# Analysis of EECS 20, Spring 2000 Realization

Edward A. Lee, Professor
eal@eecs.berkeley.edu

## Overview

We conducted a survey of the students in EECS 20 in Spring of 2000 to help determine the factors that affect preparedness for this course. There were 227 students enrolled in the class, and 176 responded to the survey. These were those students who came to their lab session in the second to last week. These students were distributed by class standing as shown below:
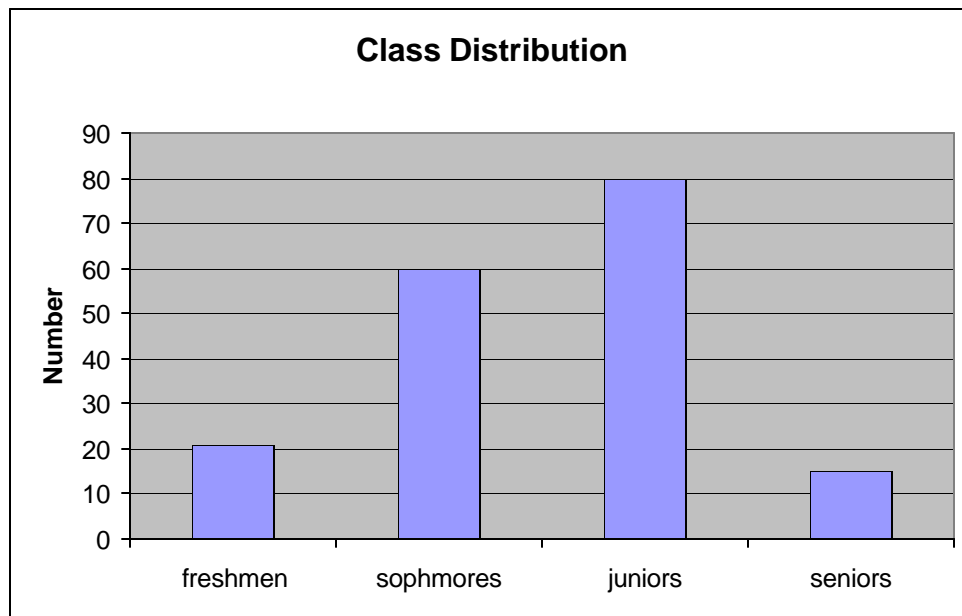


The class is dominated by juniors, probably partly because there is no comparable class offered in the junior colleges, and therefore all of the junior college transfers (who start as juniors) have to take this class.

## Analysis of the effect of class standing

The chart below shows the grade distributions by class standing, where 1 means freshman and 4 means senior.

**score vs. standing**

$y = 1.2836x + 74.719$
$R^2 = 0.0093$
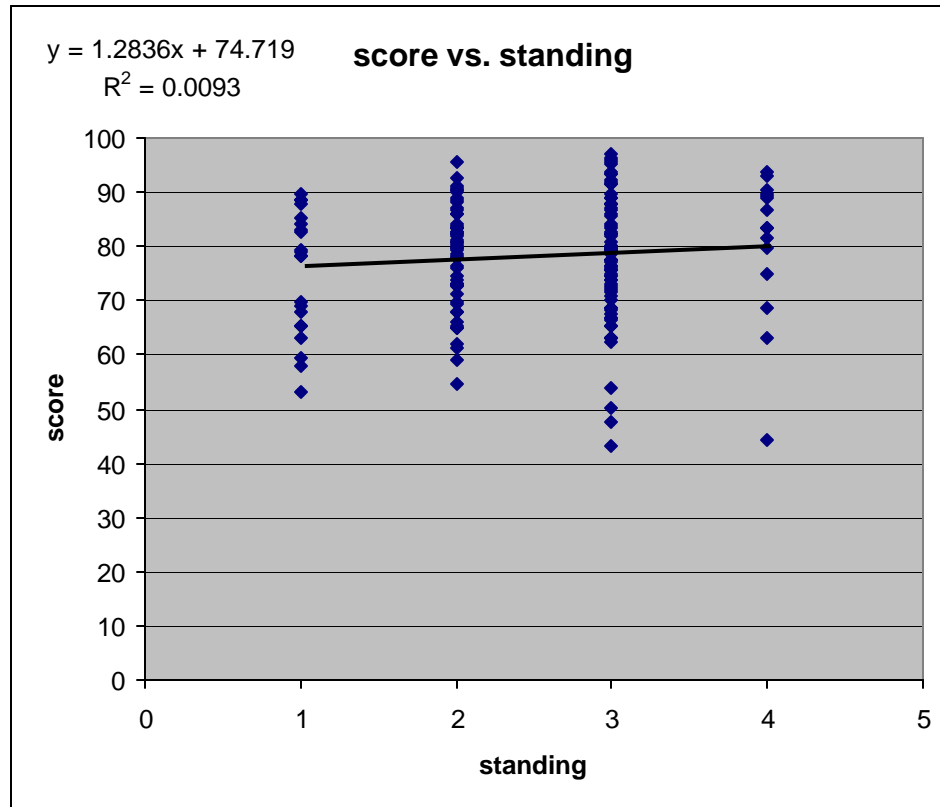
The grades here are given on a scale of 0 to 100, with key thresholds set as follows:

> 80 and above: A's (A-, A, A+)
> 63 and above: B's
> 62 and below: C's, D's, and F's

Note that these grades seem high when looking at the distribution above, but keep in mind that this represents only 176 of the 227 students in the class. As we will see below, the 51 students who did not respond to the survey did much worse in the class than the 176 who did respond.

Roughly speaking, each 6 points in the raw score translated into one grade level. E.g., a 6 point difference would change a B to a B+. An 18 point difference would change a B to an A (three grade level increments).

The line in the chart is the result of running a regression. It shows that for each year of class standing, students' grades are higher by 1.28 points (out of 100) on average, a surprisingly small number. The $R^2$ value tells us that this trend explains about 0.93 percent of the variability, essentially none of it. **Thus, class standing has little significant effect on performance.** Note that removing the outliers (senior slackers) does not change this result by much.

# Analysis of the effect of Showing Up

The data below compares the performance of students who answered the survey against those that did not.  Those that did are those that showed up for the second to last lab.

| Category | Mean | Std Dev. | P value | Difference | Number |
|---|---|---|---|---|---|
| responded | 78 | 11 | 1.27E-5 | 13 | 176 |
| did not respond | 65 | 18 | | | 51 |

The mean for those who responded was 78, vs. 65 for those who did not respond.  That is, students who showed up in the second to last lab did 13 points better (two grade levels, e.g. B to A-) than students who did not. The standard deviation is much higher for those who did not respond.  The "P value" is the result of running a t-test on the means.  Its interpretation is that the probability of getting a difference in the sample means by random chance (assuming no difference in the true means) is about $1.27 ? 10^{-5}$, or essentially nil.  I.e., the data are statistically very significant[1]. **We conclude that the respondents to the survey do not represent a random sample from the class, but rather represent the diligent subset.** These are the ones we want to be paying attention to anyway.

This class has an extensive web page, with all lecture notes available on line. One concern has been that students would not attend lecture because of the extensive facilities available on line.  However, I believe that the web page cannot substitute for attendance in lecture, so I gave pop quizzes scattered throughout the semester.  The pop quizzes counted for only 10% of the grade, but provided extra incentive for students to keep up with the class and to attend the lectures.  Analyzing the quiz results, however, yields some interesting insight.  Below is a chart showing the students' final score vs. the number of pop quizzes that they were present for:



$y = 6.7939x + 46.665$

$R^2 = 0.4006$

---
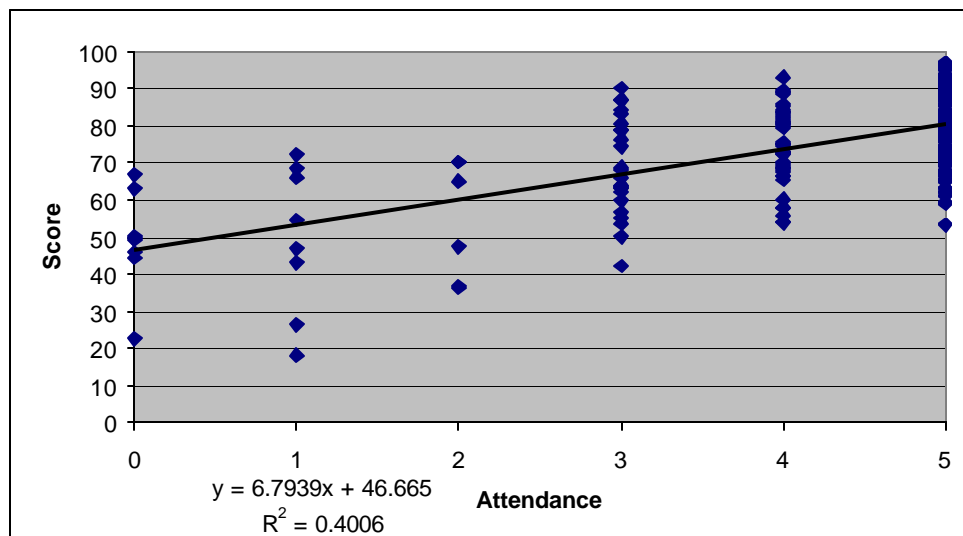
[1] Generally, statisticians consider a P-value of 5% or lower (or sometimes 1% or lower) to be "statistically significant."

There were a total of five quizzes, so students were present for anywhere from 0 to 5 of them.  I believe that this represents a reasonable measure of how much they attended class.

The conclusions are surprisingly strong.  The slope is 6.8, which says that a student who attends class 3/5 of the time vs. 2/5 of the time, on average, does one grade level better.  **A student who attends all the lectures does, on average, an astonishing 34 points better than a student who attends none of the lectures, or  almost 6 grade levels better, the difference between a C and an A!** Note that these data include the entire class, not just the diligent subset, and that attendance explains a relatively large 40% of the variability.
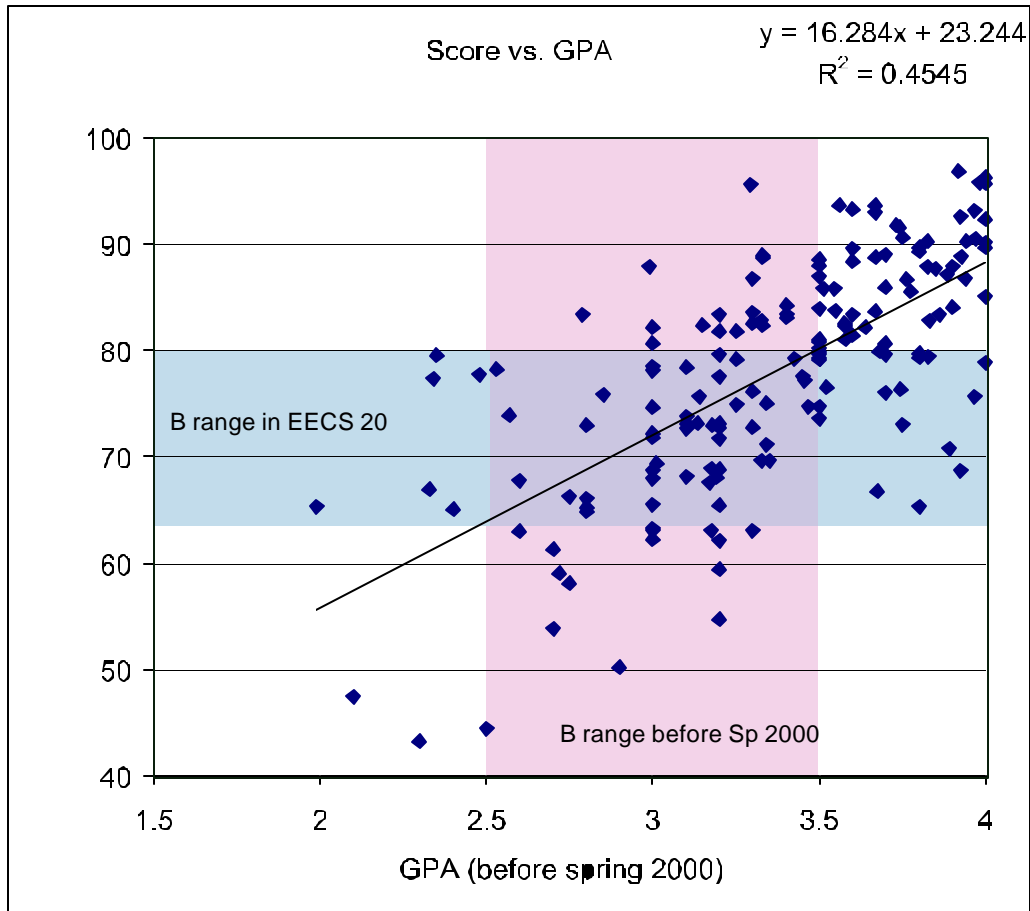
Since 10% of the grade depends on the pop quizzes, a student who took none of the quizzes had a best possible score of 90 rather than 100.  Thus, removing the effect of the quizzes, a student who attends lecture does, on average, 24 points better than a student who attends none of the lectures.  This is the difference between a C and a B+.

Of course, the students were aware that pop quizzes were being given at random times.  Thus, it may be that instead of measuring the effect of attendance we are measuring the effect of diligence.  That is, students who are diligent will attend lectures if there are pop quizzes, and students who are not diligent won't.  Of course, one also expects that students who are diligent will also do better.

## GPA Prior to taking this class

The final score of each student vs. their GPA as of the previous semester is shown on the next chart below. The shaded region horizontally indicates the scores that received a B-, B, or B+ (recall that the "diligent subset" of students responded to the survey).  The same range for the prior GPAs is shown in the vertical shaded region.

A regression trend line is shown.  The $R^2$ value indicates that about 45% of the variability can be explained by GPA.  In other words, **good students do well in this class** (surprised?). The correlation is about 0.67, which is relatively high. Moreover, the regression trend line almost perfectly forms the diagonal of the overlap of the two shaded regions.  **Thus, on average, the GPA of students was neither raised nor lowered by this class.**

**Score vs. GPA**

$y = 16.284x + 23.244$
$R^2 = 0.4545$

B range in EECS 20

B range before Sp 2000

GPA (before spring 2000)

## Analysis of the effect of Math classes

The table below compares the performance of groups of students based on which math classes they have taken.  Note that only Math 1a was formally required, so it is assumed that all students have taken that.  All of the students who responded to the survey had taken (or were concurrently taking) at least one of the math classes listed (1b, 53, 54, 55).  Some of them claimed to have not taken Math 1b, but to have not taken a more advanced math class that has Math 1b as a prerequisite (math 53 or 54).  We assumed for such students that they had taken a Math 1b equivalent somewhere else.  With this change, only 4 students had not taken Math 1b.  This is too small a number to obtain any useful data about the impact of Math 1b.  The impact of other math classes is measured by comparing them against students who have not taken them.

| Category | Mean | Std Dev. | P value | Difference | Number |
|----------|------|----------|---------|------------|--------|
| with 5x | 78 | 11 | 1.33E-02 | 8 | 166 |
| without 5x | 71 | 12 | | | 10 |
| with 53 | 78 | 11 | 6.89E-02 | 3 | 154 |
| without 53 | 75 | 11 | | | 22 |
| with 54 | 79 | 11 | 1.94E-02 | 5 | 142 |
| without 54 | 74 | 11 | | | 34 |
| with 55 | 79 | 11 | 3.26E-02 | 3 | 95 |
| without 55 | 76 | 11 | | | 81 |

The first pair compares students who have taken at least one of math 53, 54, and 55 against the few students (10 of them) who had taken none of these. A significant difference is found (8 points), and it is statistically significant (1.3 percent probability of observing a difference this big in random data). This represents slightly more than one minor grade level (e.g. B to B+).

The improvement generated by Math 53 is smaller (3 points) and (marginally) not statistically significant (almost 7% probability of observing this difference with random data).

The improvement generated by Math 54 is greater (5 points), and is statistically significant.

The improvement generated by Math 55 is smaller (3 points), but statistically significant.

**We conclude that Math 54 provides the greatest benefit, and that it provides enough benefit to affect a grade by one grade level (e.g. B+ to A-).**

The number of math classes that the students have taken might also be significant. Below is a scatter plot showing grades vs. the number of math classes taken. Taking 1 means math 1b (or in rare cases, math 55). Taking two means math 1b and one of 53, 54, and 55. Taking three means math 1b and two of 53, 54, 55.

$$y = 2.0117x + 71.5$$
$$R^2 = 0.0289$$
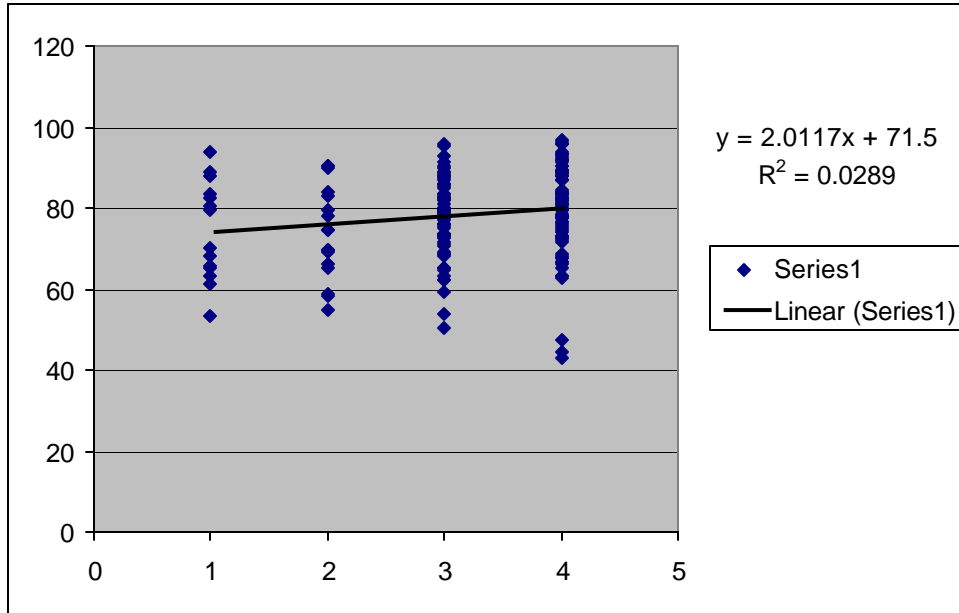
We see that each additional math class add about 2 points on average, and that only about 2.9 % of the variability is explained by this trend. **We conclude that taking more math classes helps significantly, but that many students do very well with fewer math classes.**

## Student Opinion about Math Classes

The students recommend adjusting the math prerequisites of the class as follows:



Number of students suggesting prerequisite
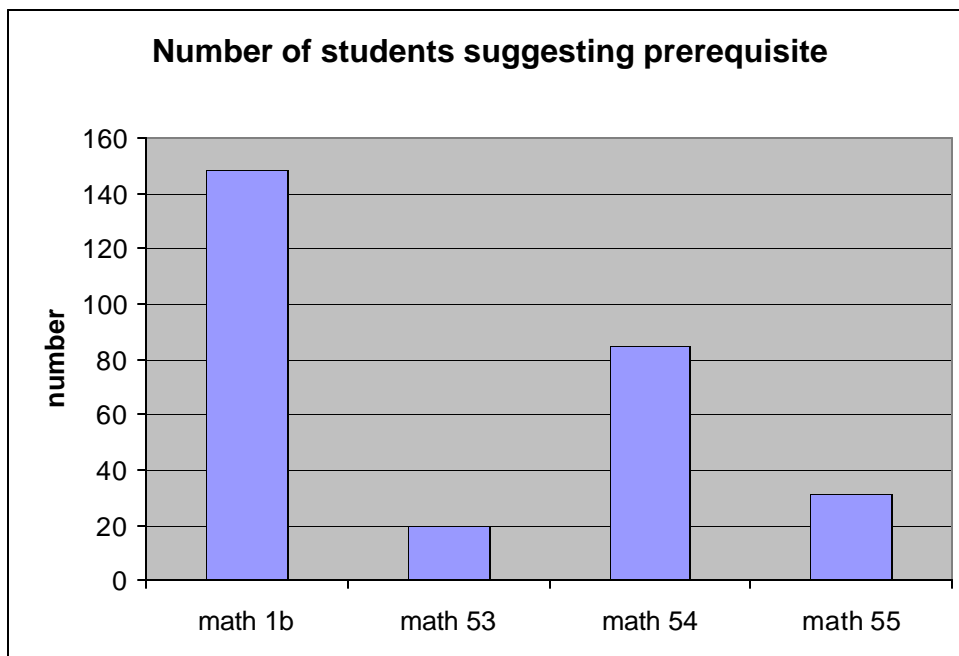
Notice that **a vast majority (148 of 176) believe that Math 1b should be a prerequisite. Almost half (85 of 176) believe that Math 54 should be a prerequisite**.

## Computing Classes Taken

The students reported having taking certain computing-related classes in numbers shown in the following chart.  The first bar, labeled "language" refers to the question: "have you taken, or are you taking a programming language class (e.g. CS 9A-Z, CS 3)."
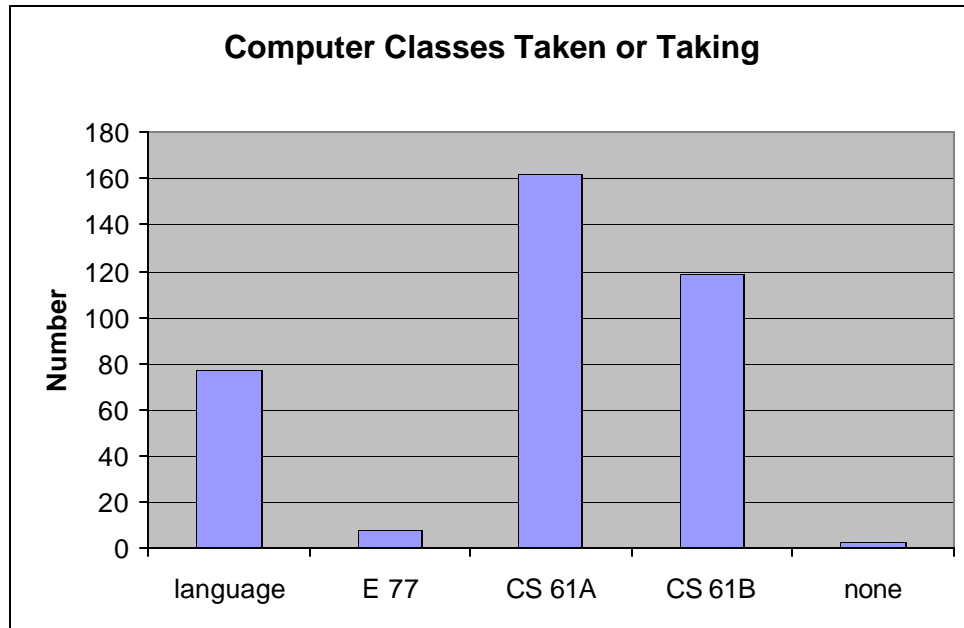


Note that very few students have taken E77, which is probably the most directly useful because of the Matlab focus, and only 4 students report having taken no computer classes. These same four students felt that there was no need for a computing prerequisite for this class (see opinion information in the section after next).
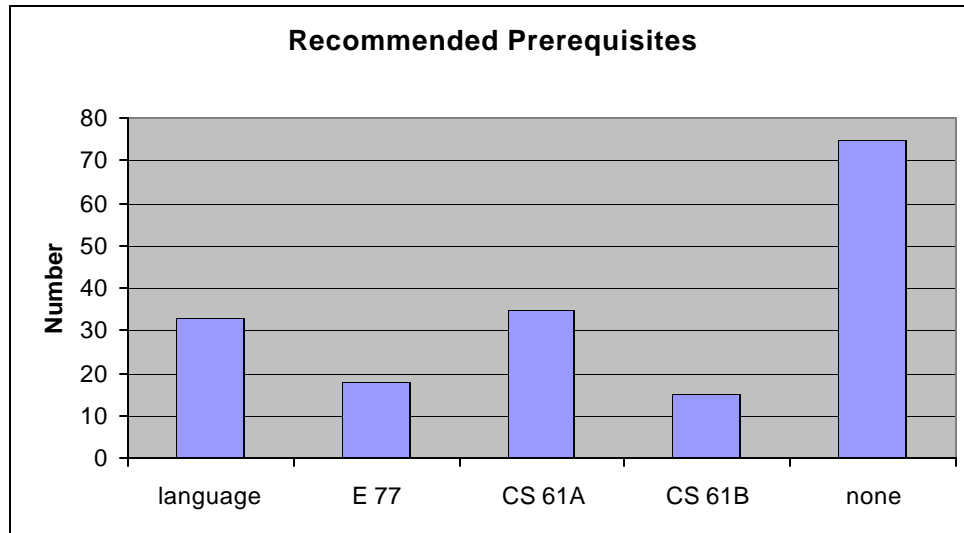
## Analysis of the Effect of CS Classes

The table below compares the performance of groups of students based on which computing classes they have taken.  Note that all but four students have taken some computing classes, so we do not attempt to determine the effect of taking some computing class.

| Category | Mean | Std Dev. | P value | Difference | Number |
|---|---|---|---|---|---|
| with E77 | 75 | 7.14 | 1.38E-1 | -4 | 8 |
| without E77 | 78 | 10.96 | | | 168 |
| with 61A | 78 | 11 | 9.06E-2 | 4 | 162 |
| without 61A | 74 | 11 | | | 14 |
| with 61B | 80 | 11 | 2.82E-1 | 1 | 115 |
| without 61B | 79 | 10 | | | 61 |

Again, the "P value" is the result of running a t-test. It gives the probability of getting at least this much difference in the mean by random chance. These P values are relatively high, so the differences are not statistically significant. **We conclude that computing classes have little effect on performance, assuming as was the case with this class that all students have had some computing classes**.

## Student Opinion about Computer Classes

The students recommend altering the prerequisites to include computer classes as follows:

**Recommended Prerequisites**

Number

| 80 | 70 | 60 | 50 | 40 | 30 | 20 | 10 | 0

language     E 77     CS 61A     CS 61B     none

## Conclusions

We summarize as follows.

**Definitions**:

- *diligent students*: those who responded to the survey.
- *grade level*: the increment from B and B+, or B+ and A-, for example.
- *class standing*: freshman, sophmore, junior, or senior.
- *taking a class*: means either concurrently or before taking EECS 20

**Conclusions**:

- Class standing has little significant effect on performance.
- On average, the GPA of students was neither raised nor lowered by this class.
- Students who attend lecture do significantly better than students who do not.
- Taking at least one of Math 53, 54, and 55 significantly helps (by slightly more than one grade level).
- Taking Math 53 or Math 55 helps by half a grade level.

- Taking Math 54 helps by almost a grade level.
- Taking more than one of Math 53, 54, or 55 helps somewhat, but many students do well without this.
- Students believe Math 1b should certainly be a prerequisite.
- Many students believe Math 54 should be a prerequisite.
- Computing classes have little effect on performance, assuming as was the case with this class, that all students have had some computing classes.
- Students do not believe a computing class prerequisite is needed.

From this, we conclude that certainly the Math prerequisite should be changed from Math 1a to Math 1b. This change is in the works. Moreover, it is worth considering making Math 54 required or recommended.

## Acknowledgements

Special thanks to Rhonda Righter for assistance with the statistical analysis.

## Appendix: Survey Contents

The text of the survey is:

What is your standing?

☐ 1. Freshman ☐ 2. Sophmore ☐ 3. Junior ☐ 4. Senior

Select the math classes that you have taken or are taking this semester.

☐ 1. Math 1b ☐ 2. Math 53 ☐ 3. Math 54 ☐ 4. Math 55

Select the math classes that you believe should be prerequisites for eecs 20 (if any).

☐ 1. Math 1b ☐ 2. Math 53 ☐ 3. Math 54 ☐ 4. Math 55

Select the computer classes that you have taken or are taking this semester.

☐ 1. Language class (e.g. CS 9A-Z, CS 3)

☐ 2. Engin 77

☐ 3. CS 61A

☐ 4. CS 61B

Select which of the following programming classes you believe should be prerequisites for eecs20 (if any).

☐ 1. Language class (e.g. CS 9A-Z, CS 3)

◯ 2. Engin 77

◯ 3. CS 61A

◯ 4. CS 61B

As of the end of 1999, what is your GPA at Berkeley? If you do not remember exactly, then please estimate it.

[                    ]