authentic and engaging the classroom learning experience is, the more students' attention is learning-focused. Job-related classwork, authentic assessments, and service-learning have the potential for more engagement.

Becoming an essential component of the modern world is machine learning (ML), with use cases pertaining to just about every industry to varying degrees—If the reader cannot find an application of ML to his or her own field, the reader is not looking hard enough! Consequently, ML education is more important than it has ever been before, and university programs in engineering and computer science are responding by developing or revising ML and related artificial intelligence courses to best prepare their graduates for employment post academia. As these beneficial measures are implemented, it is best to consider how students might eventually apply ML to their careers so they may reap the benefits a career-minded course offers, as discussed above.

1.0: Prior Work

For the reasons just described, we developed at SUNY Maritime College a course in ML tailor-made for the maritime industry. First offered in 2022, the course was taken by students in electrical, mechanical, and marine engineering, as well as naval architecture; all presumably with interest in the maritime industry. Comprising a major component of the course was a collection of several mini projects we designed [2] to focus on the maritime industry and also demonstrate most of the concepts vital to classical ML. These mini projects utilize the popular Google Colab to offer a flexible environment for a hybrid of report writing and computer programming. The mini projects are publicly available on GetHub, and solutions privately by email [2]. The initial and subsequent course offerings garnered much praise from colleagues and students, and spurred a successful undergraduate research project in ML.

1.1: Course Participants and Learning Objectives

A total of 64 undergraduate students completed our maritime-focused ML course in two course offerings during the years 2022 and 2023. Note that the course materials, instructor, pedagogy, etc. did not change across the two offerings during this timeframe. The number of students by program is shown in Table 1. Students in the electrical engineering program were required to take the course for graduation, typically during their sixth semester. No other students were required to take the course; those who registered sought technical elective credit. Course prerequisites were a basic programming class and a first course in statistics, which itself was a follow-on course to calculus.

Table 1. Number of students by program to complete our maritime-specific ML course during the years 2022 and 2023.

Electrical Eng.	Mechanical Eng.	Marine Eng.	Naval Architecture	**Total**
44	7	7	6	64

Our ML course assessed the following ABET learning outcomes, stated as follows:

The student will demonstrate an ability to

- *(Outcome 5) Function effectively on a team whose members together provide leadership, create a collaborative and inclusive environment, establish goals, plan tasks, and meet objectives.*
- *(Outcome 6) Develop and conduct appropriate experimentation, analyze, and interpret data, and use engineering judgement to draw conclusions.*

While the course assesses both of these objectives, our mini projects only target one. The mini projects were completed by each individual student and were not used for assessment of

Outcome 5. For that objective, students additionally completed a substantial group project and afterwards completed a group health assessment survey.

1.2: Related Work

To the best of our knowledge, there are no studies that report on the effects of incorporating domain-specific examples into the classroom, though related studies have shown that students greatly benefit from authentic assessments—classroom assignments and activities that closely relate to skills and activities in the real world. Service learning, one type of authentic assessment, requires students work with employees at service sites, providing the students opportunities for collaboration and real-world impact. In a project-based service-learning study [3], students reported several positive learning outcomes, including increased interest in the topic and improved communication and problem-solving skills. This type of learning, though not always possible, provides students with a deep educational experience. In another study [4], students worked with real clients as part of a course design project. Students consequently reported a deep understanding of the material upon the course's completion.

2: Data and Methods

This section details the data we collected for our study and the methods we utilized for analysis.

Like in [5], we employed the method of Difference in Differences (DiD) to estimate the impact a change in pedagogy has on students. As we will see, the DiD method requires data before and after "treatment" (i.e., the change in pedagogy). In our case, the treatment is the deployment of the maritime-specific course in 2022. "Pre-treatment" refers to a time period before 2022, and "post-treatment" 2022 and after. We defined our pre-treatment period to begin in 2019 and omitted the year 2020, which was the chaotic pandemic year that saw a mid-semester upheaval of pedagogy. Thus, for this study, the pre-treatment period was defined as 2019 and 2021—two years, symmetric with the post-treatment period 2022 and 2023.

Other than the change to maritime-specific content, the machine learning course remained the same pre- and post-treatment. The course's instructor, syllabus, objectives, expectations, content, etc. did not change from 2019 to 2023, a timeframe we refer to as the study period (omitting 2020). A total of 132 students completed the course during the study period, 68 pre-treatment and 64 post-treatment.

The DiD method can be understood as follows. If x_{pre} and x_{post} represent a metric taken from the pre- and post-treatment data, respectively, then

$$G = x_{post} - x_{pre} \quad (1)$$

is the gain of the treated data. In our case, since the course otherwise remained the same, the treated gain G in Equation (1) is due to the conversion of the course to a maritime-specific format plus inherent model variability due to unforeseen factors.

For comparison purposes, the DiD method additionally requires control data having similar qualities as the treated data. As in [5], good control data for our study utilize the same study period or participants, or are recorded from courses with similar rigor, format, pre-requisites or skillset. We can define the gain of the control data in a similar fashion as

$$\hat{G} = \hat{x}_{post} - \hat{x}_{pre}$$

Accounting for the variability of student performance occurring by unforeseen factors, this gain $\hat{G}$ may be interpreted as the gain the treated data would have seen had it not been treated.

The DiD test statistic is then

$$\beta = G - \hat{G} \quad (2)$$

By subtracting out the gain of the control data in Equation (2), one estimates the net gain resulting from the treatment, hence the name "Difference in Differences."

2.0: Final Course Grades

The final course grade a student earns is an obvious choice of metric to measure our course's impact. We chose this metric because it encompasses the student's performance over all of the course material, is convenient, and has been carefully calibrated by the seasoned instructor. The final letter grade of each of the 132 students who completed the course during the study period was recorded for analysis, and the pre- and post-treatment grade distributions are illustrated in Figure 1. No A+ or D- is awarded at the institution. For the purpose of DiD statistical analysis, the grades were converted from letters to a 4.00 grade scale based on the matrix in Table 2.

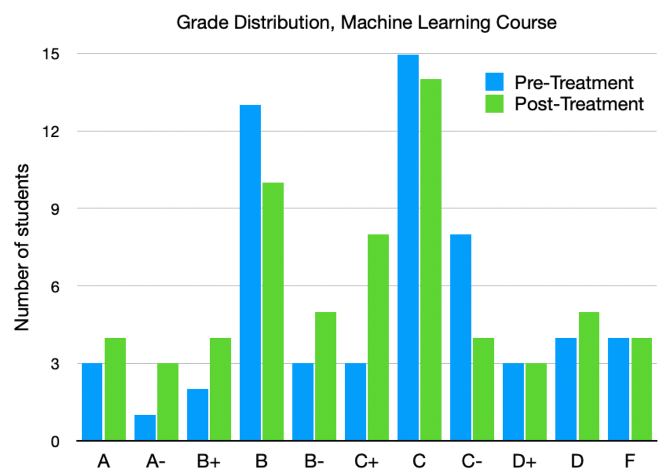

Figure 1: Student final grade distribution in Machine Learning during the years the study period.

Table 2. Matrix converting letter grades to numerical grades.

A	A-	B+	B	B-	C+	C	C-	D+	D	F
4.00	3.67	3.33	3.00	2.67	2.33	2.00	1.67	1.33	1.00	0.00

For the control data, we computed the average of the students' final course grades in the two prerequisite courses, programming and statistics. These courses are truly prerequisites; our machine learning course is heavy on programming and statistics, and very little time in the course is dedicated to reviewing them. It had been the instructor's observation that students' grades in machine learning are correlated with their grades in the two prerequisite courses. These two prerequisite courses ideally remained unchanged throughout the duration of the study period.

One prerequisite grade for four of the 132 student-subjects was not readily available, for various reasons. For these four students, we used the only prerequisite grade that was available to us, either programming or statistics, and did not take an average. Figure 2 shows a scatter plot of the 68 pre-treatment data points with the machine learning grade on the vertical axis, and the control grade the horizontal axis. The figure clearly demonstrates a strong correlation (of 0.814) before treatment began between the prerequisite grades and the machine learning grades. Thus, the prerequisite grades serve as a useful control.

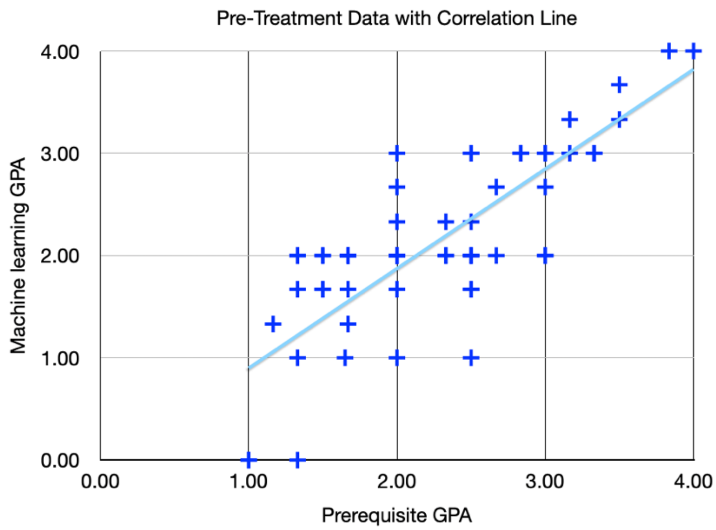

Figure 2: Pre-treatment data points scattered to show the correlation between final course grades in machine learning and the average course grades in the prerequisite courses.

2.1: ABET Learning Outcomes

The machine learning course assesses the two ABET outcomes enumerated in the introduction, and we desire to measure the impact of our course on each outcome with DiD analysis. This section briefly describes our ABET assessment strategies, and how we chose corresponding control data for DiD analysis. At our institution, ABET data are only assessed and recorded for students whose programs require the course for graduation. Therefore, data from only electrical engineering students are considered in this section. Furthermore, the data are only recorded from students who passed the course, which is standard practice for ABET assessment.

Outcomes 5 and *6* were assessed via a group project assigned near the end of each semester, which in part required students to complete Ullman's Team Health Assessment [6] based on their group experience. Their survey responses were used to assess *Outcome 5*. Each student within the group wrote a report and received an individual grade from the instructor based on their demonstrated ability to develop and conduct experimentation, interpret data, and draw conclusions. These individual grades were used to assess *Outcome 6.* Note that elaborate measures were taken to accurately assess each individual student within the group. The assessment for *Outcomes 5* and *6* remained unchanged throughout the study period.

Selecting control data for ABET learning outcomes is challenging. ABET data are aggregated by course and program but are not linked to individual students. So, it is impossible to track an individual student's progress over time. However, our electrical engineering students typically complete a required course in their fourth semester—one year before machine learning—that, conveniently, assesses both *Outcomes 5* and *6* and is taught by the same instructor as ML. So, we selected this course's ABET data to serve as our control data, and we selected the data from one year prior to each year in the study period, because we assumed that approximately the same cohort of electrical engineering students accounted for those data as in our machine learning course.

ABET data at our institution are scored as follows (in regards to satisfying the outcome):

- 1: *Not at all*
- 2: *To a limited extent*
- 3: *To a moderate extent*
- 4: *To a great extent*

The pre- and post-treatment distributions of ABET scores in machine learning are shown in Figure 3 for both *Outcomes 5* and *6*.

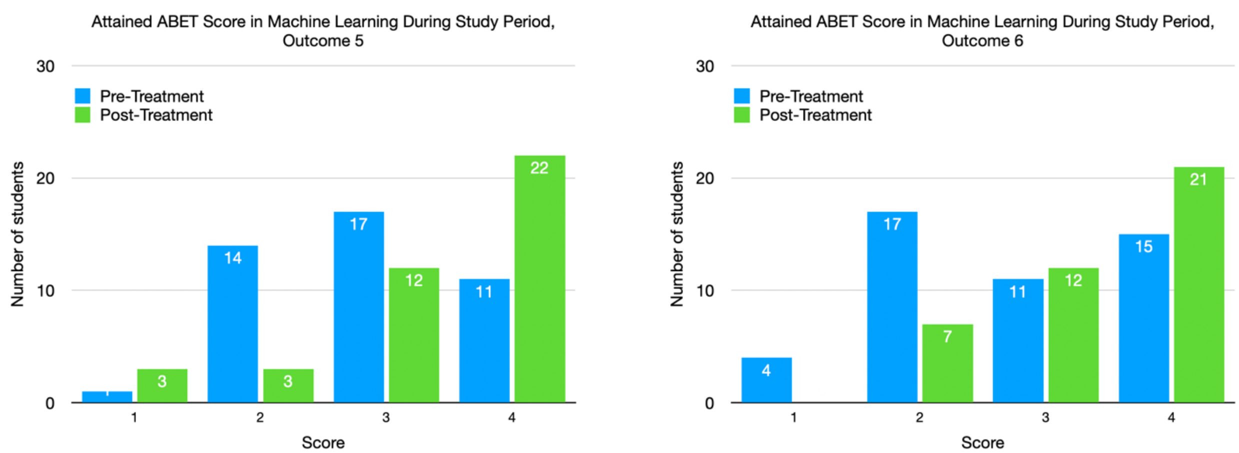

Figure 3: ABET outcome attainment in machine learning during the study period.

2.2: Course Survey

Students completed an anonymized course survey near or soon after the completion of the machine learning course every year during the study period. Two interesting prompts on the survey were

> *Prompt 1: The instructor for this course stimulated my thinking.*
>
> *Prompt 2: The instructor for this course made the course interesting.*

These prompts were selected for this study because their responses indicate the student's perspective and attitude towards the course, which interestingly can affect their own learning [5,7]. Students responded to the prompts by selecting one of the following four responses, with the numerical scores as shown:

- 1: *Strongly disagree*
- 2: *Disagree*
- 3: *Agree*
- 4: *Strongly agree*

Completing the survey was voluntary, and not all students obliged. The response distributions for both prompts and pre- and post-treatment are shown in Figure 4.

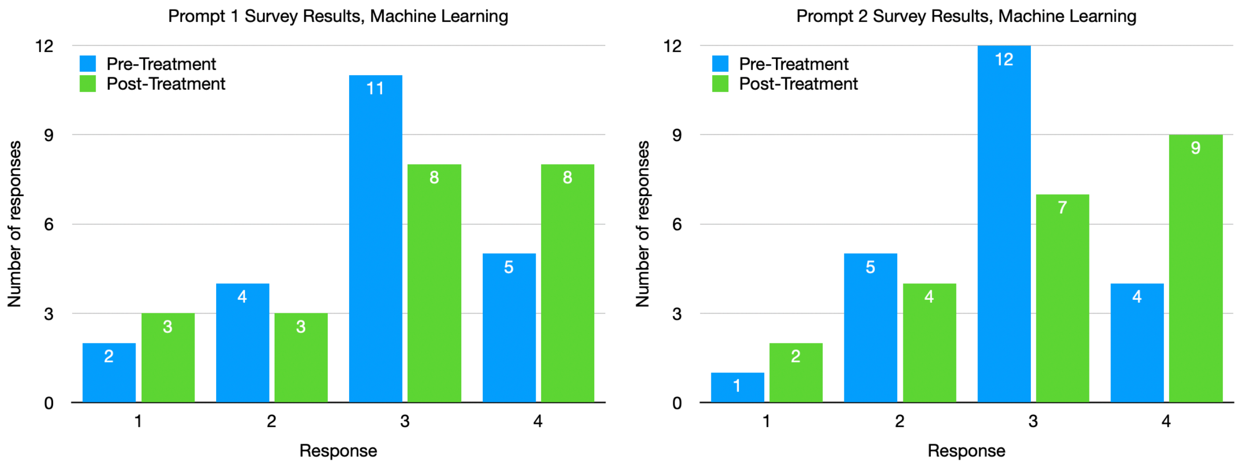

Figure 4: Survey results in machine learning during the study period.

Control data were not selected for the survey data (and DiD analysis was not performed), because we believe no other course at the institution had truly similar qualities to the machine learning and captured the essence of what a control course should. That is, we did not believe we could find a course whose survey results were correlated with ours (similar to Figure 2), taught by the same professor or by one with similar pedagogy, and with similar rigor and content that would interest students in a similar manner as machine learning. Further, survey results are kept private, so it was not feasible to inspect results from courses taught by other instructors.

3: Analysis

3.0: Impact on Final Course Grades

The average of the final course grades among the 64 students comprising the post-treatment group was $x_{post} = 2.73$, and among the 68 students in the pre-treatment group was $x_{pre} = 2.31$. Referencing Equation (1), the observed gain of the treated data is then $G = +0.42$. The control data for the same two groups of students showed $\hat{x}_{pre} = 2.57$ and $\hat{x}_{post} = 2.49$ so that $\hat{G} = -0.08$. Following from Equation (2), the DiD test statistic is $\beta = 0.42 - (-0.08) = +0.50$. That is, students saw an average net increase of +0.50—half a letter grade—to their final course grade in our maritime-specific machine learning course.

Figure 5 plots the average numerical final course grade attained by students by year during the study period. In the figure, the vertical dotted line illustrates the development of our maritime-specific course; to the right is the post-treatment period, to the left, the pre-treatment period. The blue circles—the grades attained in the machine learning course—clearly demonstrate an increase after treatment. Were the students in the post-treatment period simply better students? No, because their grades in the control courses were roughly the same as students in the pre-treatment group. In fact, the figure shows in green triangles that post-treatment students' control grades were even a bit lower.

Were these observations simply observed by chance? To answer this question, we tested for statistical significance by applying random inference [5]. We considered the null hypothesis to be that converting the course had absolutely no effect on the students' final course grade. Random inference estimates the probability of observing at least the measured test statistic $\beta = +0.50$ under the condition that the null hypothesis is true. For brevity, we do not explain the details of random inference here, but instead direct the reader to [5]. After completing 1,000 random simulations, we found a p-value of 0.004, implying that the conclusion that our course improved student grades by +0.50 GPA is statistically significant even at the 99% confidence level.

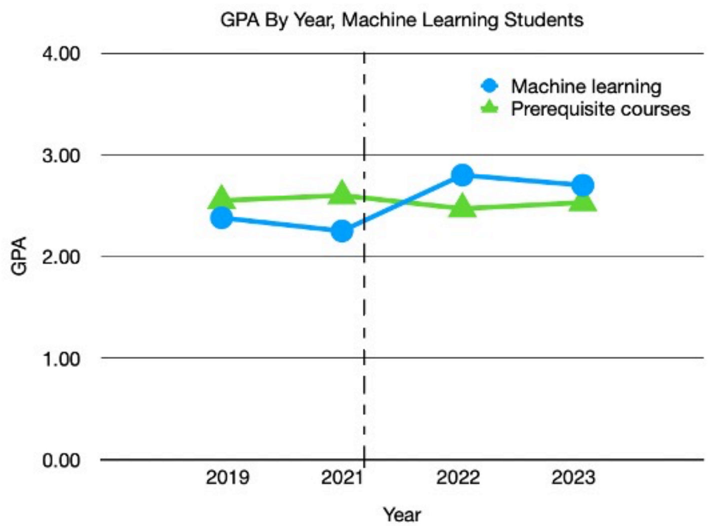

Figure 5: Numerical grades attained by machine learning students during the study period and their corresponding average grades in the prerequisites. The vertical dotted line is the development of our maritime-specific course.

3.1: Impact on ABET Learning Outcomes

We first compute the impact on ABET *Outcome* 5. The average scores of the two distributions in Figure 3, left panel, were $x_{post} = 3.33$ and $x_{pre} = 2.88$. Consequently, the observed gain in Equation (1) is $G = +0.45$. The control data for the control course one year prior showed $\hat{x}_{post} = 3.20$ and $\hat{x}_{pre} = 2.71$ so that $\hat{G} = +0.49$. The DiD test statistic in Equation (2) is $\beta = 0.45 - 0.49 = -0.04$. That is, the student attainment of ABET *Outcome 5* score was on average 0.04 (on our four-point scale) less in our maritime-specific course. We found a simulated p-value greater than 0.05, implying no statistical significance.

We next compute in a similar manner the impact on ABET *Outcome 6*. The average scores of the two distributions in the right panel of Figure 3 were $x_{post} = 3.35$ and $x_{pre} = 2.79$, and the observed gain is $G = +0.56$. The control data for the control course one year prior showed $\hat{x}_{post} = 3.03$ and $\hat{x}_{pre} = 2.95$ so that $\hat{G} = +0.08$. Therefore, the test statistic is $\beta = 0.56 - 0.08 = +0.48$, so the student attainment of ABET *Outcome 6* score was on average 0.48 higher in our maritime-specific course. With a p-value of 0.009, we found statistical significance at the 99% confidence level.

Figure 6 plots the average ABET score attained by students by year during the study period for *Outcomes 5* (left) and *6* (right). For *Outcome* 5, the figure shows with blue circles—the machine learning course—a mostly steady increase in the attained outcome excepting the decline in year 2023. From the figure, it is not clear that the development of our course (the vertical line) played any role in this increase, since there was an increase in the pre-treatment data. Moreover, the control course, which was not treated, also shows a similar pattern.

The right pane of Figure 6, *Outcome 6*, tells a different story. From the figure, we observe a sharp increase in attained scores after the development of our course, while the control data remained roughly flat. This figure clearly shows a positive impact of our course on attainment of ABET *Outcome 6*.

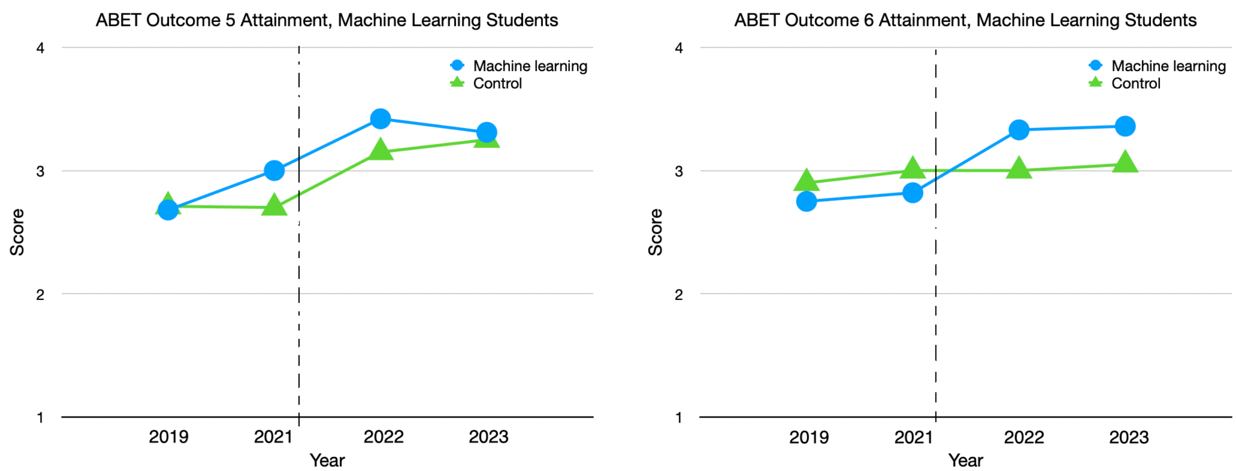

Figure 6: Average ABET scores attained by machine learning students in the machine learning course and also a control course. The vertical dotted line is the development of our maritime-specific course.

3.2: Course Survey

For the prompt *the instructor for this course stimulated my thinking*, the average response value pre-treatment was 2.86, well above *Disagree* (2), and just below *Agree* (3). The same prompt saw an average value of 2.95 for the post-treatment dataset, just about *Agree*. These average values can be observed in Figure 4 (left). Without control data, it is difficult to say whether the increase of +0.09 is due to change in pedagogy or just insignificant fluctuation brought upon by a different cohort of students. (The authors will assume the former and deny the possibility that their efforts were futile!) It is reasonable to believe that incorporating domain-specific material into the course increased student stimulation.

The prompt *the instructor for this course made the course interesting* saw similar numbers. The response data, shown in Figure 4 (right), average to 2.86 pre-treatment and 3.05 post-treatment. Like the previous prompt, this change of +0.19 in encouraging but may be by chance. However, it is again reasonable to believe that domain-specific assessments and examples increase student interest.

4: Discussion

Clearly, students enjoy positive learning effects when the instructor includes interesting examples and assessments into the course. This is made clear by our study—Final course grades and ABET outcome attainment significantly increased, and course survey results showed an encouraging increase in student interest and stimulation. Of course, this conclusion assumes students in a maritime engineering program indeed are interested in the maritime industry.

ABET *Outcome 6* attainment saw an increase of +0.48, significant at the 99% confidence level. So, we can say with 99% confidence that our change in pedagogy is the reason for this observed increase. ABET loves this. It has been our experience at two institutions and department chair at one that an ABET site visit will surely prompt the important question *What changes to courses or curriculum have been made in response to collected ABET data?* A course design (or redesign) to include domain-specific examples like in this study is an easy way to pacify the ABET team, at least for this question. *Outcome 6* is particularly affected in an ML course, since the subject matter is heavy on experimentation and data analysis. Students improved upon their ability to analyze data, and their experimental conclusions were more relevant, accurate, thorough, and logically sound.

In our study, we did not see a significant change to ABET *Outcome 5*—not a surprise, however, since our domain-specific assessments were individual assignments, not groupwork. Our assessments were not targeted to affect ABET *Outcome 5*. Nonetheless, it was interesting to perform the analysis on this outcome. The insignificant result informs us our targeting was accurate and further supports our conclusions. In our ML course, ABET *Outcome 5* attainment is measured with an end-of-semester team project, for which student-teams are wholly allowed to choose their project subject. The team project assignment remained unchanged over pre- and post-treatment periods. Presumably, students were always choosing projects that interested them, so our course redesign did not affect team dynamics.

A significant increase of +0.50 to students' final course grades was observed. While a course grade can be an imperfect measure of student understanding, it is still an important metric, with implications in scholarship, academic probation, student-life, etc. The increase was palpable throughout the semester, the instructor sensed. Quality of homework, final projects and exam scores increased, and these assessments were not typically domain specific. Anecdotally, the increased student interest in the subject matter brought forth by the maritime-related assignments promoted a level of learning that carried over to the non-maritime-related material.

5: Conclusion

This study examined the effect on maritime engineering students of incorporating maritime-related assessments in a machine learning course. The effect was positive and significant, raising final course grades and ABET outcome attainment, as well as increasing the number of positive responses in course surveys. The results are encouraging, and it would be interesting for instructors of other engineering courses in different domains to perform a similar redesign and analysis. Are the effects only observed in machine learning? Only in the maritime discipline?

Due to aggregating the data over the entire student body, one limitation of our work is that it is impossible to study *which* students benefited the most from our course redesign. For example, it might be the case that U.S. Coast Guard License students benefit more than other students.

Maybe lower-performing students (i.e., students entering the course with a lower GPA) benefit more or less than higher-performing students. The methods in this paper can be applied to answer these questions but doing so would demand a much larger class size or many years of data to create subsets large enough for statistical significance.

It is the authors' hope that instructors, students and designers will continue to download our maritime mini projects and solutions for their own study. Hopefully engineering instructors of all domains are excited by the results in this paper and use our mini projects as examples, as they look for inventive ways of incorporating their own novel domain-specific examples and assessments into their own classrooms.

6: References

[1] D. Kolb. (1984) *Experiential Learning: Experiences as The Source of Learning and Development*. Prentice-hall, 1984.

[2] V.A., Ventricelli, P.M. Kump and V.-H. Bui. (2023) "Tuning the Parameters: A Maritime-Tuned Machine Learning Course," *2023 ASEE Annual Conference*, Baltimore, MD, https://peer.asee.org/tuning-the-parameters-a-maritime-tuned-machine-learning-course

[3] M.A. McGinley (2018) "University Students Teaching Environmental Science to Primary School Students as Service-Learning: Lessons Learned. SHS Web of Conferences.

[4] E.C. Yuwono, M.D. Damajanti. (2017) "The Benefits and Constraints of Service Learning on Typography Class: Street Vendors' Banner Design Project." SHS Web Conference. 59, pp. 142-144.

[5] P.M. Kump. (2021) "How classroom flipping affects coast guard license students in engineering," *2021 ASEE Annual Conference*, Virtual meeting, https://peer.asee.org/how-classroom-flipping-affects-coast-guard-license-students-in-engineering

[6] D.G. Ullman, *The Mechanical Design Process,* 4th Ed. New York, NY: McGraw-Hill, 2010, 77.

[7] NPJ Science of Learning, "Emotions in Classrooms: The Need to Understand How Emotions Affect Learning and Education," 13 July 2017. [Online]. Available: https://npjscilearncommunity.nature.com/posts/18507-emotions-in- classrooms-the-need-to-understand-how-emotions-affect-learning-and-education.