A Positive Association Between Weekly Practice and Exam Performance in CS 1

Alexander Stennet

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2020-163
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2020/EECS-2020-163.html

August 14, 2020

Acknowledgement

Advisor: John Denero. Reader: Josh Hug

A Positive Association Between Weekly Practice and Exam Performance in CS 1

by Alexander Stennet

Research Project

Submitted to the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, in partial satisfaction of the requirements for the degree of **Master of Science, Plan II**.

Approval for the Report and Comprehensive Examination:

Committee:

Associate Teaching Professor John DeNero
Research Advisor

8/13/20
(Date)

* * * * * * *

Associate Teaching Professor Joshua Hug
Second Reader

8/13/20
(Date)

A Positive Association Between Weekly Practice and Exam Performance in CS 1

Alexander Stennet

August 13, 2020

1 Abstract

In the CS1 course CS61A at the University of California, Berkeley, lab assignments are small online programming assignments meant to be completed during a 1.5 hour lab section. These assignments target students' lecture comprehension through introductory problems; these are in addition to separate homework assignments that target students' ability to synthesize several concepts to solve challenging problems. For several semesters, the lab assignments were graded such that it was possible to forgo the completion of many of the labs and still receive a full score; this caused many students to stop completing them towards the end of the semester. The analysis in this report demonstrates there was a significant positive association between the completion of these assignments and final performance in the Fall 2019 semester. This association is demonstrated for students at all skill levels and helps support a policy change to make the assignments more strictly required.

2 Introduction

CS 61A, Structure and Interpretation of Computer Programs [1], is the first required course for the computer science major at UC Berkeley. It is also one of the largest courses on campus teaching over 1800 students during the Fall 2019 semester. This course has previously been the subject of several research projects involving tool experimentation or data analysis [3, 4, 7, 9, 11, 12].

It's important to carefully analyze the data available from a course at this scale to iterate and improve on course policy. One such example of this is a grading policy in place from Fall 2017 to Fall 2019. The points in the course were split into 4 categories: exams, projects, homework, and participation. Within the grading category of participation, a student could receive a maximum of 10 points. These could be earned by completing lab assignments or attending discussion sections (i.e. a student could attend 10 discussions and complete no lab assignments and receive full participation score). Students who completed

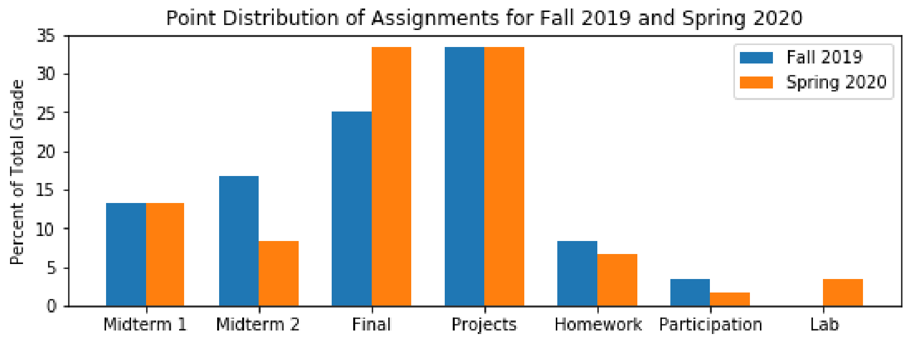


Figure 1: Plots percent of final grade assigned to each grading category. *Note: Midterm 2 and Final were lowered and raised respectively during Spring 2020 to accommodate mid-semester changes due to COVID-19*

more than 10 participation activities would instead receive exam recovery points to bump up their exam scores if they scored below 50%.

This is compared to a new policy adopted during the Spring 2020 semester, where instead, only discussion counts for participation and lab assignments are a separate category worth 10 points (each lab being worth 1pt). In addition, students are allowed up to 3 dropped lab assignments and can still receive full score. A plot comparing the point distributions for Fall 2019 and Spring 2020 can be seen in Figure 1. There are potential pros and cons to each policy. Requiring students to do more labs may help them better understand the concepts they cover; however, since there are also projects and homework, it may be more effective for them to instead spend this time better understanding the other assignments. Several studies in other subjects have attempted to measure the impact that requiring homework assignments has on student achievement [5, 6, 10]; however, they have had differing results. Additionally, it is unclear how immediately applicable their results would be as the difference between the Fall 2019 and Spring 2020 policies are more of a comparison between half-way required and completely required as opposed to completely optional compared to completely required.

The analysis in this report was done using Fall 2019 and Spring 2020 data. The data is a mixture of enrollment exports and exports from the course's online grading platform, Okpy [2].

3 Analysis

3.1 Completion Rate

First, we will analyze the completion rates of discussion and lab assignments to determine if the change in policy is warranted. If all students complete labs then little useful analysis can be made.

Q3: GCD

The greatest common divisor of two positive integers `a` and `b` is the largest integer which evenly divides both numbers (with no remainder). Euclid, a Greek mathematician in 300 B.C., realized that the greatest common divisor of `a` and `b` is one of the following:

- the smaller value if it evenly divides the larger value, or
- the greatest common divisor of the smaller value and the remainder of the larger value divided by the smaller value

In other words, if `a` is greater than `b` and `a` is not divisible by `b`, then

```
gcd(a, b) = gcd(b, a % b)
```

Write the `gcd` function recursively using Euclid's algorithm.

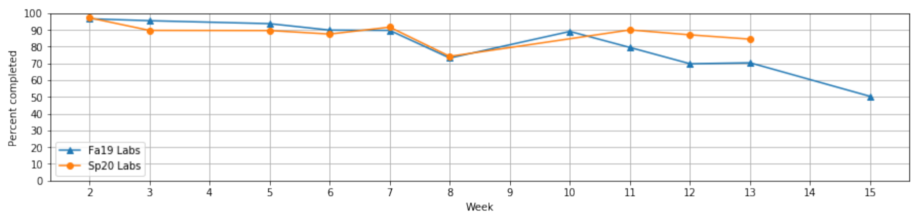
```python
def gcd(a, b):
    """Returns the greatest common divisor of a and b.
    Should be implemented using recursion.

    >>> gcd(34, 19)
    1
    >>> gcd(39, 91)
    13
    >>> gcd(20, 30)
    10
    >>> gcd(40, 40)
    40
    """
    "*** YOUR CODE HERE ***"
```

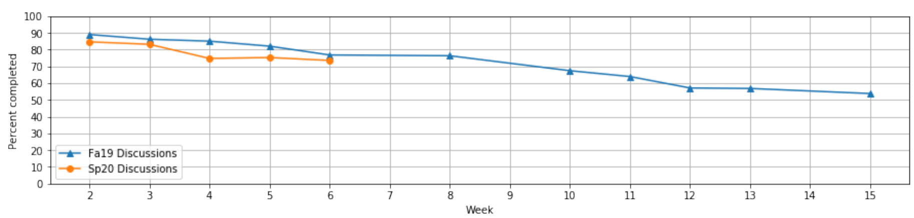
Figure 2: Example lab problem

During Fall 2019, we see from Figure 3a and Figure 3b that as the semester progresses lab completion and discussion attendance trend downwards. Specifically, week 1 starts at 96% completion for labs but drops down to 50% completion by week 15, and week 1 starts at 90% attendance for discussion but drops down to 54% attendance by week 15. Notable exceptions are week 7, where all students automatically received discussion attendance due to California PG&E outages and subsequent cancellation of in-person class as well as week 9 where the lab solutions were released with the assignment to assist in studying for the midterm.

In comparison, after the policy change was made in Spring 2020, we see from Figure 3a and Figure 3b that as the semester progresses, lab completion stays comparably higher. Specifically, week 1 starts at 97% completion for lab and only drops to 75% at a minimum for week 8. This would agree with the hypothesis that the policy change gives more of an incentive to complete labs in comparison to before; however, due to the presence of COVID-19 for the latter half of Spring 2020, it is difficult to tell if this is due to policy or due to the environmental change. The effect of COVID-19 is clearly demonstrated in the Spring 2020 discussion data due to all discussions after week 7 being made optional and credit given automatically.



(a) Lab assignment completion rates by week, Fall 2019 & Spring 2020
week 4 had a midterm so no lab assignment, week 9 had a midterm and a lab was assigned but all students automatically received credit, week 10 was spring break for Spring 2020, week 14 had an optional lab for Spring 2020 and was Thanksgiving break for Fall 2019



(b) Discussion attendance rates by week, Fall 2019 & Spring 2020
For Fall 2019, week 7 was cancelled due to city wide power outages, week 9 had a midterm so discussion was cancelled, and week 14 was Thanksgiving break for Fall 2019; for Spring 2020, weeks 7 through 14 were made optional due to COVID-19

Figure 3: Completion rates for lab/discuussion for Fa19 and Sp20 semesters
Data was pulled from Okpy [2] and scores were aggregated accross students for each assignment, filtered to only include students who received a final grade in the course (ie did not drop early)

3.2 Lab Completion Effect on Grades

Note: All analysis in the following sections were done after filtering out students who received a 0 on either Midterm 1 or the Final as they are a mix of students who were excused from the exam as well as actually received a 0.

3.2.1 Correlation

This analysis was used to primarily inform the decision to change from the original grading policy to the newer grading policy in Spring 2020. To inform this decision, we wanted to analyze how much of an impact, if any, completing lab assignments has on final grade prediction. Running a randomized trial where students had different grading policies wasn't a feasible option. The most common approach to running a randomized experiment, as seen in [5, 6, 10], is to run two concurrent courses with the only difference being a change in grading policy; however, all computer science courses at University of California, Berkeley are taught with a single lecture section, so attempting to run this experiment would cause several logistical challenges. Another potential approach is to compare two different semesters that were run similarly except for one policy change; however, due to the onset of the COVID-19 pandemic this was made impossible for the time being. Because of this, we relied on lab assignments' correlation on final grades to make a reasonable guess on whether the policy change was warranted.

A simple way to measure the relation between lab completion and final grades is to compute a simple correlation. Computing the correlation between the number of lab assignments a student completes and their final exam scores results in a correlation coefficient of 0.25 signifying a small to medium correlation. Using bootstrapping with 5000 iterations gives a 95% confidence interval of 0.21-0.30. This alone, however, isn't sufficient to warrant a policy change; it is also important to look at how this changes for students at different levels.

Prior research demonstrates that prior experience strongly predicts success in a CS 1 course [8]. Because of this, it's important that the policy change doesn't hurt low performing students while assisting high performing students. Dividing students by their midterm 1 score and then computing the correlation between their lab totals and their final score results in Table 1.

We see that no matter the score on midterm 1, there is a similar positive correlation between lab scores and final exam scores. There are two clear exceptions to this: students who scored in the 0-5 range and students who scored in the 5-10 range; however, because there are a very small number of students in both of these categories, the confidence intervals are very wide making it difficult to make any strong conclusions regarding those ranges of students. *Note: Because there were only 6 samples in the 0-5 range, bootstrap resampling occasionally selected 6 of the same points. That made it impossible to calculate the correlation coefficient, requiring the sample to be discarded and replaced.*

Midterm 1 Score	Average # of Labs Completed	Average Final Score	Correlation	95% Confidence Interval	Number of Students
0 - 5	9.67	15.67	-0.55	-0.99 - 0.51	6
5 - 10	9.67	18.92	0.060	-0.26 - 0.31	30
10 - 15	10.00	23.37	0.24	0.28 - 0.422	68
15 - 20	9.87	31.33	0.29	0.12 - 0.45	118
20 - 25	9.89	40.06	0.29	0.18 - 0.40	233
25 - 30	10.15	49.98	0.24	0.14 - 0.33	349
30 - 35	10.30	58.54	0.26	0.15 - 0.37	378
35 - 40	10.41	64.46	0.24	0.12 - 0.37	273

Table 1: Correlation between the number of lab assignments completed and the raw final exam score. Students were grouped using their raw midterm 1 score. Confidence intervals computed using bootstrapping with 5000 iterations.

3.2.2 Prediction

Another way to analyze the impact of lab assignment completion is to measure how much higher on average a student who completes most of the lab assignments scores compared to the rest of the students.

Correlation Between Midterm 1 and Final Exam Score
Doing a simple correlation between the raw point score of students' midterm 1 score and their final exam score gave a value of 0.79 demonstrating a high correlation between the two scores. Doing bootstrap sampling with 5000 iterations gives a relatively tight 95% confidence interval of 0.77 - 0.81.

Predictor Function
Because the correlation between the two exams is significant, we will create a function to predict a final exam score from a midterm 1 score. To capture the non-linear relationship between midterm 1 and final scores, we use a local regression model where the predicted value is the average final exam score of all students who were within a radius (for this analysis the chosen radius was 2 points) of their midterm 1 score. The graphed function can be seen in Figure 4. *Note: Small noise was added to students' scores to prevent exact identification of students from the graph. Calculations, however, were made using the original data.*

Using this predictor function, we can calculate the distance between a student's final exam score and their predicted score using their midterm 1 score. Comparing the average distance between the dataset predictions and the actual scores gives insight into the impact of lab completion.

Figure 5 depicts the distribution of differences between students' actual final score and their predicted final score. The students are divided into two groups: those who completed 11 or more of the 12 lab assignments and those who completed fewer than 11.

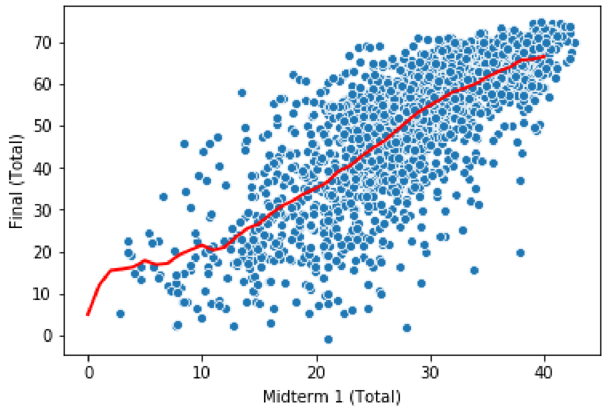


Figure 4: Red line signifies the predicted final exam score

On average, students who completed 11 or more lab assignments scored 1.9 points higher than the prediction. Students who completed fewer labs scored an average of 2.6 points lower than the prediction. An average difference of 4.5 points between these groups is worth about 6% of the final exam score for reference. Using bootstrap sampling for 5000 iterations gives a 95% confidence interval of 3.54 - 5.47 on the difference between these two groups. If we run a t-test on these distributions, we get a p-value of 1.99×10^{-19}. This implies that the difference in these distributions is statistically significant.

We can once again divide students by their midterm 1 scores to get more insight into the potential differences between differently performing students. The results of this can be seen in Table 2.

Midterm 1 Score	Average Difference Between Prediction	Number of Students	95% Confidence Interval	p-value
0 - 5	-4.74	6	-13.62 - 1.87	0.33
5 - 10	-2.91	30	-10.41 - 4.44	0.48
10 - 15	4.45	68	-1.32 - 10.02	0.13
15 - 20	6.54	118	2.03 - 10.65	0.0039
20 - 25	6.46	233	3.42 - 9.83	0.000096
25 - 30	5.09	349	2.68 - 7.37	0.000021
30 - 35	3.67	378	2.13 - 5.55	0.000031
35 - 40	2.22	273	0.65 - 3.79	0.00602

Table 2: Average distance from actual final score of 11+ lab completion and below 11

This demonstrates that even high performing students may benefit from completing lab assignments. A concern with changing the policy is that high

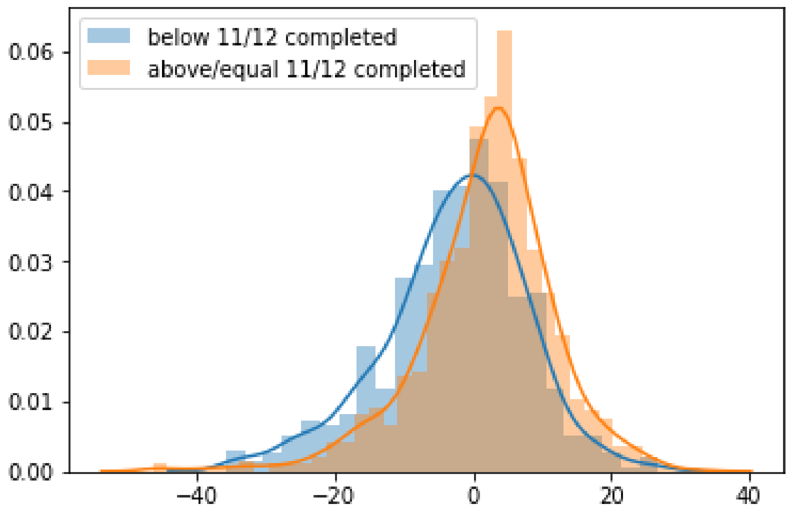


Figure 5: Distribution of Differences to Predicted Final Score

performing students would be hindered by being forced to complete lab assignments when they may have another way to study. Similarly, because of how few students fall into the 0-10 category we see a wide confidence interval making it difficult to make any strong claims about low performing students.

4 Conclusion

Preliminary results of this report were used to inform the policy change between the Fall 2019 and Spring 2020 semesters. The analysis presented supports the decision to make the grading policy change in two ways. Firstly, by demonstrating that many students were not completing the lab assignments and the policy change effectively incentivized students to complete more lab assignments, and secondly, by demonstrating that there is a statistically significant positive correlation between final exam scores and lab assignment completion.

As this is only a correlation, it's possible there are other confounding factors that could be pulled out via future experimentation. One potential factor could be a measure of student burnout. As the semester progresses, students may be unable to keep up with the work required by the course and a declining lab completion rate may be a symptom of declining course engagement. In Spring 2020, we see what appears to be a faster declining lab completion rate; however, it's possible that because the semester was moved online, due to COVID-19, students found it harder to engage and burned out quicker. In future semesters, it may be possible that the semesters will be similar enough to make a policy comparision analysis possible to better understand whether this is simply correlated or if there is a causal link. Additionally, work can be done to extend this analysis to previous semesters to see how generalizable these results are.

5 References

References

[1] CS 61A: Structure and Interpretation of Computer Programs. https://cs61a.org/. Accessed: 2020-07-24.

[2] OK. https://oky.org/. Accessed: 2020-08-02.

[3] Soumya Basu, Albert Wu, Brian Hou, and John DeNero. Problems before solutions: Automated problem clarification at scale. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale*, L@S '15, page 205–213, New York, NY, USA, 2015. Association for Computing Machinery.

[4] John DeNero and Stephen Martinis. Teaching composition quality at scale: Human judgment in the age of autograders. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education*, SIGCSE '14, page 421–426, New York, NY, USA, 2014. Association for Computing Machinery.

[5] Tisha LN Emerson and Kimberly D Mencken. Homework: To require or not? online graded homework and student achievement. *Perspectives on Economic Education Research*, 7(1):20–42, 2011.

[6] Borislava Gutarts and Franque Bains. Does mandatory homework have a positive effect on student achievement for college students studying calculus? *Mathematics and Computer Education*, 44(3):232–244, Fall 2010. Copyright - Copyright Mathematics and Computer Education Fall 2010; Document feature - Tables; ; Last updated - 2017-10-31; CODEN - MCEDDA.

[7] Andrew Head, Elena Glassman, Gustavo Soares, Ryo Suzuki, Lucas Figueredo, Loris D'Antoni, and Björn Hartmann. Writing reusable code feedback at scale with mixed-initiative program synthesis. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*, L@S '17, page 89–98, New York, NY, USA, 2017. Association for Computing Machinery.

[8] Lisa Lacher, Albert Xin Jiang, Yu Zhang, and Mark Lewis. Aptitude and Previous Experience in CS1 Classes. 07 2017.

[9] R. Rolim, G. Soares, L. D'Antoni, O. Polozov, S. Gulwani, R. Gheyi, R. Suzuki, and B. Hartmann. Learning syntactic program transformations from examples. In *2017 IEEE/ACM 39th International Conference on Software Engineering (ICSE)*, pages 404–415, 2017.

[10] Carolyn S Ryan and Nancy S Hemmes. Effects of the contingency for homework submission on homework submission and quiz performance in a college course. *Journal of Applied Behavior Analysis*, 38(1):79–88, 2005.

[11] Sumukh Sridhara, Brian Hou, Jeffrey Lu, and John DeNero. Fuzz testing projects in massive courses. In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*, L@S '16, page 361–367, New York, NY, USA, 2016. Association for Computing Machinery.

[12] Kristin Stephens-Martinez, An Ju, Colin Schoen, John DeNero, and Armando Fox. Identifying student misunderstandings using constructed responses. In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*, L@S '16, page 153–156, New York, NY, USA, 2016. Association for Computing Machinery.