Analysis of Factors and Interventions Relating to Student Performance in CS1 and CS2

Allen Guo

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2020-22

http://www2.eecs.berkeley.edu/Pubs/TechRpts/2020/EECS-2020-22.html

May 1, 2020

Analysis of Factors and Interventions Relating to Student Performance in CS1 and CS2

Allen Guo

University of California, Berkeley
Berkeley, CA, USA
allenguo@berkeley.edu

Abstract

In this report, we describe how student performance relates to background factors, prior grades, and utilization of interventions in the context of UC Berkeley's CS1 and CS2 courses.

1 Introduction

As educators, we are continually haunted by a grand and terrible question: why do some students succeed and others fail?

This inquiry is far from academic. By better understanding the factors related to student performance, we can optimize how we apply limited resources toward helping students succeed, we can give students more targeted and more accurate advice, and we can ensure that departmental policies on grading and major declaration are reasonable and fair.

There has been substantial prior work (dating back to the 1960s) exploring how various factors relate to performance in introductory CS courses. Background factors that have been linked to differences in performance or retention include gender [6, 12], age [2], prior programming experience [8, 9, 17, 19, 21], math background [13, 20], science aptitude [5], standardized test scores [4], attributional style [7], and personality type [1]. Other studies have analyzed factors measured during the course itself, such as comfort level [3, 11, 22], online discussion forum utilization [16], time spent coding [14], and other kinds of programming process data [10].

In spite of all of this, there have been few clear results. Effect sizes and sample sizes are typically small, and results are often contradicting. This has motivated us to conduct our own analysis, using student data from Berkeley's CS1 and CS2 courses. Our dataset included grades, survey responses, and data about how students utilized supplementary course resources, or *interventions*, such as tutoring and extra discussion sections. Our research questions were as follows:

1. How do student background factors like age, gender, and programming background relate to student performance?
2. How does CS1 performance relate to CS2 performance?
3. Are the interventions that are currently offered beneficial?

2 Data

We analyzed data from the following two courses:

- **CS 61A** ("Structure and Interpretation of Computer Programs") is Berkeley's CS1 course. Taught primarily in Python, 61A covers basic programming, recursion, object-oriented programming, and interpreters.
- **CS 61B** ("Data Structures") is Berkeley's CS2 course, taught in Java. It covers data structures and algorithms at an introductory level.

There are two computing-related majors at Berkeley: Computer Science (LSCS), in the College of Letters and Sciences, and Electrical Engineering and Computer Sciences (EECS), in the College of Engineering. All EECS majors and students intending to major in LSCS must take both 61A and 61B.

For our analyses, we only considered students who received letter grades (ranging between F and A+). Students who dropped, did not complete the course, or took the course pass-fail (P/NP or S/U) were excluded.

Our 61A data is from the Fall 2017 offering of the course, taught by John DeNero and Paul Hilfinger. 1473 students completed the course.

Our 61B data is from the Spring 2017 offering of the course, taught by Josh Hug. 1262 students completed the course.

Note that the 61B offering occurred *before* the 61A offering—this was not a longitudinal analysis tracking the same students across both courses. Only two students appear in both the 61A and 61B datasets.

2.1 Grades

We used grade data from the final gradebooks for 61A and 61B.

Both 61A and 61B have fixed, predetermined boundaries for assigning letter grades—students are not graded on a curve. Point values for all assignments and assessments are also predetermined.

In 61A, there were 300 points possible: 125 from programming assignments (homeworks and projects, primarily autograded), 10 points from additional smaller assignments, 40 points from Midterm 1, 50 points from Midterm 2, and 75 points from the final. There were also 10 points of extra credit possible, plus a recovery policy for students who underperformed on exams.

In 61B, there were 768 points possible: 368 from programming assignments (homeworks, labs, and projects, primarily autograded), 80 points from Midterm 1, 120 points from Midterm 2, and 200 points from the final. There were also 16 points of extra credit possible, plus additional recovery policies for struggling students.

For readability, we have rescaled all point values from each course by dividing by the total number of possible points and multiplying by 100. Henceforth, all points values will be given as scaled values.

2.1.1 Grade Bins

For this report, we define one grade bin as the distance between adjacent letter grades, including "plus" and "minus" grades. For example, an A is two grade bins away from a B+.

The "typical" 61A grade bin was 6.40 points wide. This was calculated by finding, for each student, the width of the bin corresponding to the grade they received, then averaging across these values.

The typical 61B grade bin was 5.92 points wide.

2.2 Surveys

Students were asked to fill out surveys at the beginning and end of each course, in exchange for a nominal amount of extra credit.

For 61A, we used responses from the start-of-semester survey, released during Week 1. Of the 1473 students who completed the course, 1442 (97.9%) filled out the survey.

For 61B, we used responses from the end-of-semester survey, released a week before the final exam. Of the 1262 students who completed the course, 1139 (90.3%) filled out the survey.

The text and response rates for individual questions are presented in Appendix A.

2.3 Interventions

In both 61A and 61B, select students attended supplementary small-group discussion sections led by student volunteers from Computer Science Mentors (CSM), a student-run organization. These sections were voluntary and open to all students. In 61B, of the 1262 students who completed the course, 295 signed up to attend CSM sections. (We did not use data pertaining to CSM sections for 61A.)

61B also offered weekly group tutoring sessions led by course staff members, which were similar in format to CSM sections. Due to high demand, students interested in attending group tutoring sessions were required to submit applications. Students were admitted primarily based on their performance thus far in the course. This was particularly true in the wake of Midterm 1, when TAs were directed to actively encourage students with low midterm scores to apply. In total, 86 students attended at least one group tutoring session.

Also available to students were weekly one-on-one tutoring sessions. These sessions began two weeks after Midterm 1 and ran for eight weeks total. The tutors for these sessions were students in CS 370, a course on CS pedagogy. Tutor availability was limited—tutees were accepted on a first-come, first-served basis each week. Tutor-tutee pairings were not guaranteed to be consistent across weeks. In total, 116 students attended at least one one-on-one tutoring session, according to the end-of-semester survey.

Finally, throughout the semester, the course staff hosted *guerrilla sections*, weekend events where students worked on worksheet problems in groups. Guerrilla sections are unique in that students are asked to obey the following rule: "You may not proceed to the next question until *everyone* in your group completely understands the answer." In total, 377 students attended at least one guerrilla section, according to the end-of-semester survey.

3 Background Factors

We started by measuring how various student background factors relate to performance.

The background factors we considered were gender, start age (the age at which the student started programming), URM status (whether the student identified as an underrepresented minority), major, transfer status, high school CS experience, sentiment on taking CS 10 (how strongly the student considered taking CS 10, Berkeley's CS0 course), and prior experience in the primary programming language for the course.

Our chosen metric for performance (i.e., the response variable) was total score—the total number of points the student received in the entire course. See Subsection 2.1 for details.

We performed this analysis separately for 61A and 61B.

3.1 61A

Our data on student background factors came from the start-of-semester survey. Appendix A.1 describes how this data was preprocessed.

Table 1 shows the size and mean total score of each group, along with other descriptive statistics.

First, we analyzed each factor individually using a one-way ANOVA, followed by a post-hoc Tukey HSD test.[1] For each factor, our null hypothesis was that mean performance (total score) did not differ for different levels of the factor. Our significance level was $\alpha = .05$ throughout.

- **Gender**: We found a significant effect, $F(1, 1420) = 57.84$, $p < .001$, $\eta^2 = .04$. Males performed significantly better than females, $t(771.93) = 7.65$, $p < .001$. The mean difference was 5.36 points.
- **Start age**: We found a significant effect, $F(5, 1380) = 42.61$, $p < .001$, $\eta^2 = .13$. The post-hoc analysis results are shown in Table 2. Generally, students who started programming earlier performed better than those who didn't.
- **URM status**: We found a significant effect, $F(2, 1388) = 55.63$, $p < .001$, $\eta^2 = .07$. Students who responded "No" performed significantly better than those who responded "Yes" or "Don't know". Students who responded "Don't know" performed significantly better than those who responded "Yes".
- **Major (EECS vs. LSCS vs. Other)**: We found a significant effect, $F(2, 1439) = 73.92$, $p < .001$, $\eta^2 = .09$. LSCS students and EECS students performed significantly better than Other students, but LSCS and EECS students did not differ significantly.
- **Major and transfer status (EECS non-transfer vs. LSCS non-transfer vs. EECS transfer vs. LSCS transfer)**:[2] We found a significant effect, $F(3, 1013) = 3.88$, $p < .01$, $\eta^2 = .01$. EECS non-transfer students performed significantly better than EECS transfer students. No other pair of groups differed significantly.
- **High school CS experience**: We found a significant effect, $F(1, 1253) = 75.84$, $p < .001$, $\eta^2 = .06$. Students with high school CS experience performed significantly better than those without, $t(757.72) = 8.16$, $p < .001$. The mean difference was 5.97 points.
- **Sentiment on taking CS 10**: We found a significant effect, $F(4, 1412) = 24.86$, $p < .001$, $\eta^2 = .07$. Students who took CS 10 or strongly considered taking CS 10 performed significantly

[1] For statistical testing, we used the *statsmodels* library [18].

[2] We chose to study this particular combination of factors because a professor we spoke with had hypothesized that LSCS non-transfer students perform worse than EECS non-transfer students.

Table 1: Descriptive statistics of the 61A total scores for various groups.

	count	mean	std	min	25%	50%	75%	max
Gender								
Female	412.0	80.61	11.94	6.0	74.98	81.14	88.67	103.83
Male	1010.0	85.96	12.10	0.0	79.02	87.67	94.83	105.67
Start Age								
12 years old or younger	123.0	91.98	8.67	69.80	86.92	94.17	98.92	105.67
13-14 years old	251.0	89.49	10.70	38.31	83.42	92.00	97.42	104.00
15-16 years old	419.0	86.22	9.54	47.67	79.75	87.00	93.83	103.50
17-18 years old	377.0	81.69	12.47	0.00	76.25	82.50	89.83	102.83
19-20 years old	162.0	78.96	11.30	33.21	71.38	78.49	86.33	99.17
21-25 years old	54.0	76.25	15.27	17.51	70.43	77.33	84.44	102.17
URM Status								
Don't know	186.0	82.44	11.59	36.47	75.38	83.92	90.40	102.17
No	1040.0	86.05	10.98	0.00	79.00	87.25	94.67	105.67
Yes	165.0	75.81	16.75	2.00	67.67	78.17	86.67	102.83
Major								
EECS	402.0	87.26	10.06	38.55	80.33	88.50	95.17	105.67
LSCS	618.0	86.50	11.10	2.00	80.45	87.50	94.83	104.00
Other	422.0	78.60	13.77	0.00	72.23	78.75	88.00	103.83
Major and Transfer Status								
EECS non-transfer	323.0	88.01	9.92	50.83	81.71	89.50	95.58	105.67
EECS transfer	78.0	84.28	10.10	38.55	78.44	83.83	92.12	102.17
LSCS non-transfer	557.0	86.70	10.88	2.00	80.50	87.67	94.83	104.00
LSCS transfer	59.0	84.37	13.01	36.47	79.11	86.17	94.92	100.56
High School CS Experience								
False	445.0	81.34	13.28	0.0	75.17	82.33	90.67	102.17
True	810.0	87.31	10.59	2.0	80.94	88.50	95.33	105.67
Sentiment on Taking CS 10								
Considered briefly	488.0	83.14	11.82	0.00	76.43	84.42	91.71	103.83
Considered strongly	55.0	77.47	9.77	33.21	72.66	78.17	82.21	95.67
Don't know about CS 10	96.0	83.45	13.89	17.51	76.23	86.17	92.58	102.50
Never considered	724.0	86.85	11.27	2.00	79.83	88.33	95.33	105.67
Took CS 10	54.0	73.91	15.24	6.00	68.92	75.33	82.17	99.17
Python Experience								
False	733.0	82.05	12.65	0.0	75.33	83.33	90.50	103.83
True	687.0	87.26	10.98	6.0	80.21	89.00	95.83	105.67

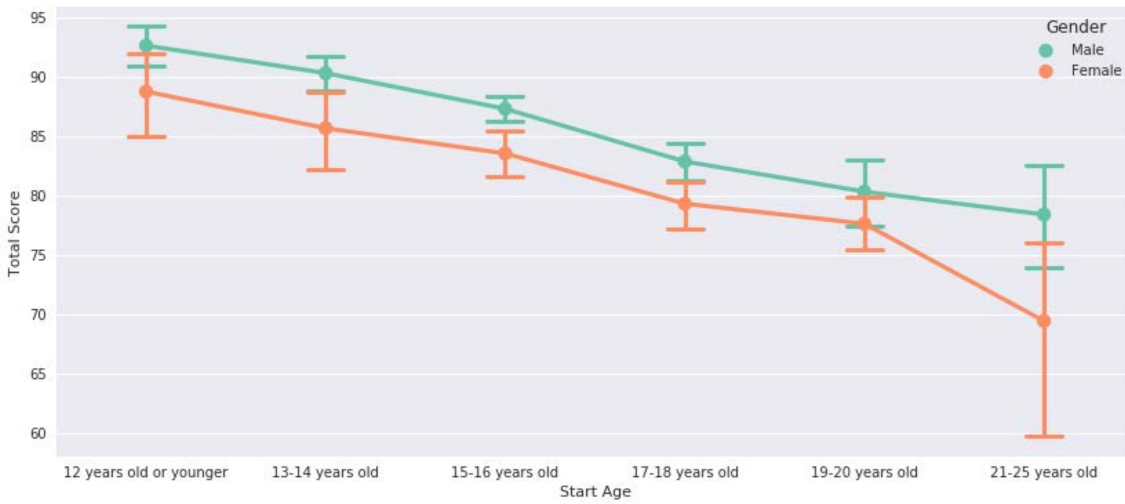

Figure 1: Interaction plot for gender and start age (61A). An orange or green dot indicates the mean total score of all students with the corresponding gender and start age. Error bars indicate 95% confidence intervals.

worse than students in the other three groups. The full post-hoc analysis results are shown in Table 3.

- **Python experience**: We found a significant effect, $F(1,1418) = 68.26$, $p < .001$, $\eta^2 = .05$. Students with Python experience performed significantly better than those without, $t(1409.81) = 8.30$, $p < .001$.

We then looked at how gender, URM status, and major relate to performance, controlling for start age. To do this, we performed a one-way ANCOVA for each factor, with start age (treated as a continuous variable) as the covariate. Each model included an interaction term, to determine whether there exists a significant interaction between the factor and the covariate.

- **Gender**: We found significant main effects for gender, $F(1,1365) = 33.81$, $p < .001$, $\eta^2 = .02$, and start age, $F(1,1365) = 175.82$, $p < .001$, $\eta^2 = .11$, but the interaction between gender and start age was not significant, $F(1,1365) = .003$, $p = .96$, $\eta^2 < .001$. Indeed, we found this model to be surprisingly additive: as shown in Figure 1, the gap between the male and female group means is fairly consistent across start age groups.

- **URM status**: We found significant main effects for URM status, $F(2,1365) = 36.04$, $p < .001$, $\eta^2 = .05$, and start age, $F(1,1365) = 180.82$, $p < .001$, $\eta^2 = .11$. The interaction was not significant, $F(2,1365) = 2.64$, $p = .07$, $\eta^2 < .01$.

- **Major (EECS vs. LSCS vs. Other)**: We found significant main effects for major, $F(2,1380) = 30.37$, $p < .001$, $\eta^2 = .04$, and start age, $F(1,1365) = 116.48$, $p < .001$, $\eta^2 = .07$. The interaction was not significant, $F(2,1365) = 1.67$, $p = .18$, $\eta^2 < .01$.

Table 2: Start age post-hoc analysis (Tukey HSD test, FWER $= 0.05$) for 61A. *meandiff* is the difference between group means (*group2* mean minus *group1* mean). *lower* and *upper* are the bounds for a 95% confidence interval (adjusted for family-wise error rate) for *meandiff*. *reject* indicates whether the mean difference is significantly different from zero.

group1	group2	meandiff	lower	upper	reject
12 years old or younger	13-14 years old	-2.4902	-5.9481	0.9677	False
12 years old or younger	15-16 years old	-5.7591	-8.9809	-2.5372	True
12 years old or younger	17-18 years old	-10.2873	-13.5496	-7.025	True
12 years old or younger	19-20 years old	-13.0214	-16.7787	-9.2641	True
12 years old or younger	21-25 years old	-15.733	-20.8617	-10.6044	True
13-14 years old	15-16 years old	-3.2689	-5.7765	-0.7613	True
13-14 years old	17-18 years old	-7.7971	-10.3565	-5.2377	True
13-14 years old	19-20 years old	-10.5312	-13.6974	-7.3649	True
13-14 years old	21-25 years old	-13.2428	-17.9556	-8.53	True
15-16 years old	17-18 years old	-4.5282	-6.7584	-2.298	True
15-16 years old	19-20 years old	-7.2623	-10.1689	-4.3557	True
15-16 years old	21-25 years old	-9.974	-14.5164	-5.4315	True
17-18 years old	19-20 years old	-2.7341	-5.6855	0.2173	False
17-18 years old	21-25 years old	-5.4457	-10.017	-0.8745	True
19-20 years old	21-25 years old	-2.7117	-7.6483	2.225	False

Table 3: "Sentiment on taking CS 10" post-hoc analysis (Tukey HSD test, FWER $= 0.05$) for 61A.

group1	group2	meandiff	lower	upper	reject
Considered briefly	Considered strongly	-5.6754	-10.2501	-1.1007	True
Considered briefly	Don't know about CS 10	0.3093	-3.2817	3.9003	False
Considered briefly	Never considered	3.7116	1.8279	5.5954	True
Considered briefly	Took CS 10	-9.2278	-13.8404	-4.6152	True
Considered strongly	Don't know about CS 10	5.9847	0.5456	11.4237	True
Considered strongly	Never considered	9.387	4.8885	13.8855	True
Considered strongly	Took CS 10	-3.5524	-9.7139	2.6091	False
Don't know about CS 10	Never considered	3.4023	-0.0911	6.8958	False
Don't know about CS 10	Took CS 10	-9.5371	-15.0081	-4.0661	True
Never considered	Took CS 10	-12.9394	-17.4765	-8.4023	True

3.2 61B

We repeated the above procedures for 61B.

Our data on student background factors came from the end-of-semester survey. Appendix A.2 describes how this data was preprocessed.

Table 4 shows the size and mean total score of each group, along with other descriptive statistics.

Our one-way ANOVA results were as follows:

- **Gender**: We found a significant effect, $F(1, 1082) = 18.77$, $p < .001$, $\eta^2 = .02$. Males performed significantly better than females, $t(499.42) = 4.36$, $p < .001$. The mean difference was 3.06 points.
- **Start age**: We found a significant effect, $F(5, 1124) = 25.56$, $p < .001$, $\eta^2 = .10$. The post-hoc analysis results are shown in Table 5. Generally, students who started programming earlier performed better than those who didn't.
- **URM status**: We found a significant effect, $F(2, 1103) = 28.40$, $p < .001$, $\eta^2 = .05$. Students who responded "No" performed significantly better than those who responded "Yes" or "I don't know", but the "Yes" and "I don't know" groups did not differ significantly.
- **Major (EECS vs. LSCS vs. Other)**: We found a significant effect, $F(2, 1136) = 36.99$, $p < .001$, $\eta^2 = .06$. LSCS students and EECS students performed significantly better than Other students, but LSCS and EECS students did not differ significantly.
- **Major and transfer status (EECS non-transfer vs. LSCS non-transfer vs. EECS transfer vs. LSCS transfer)**: We found a significant effect, $F(3, 862) = 5.45$, $p < .01$, $\eta^2 = .02$. EECS non-transfer students performed significantly better than EECS transfer students and LSCS transfer students. No other pair of groups differed significantly.
- **High school CS experience**: We found a significant effect, $F(1, 1137) = 36.75$, $p < .001$, $\eta^2 = .03$. Students with high school CS experience performed significantly better than those without, $t(1110.90) = 6.10$, $p < .001$. The mean difference was 3.91 points.
- **Sentiment on taking CS 10**: We found a significant effect, $F(4, 1134) = 19.94$, $p < .001$, $\eta^2 = .07$. The post-hoc analysis results are shown in Table 6.
- **Java experience**: We found a significant effect, $F(1, 1137) = 46.74$, $p < .001$, $\eta^2 = .04$. Students with Java experience performed significantly better than those without, $t(790.93) = 6.83$, $p < .001$. The mean difference was 3.91 points.

The results for our ANCOVAs with start age as the covariate were as follows:[3]

- **Gender**: We found significant main effects for gender, $F(1, 1022) = 5.68$, $p = .02$, $\eta^2 < .01$, and start age, $F(1, 1022) = 140.41$, $p < .001$, $\eta^2 = .12$. The interaction was not significant, $F(1, 1022) = 2.13$, $p = .15$, $\eta^2 < .01$. We did not see the high level of additivity we saw for 61A.

[3] As before, we treated start age as a continuous variable. We also excluded the "21-25 years old" group to ensure that performance and start age were linearly related.

Table 4: Descriptive statistics of the 61B total scores for various groups.

	count	mean	std	min	25%	50%	75%	max
Gender								
Female	283.0	81.10	10.12	18.66	76.60	82.92	88.13	98.18
Male	801.0	84.16	10.23	1.95	79.83	86.27	90.84	99.71
Start Age								
12 years old or younger	73.0	89.01	12.10	1.17	86.54	91.34	94.95	99.05
13-14 years old	147.0	87.95	8.42	32.89	85.05	89.41	93.23	99.71
15-16 years old	293.0	85.09	9.32	1.17	81.98	86.85	90.81	98.87
17-18 years old	428.0	81.37	11.03	1.95	77.55	83.09	88.60	99.32
19-20 years old	139.0	77.13	11.78	18.66	70.52	80.49	84.82	98.18
21-25 years old	50.0	80.03	9.80	46.41	76.22	81.67	85.99	96.44
URM Status								
I don't know	116.0	78.95	11.30	32.56	74.20	81.60	86.20	96.44
No	846.0	84.43	10.78	1.17	80.63	86.49	91.15	99.71
Yes	144.0	78.45	10.46	18.66	74.84	80.88	85.35	96.41
Major								
EECS	328.0	85.07	9.52	34.27	80.72	86.99	91.93	99.45
LSCS	538.0	84.25	10.45	1.17	81.22	85.76	90.17	99.35
Other	273.0	78.25	12.28	1.17	71.17	79.30	87.27	99.71
Major and Transfer Status								
EECS non-transfer	249.0	86.08	9.08	34.27	81.76	88.05	92.51	99.45
EECS transfer	79.0	81.89	10.21	46.41	77.32	84.32	87.87	99.05
LSCS non-transfer	500.0	84.51	10.07	1.17	81.37	86.00	90.35	99.35
LSCS transfer	38.0	80.92	14.28	13.79	79.73	83.30	86.83	97.25
High School CS Experience								
False	631.0	81.30	11.13	1.17	77.06	83.43	88.50	99.32
True	508.0	85.22	10.44	1.17	81.38	87.30	91.65	99.71
Sentiment on Taking CS 10								
Considered briefly	226.0	80.84	10.33	32.56	76.46	83.25	88.12	97.64
Considered strongly	27.0	78.24	13.08	34.27	75.28	81.43	86.65	90.70
Don't know about CS 10	148.0	83.35	8.64	57.99	78.89	84.26	89.08	99.71
Never considered	688.0	84.66	10.75	1.17	80.83	86.71	91.33	99.45
Took CS 10	50.0	72.55	14.13	18.66	64.89	74.41	81.84	95.22
Java Experience								
False	392.0	80.03	10.82	18.66	75.16	81.65	87.21	99.32
True	747.0	84.63	10.77	1.17	80.95	86.71	91.18	99.71

Table 5: Start age post-hoc analysis (Tukey HSD test, FWER $= 0.05$) for 61B.

group1	group2	meandiff	lower	upper	reject
12 years old or younger	13-14 years old	-1.0551	-5.3139	3.2036	False
12 years old or younger	15-16 years old	-3.9147	-7.8055	-0.0239	True
12 years old or younger	17-18 years old	-7.6374	-11.4038	-3.871	True
12 years old or younger	19-20 years old	-11.8755	-16.1748	-7.5763	True
12 years old or younger	21-25 years old	-8.9773	-14.4373	-3.5172	True
13-14 years old	15-16 years old	-2.8595	-5.8658	0.1467	False
13-14 years old	17-18 years old	-6.5822	-9.4257	-3.7388	True
13-14 years old	19-20 years old	-10.8204	-14.3393	-7.3015	True
13-14 years old	21-25 years old	-7.9222	-12.7916	-3.0527	True
15-16 years old	17-18 years old	-3.7227	-5.978	-1.4674	True
15-16 years old	19-20 years old	-7.9609	-11.0242	-4.8975	True
15-16 years old	21-25 years old	-5.0626	-9.6137	-0.5115	True
17-18 years old	19-20 years old	-4.2381	-7.1419	-1.3344	True
17-18 years old	21-25 years old	-1.3399	-5.7852	3.1054	False
19-20 years old	21-25 years old	2.8982	-2.0066	7.8031	False

Table 6: "Sentiment on taking CS 10" post-hoc analysis (Tukey HSD test, FWER $= 0.05$) for 61B.

group1	group2	meandiff	lower	upper	reject
Considered briefly	Considered strongly	-2.5996	-8.5246	3.3254	False
Considered briefly	Don't know about CS 10	2.5063	-0.5706	5.5832	False
Considered briefly	Never considered	3.8209	1.59	6.0519	True
Considered briefly	Took CS 10	-8.2953	-12.8429	-3.7477	True
Considered strongly	Don't know about CS 10	5.1059	-0.9834	11.1953	False
Considered strongly	Never considered	6.4205	0.7118	12.1293	True
Considered strongly	Took CS 10	-5.6957	-12.6451	1.2536	False
Don't know about CS 10	Never considered	1.3146	-1.322	3.9512	False
Don't know about CS 10	Took CS 10	-10.8016	-15.5613	-6.0419	True
Never considered	Took CS 10	-12.1162	-16.3782	-7.8542	True

- **URM status**: We found significant main effects for URM status, $F(2, 1041) = 15.28$, $p < .001$, $\eta^2 = .03$, and start age, $F(1, 1041) = 93.59$, $p < .001$, $\eta^2 = .08$. The interaction was not significant, $F(2, 1041) < .001$, $p > .999$, $\eta^2 < .001$.
- **Major (EECS vs. LSCS vs. Other)**: We found significant main effects for major, $F(2, 1074) = 19.98$, $p < .001$, $\eta^2 = .03$, and start age, $F(1, 1074) = 75.40$, $p < .001$, $\eta^2 = .06$. The interaction was not significant, $F(2, 1074) = 2.08$, $p = .12$, $\eta^2 < .01$.

3.3 Discussion

To sum up, there was a statistically significant difference in mean performance between groups for every factor we tested, in both 61A and 61B. Generally speaking, stronger performance was associated with being male, having started programming at an early age, not self-identifying as an underpresented minority, being either an LSCS or EECS student, not being a transfer student, having high school CS experience, not having considered taking CS 10, and having prior experience in the primary programming language for the course.

These results were largely in line with what we expected based on our prior experience as instructors, but it contradicts results at other institutions that found, for instance, no significant effect for prior programming experience in CS2 [8] or no significant effect for gender in CS1 [5, 11].

Despite popular claims to the contrary, we did not find a statistically significant difference in mean performance between LSCS students and EECS students in either 61A or 61B. This is true even when looking at only transfer students or only non-transfer students.

All effect sizes (as measured by eta-squared) were small or moderate. In both courses, start age explained the largest proportion of variance in performance (13% in 61A and 10% in 61B). Note that because factors have varying group sizes and numbers of groups, it is difficult to directly compare effect sizes between factors.

Generally, for a given factor, the 61A effect size was larger than the 61B effect size. This suggests that 61B performance depends less on background factors than does 61A performance, perhaps due to self-selection.

In both courses, gender, URM status, and major all had significant effects even when controlling for start age. This suggests that the performance gaps associated with those factors cannot be fully explained by differences in prior programming experience.

4 Prior Grades

On the 61B start-of-semester survey, we asked students about their prior grade—the letter grade they had received in 61A. Table 7 shows the size and mean total score of each prior grade group, along with other descriptive statistics.[4]

We had the following questions:

1. How do background factors relate to performance, controlling for prior grade?
2. How strongly does prior grade predict performance?

[4]Students who took an introductory programming course that is not in the EECS/LSCS curriculum, such as Engineering 7, are included in the "Didn't take" group.

Table 7: Descriptive statistics of the 61B total scores for various levels of prior grade.

Prior Grade	count	mean	std	min	25%	50%	75%	max
A+	40.0	95.13	3.50	85.69	93.96	95.99	97.15	99.71
A	211.0	90.08	11.23	1.17	89.09	92.10	94.30	99.45
A-	246.0	87.56	4.59	59.04	85.48	88.27	90.58	95.75
B+	284.0	83.04	6.21	51.53	80.45	84.00	86.86	96.19
B	218.0	75.51	10.03	1.95	72.39	77.72	81.48	88.60
B-	52.0	67.71	10.40	32.56	63.37	70.24	75.15	82.13
C+	10.0	64.45	8.76	50.70	57.29	64.94	72.52	75.16
C	8.0	59.18	11.41	40.12	54.84	60.77	65.10	75.41
C-	1.0	54.17	NaN	54.17	54.17	54.17	54.17	54.17
Didn't take	49.0	82.20	10.36	55.96	77.34	83.38	89.65	98.18
Decline to state	20.0	73.93	14.66	32.89	67.01	76.79	85.03	92.21

4.1 Prior Grade as a Controlling Factor

We performed a one-way ANCOVA for each factor, with prior grade as the covariate.[5] Each model included an interaction term, to determine whether there is a significant interaction between the factor and the covariate.

- **Gender**: We did not find a significant main effect for gender, $F(1, 996) = 0.53$, $p = .47$, $\eta^2 < .001$. We did find a significant main effect for prior grade, $F(1, 996) = 844.60$, $p < .001$, $\eta^2 = .46$. The interaction was not significant, $F(1, 996) = .21$, $p = .64$, $\eta^2 < .001$. See Figure 2 for visualization.
- **Start age**: We found significant main effects for start age, $F(5, 1034) = 5.28$, $p < .001$, $\eta^2 = .02$, and prior grade, $F(1, 1034) = 446.98$, $p < .001$, $\eta^2 = .29$. The interaction was not significant, $F(5, 1034) = 1.70$, $p = .13$, $\eta^2 < .01$.
- **URM status**: We found significant main effects for URM status, $F(2, 1016) = 5.75$, $p < .01$, $\eta^2 < .01$, and prior grade, $F(1, 1016) = 508.11$, $p < .001$, $\eta^2 = .33$. The interaction was not significant, $F(2, 1016) = .71$, $p = .49$, $\eta^2 < .001$.
- **Major (EECS vs. LSCS vs. Other)**: We found significant main effects for major, $F(2, 1045) = 12.91$, $p < .001$, $\eta^2 = .02$, and prior grade, $F(1, 1045) = 552.86$, $p < .001$, $\eta^2 = .34$. The interaction was not significant, $F(2, 1045) = 2.99$, $p = .05$, $\eta^2 < .01$.
- **Major and transfer status (EECS non-transfer vs. LSCS non-transfer vs. EECS transfer vs. LSCS transfer)**: We found a significant interaction, $F(3, 825) = 6.52$, $p < .001$, $\eta^2 = .02$. The interaction appears to be disordinal in nature: the performance of LSCS transfer students remains relatively consistant across levels of prior grade, whereas for the other three groups, performance increases as prior grade improves. Because the interaction was significant, we did not interpret the main effects.

[5] We treated prior grade as a continuous variable, using only the six grade bins between B- and A+. (The C-range grade bins contained too few students.)

Table 8: Contingency table relating 61B grade and prior grade.

61B Grade Prior Grade	A+	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	F	All
A+	16	20	3	1	0	0	0	0	0	0	0	0	0	40
A	14	109	65	16	3	0	0	1	0	0	0	0	3	211
A-	0	52	125	57	10	0	1	1	0	0	0	0	0	246
B+	0	14	87	132	33	10	4	2	2	0	0	0	0	284
B	0	0	7	87	64	21	19	16	0	1	0	1	2	218
B-	0	0	0	5	14	10	9	7	4	2	0	1	0	52
C+	0	0	0	0	3	1	1	3	2	0	0	0	0	10
C	0	0	0	0	1	0	1	4	0	1	1	0	0	8
C-	0	0	0	0	0	0	0	0	1	0	0	0	0	1
Didn't take	2	8	10	15	9	0	1	3	1	0	0	0	0	49
Decline to state	0	2	2	4	3	3	2	2	1	0	0	1	0	20
All	32	205	299	317	140	45	38	39	11	4	1	3	5	1139

- **High school CS experience**: We did not find a significant main effect for high school CS experience, $F(1, 1047) = 3.57$, $p = .06$, $\eta^2 < .01$. We did find a significant main effect for prior grade, $F(1, 1047) = 555.58$, $p < .001$, $\eta^2 = .35$. The interaction was not significant, $F(1, 1047) = .68$, $p = .41$, $\eta^2 < .001$. See Figure 3 for visualization.

- **Sentiment on taking CS 10**: We found significant main effects for sentiment on taking CS 10, $F(4, 1041) = 4.38$, $p < .01$, $\eta^2 = .01$, and prior grade, $F(1, 1041) = 525.20$, $p < .001$, $\eta^2 = .33$. The interaction was not significant, $F(4, 1041) = 1.04$, $p = .38$, $\eta^2 < .01$.

- **Java experience**: We found significant main effects for Java experience, $F(1, 1047) = 5.16$, $p = .02$, $\eta^2 < .01$, and prior grade, $F(1, 1047) = 543.83$, $p < .001$, $\eta^2 = .34$. The interaction was not significant, $F(1, 1047) = 1.26$, $p = .26$, $\eta^2 < .001$.

4.2 Prior Grade as a Predictor

Table 8 describes the multivariate frequency distribution of 61A and 61B letter grades. It is clear from this table alone that prior grade is a *very* strong predictor of 61B performance. Most students lie near the main diagonal of the table (shaded), indicating only a small difference between their 61A and 61B letter grades. Indeed, 86.5% of 61B students received letter grades within one bin of their 61A grades. Only one single student moved up three or more grade bins (from a C to a B).

To more accurately measure the predictive ability of prior grade, we trained a random forest regression model[6] to predict performance, as measured by total score, from prior grade (dummy-encoded). We measured the model's accuracy using 10-fold cross-validation: on average, the model attained a mean absolute error (MAE) of 6.17. This is approximately the width of the typical 61B grade bin (5.92 points).

[6]This procedure is implemented in the *scikit-learn* library [15] as `sklearn.ensemble.RandomForestRegressor`. We used the default hyperparameters, e.g., our forest contained 10 trees.

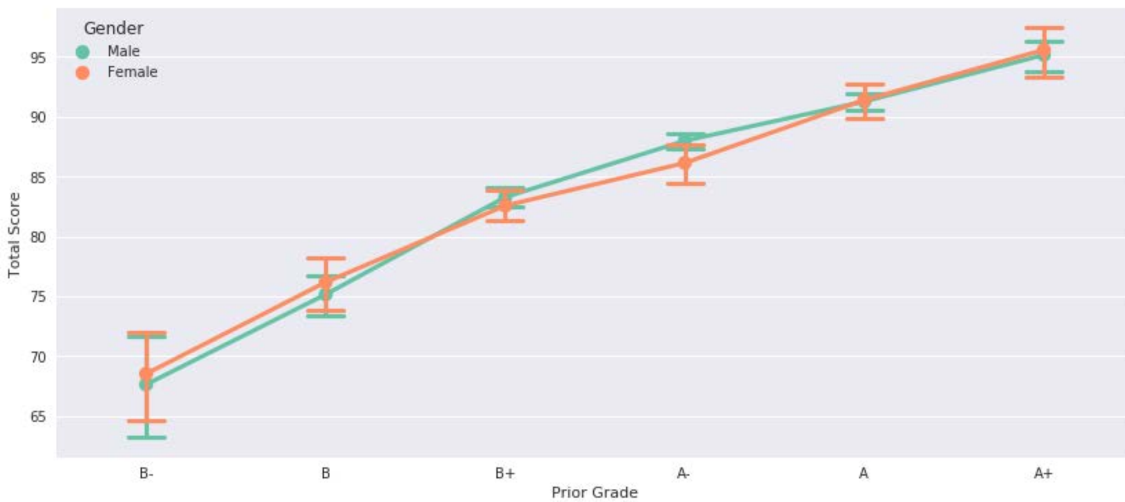

Figure 2: Interaction plot for gender and prior grade (61B).

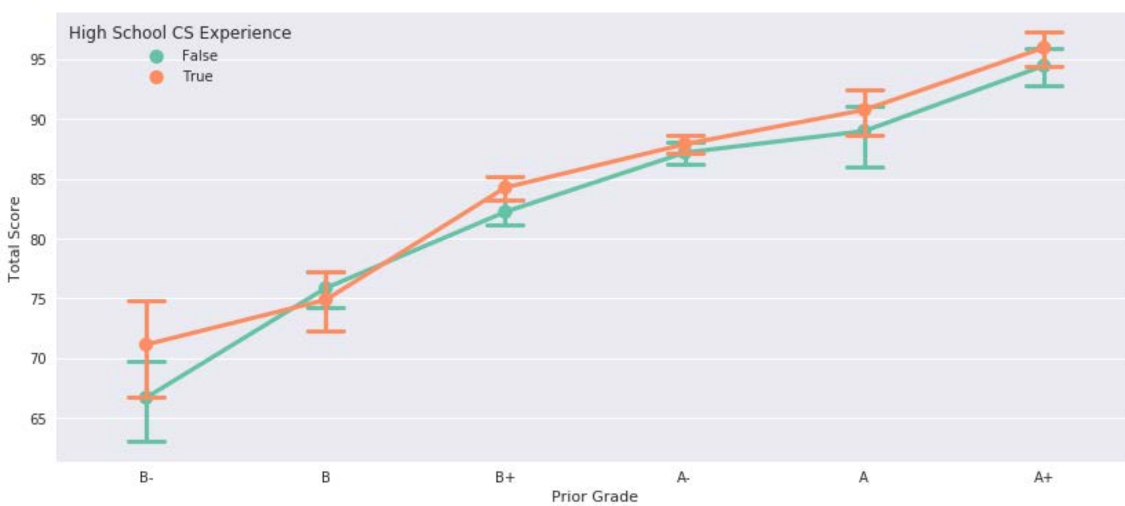

Figure 3: Interaction plot for high school CS experience and prior grade (61B).

For comparison, we trained another model on the combined background factors from Subsection 3.2[7] (also dummy-encoded). This model attained a 10-fold cross-validation MAE of 8.26. This suggests that prior grade significantly outperforms the background factors in predicting performance.

4.3 Discussion

Prior grade was a highly effective covariate, explaining away a substantial portion of the variability in performance. Controlling for prior grade led to a substantial drop in effect size for each of the 61B background factors (relative to the effect sizes reported in Subsection 3.2).

We were surprised to find that, controlling for prior grade, neither gender nor high school CS experience had a significant main effect on performance. In other words, we should expect a man and a woman who performed equally well in 61A to perform equally well in 61B. And we should expect two students—one with high school CS experience and the other without—who performed equally well in 61A to perform equally well in 61B.

It is also clear that prior grade is a strong predictive factor for performance. Specifically, using only a student's 61A grade, we can expect to predict their total score in 61B with an error of approximately one grade bin.

5 Interventions

Our final goal was to measure the benefit (increase in performance) attributable to each of the four interventions described in Subsection 2.3: CSM sections, group tutoring, one-on-one tutoring, and guerrilla sections.

For each intervention, our procedure was as follows:

1. Based only on the students who did *not* receive the intervention, we fit a linear regression model predicting final exam scores from Midterm 1 scores.

2. For each student who *did* receive the intervention, we used the model to predict their final exam score from their actual Midterm 1 score. (This is our best estimate of what they would have received on the final exam had they not received the intervention.)

3. We then subtracted their predicted final exam score from their actual final exam score to compute the "improvement" attributable to the intervention.

This approach is reasonable because the interventions we considered were primarily administered between Midterm 1 (our "pretest") and the final exam (our "posttest"). We chose linear regression as our model because we had strong prior expectations that final exam scores and Midterm 1 scores were linearly related. We found this to be true in practice.[8]

Our results were as follows:

[7] Instead of using the"major and transfer status" factor from Subsection 3.2, we used we used a single dichotomous (true/false) factor representing only transfer status.

[8] Each linearly regression model we fit had $r^2 > 0.45$.

- **CSM sections**: Our intervention group consisted of all students who signed up to attend CSM sections, of which there were 291.[9] The mean improvement was -0.59 points.[10] The mean improvement was significantly different from zero, $t(290) = 3.83$, $p < .001$.
- **Group tutoring**: Our intervention group consisted of all students who attended at least one group tutoring session, of which there were 84. The mean improvement was 0.10 points. The mean improvement was not significantly different from zero, $t(83) = .33$, $p = .74$.
- **One-on-one tutoring**: Our intervention group consisted of all students who attended at least one one-on-one tutoring session, of which there were 116. The mean improvement was -0.25 points. The mean improvement was not significantly different from zero, $t(115) = 1.00$, $p = .32$.[11]
- **Guerrilla sections**: Our intervention group consisted of all students who attended at least three guerrilla sections,[12] of which there were 129. The mean improvement was -0.35 points. The mean improvement was not significantly different from zero, $t(128) = 1.56$, $p = .12$.

5.1 Discussion

Of the interventions we considered, none were associated with a statistically significant mean increase in performance. On the contrary, CSM sections were associated with a slight *decrease* in performance.

Because the interventions were self-selected, not randomly assigned, interpreting these results is difficult. In particular, it would be wrong to conclude any kind of direct causal relationship.

Prior to conducting this analysis, we were unsure about the direction in which self-selection biases the data: Should we expect students who opt into interventions to perform *better* than their peers because they are more engaged in the course and more concerned about their performance? Or should we expect them to perform *worse* than their peers because they are aware of their own weaknesses (including those not measured by Midterm 1) and are seeking help? Assuming that the interventions do not actually cause a decrease in performance, it appears that the latter effect (downward bias) is dominant.

Although these results are somewhat disappointing, we are optimistic about the amount of benefit interventions could theoretically provide. During the Spring 2016 offering of 61B, the professor (Josh Hug) gave students the option to receive a failing grade in the course, conditional on their performance, so that they could retake 61B in a later semester. 34 students received failing grades through this process. 28 out of those 34 subsequently retook and completed 61B between Summer 2016 and Spring 2017. The average difference between their new grades and the grades they would have received in Spring 2016 was +2.54 grade bins—a remarkable improvement.[13] 14 of the 28 students improved by three or more grade bins. Moreover, 4 out of 28 saw an increase in

[9] For each of these analyses, we excluded students who did not take both Midterm 1 and the final.

[10] The final was worth 26.04 (scaled) points.

[11] We repeated this analysis with the 43 students who attended at least *three* one-on-one tutoring sessions as our intervention group. The results were similar: the mean improvement was -0.24 points and the mean improvement was not significantly different from zero, $t(42) = .54$, $p = .59$.

[12] We chose this threshold because there were two guerrilla sections prior to Midterm 1, so each student in our intervention group attended at least one guerrilla section after Midterm 1.

[13] To our knowledge, the difficulty of the course does not vary significantly from semester to semester.

three or more grade bins relative to their self-reported 61A letter grade.[14] By comparison, only 1 out of 573 students made this jump in Spring 2017 (see Table 8).[15]

Our takeaway from this experience is that a student who does not do well in 61B initially is far from doomed: with additional time and resources, they too can succeed. The interventions we provide should seek to provide as much of this benefit as possible within the timeframe of a single semester.

6 Conclusion

In this work, we described the relationships between various student background factors and performance in 61A and 61B. We were surprised by the strength of prior grade as a predictive factor. And we assessed the benefit associated with interventions like CSM sections and group tutoring, with inconclusive results.

Practically speaking, our results are not encouraging. Some students perform worse than others, in both 61A and 61B. Moreover, students who perform poorly in 61A are likely to continue performing poorly in 61B, and it is unclear whether the provided interventions lead to substantial increases in performance.

Despite all of this, the Spring 2016 experiment mentioned above gives us hope. By providing the right resources to the right students, we may yet see significant, measurable gains in performance.

A natural next step is to perform randomized controlled experiments to assess the effectiveness of interventions under minimal-bias conditions. While there are ethical concerns associated with implementing such experiments in their most direct form—randomly selecting students to either receive or be denied an intervention—channeling departmental resources toward potentially ineffective interventions would be an even greater tragedy.

[14] Self-reported 61A letter grades were collected in the end-of-semester survey. The wording was the same as what is shown in Appendix A.2.

[15] 573 is the number of students with self-reported 61A grades between C- and B+, i.e., it excludes students who could not possibly have improved by three or more grade bins.

Bibliography

[1] Carol Ann Alspaugh. Identification of some components of computer programming aptitude. *Journal for Research in Mathematics Education*, 3(2):89–98, 1972. ISSN 00218251, 19452306. URL `http://www.jstor.org/stable/748667`.

[2] Francisco Araque, Concepción Roldán, and Alberto Salguero. Factors influencing university drop out rates. *Comput. Educ.*, 53(3):563–574, November 2009. ISSN 0360-1315. doi: 10.1016/j.compedu.2009.03.013. URL `http://dx.doi.org/10.1016/j.compedu.2009.03.013`.

[3] Susan Bergin and Ronan Reilly. Programming: Factors that influence success. *SIGCSE Bull.*, 37(1):411–415, February 2005. ISSN 0097-8418. doi: 10.1145/1047124.1047480. URL `http://doi.acm.org/10.1145/1047124.1047480`.

[4] D. F. Butcher and W. A. Muth. Predicting performance in an introductory computer science course. *Commun. ACM*, 28(3):263–268, March 1985. ISSN 0001-0782. doi: 10.1145/3166.3167. URL `http://doi.acm.org/10.1145/3166.3167`.

[5] Pat Byrne and Gerry Lyons. The effect of student attributes on success in programming. *SIGCSE Bull.*, 33(3):49–52, June 2001. ISSN 0097-8418. doi: 10.1145/507758.377467. URL `http://doi.acm.org/10.1145/507758.377467`.

[6] Annagret Goold and Russell Rimmer. Factors affecting performance in first-year computing. *SIGCSE Bull.*, 32(2):39–43, June 2000. ISSN 0097-8418. doi: 10.1145/355354.355369. URL `http://doi.acm.org/10.1145/355354.355369`.

[7] John W. Henry, Mark J. Martinko, and Margaret Anne Pierce. Attributional style as a predictor of success in a first computer science course. *Computers in Human Behavior*, 9(4): 341 – 352, 1993. ISSN 0747-5632. doi: https://doi.org/10.1016/0747-5632(93)90027-P. URL `http://www.sciencedirect.com/science/article/pii/074756329390027P`.

[8] Edward Holden and Elissa Weeden. The experience factor in early programming education. In *Proceedings of the 5th Conference on Information Technology Education*, CITC5 '04, pages 211–218, New York, NY, USA, 2004. ACM. ISBN 1-58113-936-5. doi: 10.1145/1029533.1029585. URL `http://doi.acm.org/10.1145/1029533.1029585`.

[9] Charles P. Howerton. The impact of precollege computer exposure on student achievement in introductory computer programming courses. *Computer Science Education*, 1(1):73–84, 1988. doi: 10.1080/0899340880010107. URL `https://doi.org/10.1080/0899340880010107`.

[10] Petri Ihantola, Arto Vihavainen, Alireza Ahadi, Matthew Butler, Jürgen Börstler, Stephen H. Edwards, Essi Isohanni, Ari Korhonen, Andrew Petersen, Kelly Rivers, Miguel Ángel Rubio, Judy Sheard, Bronius Skupas, Jaime Spacco, Claudia Szabo, and Daniel Toll. Educational data mining and learning analytics in programming: Literature review and case studies. In *Proceedings of the 2015 ITiCSE on Working Group Reports*, ITICSE-WGR '15, pages 41–63, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-4146-2. doi: 10.1145/2858796.2858798. URL `http://doi.acm.org/10.1145/2858796.2858798`.

[11] Philip R. Ventura Jr. Identifying predictors of success for an objects-first cs1. *Computer Science Education*, 15(3):223–243, 2005. doi: 10.1080/08993400500224419. URL `https://doi.org/10.1080/08993400500224419`.

[12] Sandra Katz, David Allbritton, John Aronis, Christine Wilson, and Mary Lou Soffa. Gender, achievement, and persistence in an undergraduate computer science program. *SIGMIS Database*, 37(4):42–57, November 2006. ISSN 0095-0033. doi: 10.1145/1185335.1185344. URL `http://doi.acm.org/10.1145/1185335.1185344`.

[13] John Konvalina, Stanley A. Wileman, and Larry J. Stephens. Math proficiency: A key to success for computer science students. *Commun. ACM*, 26(5):377–382, May 1983. ISSN 0001-0782. doi: 10.1145/69586.358140. URL `http://doi.acm.org/10.1145/69586.358140`.

[14] Jonathan P. Munson and Joshua P. Zitovsky. Models for early identification of struggling novice programmers. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, SIGCSE '18, pages 699–704, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-5103-4. doi: 10.1145/3159450.3159476. URL `http://doi.acm.org/10.1145/3159450.3159476`.

[15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[16] Cristóbal Romero, Manuel-Ignacio López, Jose-María Luna, and Sebastián Ventura. Predicting students' final performance from participation in on-line discussion forums. *Comput. Educ.*, 68:458–472, October 2013. ISSN 0360-1315. doi: 10.1016/j.compedu.2013.06.009. URL `http://dx.doi.org/10.1016/j.compedu.2013.06.009`.

[17] Marian Gunsher Sackrowitz and Ann Parker Parelius. An unlevel playing field: Women in the introductory computer science courses. *SIGCSE Bull.*, 28(1):37–41, March 1996. ISSN 0097-8418. doi: 10.1145/236462.236488. URL `http://doi.acm.org/10.1145/236462.236488`.

[18] Skipper Seabold and Josef Perktold. Statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

[19] Anya Tafliovich, Jennifer Campbell, and Andrew Petersen. A student perspective on prior experience in cs1. In *Proceeding of the 44th ACM Technical Symposium on Computer Science Education*, SIGCSE '13, pages 239–244, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1868-6. doi: 10.1145/2445196.2445270. URL `http://doi.acm.org/10.1145/2445196.2445270`.

[20] Laurie Honour Werth. Predicting student performance in a beginning computer science class. *SIGCSE Bull.*, 18(1):138–143, February 1986. ISSN 0097-8418. doi: 10.1145/953055.5701. URL `http://doi.acm.org/10.1145/953055.5701`.

[21] Chris Wilcox and Albert Lionelle. Quantifying the benefits of prior programming experience in an introductory computer science course. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, SIGCSE '18, pages 80–85, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-5103-4. doi: 10.1145/3159450.3159480. URL `http://doi.acm.org/10.1145/3159450.3159480`.

[22] Brenda Cantwell Wilson and Sharon Shrock. Contributing to success in an introductory computer science course: A study of twelve factors. *SIGCSE Bull.*, 33(1):184–188, February 2001. ISSN 0097-8418. doi: 10.1145/366413.364581. URL `http://doi.acm.org/10.1145/366413.364581`.

A Survey Questions

A.1 61A Start-of-Semester Survey

Factor	Relevant Survey Question(s)	Preprocessing
Gender	Gender • Male • Female • Trans • Non-binary • Decline to state • Other (custom response)	Excluded all students who did not respond either “Male” or “Female” (20 total).
Start age	When did you start programming? If you’re not sure, your best guess is fine. • 12 years old or younger • 13-14 years old • 15-16 years old • 17-18 years old • 19-20 years old • 21-25 years old • 26+ years old • Decline to state	Excluded all students who responded “26+ years old” (18 students) or “Decline to state” (38).
URM status	Ethnic or Racial Minority • Yes • No • Don’t know • Decline to state [Help text: Do you consider yourself to be an underrepresented ethnic or racial minority within computer science courses?]	Excluded all students who responded “Decline to state” (51 students).
Major	What is your major (or majors)? [Can select multiple options.] • Undeclared and don’t know yet • L&S, intending to declare CS • L&S CS (already officially declared) • EECS • Applied Mathematics • Architecture • [25 more options...] • Decline to state • Other (custom response) [Help text: If you are undeclared, select your most likely choice(s).]	Mapped all students who selected LSCS as one of their majors to “LSCS”, all students who selected EECS as one of their majors to “EECS”, and all other students to “Other”.

Factor	Relevant Survey Question(s)	Preprocessing
Transfer status	Are you a transfer student? • Yes, from a community college • Yes, from another four year college • No • Decline to state	Excluded all students who responded "Decline to state" (7 students). Mapped both affirmative groups to "Yes".
Major and transfer status	[This was a combination of the previous two factors.]	We only wanted to consider LSCS and EECS students for this factor, so all students in the "Other" group for major were also excluded.
High school CS experience	Which of the courses below have you taken? [Can select multiple options.] • CS 10 at UC Berkeley • DS 8 at UC Berkeley • E 7 at UC Berkeley • At least one high school computer science course • An introductory programming course at another college or university • At least one online programming course (e.g. Coursera) [Help text: If considering online courses, check the box only if you completed at least half of the class.]	Excluded all non-responders (187 students), then mapped all students who selected "At least one high school computer science course" to "True" and all other students to "False".
Sentiment on taking CS 10	Did you consider taking CS 10? • Yes, and I took CS 10 • I strongly considered taking CS 10, but did not • I briefly considered taking CS 10, but did not • I never considered taking CS 10 • I don't know anything about CS 10	Excluded all non-responders (25 students), then shortened group names.
Python experience	In which of the following have you written programs? [Can select multiple options.] • C • C++ • Java • Matlab • Python • Scheme • Scratch/Snap! • None of the above	Excluded all non-responders (25 students), then mapped all students who selected "Python" to "True" and all other students to "False".

A.2 61B End-of-Semester Survey

Factor	Relevant Survey Question(s)	Preprocessing
Gender	What is your gender? • Male • Female • Other	Excluded all students who responded "Other" (55 total).
Start age	When did you start programming? • 12 years old or younger • 13-14 years old • 15-16 years old • 17-18 years old • 19-20 years old • 21-25 years old • 26+ years old • Decline to state [Help text: If you don't remember exactly, your best guess is fine.]	Excluded all students who responded "26+ years old" (9 students).
URM status	Do you consider yourself to be a member of an underrepresented ethnic or racial minority within computer science? • Yes • No • I don't know	Excluded all non-responders (33 students).
Major	What is your major? • L&S Computer Science • EECS • Applied Mathematics • Architecture • [24 more options...] • Other (custom response) [Help text: If you are undeclared, choose the major category that you will most likely pursue. If you have multiple majors, choose the first that applies.]	Mapped all students who selected LSCS to "LSCS", all students who selected EECS to "EECS", and all other students to "Other".
Transfer status	Were you enrolled in another college or university before UC Berkeley? • Yes • No [Help text: For example, are you a junior transfer student?]	

Factor	Relevant Survey Question(s)	Preprocessing
Major and transfer status	[This was a combination of the previous two factors.]	We only wanted to consider LSCS and EECS students for this factor, so all students in the "Other" group for major were also excluded.
High school CS experience	Which of the following computers science courses have you taken? [Can select multiple options.] • CS61A at Berkeley • CS61AS at Berkeley • CS10 at Berkeley • E7 at Berkeley • CS61C at Berkeley • CS C8 (Data 8) at Berkeley • At least one high school computer science course. • At least one post-high school computer science course other than those listed above. • I had not taken any official CS courses before 61B. • Other (custom response) [Help text: Please only include courses taken for academic credit.]	Mapped all students who selected "At least one high school computer science course." to "True" and all other students to "False".
Sentiment on taking CS 10	Have you taken or considered taking CS 10? • Yes, I took CS 10 • I strongly considered taking CS 10, but did not • I briefly considered taking CS 10, but did not • I never considered taking CS 10 • I don't know anything about CS 10	Shortened group names.
Java experience	How much Java experience did you have before this course? • 1: None • 2 • 3 • 4 • 5: As much (or more) than is covered in 61B.	Mapped all students who selected "1: None" to "False" and all other students to "True".

Factor	Relevant Survey Question(s)	Preprocessing
Prior grade	What grade did you earn in 61A or 61AS? • I did not take 61A or 61AS. • I decline to state. • D+ or lower. • C- • C • [6 more options...] • A+	Shortened "I decline to state." to "Decline to state" and "I did not take 61A or 61AS." to "Didn't take". No students responded "D+ or lower." (CS 61AS is a 61A alternative that was last offered in Spring 2016. We believe the number of students in our dataset who took 61AS instead of 61A is negligible.)
One-on-one tutoring attendance	How many times did you attend 1-on-1 tutoring? • 0 • 1 • 2 • 3 • 4 • 5 • 6 • 7 • 8 [Help text: Please only include tutoring offered as part of this course (on Piazza)]	
Guerrilla section attendance	How many optional guerrilla sections did you attend? • 0 • 1 • 2 • 3 • 4 • 5 • 6 • 7	