2018 ASEE Annual Conference & Exposition, June 24–27, 2018, Salt Palace Convention Center, Salt Lake City, UT

Paper ID #22083

Classifying the Engineering Mathematics Student: An Investigation of Trends in Learning

Dr. Paul L. Goethals, United States Military Academy

Dr. Paul Goethals is currently an Academy Professor in the Department of Mathematical Sciences at the United States Military Academy (USMA). He holds a B.A. in Chemistry from Indiana University, an M.S. in Mathematics from Florida State University, and a Ph.D. in Industrial Engineering from Clemson University. Dr. Goethals' research interests include experimental design, process optimization, and work with applications in national defense.

Karoline Hood, United States Military Academy

Classifying the Engineering Mathematics Student: An Investigation of Trends in Learning

Department of Mathematical Sciences
United States Military Academy, West Point, NY 10996

In order to advance our educational methods and ensure that our practices result in the attainment of learning outcomes, we rely heavily on teacher feedback through course assessments. These evaluations may enable us to pinpoint teaching techniques that assure a higher level of student conceptual understanding with improved performance. Unfortunately, however, given the nature of the course assessment, this input can be inherently biased and without a true scientific foundation. A number of researchers examining the educational challenges of the assessment process have come to this conclusion. As part of the assessment process for meeting student learning outcomes, the Department of Mathematical Sciences at the United States Military Academy performed a preliminary study in 2017 with its foundation course, Engineering Mathematics – the intent was to identify alternative measures of assessment for student performance. An investigation of trends was conducted to examine the varying degrees of student confidence in acquiring conceptual knowledge of vector calculus, the application of ordinary differential equations, and partial differential equations. A pre- and post-assessment given with graded events enabled the faculty to classify the students into one of several groups and make inferences as to their ability to achieve specific objectives. This ongoing work, which is to be expanded in scope for future terms, may provide insights for identifying trends in learning, specifically with regard to an engineering mathematics program.

1. Introduction

In some philosophical discussions, it is recognized as the Socratic Paradox, i.e. “knowing what you do not know,” – in this light, an individual is considered ‘better off’ knowing that and what they do not know, versus knowing that they do not know [1]. This brings about a number of pedagogical questions for the classroom, some of which may lead to insights in how student confidence relates to achieving student learning outcomes. In 2017, an internal assessment study was conducted in the Engineering Mathematics classroom at the United States Military Academy (USMA), specifically to investigate the effects of confidence on performance. The intent of the experience was to determine if one may identify alternative measures of assessment for student performance. Several student categories were identified analogous to the groups described in the Socratic Paradox, i.e. based upon a student’s assessment on whether they know or do not know the material. In the following section, background information on the course and the student environment are given. Then, in section 3, a review of the pedagogical research on this topic is provided. Finally, upon describing the methodology for the study in section 4, the results of the study are discussed in section 5, followed by conclusions from the assessment.

2. Background: The Learning Environment

As an undergraduate institution, the United States Military Academy offers an array of disciplines for students to major in over their four-year journey to becoming officers in the Army. The engineering sciences make up a very large part of the academic offerings – students may choose to pursue degrees in chemical, civil, mechanical, electrical, environmental, nuclear, or systems engineering. A series of four to five courses in the mathematical sciences, depending upon their selection of major, are taken during their Academy experience, in order to prepare the students for further study in their major programs. The culminating 3-credit hour math course is Engineering Mathematics, which provides a basis of knowledge in vector calculus, the application of ordinary differential equations, and partial differential equations such as the heat and wave equation.

In accordance with the USMA Dean's Policy Memorandums, student learning outcomes are defined as "broad statements that describe the knowledge, skills, behaviors, and/or attitudes that cadets should attain by the time they complete the developmental experience." These outcomes support the academic program goals and the outcomes of the West Point Leader Development System. For Engineering Mathematics, the objective of the course is to provide a solid foundation in the mathematics necessary for continued study and success in the student's engineering discipline. This foundation includes four focus areas: (i) modeling with, solving, and interpreting the solutions of ordinary differential equations, (ii) applying vector calculus to model and solve problems related to vector fields and vector differential and integral equations, (iii) modeling with and solving first- and second-order linear equations, applying linear transformations, and interpreting their solutions, and (iv) modeling with, solving, and interpreting the solutions of classical partial differential equations, such as the heat and wave equations. Within each of the focus areas, supporting goals exist that are further nested within each of these course objectives.

During the academic year in 2017, 51 engineering major students completed Engineering Mathematics during their educational experience. For the course, a student could earn up to 2000 points in graded material, consisting of quizzes, problem sets, a course-wide project, and examinations. The course was taught in four sections by three different instructors, enabling the student-teacher ratio to remain small at roughly 14:1. A typical class consisted of a 20-25 minute period of instruction on a subject with another 30 minutes of the students working problems on the boards (see Figure 1). The low student-teacher ratio facilitated more interaction with the students and a better awareness or assessment of student skills. To ensure the course objectives were being met, the instructors met on a weekly basis as a means to compare and contrast teaching experiences. Emphasis for evaluation was placed on those concepts that students generally fared poorly on with respect to class observations or initial assessments.

Figure 1: The Classroom Environment (no more than 18 students in a class)

3. A Brief Review of Confidence Study Research

Pedagogical studies involving the role of student confidence and its relationship with cognitive performance are prevalent within the educational literature. Early researchers noted a correlation between recognizing how much one knows and their level of confidence [2]. The evaluation approaches used, however, among the many research evaluations are quite diverse. One of the dominant approaches to studying the relationship between confidence and cognitive performance is the ecological approach, whereby the difference between what we know and how much we know is determined by the difficulty level of the tests themselves rather than the individual [3-4]. Another central approach used for the studies is the heuristics and biases technique, where the individual's experience is what determines a lack of calibration with confidence [5-6].

Despite the breadth of confidence studies in the literature, the research is also extremely varied in terms of the subjects, the focal group, and the measurement tool used. Some investigators chose to examine differences in gender [7], ethnicities [8], and even language comprehension [9-11]. A large number of the studies assessed performance in standardized testing, while others specifically investigated confidence in writing or in speaking [8]. The results also put special emphasis on a breadth of attributes, such as personality, metacognition, and cognitive abilities [12-14]. Finally, the methods of evaluation used by researchers range from statistical hypothesis testing, to confidence metrics, scores, and correlation matrices.

In order to gather the data necessary to measure confidence, a typical study requests that the participants provide a rating expressed in terms of a percentage, corresponding to their confidence level with their response [15-16]. This is primarily done after the response to a particular question is made or even after an entire test or examination is taken. One recent study, which is the impetus for this research, instead performed a pre- and post-assessment at the beginning and end of a course to determine if student learning outcomes were being met [17]. The researchers then used a five-item Likert scale to calculate confidence scores using a scale of 1 = "not confident at all," 2 = "slightly confident," 3 = "somewhat confident," 4 = "confident," and 5 = "very confident." The study found that confidence was correlated in some way with academic performance.

It is important to note that although researchers generally recognize the benefits of student self-assessment [18], there are some researchers that point out its inaccuracies in measurement at times [19-21]. A recent study found the literature analyzing student self-assessment to be very limited; specifically, it was found to contain a number of contradictory results, with a large number of researchers in favor of its use and an equally large number of researchers against its use in assessment [22].

4. The Assessment Study

Looking at previous research on these studies, it is apparent that there is no widely used rubric for assessing confidence – a myriad of techniques and methodologies exist. Hence, we did not feel obligated to use one particular approach over another. For our investigation, we sought to perform a pre- and post-assessment not at the beginning and end of the course, but rather at the beginning and end of each in-class graded event, such as a quiz. Prior to receiving the questions for the graded event, the students were asked two questions, "*What do you believe your final score out of #points will be?*" and "*Why did you assess your score at that level?*" Then, immediately after taking the graded event, the students were asked one question, "*Now having taking the (graded event), what do you believe your final score out of #points will be?*" These responses were recorded throughout the academic semester across three different sections of Engineering Mathematics.

Recording the responses before and after the graded events enabled us to make several interpretations with respect to confidence and evaluate different hypotheses. We anticipated being able to classify the students into one of four groups, depending upon their pre- and post-assessment responses (see Table 1) in comparison to their performance. "High performers" were classified as individuals who both (i) scored above the average on the graded event, and (ii) whose pre-test estimate was within one letter grade of their actual performance. "Low performers" (i) scored below the average on the event, and (ii) had a pre-test estimate within one letter grade of their actual performance. "Overconfident performers" were individuals who both (i) scored less than their estimate, and (ii) whose pre-test estimate was more than a one letter grade difference from their actual performance. Finally, "underconfident performers" were individuals who both (i) scored greater than their estimate, and (ii) whose pre-test estimate was more than a one letter grade difference from their actual performance. For our study, we sought to determine relationships between these groups and their performance.

Category	**Short Description**	**Description**
High Performers	*Know what they don't know*	These students are confident they know the material and perform as if they do know the material.
Low Performers	*Know what they don't know*	These students are confident they do not know the material and perform as if they do not know the material.
Overconfident Performers	*Don't know what they don't know*	These students are confident they know the material but perform as if they do not know the material.
Underconfident Performers	*Don't know what they know*	These students are not confident they know the material but perform as if they do know the material.

Table 1: Student Category Classifications

An initial framework of research questions, listed below, were formulated to help guide our analysis of the responses:

- *What is the relationship between the categories and their overall performance in the course?*
- *What percentage of the students are overconfident and underconfident performers and does this change over time (i.e. do the students become attuned to the difficulty of the graded events)?*
- *Do the students generally remain in their same category over time (i.e. could their confidence choices be related in some way to a character trait or their personality, or is it more about knowledge retention)?*
- *Does the relationship between confidence and student performance change when we consider the post-assessment rather than the pre-assessment?*
- *What do the qualitative remarks on the pre-assessment with respect to why they reported that particular confidence level tell us about the outcomes?*

Classifying the students according to different categories enabled us to further assess the role of confidence in performance. Students classified as high or low performers either knew that they were going to perform well and did or knew that they were going to perform poorly and did. Their pre-test estimate before seeing the graded event aligned with their actual performance score. Overconfident students, in contrast, suggested a high performance before taking the graded event but then performed poorly – these individuals are assessed as not knowing what they do not know. Finally, underconfident students suggested a low performance before taking the graded event, but then performed well. These students were classified as not knowing what they know.

For the assessment, three of the four sections in Engineering Mathematics participated in the study, resulting in a total sample size of 36 students. While a much larger sample size is desirable, this quantity enables us to infer some trends in the data through visualization techniques. While it is recognized that typical approaches to comparing groups of data include statistical testing and that there are hypotheses tests specifically for small sample sizes, the intent of this study is solely to explore whether a trend appears to exist when using categories of students such as that defined. If it is determined that a trend appears to exist using this framework, a follow-on investigation would continue with future classes whereby a greater sample size is acquired.

A total of five graded events were used to gather the data, resulting in a total of 360 observations over the duration of the course. It is important to note there are several assumptions made with our study. First, we assume that the process of scoring grades on the events among different faculty members, as well as the instruction received on the various topics, is uniform across the different sections. We also assume that the testing environment remains unchanged throughout the sequence of graded events. Finally, we assume that performance on the graded events relates to the successful learning of the student outcomes.

5. Results and Discussion

Upon gathering the data for the study, the students were sorted by overall performance, best to worst, and further identified according to their quartile, the first quartile corresponding to the top 9 student performances and the fourth quartile corresponding to the lowest 9 student performances. Figure 1 depicts the population, by category, for each of these quartiles.

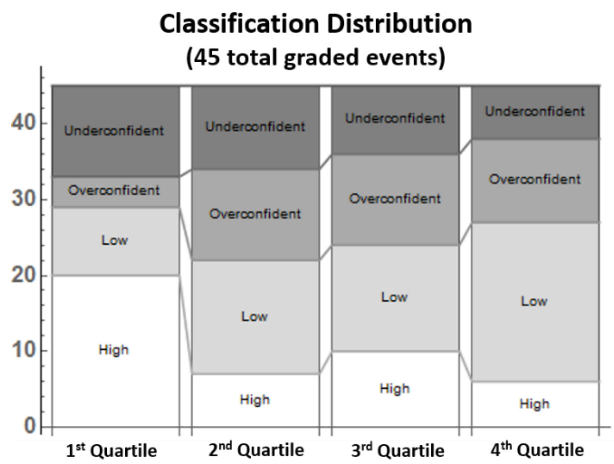

Figure 1: Classification by Quartile (using the Pre-Assessment and Actual Performance)
(1st quartile = top performers, to 4th quartile = bottom performers)

5.1 *What is the relationship between the categories and their overall performance in the course?*

There are several trends that can be observed when considering the distribution of categories by quartile. Intuitively, we would expect that we would have a larger number of high performers in the first quartile and a larger population of low performers in the fourth quartile, and that is what we observe. What is not as intuitive, however, is the decrease in the population of underconfident students as one moves from the first to the fourth quartile. This group of students is not confident that they know the material but perform as if they do know the material. While one would anticipate the top students scoring high, one would not expect them to necessarily pre-assess themselves as more than a letter grade lower than a top grade. This trend could suggest a number of different reasons, one of which is that the quality of underconfidence occurs with all students, but top students are just more capable of overcoming the assessment differential.

Another observation from the classification by quartile is the absence of overconfident students from the first quartile, where the population of overconfident students generally remains unchanged from the second to fourth quartiles. This may suggest that a threshold exists between the top student and other students whereby the quality of overconfidence is more prevalent.

5.2 *What percentage of the students are overconfident and underconfident performers and does this change over time (i.e. do the students become attuned to the difficulty of the graded events)?*

If the students are becoming attuned to the graded events over time, i.e. they gain a greater understanding of their level of complexity, we would anticipate that the population of overconfident and underconfident students would decrease over time. Students would be able to more closely align their pre- and post-assessment scores. Figure II depicts the distribution of categories over time.

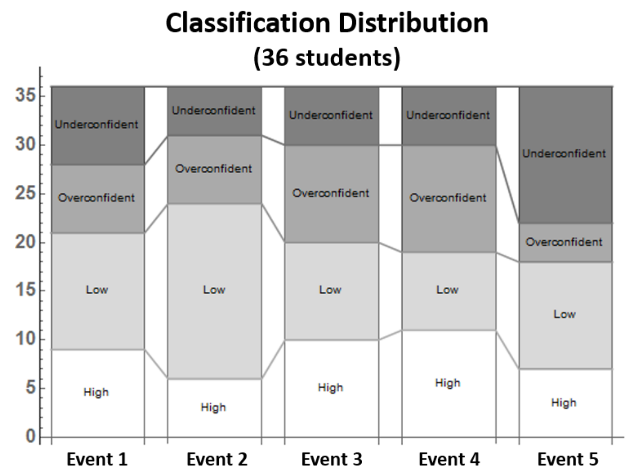

Figure 2: Classification by Graded Event

Roughly one-third of all of the students fall within the overconfident and underconfident student criteria. As the course progresses, we observe a general trend toward a slightly greater population of these students, which goes against our intuition prior to the study. There does not appear to be any indication that students are becoming attuned to the graded events. A greater sample size would have to be acquired, however, in order to identify clear patterns and definitively suggest these population trends.

5.3 *Do the students generally align with one category (i.e. could their confidence choices be related in some way to a character trait or their personality, or is it more about knowledge retention)?*

With five graded events and four categories, every student will have at least one category repeated among the five events. Perhaps a small number of students accounts for the bulk of one classification distribution. Table 2 depicts the order of merit list for student performance and the corresponding matrix of matches per category.

An interesting data point is the fact that 89% of the fourth quartile has 3 matching categories per student, while only 45% of the first quartile exhibits this same pattern. Looking at the matches selected for the fourth quartile, more than three-quarters of them are low and overconfident performers. While it is intuitive that these individuals would perform as if they do not know the material, it is less intuitive that we would observe overconfident qualities at the very bottom of the student population. In fact, in the bottom one-third of the OML, we find that 84% of the students have at least one experience as an overconfident performer across the five graded events. This percentage drops to 67% for the middle one-third and to 41% for the top one-third of the students. This does seem to suggest that overconfidence is related in some way to performance. A greater sample size, however is needed to provide further evidence of this relationship.

	x1				x2				x3				x4			
OML \ Category	H	L	O	U	H	L	O	U	H	L	O	U	H	L	O	U
1																
2																
3																
4																
5																
6																
7																
8																
9																
10																
11																
12																
13																
14																
15																
16																
17																
18																
19																
20																
21																
22																
23																
24																
25																
26																
27																
28																
29																
30																
31																
32																
33																
34																
35																
36																

Table 2: Student Performance Order of Merit List (with Category classification, by quantity) (H = High, L = Low, O = Overconfident, and U = Underconfident Performers)

5.4 *Does the relationship between confidence and student performance change when we consider the post-assessment rather than the pre-assessment?*

Immediately upon completion of each graded event, the students were again asked to report the number of points that they believe they earned. This data point gives us the opportunity to compare and contrast confidence levels given an environment of uncertainty and certainty, i.e. imperfect and perfect knowledge of the graded event questions. Since the pre-assessment was conducted prior to passing out the graded event, a response was collected from every student. For the post-assessment, however, the last question on the quiz captured the student responses.

As a result of time constraints for a small percentage of the students, they failed to provide a response to the post-assessment – these are identified as “N/A” in the population distribution.

Intuitively, we would anticipate a greater percentage of high and low Performers when considering the post-assessment versus the pre-assessment. Figure 3 captures the classification population, by category, for each of the quartiles given that the post-assessment is considered.

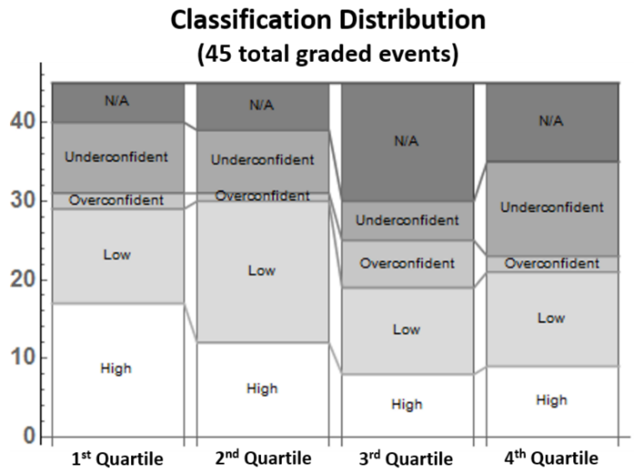

Figure 3: Classification by Quartile (using the Post-Assessment and Actual Performance)
(1st quartile = top performers, to 4th quartile = bottom performers)

With the 5-15 individuals per quartile electing not to complete the post-assessment (N/A), a comparison of the pre- and post-assessment distributions is challenged with respect to high and low performers. There is one trend, however, that we can see in comparing the distributions and that is the considerable decrease in the overconfident performer population at each quartile. The population of these students is generally greater for the 3rd and 4th quartile than that of the upper quartiles, in line with our observations of the pre-assessment data. One would expect the populations of overconfident and underconfident students to decrease with a post-assessment; however, we don’t necessarily see this in the results with underconfident students.

When we consider the classification distribution over time, by graded event rather than quartile (Figure 4), it is apparent that the population of overconfident and underconfident students are increasing with the post-assessment for at least three of the five events. Finally, when we consider the order of merit list for student performance and the corresponding matrix of matches per category (Figure 5), there are some notable differences between the pre- and post-assessment distributions. For instance, the number of individuals within the fourth quartile with 3 matching categories decreases from 89% with the pre-assessment to 33% with the post-assessment. In addition, the individuals in the bottom one-third of the OML that have at least one experience as an Overconfident Performer reduces from 84% to 41% between pre- and post assessment. Moreover, although there are a number of unqualified responses (N/A), the total number of individuals with 3 or more matching categories for the post-assessment is less than one-half of that for the pre-assessment. It appears that a considerable amount of these matches are from the population of overconfident and underconfident students, which reduces from ten to four between pre- and post-assessment.

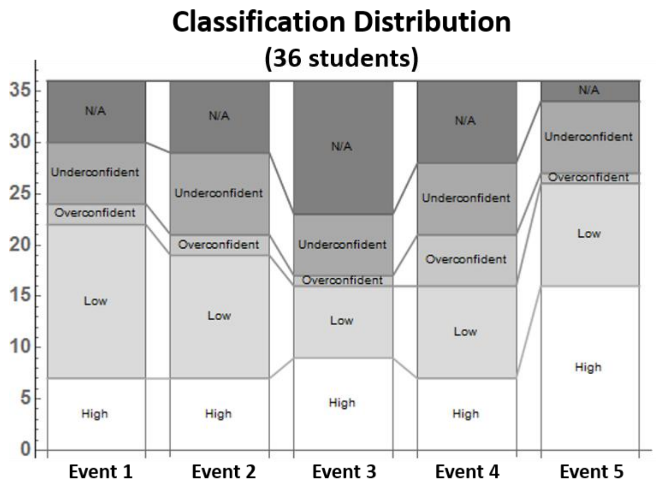

Figure 4: Classification by Graded Event

5.5 *What do the qualitative remarks on the pre-assessment with respect to why they reported that particular confidence level tell us about the outcomes?*

Each student provided a response to the question "*Why did you assess your score at that level?*" prior to taking the graded event. Most all of the comments for overconfident and underconfident students provided some affirmation on how prepared or unprepared, respectively, the individual was for the event. The most interesting comments are those from the overconfident and underconfident students. As a large sample of the comments for these students indicates in Table 4, the individuals seem to recognize and confirm these qualities in their responses. The comments provide qualitative feedback that supports the comparison of the quantitative results.

Category	**Short Description**	**"*Why did you assess your score at that level?*"**
Overconfident Performers	*Don't know what they don't know*	1. I typically do well on quizzes. 2. I always believe I can excel at everything, even if I'm unprepared. 3. I think I understand the material, but am not so good at integrating by hand or doing cross products. 4. I'm ready. 5. I have been working through a lot of example problems from the textbook to prepare for class. 6. Not too much material and I have some prior knowledge on it. 7. I'm ready. 8. Even though the material has more topics per se, I feel like I am able to follow it better. 9. I know the material well and feel confident. 10. I am prepared. 11. Feel ready. 12. I feel comfortable with these sections in the book. 13. I've looked over almost every problem and am pretty aware of procedures. 14. Feel confident. 15. Positive mental attitude.

Category	Short Description	*"Why did you assess your score at that level?*
Underconfident Performers	*Don't know what they know*	1. I can't remember all of the equations. 2. I'm not tracking this material too well. 3. I am confident with what we did in class; however, I did not work on the suggested problems. 4. Reliance on references; I should have done the suggested problems. 5. Did not spend adequate time preparing for the quiz. 6. I practiced all of the problems and while working I needed assistance throughout. 7. I will probably run out of time or mess up somehow. 8. I am really struggling to understand this material. 9. My math skills are weak. 10. I will mess up algebraically or forget a formula. 11. I did not study. 12. Still unsure about the material. 13. This class is hard. 14. Not as prepared as I could be. 15. Low confidence on the material.

Table 4: Overconfident and Underconfident Population Qualitative Responses, Pre-Assessment

6. Conclusion

The intent of this preliminary study was to seek alternative measures of assessment for student performance. By classifying the students into groups based upon their pre- and post-assessment scores, we were able to establish a greater association between confidence and performance with each graded event. As a preliminary study, we observed several trends which should be investigated further with a greater sample size of participants. In particular:

(i) The quantity of overconfident performers appears to be significantly higher in the lower quartiles than that of the upper quartiles.

(ii) The differences among the quartiles when considering the pre- and post-assessments is apparent, especially with regard to multiple category matches.

(iii) Under some conditions, the quantity of high and low performers appears to increase from the pre- to post-assessment.

(iv) With regard to the pre-assessment, the average difference between the estimate of the overconfident performers and their actual score was 29% with a range from 12 to 66%. The average difference between the estimate of the underconfident performers and their actual score was 22% with a range from 12 to 68%.

(v) For the post-assessment, the average difference between the estimate of the overconfident performers and their actual score was 21% with a range from 12 to 32%. The average difference between the estimate of the underconfident performers and their actual score was 26% with a range from 12 to 48%.

In addition, there are several insights from the study that should be examined further:

(i) The qualitative feedback from the overconfident and underconfident performers frequently referred to preparedness in their responses. Their inability to estimate their score within a letter grade of their actual score may indicate the students are not aware of how they should prepare for the mathematical graded events.

(ii) Intuitively, one would think that students would become more accustomed to the graded events within a course over time, as well as the grading tendencies of the instructor, and be able to more readily predict their actual score as a post-assessment toward the latter part of the course. The results depicted in Figure 4 do not appear to support this inference. A larger array of graded events, however, is needed to definitively suggest a trend.

(iii) For overconfident performers, the difference between the estimated score and the actual score decreased from pre- to post-assessment, whereas for underconfident performers, the difference increased. A greater sample size may provide more information regarding this insight.

	x1					x2					x3					x4				
OML \ Category	H	L	O	U	N/A	H	L	O	U	N/A	H	L	O	U	N/A	H	L	O	U	N/A
1	■						■			■										
2		■														■				
3			■			■				■										
4									■			■								
5	■						■		■											
6		■														■				
7	■						■		■											
8			■	■							■									
9	■	■			■				■											
10	■	■			■				■											
11						■						■								
12		■		■							■									
13				■		■	■													
14	■			■								■								
15						■						■								
16							■							■						
17		■	■												■					
18	■						■			■										
19	■			■	■		■													
20		■		■	■	■														
21					■															■
22	■			■	■		■													
23	■	■			■			■												
24		■		■											■					
25	■		■		■		■													
26		■		■	■	■														
27		■			■								■							
28							■							■						
29		■																		■
30		■			■									■						
31		■		■	■	■														
32	■						■		■											
33										■	■									
34	■						■	■												
35	■			■	■		■													
36	■	■			■				■											

Table 3: Student Performance Order of Merit List (with Category classification, by quantity)
(H = High, L = Low, O = Overconfident, and U = Underconfident Performers)

In summary, this ongoing study sought to examine the relationship between student confidence and student performance. With several insights, the intent is to expand the breadth of the study, so that implications for teaching in general may be more easily identified. There are also a number of research extensions for this confidence study. Given a greater sample size with similar observations, one might introduce a probabilistic model for forecasting grades. This study put greater emphasis on analyzing the quantitative information with qualitative comments in support. With present-day natural language processing algorithms, qualitative assessments may have a greater role in the interpretation of the outcomes. A study to compare and contrast preparedness for math graded events versus that of other subjects would also be useful. Asking similar-type questions before and after graded events that are more specific to confidence and preparedness would also provide a beneficial perspective. Finally, alternative classification schemes may be examined to identify other patterns between confidence and academic performance.

7. Acknowledgement: The views expressed in this paper are those of the authors, and do not represent the official policy or position of the United States Army, the Department of Defense, or the United States Military Academy.

8. References

[1] S. Austin, The paradox of Socratic ignorance (how to know that you don't know), *Philosophical Topics*, vol. XV, no. 2, pp. 23-34, 1987.
[2] S. Lichtenstein, B. Fischhoff, Do those who know more also know more about how much they know?, *Organizational Behavior and Human Performance*, vol. 20, no. 2, pp. 159-183, 1977.
[3] G. Gigerenzer, U. Hoffrage, H. Kleinbolting, Probabilistic mental models: a brunswikian theory of confidence, *Psychological Review*, vol. 98, pp. 506-528.
[4] P. Juslin, H. Olsson, Thurstonian and bruswikian origins of uncertainty in judgment: a sampling model of confidence in sensory discrimination, Psychological Review, vol. 10, pp. 344-366.
[5] D. Kahneman, P. Slovic, A. Tversky, Judgments under uncertainty: heuristics and biases, *Cambridge University Press*, Cambridge, England, 1982.
[6] J.B. Soll, Determinants of overconfidence and miscalibration: the roles of random error and ecological structure, *Organizational Behavior and Human Decision Processes*, vol. 65, pp. 117-137, 1996.
[7] G. Pallier, R. Wilkinson, V. Danthiir, S. Kleitman, G. Knezevic, L. Stankov, Individual differences in the realism of confidence judgments, Journal of General Psychology, vol. 129, pp. 257-300, 2002.
[8] L. Stankov, J. Lee, Confidence and cognitive test performance, *Journal of Educational Psychology*, vol. 100, no. 4, pp. 961-976, 2008.
[9] J.C. Angelis, S.S. Swinton, W.R. Cowell, The performance of non-native speakers of English on TOEFL and verbal aptitude tests, Educational Testing Service, Princeton, New Jersey, 1979.
[10] D.C. Johnson, The TOEFL and domestic students: conclusively inappropriate, *TESOL Quarterly*, vol. 11, pp. 79-86, 1977.
[11] L.J. Stricker, The performance of native speakers of English and ESL speakers on the computer-based TOEFL and GRE general test, *Language Testing*, vol. 21, pp. 146-173.

[12] S. Kleitman, L. Stankov, Ecological and person-oriented aspects of metacognitive processes in test-taking, *Applied Cognitive Psychology*, vol. 15, pp. 321-341, 2001.
[13] L. Stankov, Calibration curves, scatterplots and the distinction between general knowledge and perceptual tasks, Learning and Individual Differences, vol. 10, pp. 29-50, 1998.
[14] S. Tobias, H.T. Everson, Issues in the measurement of metacognition: assessing metacognitive knowledge monitoring, *Buros Institute of Mental Measurements*, Lincoln, Nebraska, pp. 147-222, 2000.
[15] J. Crawford, L. Stankov, Age differences in the realism of confidence judgments: a calibration study using tests of fluid and crystallized intelligence, *Learning and Individual Differences*, vol. 6, pp. 84-103.
[16] N. Harvey, *Confidence in judgment*, Trends in Cognitive Sciences, vol. 1, pp. 78-82.
[17] C.A. Shoemaker, Student confidence as a measure of learning in an undergraduate principles of horticultural science course, *HorTechnology*, vol. 20, no. 4, pp. 683-688, 2010.
[18] L. De Grez, M. Valcke, I. Roozen, How effective are self- and peer assessment of oral presentation skills compared with teachers' assessments? *Active Learning in Higher Education*, vol. 13, no. 2, pp. 129-142, 2012.
[19] D. Boud, The role of self-assessment in student grading. *Assessment and Evaluation in Higher Education*, vol. 15, pp. 20-30, 1989.
[20] D. Boud, N. Falchikov, Quantitative studies of student self-assessment in higher education: a critical analysis of findings. *Higher Education*, vol. 18, pp. 529-549, 1989.
[21] M. Taras, Student self-assessment: processes and consequences. *Teaching in Higher Education*, vol. 15, no. 2, pp. 199-209, 2010.
[22] A. Bolivar-Cruz, D. Verano-Tacoronte, S.M. Gonzalez-Betancor, Chapter 2: Is university students' self-assessment accurate? Sustainable learning in higher education: developing competencies for the global marketplace. Edited by M. Peris-Ortiz and M. Lindahl. *Springer*, 2015.