The Impact of Periodic Low-Stakes Testing on Environmental Engineering Education

Kyle Murray[1] and Nathaniel Sheehan[1]

[1]United States Military Academy

Abstract

Perhaps the most fundamental goal in education is achieving student retention of new information in order to accomplish learning outcomes. Scores of educational studies coupled with a plethora of books and papers over recent decades attempt to identify ideal educational practices which enable students to better retain material taught in the classroom. From these, numerous conclusions further support a range of influential factors that impact student retention, such as the method of instruction, classroom environmental conditions, relationship dynamics between students and teachers, and assessment or testing patterns. The focus of this study centers on this last factor, testing patterns, and how student retention can be improved in order to better accomplish learning outcomes. In our study, we examine and compare the overall value of periodic open and closed-note formative assessments in an engineering course. We analyzed student performance for 202 students enrolled in an undergraduate environmental engineering course that covers several major topics to include environmental engineering fundamentals and chemistry, water and wastewater treatment methods and design, as well as air pollution modeling and control. This semester-long study included a comparison of student performance on major graded events, including midterm and course-end comprehensive examinations along with final term grades between class sections completing open-note, closed-note, and no additional formative assessments. Our study indicates that the type of testing, open versus closed-note, did not have a statistically different impact on overall course performance. However, the study did show a statistically significant increase of 2.97-4.87% in course performance and final averages between sections completing either type of periodic testing, versus sections completing no additional testing. This finding suggests that formative assessments not only serve to achieve better retention in an environmental engineering course, but further support current academic literature asserting that testing in the classroom generally results in improved student performance.

Keywords

Environmental engineering education, engineering education, learning outcomes, retention, periodic testing, low-impact testing, formative assessment

Background and Introduction

Students seek out education to acquire knowledge, develop understanding, and then apply that understanding to the world around them in a thoughtful and constructive way. Society champions the importance of education at a young age to help individuals with limited life experience learn about and understand the world in which they exist. In recent decades, educators continue to focus efforts on defining the “best” method to instruct students to achieve course objectives and outcomes. Many institutions require their instructors to provide tangible evidence that shows students are achieving these course objectives [1]. Detailed course assessments are becoming more

and more common, hoping to illuminate areas for improvement to make courses more effective at educating students. Scores of educational studies coupled with a plethora of books and papers attempt to identify ideal educational practices which enable students to better retain material taught in the classroom. Numerous conclusions found in these works further support a range of influential factors that impact student retention, such as the method of instruction, classroom environmental conditions, relationship dynamics between students and teachers, and assessment or testing patterns. Past educational research and studies support the notion that student learning improves with testing and that variables such as frequency of testing, feedback from testing, and delays in recalling information have measurable impacts on student retention and performance [2, 3, 4]. We sought to measure and harness the positive impacts of testing within an integrated, multidisciplinary environmental engineering program at our institution.

The main focus of this study is to assess parameters involved in student testing and evaluate the subsequent impact on overall student retention of and performance with the course material throughout a single semester. Impacts from the following student testing parameters were assessed: open versus closed-note testing and announced versus unannounced testing. This study aims to examine the value of periodic low-stakes testing among non-engineering students enrolled in an environmental engineering course as part of a required engineering sequence at the United States Military Academy (West Point) [5, 6].

Lastly, the faculty at West Point seek to be accessible to their students [7] and improve student learning in the classroom. We are encouraged to achieve both of these fundamentals to enhance the learning experience of our students. Part of the motivation for this study stems from our desire to improve the learning experience in our own classrooms. This study illustrates our desire to assess student testing in order to improve overall student retention and performance.

Methods

Design Parameters of the Study

The assessed student population consisted purely of non-engineering majors (e.g. English, Math, Leadership, History, etc.) who must each complete the three-course environmental engineering sequence, taken during their junior and senior years. At West Point, all students, regardless of major, must either major in an engineering field or take a three-course engineering sequence in order to graduate. The course, EV350: Environmental Engineering Technologies, that was chosen for this study is the second in that three-course engineering sequence and is considered a high-enrollment course with 202 students enrolled. At this point of the environmental engineering sequence, students have already learned basic environmental science fundamentals to which they can now apply in designing solutions to complex environmental problems dealing with water and air pollution in developed societies. We provided additional testing to four out of the twelve sections across the course, consisting of a total of 67 students. We were subsequently able to compare between sections with different types of testing and sections with no additional testing at all.

The Formative Assessment

We chose to evaluate and assess student retention using a series of ten formative assessments administered throughout the semester. Formative assessments are an effective tool to both assess

student understanding and retention of course material, while also providing a means to transmit feedback to the student on the overall quality of learning taking place. Formative assessments specifically targeted on individual topics help clarify to students what they should be learning and grasping from the course [8]. The formative assessments given as part of this study were focused on key topic areas that students are tested on during the semester as part of a series of comprehensive exams throughout the course. Furthermore, we designed our formative assessments to be simple, low-stake quizzes given every four to five lessons, on average. A 1991 study [9] determined that students given at least one test over an average 15-week semester scored better on examinations than students who were not tested during the semester. These quizzes would focus on material already covered to ensure recall of information previously learned in the course. The assessments themselves consisted of an approximately even distribution of multiple choice, true-false, matching, and fill-in-the-blank questions, with one short answer question requiring students to describe a specific concept or list out fundamental principles/aspects of a theory or process. These formative assessments were designed to be taken within 10 minutes at the beginning or end of class and were worth a total of 7 points each for a total of 70 points across the semester. This equated to 7% of the course grade, aligning well with the concept of frequent, low-stakes testing. To maximize the feedback benefits of formative assessments, instructors would go over each of the answers with the class as a whole, addressing common mistakes trending across the section and then answer questions for individual students regarding the assessment.

Varying Parameters

As part of this study, we also evaluated the impacts of administering open-note quizzes for two sections and closed-note quizzes for the remaining two sections throughout the entire semester. We sought to better assess the value of open versus closed-note testing in the classroom. Student study habits and preparation for class may vary if they know they are allowed to use references or not. Sections given open-note quizzes were allowed to use hardcopy notes only. Textbooks and digital notes were not allowed. Closed-note quizzes did not allow any references of any kind.

Additionally, we administered announced and unannounced quizzes as part of this study. An equal number of quizzes were announced and unannounced for all four student sections. This was done to assess the impact of "pop" quizzes versus known quizzes from the student's perspective. Similar to open and closed-note testing, we were able to evaluate the overall value of administering announced and unannounced quizzes.

Results

Overall, our study found that students who were tested, regardless of open versus closed-note or announced versus unannounced formats, scored on average 2.97 - 4.87 % higher on major graded events versus students who were not tested throughout the semester. Figure 1 displays the difference in student performance on 4 major graded events – Written Partial Review (WPR) 1, 2, and 3, and the Term End Exam (TEE) – and includes the compared course end average as a result of testing or not testing students. Using a standard t-test, our statistical analysis determined that a significant difference exists between student performance from sections being tested and not being tested. Our t-test results produced t-values that were greater than the critical t-values and p-values less than 0.05 for all five performance events in Figure 1. This indicates a statically significant difference between the performance event mean-values of the two data groups assessed, sections tested versus sections not tested. The following are the individual mean differences for each

student performance event, with academic performance in sections tested being higher than sections not tested – WPR 1 (4.87%), WPR 2 (3.40%), WPR 3 (4.05%), TEE (3.43%), and Course End Average (2.97%). See Table 1 for a brief summary of our t-test statistical analysis. This table includes the distribution of students receiving periodic quizzes and those who did not.

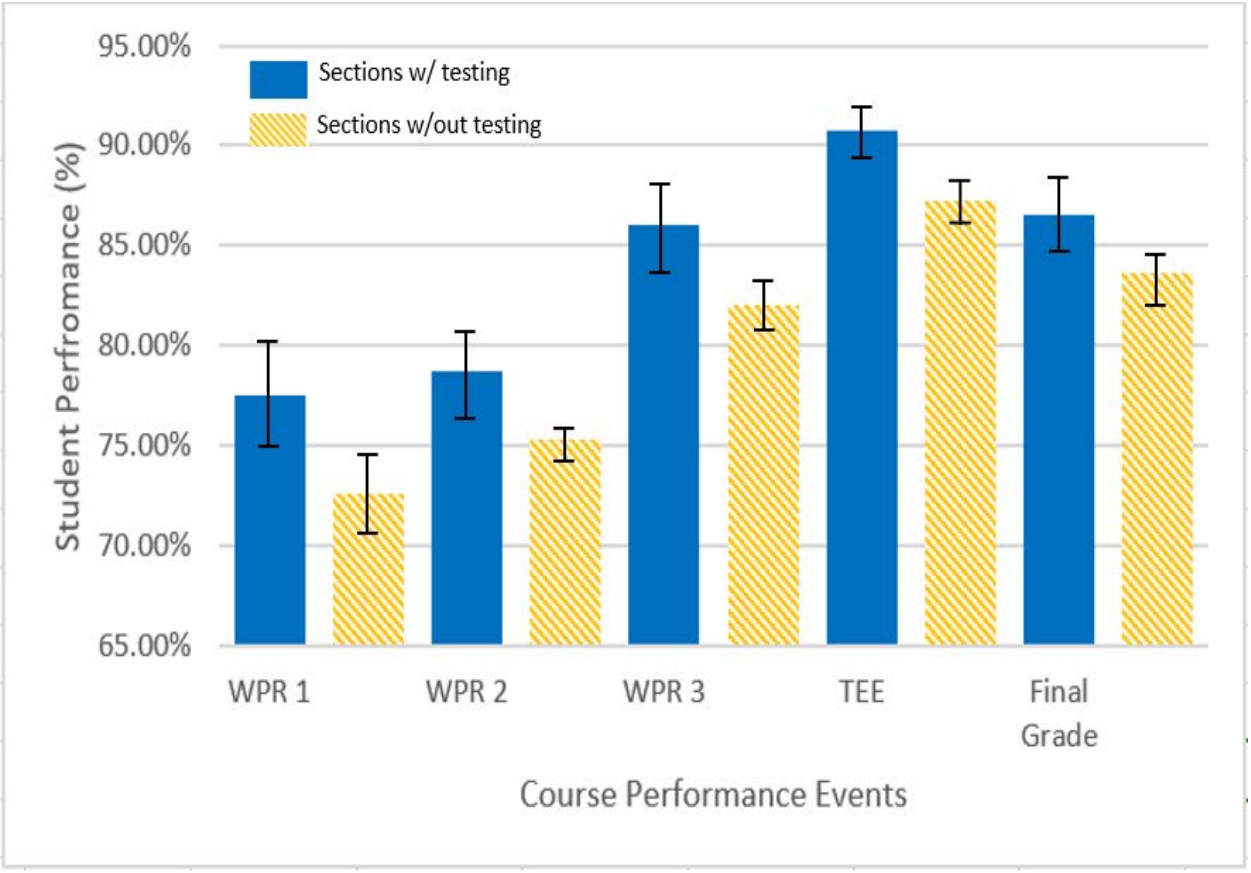

Figure 1: Course and Major Graded Event Performance between Sections with Additional Testing and Sections without Additional Testing. Blue represents sections where students received periodic low stakes testing. Orange stripes represents sections that did not receive the additional, low stakes testing. Performance on every major event and overall course performance were statistically significant. Bars shown represent 95% Confidence Intervals determined by statistical analysis using standard t-test. (Testing: n = 67 and No Testing: n = 135)

Our subsequent evaluation of administering open versus closed-note quizzes yielded results slightly favoring closed-note testing, but ultimately is inconclusive due to the minimal difference between student performance averages. These results are presented in Figure 2. Our statistical analysis using the t-test indicated a statistically significant difference of 3.87% in only WPR 1 mean values favoring closed-note testing. Statistical analysis of the remaining student performance events produced mixed results with some slightly favoring open-note testing, while others slightly favored closed-note testing. Other than WPR 1, the remaining performance events did not yield statistically significant differences between the mean values of the data groups assessed. Likewise, our assessment on the impact of announcing some quizzes while not announcing others yielded slightly better results for the announced quizzes but remains mainly inconclusive. One would certainly expect students to prepare differently when they are aware an exam is coming. Additionally, we found no statistically significant difference in student

performance between sections taught by different instructors or at different times of day as it pertains to conducting open versus closed-note or announced versus unannounced testing.

Table 1: Summary of t-test Statistical Analysis. This summary shows the following for each assessed performance event: mean averages, mean difference, the 95% confidence intervals, and pertinent t-Test values (t-Value, Critical t-Value, and P-Value).

Performance Event	Mean Performance Averages		Mean Difference	95% Confidence Intervals	t-Test Analysis		
					t-Value	Critical t-Value	P-Value
WPR 1	Sections with Testing (n = 67)	77.50%	4.87%	74.08 - 80.92%	2.1876	1.9719	0.0299
	Sections Without Testing (n = 135)	72.63%		70.06 - 75.19%			
WPR 2	Sections with Testing (n = 67)	78.66%	3.40%	76.02 - 81.31%	1.9916	1.9719	0.0478
	Sections Without Testing (n = 135)	75.26%		73.29 - 77.22%			
WPR 3	Sections with Testing (n = 67)	86.07%	4.05%	83.56 - 88.58%	2.3041	1.9719	0.0222
	Sections Without Testing (n = 135)	82.02%		79.94 - 84.10%			
TEE	Sections with Testing (n = 67)	90.71%	3.43%	88.97 - 92.46%	2.6452	1.9719	0.0088
	Sections Without Testing (n = 135)	87.28%		85.72 - 88.85%			
Course End Average	Sections with Testing (n = 67)	86.56%	2.97%	84.92 - 88.19%	2.7091	1.9719	0.0073
	Sections Without Testing (n = 135)	83.59%		82.32 - 84.87%			

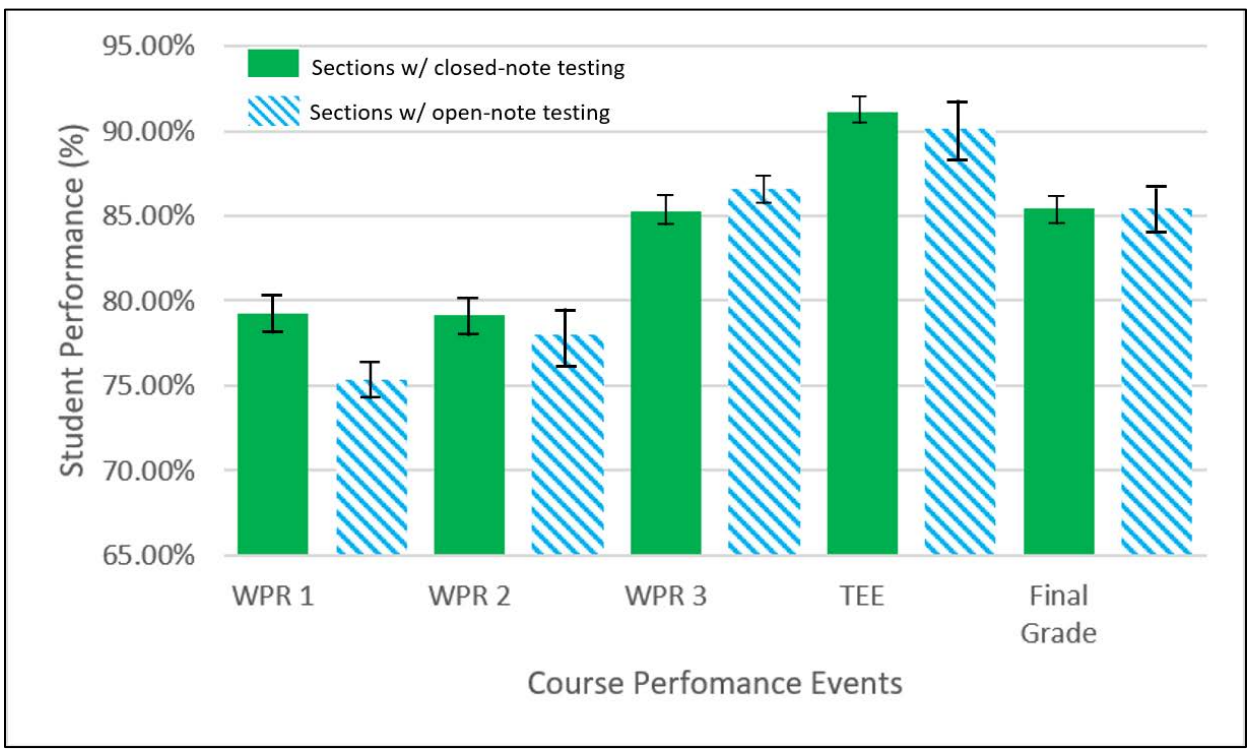

Figure 2: Course and Major Graded Event Performance between Sections with Open-Note Additional Testing and Sections with Closed-Note Additional Testing. Green represents sections where students received closed note testing. Blue stripes represents sections that received open note testing. Performance on every major event and overall course performance were not statistically different. Bars shown represent 95% Confidence Intervals determined by statistical analysis using standard t-test. (Open-Notes Testing: n = 34, Closed-Notes Testing: n = 33)

Discussion

While the overall results of this study may be surprisingly positive to some, they were in the direction we generally expected. One might assume that students majoring in non-engineering fields of study would be dissuaded by taking additional tests in the classroom. However, as our results show, the four sections that conducted additional, low stakes testing throughout the semester, achieved statistically significant higher end of course averages, as opposed to the eight course sections that did not receive the additional testing. These results align with the current literature that shows testing as a powerful tool often employed to increase student recall and performance in the classroom. These students likely gained both experience and competence in the subject matter due to the realistic recall required by the additional quizzing. In all cases, the students were given feedback on their performance following each quiz. This feedback reinforced student assessment of individual performance with the course material and gave students areas to focus their studying.

The second part of our experiment sought to determine whether the expected positive results of additional, low stakes testing, would be positively influenced if the students were able to use their notes or not. Our hypothesis asserts that testing conducted without notes would better solidify the classroom material for the students learning since they would be expected to recall the matter without the aid of notes. The sections not allowed to use notes did score slightly higher overall, but the results were not significantly different than the sections allowed to use notes on the quizzes. The same can be said regarding the impact of announcing quizzes ahead of time. Overall, student performance was slightly better in most cases when quizzes were announced versus unannounced, which is reasonable expected because the average student will likely prepare for an exam they know is coming. This was valuable to understand because our relatively inconclusive results for adjusting testing parameters – open versus closed-note quizzing or quiz announcement – further supports the notion that it is the act of additional testing, and not simply adjusting specific testing conditions, that more greatly impacts student performance. In this case, improved student performance resulted far more from whether additional testing was conducted or not than whether this testing was announced or conducted with or without the aid of student notes.

This study illuminated additional areas of value resulting from the periodic, low stakes testing, that may have helped increase student learning. First, instructors noted if a significant portion of a section incorrectly answered certain questions or seemed to miss specific concepts on a quiz. As a result, these topics could be quickly reviewed by the instructor in the following class to help ensure that students properly understood the material before moving forward in the curriculum. Deficiencies in learning may not have otherwise been left to exist until after the next major graded assignment. This may have resulted in more course points lost and lower end of course averages in the sections being tested. Second, the additional quizzes may have assisted the students by providing additional material to study, and perhaps teaching the students how questions could be asked or worded on an exam, thus guiding the way they studied for future next graded events. This may have been an inadvertent advantage given to the sections who were tested versus those sections not tested.

Overall, our inclusion of low-stakes quizzing throughout the semester, in addition to the regular course instruction, was considered significantly valuable to increasing student learning and performance in the classroom. These periodic quizzes took minimal class time, while serving to

further solidify material retention for students and allowing the instructor to gain feedback on the degree to which certain course topics and material were retained in the classroom. The inclusion of low stakes quizzing in the classroom should be considered for use to achieve the tangible positive results we obtained.

Conclusions

Our results suggest that the inclusion of periodic, low stakes quizzing helps to increase overall classroom performance of non-engineering students taking an environmental engineering course. While no statistically significant differences were found between student sections allowed to use notes and those not allowed to use notes on quizzes, there did exist a statistical difference in student performance between student sections given periodic quizzes at all and those not given any additional quizzes over the semester. Therefore, our overall conclusion from this study is that the act of periodic testing in the classroom by itself positively impacts student learning more than adjusting specific testing conditions. We acknowledge that testing conditions surely impact student performance during testing, but we insist that the act of testing itself has a more significant impact. Due to the cumulative nature of major graded events in the course examined in this study, we further conclude that student retention on key topic areas must also have increased resulting in better performance throughout the semester. Additionally, this study helped us identify and consider other indirect effects that additional classroom testing may have had on achieving better student performance.

How we can improve student retention and learning in the classroom is the idea that seems to be under constant review and study by academic professionals in nearly every field. Future areas of research regarding the value of periodic, low stakes quizzing should seek to address the impacts of adjusting other testing conditions on student performance. In addition to looking at the impact of additional testing by itself in the classroom, this study was limited to adjusting two other testing conditions or parameters, the announcement of quizzes and the use of notes during those quizzes. Other common testing parameters that should be analyzed and assessed include quiz formats and layout in relation to the types of material being evaluated, the time at which quizzes are given following the material being taught in the classroom, and the time at which quizzes are given during a classroom period. For instance, are quizzes completed at the beginning, middle, or end of the class period? Additionally, evaluating adjustments to the physical environment in the classroom could be assessed, such as temperature, audio-visual conditions, and classroom design to include student seating placement and layout.

In the future, the overall purpose of similar educational studies should continue centering on improving student retention and learning in the classroom. While grades and averages are important measures of student performance, retention and learning is our long-term goal for our students to be successful over their careers in their chosen profession. As educators, we have the inherent responsibility to provide our students the best opportunities for them to grow and learn in our classrooms.

References

[1] R. B. Barr, and J. Tagg, "From teaching to learning: A new paradigm for undergraduate education," *Change*, vol. 27, pp. 12-25, November/December 1995.

[2] S. D. Downs, "Testing in the College Classroom: Do Testing and Feedback Influence Grades Throughout an Entire Semester?" *Scholarship of Teaching and Learning in Psychology*, vol. 1, no. 2, pp. 172-181, Apr. 2015.

[3] J. Cranney, M. Anh, R. McKinnon, S. Morris, and K. Watts, "The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting," *European Journal of Cognitive Psychology*, vol. 21, pp. 919-940, Aug. 2009.

[4] R. P. Phelps, "The effects of testing on student achievement, 1910-2010," *International Journal of Testing*, vol. 12, pp. 21-43, Jan. 2012.

[5] E. Martinez, C. Ouellette, L. Plante, B. Wallen, and J. Starke, *An Environmental Engineering Sequence: Deliberately Addressing and Evaluating Attitudes and Knowledge*, American Society for Engineering Education Middle Atlantic States Fall Conference, October 6-7, 2017, Penn State University - Berks, Reading, PA, 2017.

[6] C. Ouellette, L. Plante, E. Martinez, B. Wallen, and J. Starke, *Informing an Environmental Ethic in Future Leaders Through an Environmental Engineering Sequence*, American Society for Engineering Education Annual Conference & Exposition, June 24-27, 2018, Salt Lake City, UT, 2018.

[7] "Most Accessible Professors," The Princeton Review, available from https://www.princetonreview.com/college-rankings?rankings=most-accessible-professors, accessed 23 August 2018.

[8] N. Bakula, "The Benefits of Formative Assessments for Teaching and Learning," *Science Scope*, vol. 34, no. 1, pp. 37-43, Sep. 2010.

[9] R. L. Bangert-Drowns, J. A. Kulik, and C. C. Kulik, "Effects of frequent classroom testing," *The Journal of Educational Research*, vol. 35, no. 2, pp. 117-120, November/December 1991.

Author Biographies

Kyle Murray, PE, United States Military Academy

Kyle Murray is a Major and an Aviation Officer in the United States Army and an Instructor in the Department of Geography and Environmental Engineering at the United States Military Academy. He is a 2007 graduate of the United States Military Academy with a B.S. in Civil Engineering and graduated from the University of California – Los Angeles (UCLA) with an M.S. in Environmental Engineering in 2017. His research interests include wastewater treatment technologies and microbiology as well as air pollution mitigation and control technologies. His current research focuses on engineering education in the field of environmental engineering. He teaches courses in Environmental Biological Systems, Environmental Science, Environmental Engineering Technologies, and Environmental Chemistry. He is also a licensed Professional Engineer in the state of Delaware.

Nathaniel P. Sheehan, PE, PMP, REHS/RS, United States Military Academy

Nathaniel Sheehan is a Captain in the United States Army and an Assistant Professor in the Department of Geography and Environmental Engineering at the United States Military Academy. He is a 2010 graduate of the United States Military Academy with a B.S. in Environmental Engineering and a 2013 Graduate from the University of Arkansas - Fayetteville with an M.S. in Engineering. He teaches Physical and Chemical Treatment, Environmental Science, and Environmental Engineering Technologies. He is also a licensed Professional Engineer in the state of Delaware.