Engineering Educators Bringing the World Together

2025 ASEE Annual Conference & Exposition

Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025

ASEE

Paper ID #45713

A comparison of students expected grades and their actual quiz performance

Prof. Erik Hurlen, University of Washington

Dr. Erik Hurlen received his Ph.D. in Engineering Sciences (Aerospace Engineering) from the University of California San Diego in 2006. He spent a few years in industry before returning to academia to teach Math, Physics, and Engineering at many community colleges in the San Diego Area. Dr. Hurlen was an Instructional Assistant Professor in the Mechanical Engineering Department at the University of Mississippi from 2014 to 2019. In 2019, Dr. Hurlen returned to San Diego as a Lecturer in the Mechanical Engineering Department at San Diego State University for the 2019-2020 academic year. He is currently an Associate Teaching Professor in the department of Aeronautics and Astronautics at the University of Washington.

A comparison of students' expected grades to their actual quiz performances

Abstract

Students can sometimes either overestimate or underestimate their abilities. This can be particularly true of higher stakes assessments such as quizzes and exams. Students with lower confidence often struggle with retention and performance and could possibly lose interest in engineering. This is particularly true of underrepresented students in engineering and therefore it is crucial to understand any demographic discrepancies that may exist. This paper examines the confidence of students in two second year Engineering classes by having them predict their scores both before and after quizzes and then compares those predictions to their actual performance. This is then broken down by student reported demographic data to support previous research and to determine any new emerging trends. The data suggested that students with lower grades tended to overestimate their performance, while higher achieving students tended to underestimate their abilities. This lower confidence was particularly true for non-male and older students.

1. Introduction

Confidence and self-efficacy beliefs are linked to student’s persistence, achievement, and interest [4]. Therefore, it is crucial to understand student’s confidence levels and the reasons behind them. Some students tend to overestimate their knowledge and abilities while others tend to underestimate them. Previous research has shown evidence of females and other underrepresented minorities tending to be less confident in their abilities compared to other populations [3,4,5]. Additionally, students with lower grades tend to be overconfident in their abilities whereas higher achieving students tend to underestimate their performance [2].

The purpose of this study is to determine whether students tend to overestimate or underestimate their quiz performance in two second year engineering courses, Statics and Thermodynamics. The data are analyzed both for the students as a whole and broken down into groups based on demographics.

2. Motivation and Previous Research

Student confidence levels have been linked to retention and a sense of belonging in engineering programs. Many previous studies have focused on self-efficacy rather than simply confidence. The difference being that self-efficacy also takes into account the ability to perform at a high level, not just the belief of being able to do so. Bandura [1] defines four sources which contribute to self-efficacy. The strongest of which is performance (success raises self-efficacy). The others are vicarious experiences (comparison with peers), verbal persuasions (feedback from others), and physiological and affective states (anxiety etc.).

There have been many studies on gender differences. Jones [5] studied how gender differences affect student motivation constructs, achievement, and career plans, along with the interactions among them. Jones found that female students did have lower self-efficacy, but not necessarily to a statistically significant level. In Hutchison et.al. [4], it was found that females who persist in STEM have lower self- efficacy perceptions than their male colleagues. These were also broken down by race and ethnicity. Fraley et.al. [3] assessed confidence and competency of first year engineering students. It was found that many students enter class with a pre-conceived notion of already knowing the material. It was also found that female students predicted their skill level much more accurately than the male students. While at the same time, less confidence was found in the female students, even with similar levels of correctness.

Dunning [2] introduced the Dunning-Kruger effect, which defines a sense of over confidence in low performers. It was found that students tend to overestimate their abilities as a whole. It was found that lower achieving students did at least predict lower scores than the higher achieving students, however the magnitude of the discrepancy was much larger for the lower achievers. It was also found that the highest achieving group actually underestimated their abilities. There is also mention of previous research indicating the lower self-confidence experienced by female students compared to males despite actual performance being equal. There is also the observation that high performing students became more accurate in predicting their test performance over time while lower achieving students remained “stubbornly optimistic”.

Parsons et.al. [7] found that students with higher confidence in mathematics achieved higher grades in first year engineering mathematics. Here it was also found that age, dyslexia and the time spent working outside lectures did not have any significant effect on grades. Reed [8] developed a grading scheme that incorporated their confidence levels, which slightly biased the students towards selecting not confident. It included binary scoring (confident/not, right/wrong) as opposed to allowing for partial credit. It was found here that students have a good knowledge of their abilities regardless of their mastery of the material. Kribs [6] found that self-assessment scores were a good predictor of student performance. This study included other factors such as sleep the night before. It was also found that student assessment scores increased between pre-test and post-test.

3. Hypotheses

Based on previous research, it is expected that the female and other underrepresented minority students will be less confident in their abilities. Second, it is expected that older students will have a more realistic expectation of their abilities and therefore predict their scores more accurately than younger students. Similarly, it is expected that students with lower grades tend to overestimate their abilities.

It is also anticipated that students will be better at predicting their scores after the quiz than before. This is due to the fact that they have not only seen the actual question but also should have some idea of how well they did. It is also expected that students will get better at predicting their scores later in the course as they get familiar with the format and style of questions on the quizzes as well as the grading habits and rigor expected from the instructor. This has also been observed in [2].

Therefore, this study will aim to validate the following five hypotheses:

1. Students will be better at predictions in the post-quiz surveys compared to the pre-quiz surveys
2. Students will get better at predicting their scores throughout the course
3. Older students will be better at predicting their abilities than younger students
4. Female students will be less confident in their abilities
5. Students with lower grades will tend to overestimate their abilities

4. Procedure

Students in two sophomore level engineering classes (Statics and Thermodynamics) were surveyed both before and after each quiz during the summer quarter. Each course had a 15 minute quiz after each lecture, generally consisting of a single question testing the students on their knowledge of the topic covered. The pre-quiz survey asked the students what score they anticipated on the upcoming quiz as well as a brief description of the reasons behind their answer. The post-quiz survey asked the students what score they anticipated on the quiz they just took, and why. This resulted in a possible 48 surveys per student throughout the quarter.

A demographic survey was also provided at the beginning of the quarter so student scores could be aggregated by their self-reported data. All surveys were optional, and were approved by the university's institutional research board. Almost two thirds of the students participated in at least one of the surveys throughout the quarter, while about half of the class participated in the majority of them. Total enrollment in both classes combined was 49 students, and the survey participation was as follows:

- 21 students completed the demographic survey
- 29 students filled out at least one pre- or post-quiz survey
- 20 students did both (1 only did the demographic survey, 9 students filled out some pre/post-quizzes without doing the demographic survey)
- 17 students did the demographic survey and the majority of the pre/post-quiz surveys, with many doing >90% (this means that only 3 students who did the demographic survey only filled out a few of the pre/post-quizzes)

For the demographic aggregation of the data, students who did not fill out the demographic survey had to be removed from consideration. Many of these students only filled out a small number of the pre/post-quizzes so this only resulted in a small loss of survey data.

5. Results and Discussion

5.1 Overall Data

Figures 1 through 4 show data for the overall student population on all of the quizzes administered throughout the course. There was a total of 413 score predictions from the pre-quiz surveys, 382 for the post-quiz surveys, and 423 actual quiz scores. Note this last value is just the quizzes where a student filled our either the pre- or post-quiz survey (or both) since these are the only ones that could be compared to predictions.

From the overall averages seen in table 1, the students as a whole underestimated their abilities in both the pre-quiz and post-quiz surveys. It can also be seen that their predictions on the post-quiz surveys were less accurate than those in the pre-quiz survey.

Table 1: Average scores (out of 10) for all of the data collected in the pre-quiz surveys, actual quiz scores, and post-quiz surveys.

Pre-quiz survey	7.53
Actual quiz	7.84
Post quiz survey	7.37

Percentages of each quiz score (integer values between 0 and 10 inclusive) are shown in figure 1. It can be seen that the highest achieved value was a 10 (approximately 35% of quizzes). However, the percentage of students who actually predicted this score value was lower for both the post-quiz survey (27%) and in the pre-quiz survey (20%). A perfect score of 10 was the most popular response in the post-quiz survey, followed by a score of 8 (18%). For the pre-quiz surveys, scores of 7 and 8 were the most popular response (24% and 23% respectively). Scores of 9 were not predicted all that often (under 10% of the time). This could either come from students feeling that their mistakes had a larger impact, or that they had a harder time realizing they had made small mistakes. It is also interesting in this context that the number of actual quiz scores of 9 is lower than those of 7, 8, and 10, so generally students often made more fundamental mistakes, and perhaps they are aware of this. At the other end of the spectrum, there were hardly any predicted scores of 1/10 and 2/10. Rather, it seems that when the students were not at all confident in their abilities, they chose a score of zero. It can be seen that actual quiz scores of 2/10 were much more numerous than the predicted values.

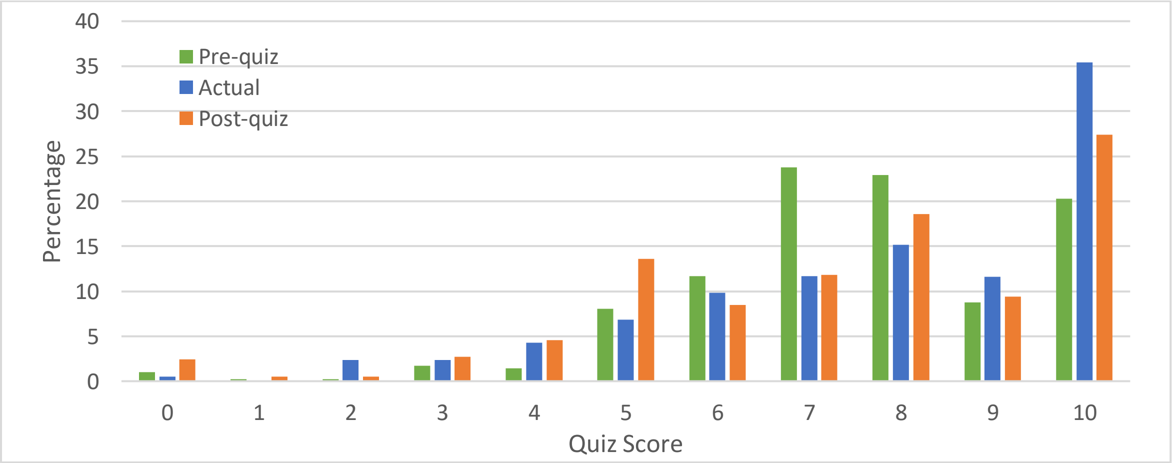

Figure 1: Percentages of student scores for all quizzes administered throughout the course.

The difference in the students predicted scores versus their actual performance is shown in figure 2. A value of 0 means the students correctly predicted their score, while a value of +10 would indicate, for example, that the student predicted a 10 but in reality, received a 0. It can be seen that the most likely value, around 25% of the time, was that the students predicted their score accurately. Also, they were within +/- 2 points another almost 50% of the time. This demonstrates that student ability to accurately predict their scores, both before and after the quizzes, is fairly high.

The other interesting trend that can be seen in figure 2 is that the negative values are almost always higher than their positive counterparts. This indicates that the students as a whole tend to underestimate their abilities more than them overestimating.

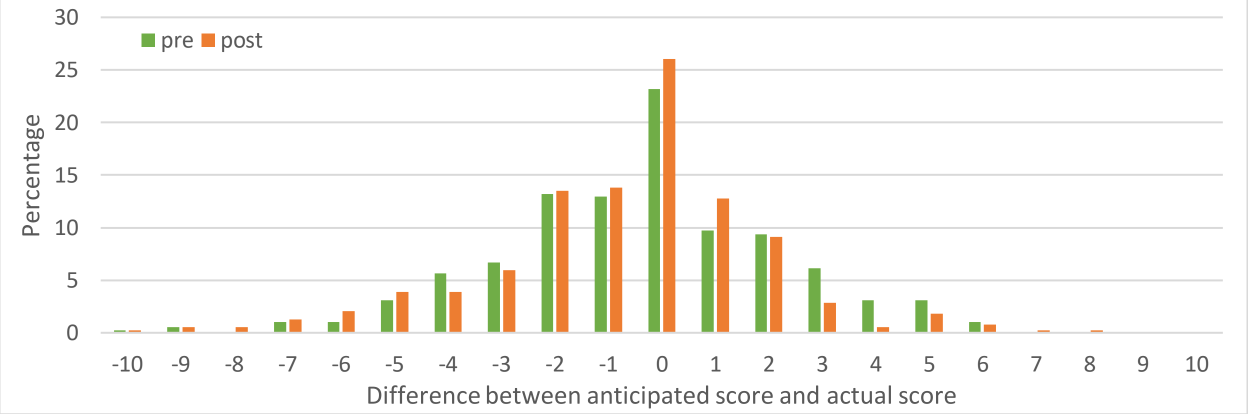

Figure 2: Difference between the predicted score by the students and their actual score for all quizzes administered throughout the course.

Figure 3 shows a chronological performance metric of the student averages by quiz number. Similarly, figure 4 shows the absolute value of the difference in predicted versus achieved values. A slight downward trend in both the predicted and achieved values can be seen in figure 3. Although, this is somewhat expected as the course material gets more difficult as the course progresses. In figure 4, there is a slight downward trend in the pre-quiz prediction values, indicating that the students do get better at predicting their scores as the course progresses. However, this cannot be seen in the post-quiz predictions. Also, the regression coefficient (R^2 value) is less than 0.05 in both cases, indicating that this is not a strong trend in either case.

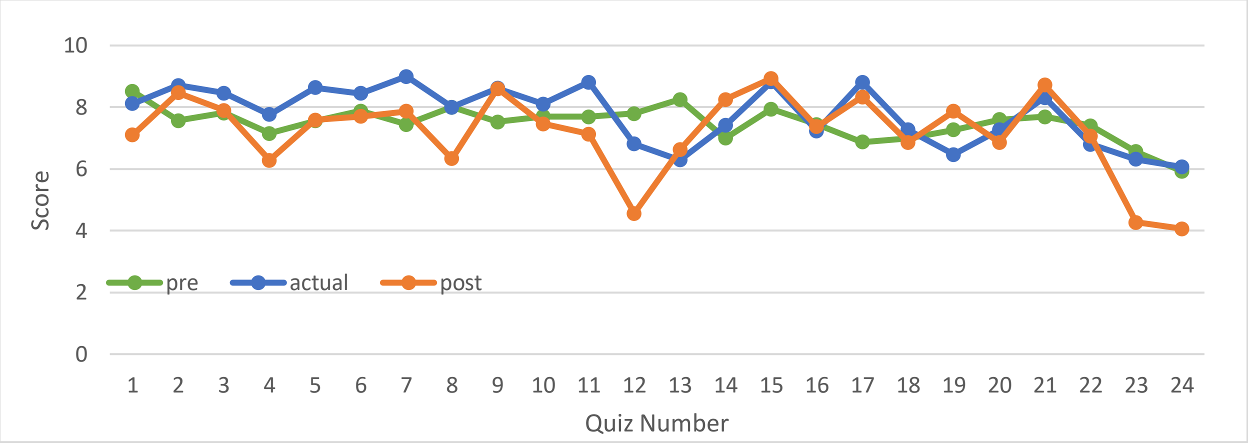

Figure 3: Average performance and predicted values for the students as a function of quiz number.

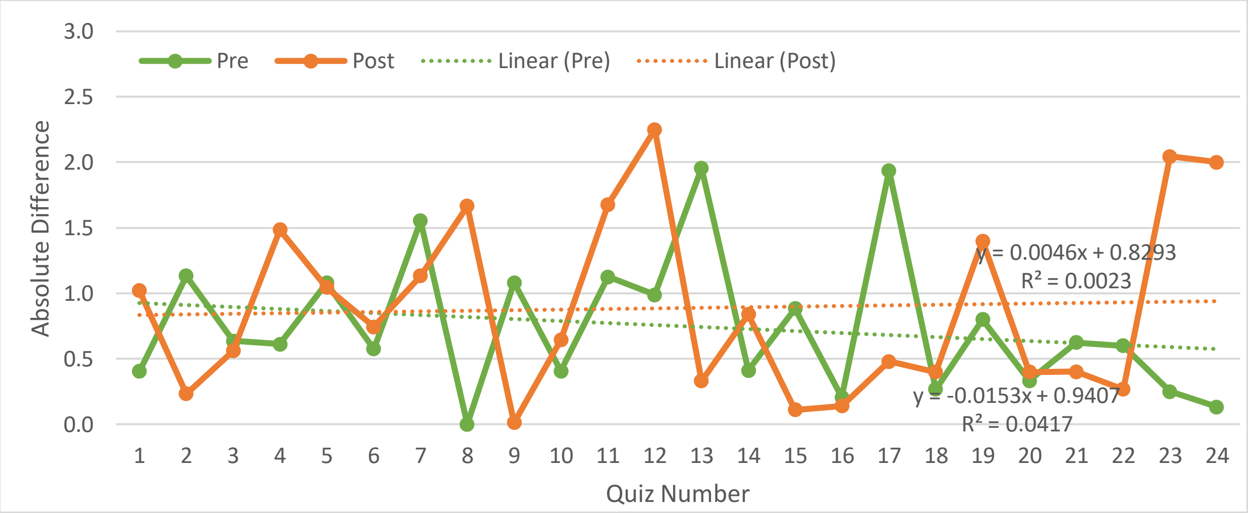

Figure 4: Absolute value of the difference in predicted score versus actual score for the students as a function of quiz number.

5.2 Variations by Populations - Age

The first demographic category analyzed was by the age of the students. Data for averages and differences are shown below in tables 2 and 3, while plots of the difference between expected and actual scores are shown as a function of age in figure 5.

The data that are grouped by demographics can be analyzed in two forms. First, there is the data from each individual student. In this case, each student is defined as a single, distinct data point reflecting the average quiz/survey scores of that student. This is shown in table 2 and the left side of figure 5. The second option is to use each survey/quiz score as a data point, as in table 3 and the right side of figure 5. The advantage of this is that there are many more data points. The disadvantage is that they are not necessarily independent data points, as personal trends influence some data more than others. This is most relevant with the 19 year old category. Of the seven students, three did nearly all of the surveys, one did a little more than half, while the other three students did very few between them (these were the students identified in the procedure section that brought the number from 20 to 17). This significantly alters the values for this category as the three who did many surveys tended to underestimate their abilities much more than those who only did one or two. There are also slightly less data points in the demographic analysis (374 pre-quiz surveys and 353 post-quiz) as the students who did not fill out the demographic survey could not be categorized.

Table 2: Data based on students: The average scores for each student averaged for each age category.

Age	# Students	Metric	Average	Difference	Standard Deviation
19	7	Pre-Quiz	7.742	0.101	2.077
		Actual Quiz	7.642		
		Post-Quiz	6.629	-1.012	3.523
20	7	Pre-Quiz	7.524	0.172	1.011
		Actual Quiz	7.351		
		Post-Quiz	7.370	0.018	0.795
22	2	Pre-Quiz	8.271	-0.740	2.269
		Actual Quiz	9.010		
		Post-Quiz	7.255	-1.756	1.467
27	2	Pre-Quiz	6.333	-2.042	1.061
		Actual Quiz	8.375		
		Post-Quiz	6.739	-1.636	0.369
34	2	Pre-Quiz	9.186	1.284	0.327
		Actual Quiz	7.901		
		Post-Quiz	8.245	0.343	0.699

Table 3: Data based on each quiz score: The difference of scores for each quiz/survey averaged for each age represented.

Age	# Data Points	Metric	Difference	Standard Deviation
19	85	Pre-Quiz	-0.959	2.221
	78	Post-Quiz	-1.615	1.829
20	147	Pre-Quiz	0.167	1.339
	138	Post-Quiz	0.141	1.287
22	48	Pre-Quiz	-0.740	1.458
	47	Post-Quiz	-1.755	0.670
27	48	Pre-Quiz	-2.042	0.530
	47	Post-Quiz	-1.638	0.922
34	46	Pre-Quiz	1.326	0.667
	43	Post-Quiz	0.395	0.204

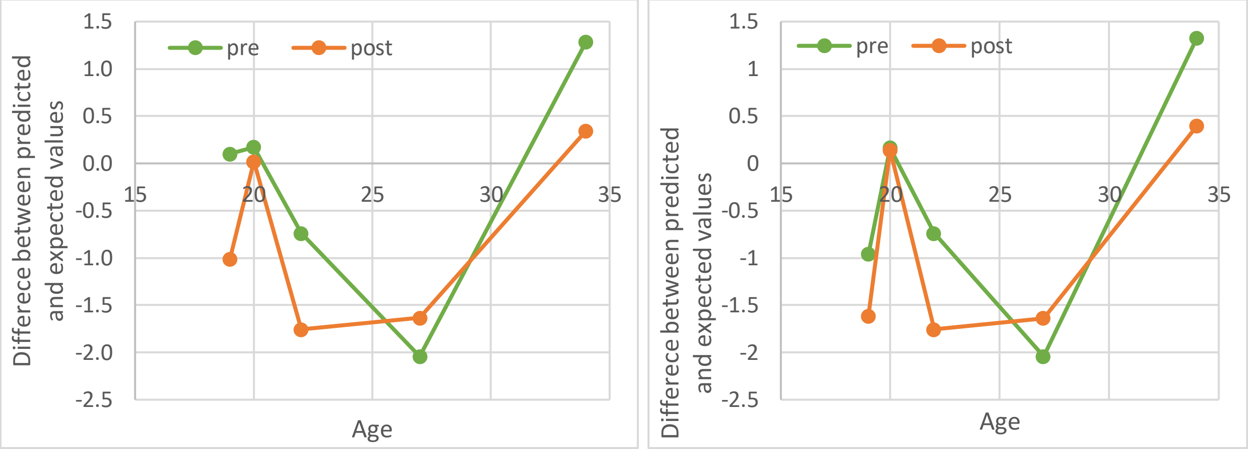

Figure 5: Differences in expected scores versus actual performance as a function of age. Left: Student averaged values (average of each student's average). Right: Data averaged values (average of all data points for each age).

It can be seen in the tables and/or figure that younger students are, in general, more confident in their abilities, with mid age students (those in their 20s) being less confident. It is also interesting that the 34 year old students surveyed were the most confident. However, with only two students it is debatable how significant this result may be.

In terms of statistical significance, a t-test was used to check for any differences between age populations. Treating each student as a single data point (Table 2 values), the only grouping to show a significant difference (at the 0.05 level) was between 20 and 27 year olds. Here the critical t value is 1.895, where the pre-test score difference t value is 2.220 while the post-test value is 1.795, for an average of 2.01. Using the data point analysis (Table 3), many more differences can be considered statistically significant due to the increased number of data points. For instance, for 20 year olds compared to 27 year olds, the critical t value is 1.65 whereas the computed t value is 4.65. However, these comparisons are not as strong due to the lack of independence in the data.

5.3 Variations by Populations - Gender

The second demographic comparisons to be made is by gender. The data are shown in tables 4 (treating individual students as data points) and 5 (each quiz as a data point). There are two factors which heavily influence the analysis here. The first is the large discrepancy in numbers between male (17) and female (2), or even non-male (3) students. The second is the fact that one of the female students only completed one survey, a pre-quiz survey where she overestimated her score by 1 point. This increases the perceived confidence of female students when treating each student individually (table 4).

Looking at the data on a survey specific basis, a much larger difference can be seen in male and non-male (female and non-binary) students (table 5). Here the computed t values for the pre-quiz and post-quiz surveys are 1.437 and 2.544 respectively. Compared to the critical value of

1.65 for a significance level of 0.05, we can see that there may be a statistically significant difference here. However, caution is needed as the data only included multiple survey results from 1 female and 1 non-binary student.

Table 4: Data based on students: The average scores for each student averaged for each gender category.

Gender	# Students	Metric	Average	Difference	Standard Deviation
Male	17	Pre-Quiz	7.695	-0.018	1.520
		Actual Quiz	7.714		
		Post-Quiz	7.272	-0.442	1.162
Female	2	Pre-Quiz	8.477	0.499	2.153
		Actual Quiz	7.978		
		Post-Quiz	6.000	-1.978	--
Non-Binary	1	Pre-Quiz	6.667	-1.771	--
		Actual Quiz	8.438		
		Post-Quiz	6.217	-2.220	--

Table 5: Data based on each quiz score: The difference of scores for each quiz/survey averaged for each gender category.

Age	# Data Points	Metric	Difference
Male	327	Pre-Quiz	-0.272
	309	Post-Quiz	-0.578
Non-male	47	Pre-Quiz	-0.862
	44	Post-Quiz	-1.602

5.4 Variations by Populations - Other

Other demographic comparisons were made by race, socioeconomic status, and major. None of these resulted in any observed significant differences between the various populations.

5.5 Variations by Grades

The last comparison was made by final grade received in the course. The students were separated into categories of A's (3.6-4.0), B+ (3.1-3.5), B- (2.6-3), C+ (2-2.5), C-/D (1-2), and F (0-1) as shown in table 6. The average student quiz scores and survey predictions, along with the average differences between them are also shown compared with final grades in figure 6 (left) as well as the average difference between predicted scores and actual scores for each grade category (on the right). It can be seen in these graphs that the perceived abilities of the lower achieving students are a lot higher than their actual performance, while the higher achieving students tend to underestimate their abilities, as expected.

Table 6: Data for students based on assigned grade in the course. Averages shown are student averaged data (students as data points as opposed to average of all quizzes).

Grade	# Students	# Data points	Metric	Average	Difference	Standard Deviation
3.6-4 (A)	10	153	Pre-Quiz	8.413	-0.403	1.786
		154	Actual Quiz	8.816		
		141	Post Quiz	7.755	-1.061	2.880
3.1-3.5 (B+)	10	171	Pre-Quiz	7.233	-0.686	1.436
		174	Actual Quiz	7.919		
		162	Post Quiz	7.551	-0.368	2.724
2.6-3 (B-)	3	4	Pre-Quiz	6.667	-1.500	0.943
		4	Actual Quiz	8.167		
		1	Post Quiz	8.000	-0.167	--
2-2.5 (C+)	4	56	Pre-Quiz	7.013	0.534	0.890
		58	Actual Quiz	6.479		
		50	Post Quiz	6.534	0.056	1.335
1-2 (C-/D)	1	7	Pre-Quiz	7.714	2.381	1.254
		9	Actual Quiz	5.333		
		7	Post Quiz	6.000	0.667	2.309
0-1 (F)	1	22	Pre-Quiz	6.773	2.314	1.412
		24	Actual Quiz	4.458		
		21	Post Quiz	6.619	2.161	2.439

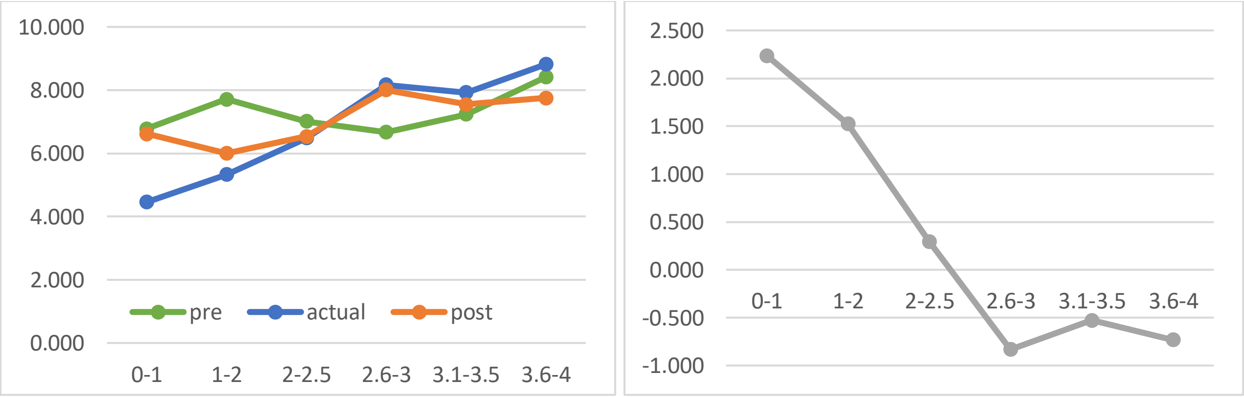

Figure 6: Left: Average student scores vs final assigned grade in the course. Right: Average difference between predicted scores and actual quiz scores by final grade.

The one possible outlier in this trend is the B- category, where the students outperform their anticipated scores by more than the B+ category. Like the majority of the outlying data points found in this study, this corresponds to the category which contains the majority of students who only did a few surveys.

A t-test separating A and B students from C, D, and F students yields t values of 2.315 and 0.912 respectively for the pre-quiz and post-quiz score differences. The average of these two is 1.61. The critical t value in this case is 1.708 for a significance level of 0.05. So, it can be seen that the difference in the two populations is close to being statistically significant in this case. This is for the student averaged values (shown in table 6) as opposed to the individual data points. If the analysis used all of the individual data points, the t value (5.13 in this case) would easily be over the critical value.

6. Conclusion

As a whole, most students tended to underestimate their abilities both before and after the quizzes. Their predicted scores were lower on the post-quiz surveys than on the pre-quiz surveys. This suggests that students' post-quiz predictions were actually less accurate than their pre-quiz predictions, negating the first hypothesis. It would be interesting in future to investigate if students tend to feel worse about their performance after the quiz than before, regardless of whether their prediction was above or below their actual scores. Perhaps they tend to be more optimistic just before the quiz but more pessimistic just after. This is backed up by the data that only three students had an average post-quiz prediction that was higher than their pre-quiz prediction. It is also interesting to note that these three students tended to highly underestimate their abilities (their difference between predicted and actual scores averaged over 1 point lower on predicted score). There could also be investigations into whether students experience post-quiz anxiety or cognitive biases that influence their self-assessments.

The pre-quiz predictions of the students tended to better at the end of the course compared to the beginning. However, this was not the case for the post-quiz surveys. Also, the regression coefficient was fairly low, indicating only a small correlation. So, although there is a slight trend here, there is little evidence to support the second hypothesis of students being better at predictions at the end of the course.

It was found that students in their early and mid 20's tended to be the ones who underestimated their abilities the most, while 19 and 20 year olds were more confident. Interestingly 34 year olds surveyed as the most confident, however this was affected by the small number of data points available. So, even though hypothesis number 3 was supported, it cannot be concluded with high reliability in this study.

There was evidence that female and non-binary students tended to underestimate their scores much more than male students. But again, the small number of data points available for these underrepresented groups made it hard to say with any certainty that hypothesis 4 should be validated.

Finally, students with lower final grades tended to overestimate their quiz performance while students with higher final grades tended to underestimate their scores. Separating students into high grades (A's and B's) and low grades (C's, D's, and F's) yielded an obvious discrepancy, which is very close to being statistically significant. Therefore, hypothesis 5 is close to being verified by the data collected in this study.

Although there are at least some data supporting all of the five hypotheses, the low number of students in certain demographic categories makes it hard to say with any certainty that these hypotheses are supported with any statistical significance. The strongest two cases are for students with low grades overestimating their abilities (hypothesis 5) and younger students being more confident in their abilities (hypothesis 3).

Another trend that was noticed was that students seem to overestimate the impact of their shortcomings, both with not many 9's predicted (compared to 7's and 8's) and not many 1's and 2's (but rather 0's) chosen.

Summer classes tend to have lower enrollments than classes throughout the academic year. There are future research plans to expand the data set here using these larger classes. The data collected will also be expanded to ask students about other aspects which may affect their abilities and confidence such as study habits. It is expected that the trends found in this study will continue. With more data points available it may be possible to more confidently validate the hypotheses provided here.

There are strategies for increasing student confidence such as fostering a growth mindset and providing students with more positive feedback and encouragement. These could be emphasized with students who are identified as having lower confidence. By identifying the reasons behind lower confidence in students, more effective differentiated instructional methods could be used to help those students persevere and succeed in engineering courses.

7. References

1. A. Bandura, *Self-Efficacy the Exercise of Control*, New York: W. H. Freeman and Company, 2000.
2. D. Dunning, "The Dunning-Kruger effect. On being ignorant of one's own ignorance", In J. M. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology*, Vol. 44, pp. 247-296, Elsevier Inc., 2011.
3. M. A. Fraley, A. Kemppainen, A. J. Hamlin, and G. L. Hein, "Confidence in Computational Problem-Solving Skills of First-Year Engineering Students", *2016 ASEE Annual Conference and Exposition, New Orleans, LA, June 26-29*, 2016.
4. M. A. Hutchison, D. K. Follman, M. Sumpter, and G. M. Bodner, "Factors influencing the self-efficacy beliefs of first-year engineering students", *Journal of Engineering Education* 95 (1): 39-47, 2006.
5. D. J. Jones, M. C. Paretti, S. F. Hein, and T. W. Knott, "An Analysis of Motivation Constructs with First-Year Engineering Students: Relationships Among Expectancies, Values, Achievement, and Career Plans", *Journal of Engineering Education* 99 (4): 319-336, 2010.
6. J. D. Kribs, "Student Learning and Confidence in a Technology Management Graduate Statistics Course", *2022 ASEE Annual Conference and Exposition, Minneapolis, MN, June 26-29*, 2022.

7. S. Parsons, T. Croft, and M. Harrison, “Does students’ confidence in their ability in mathematics matter?”, *Teaching Mathematics and its Applications: An International Journal of the IMA,* Volume 28, Issue 2, Pages 53–68, June 2009.

8. K. Reed, “Assessment of Student’s Confidence of Learned Knowledge”, *2012 ASEE Annual Conference, San Antonio, TX, June 10-13*, 2012.