# Technical Note on UTeach Computer Science Principles and Underrepresented Students

Michael Marder, Executive Director of UTeach and Department of Physics, UT Austin

*The AP Computer Science scores of underrepresented students who took UTeach CS Principles were significantly better than national averages. This remains true after controlling for the demographic composition of the schools using the UTeach curriculum.*

In the Spring 2017 administration of the Advanced Placement Computer Science Principles exam, classrooms that made use of the UTeach CSP curriculum did well. Overall 82.8% passed versus 74% nationally. In addition, 70% of Hispanic students who took the exam passed versus 57% nationally, 55% of Black students passed versus 42% nationally, and 82% of women passed versus 70% nationally.

To know whether these differences are significant, one must check two things. First, was the sample large enough to ensure that the results are not due to chance? Second, could the results be due to UTeach having offered the course in schools with privileged student populations?

The answer is that uncertainties due to sampling are too small to explain the difference between the UTeach results and those of the national sample. This is true even after correcting for the demographic characteristics of the schools where the UTeach course was offered.

*Table 1: Results for AP CS Principles by subgroup, overall for the nation, adjusted for the demographics of schools offering the UTeach course, and compared with the UTeach results.* $p < .1$, .; $p < .05$, $*$; $p < .01$, $**$; $p < .001$, $***$

| Group | Percentage passing CSP | Percentage *expected* to pass CSP, adjusted for UTeach CSP student demographics | Percentage *actually* passing UTeach CSP (SEM) | $p$ |
|---|---|---|---|---|
| All | 74% | 78% | 83% (0.9%) | *** |
| White | 82% | 83% | 83% (1.5%) | |
| Black | 42% | 47 % | 55% (5.7%) | . |
| Hispanic | 57% | 61% | 70% (2.5%) | *** |
| Asian | 84% | 86% | 91% (1.1%) | *** |
| Male | 76% | 80% | 83% (1.1%) | ** |
| Female | 70% | 74% | 82% (1.6%) | *** |

The main results appear in Table 1. The second column shows the percentage of students passing AP CSP nationally in 2017 overall and by subgroup. Schools that offered UTeach CSP

had somewhat lower levels of economic need than U.S. high schools overall. The third column shows how the national results would be expected to change by adjusting from the poverty levels and ethnic composition of U.S. schools overall to those that offered UTeach CSP. The fourth column shows the actual results in the schools offering UTeach CSP. They are considerably higher than the expectations from adjusted national results. The fourth column also lists the uncertainty in the passing percentage of UTeach CSP students due to sampling a finite population. The fifth column indicates the statistical significance of the differences between the UTeach CSP results and the adjusted national averages.

The information used to obtain these results was more limited than is customary for such analysis. Therefore, I made assumptions that will be described below. **The assumptions were generally chosen to diminish rather than increase the likelihood that the UTeach results would appear significant.**

Sampling uncertainty was determined from a standard expression. When one makes $N$ measurements that fall in a bin with probability $p$, then the uncertainty associated with the estimate of $p$ is $\sqrt{p(1-p)/N}$. The fourth column of Table 1 shows that uncertainties for subgroups were less than 2% except for the Hispanic and Black students.

I estimated a binomial logistic regression model. This means assuming the probability of passing the CS Principles exam is of the form

$$P(S|MEG) = \frac{1}{1+\exp(-\{A_0 + A_M M + A_E E + A_G G\})} . \quad (1)$$

.

Here, $P(S|MEG)$ is the probability of passing ($S = 1$) the CS Principles exam (score of 3 or greater) given a student's race and ethnicity $M$, Free and Reduced-Price Lunch (economically disadvantaged) status $E$, and gender $G$. As shown in Table 1, the College Board has released data on passage rates as a function of minority status and gender. This means we know $P(S|M)$ and $P(S|G)$. However, the College Board did not provide information on how passing rates depend upon Free and Reduced-Price Lunch status $E$.

I have computed many different regressions of this type in cases where both race and ethnicity and economic status are known. Typically, the model coefficients for Equation (1) describing Black and Hispanic students are close to each other and close to the coefficient for economic disadvantage. Thus, I assume that

$$A_E = (A_{M=H} + A_{M=B}\ )/2. \quad (2)$$

That is, the effect of being economically disadvantaged adds to the effect of being either Black or Hispanic (these are usually treated as mutually exclusive although technically they are not), and the amount by which it does so is the average of the coefficients for Black and Hispanic students.

This is the main assumption needed to get an estimate of the joint probability distribution $P(S|MEG)$. Also assume that gender is not correlated with race/ethnicity or economic need. This means that the fraction of women taking the exam is constant no matter the demographic makeup of the school. (Additional College Board data could improve the analysis at this point.) With this assumption, one can derive the identities

$$P(S|M) = \sum_{\substack{E=0,1 \\ G=\text{Male,Female}}} P(S|MEG)P(E|M)P(G). \qquad (3a)$$

$$P(S|G) = \sum_{\substack{E=0,1 \\ M=\text{W,B,H,A}}} P(S|MEG)P(E|M)P(M). \qquad (3b)$$

The probability $P(E|M)$ is the probability that an AP student from a particular racial or ethnic group is eligible for Free and Reduced-Price Lunch. This can be estimated for any particular collection of schools from the National Center for Educational Statistics (NCES) Common Core of educational data. Doing so requires an additional assumption, which is that students taking the AP CS Principles exam in each school are demographically similar to the full school population. The population taking AP exams at any given school is likely to be economically privileged, so the probability of being economically disadvantaged is in general likely overstated. Because I do not have access to the identifiers for all the schools in which students took the AP CS Principles exam, I used data on minority status and economic need for all the public and private high schools in the NCES dataset for 2014–2015. The schools offering AP CS Principles are likely to be more privileged than U.S. schools overall. (Additional College Board data could improve the analysis at this point.)

*Table 2: Probability of passing CS Principles for students of different race/ethnicity. Source: College Board report on AP CSP for all U.S. schools*

| **Race/Ethnicity** $\boldsymbol{M}$ | **Passing Probability** $\boldsymbol{P(S=1|M)}$ |
|---|---|
| White | 0.82 |
| Black | 0.42 |
| Hispanic | 0.57 |
| Asian | 0.84 |

*Table 3: Gender breakdown of CS Principles students*

| **Gender** $\boldsymbol{G}$ | **Passing Probability** $\boldsymbol{P(S|G)}$ | **Probability of Gender** $\boldsymbol{P(G)}$ |
|---|---|---|
| Male | 76% | 70% |
| Female | 70% | 30% |

Using these probabilities in Equation (3), one has six equations in six unknowns; without loss of generality, the coefficients for White students $A_W$, male students $A_{\text{Male}}$ , and students not economically disadvantaged can be set to zero. The equations are nonlinear. I found the unknowns using the lmfit package in Python.

*Table 4: Coefficients for conditional probability model in Equation (1). One coefficient can be set to zero as a reference value, and this choice is made for $A_{M=W}$, the coefficient for White students. The coefficient for economic disadvantage is determined from Equation (2).*

| Model coefficient | Value |
|---|---|
| $\boldsymbol{A_0}$ | 2.19 |
| $\boldsymbol{A_{M=W}}$ | 0.00 |
| $\boldsymbol{A_{M=B}}$ | −1.59 |
| $\boldsymbol{A_{M=H}}$ | −0.95 |
| $\boldsymbol{A_{M=A}}$ | 0.21 |
| $\boldsymbol{A_E}$ | −1.27 |
| $\boldsymbol{A_{G=\text{Male}}}$ | 0.00 |
| $\boldsymbol{A_{G=\text{Female}}}$ | −0.35 |

With these coefficients in Equation (1), we have an explicit expression for the probability of passing the AP CS exam for students with different characteristics. We can take up the question of how results for UTeach are affected by the demographics of the schools that selected the UTeach curriculum. We know the number of separate class sections of UTeach CSP in each school, and we assume all the classes were of the same size for the purpose of computing the weighted average of economically disadvantaged students.

*Table 5: Probability that a student is eligible for Free and Reduced-Price Lunch (economically disadvantaged) for different ethnic/racial groups, comparing UTeach CS Principles students with those in all U.S. high schools. The estimates come from NCES Common Core of Data for which the racial/ethnic proportions and proportions of economic need are available. The computations assume that in each school M and E are uncorrelated. The UTeach computations are weighted by the number of class sections taught in each school, assuming all class sizes are the same.*

| Race/Ethnicity $\boldsymbol{M}$ | Economic Disadvantage Probability $\boldsymbol{P(E=1\|M)}$ All U.S. High Schools | Economic Disadvantage Probability $\boldsymbol{P(E=1\|M)}$ UTeach CS Principles High Schools |
|---|---|---|
| White | 37% | 29% |
| Black | 65% | 51% |
| Hispanic | 65% | 52% |
| Asian | 38% | 29% |

*Table 6: Probability of belonging to racial/ethnic groups M. Some groups, including Two or More Races, were excluded from analysis. Source: College Board.*

| Race/Ethnicity $M$ | Probability $P(M)$ All U.S. CS Principles high schools | Probability $P(M)$ UTeach CS Principles high schools |
|---|---|---|
| White | 49% | 36% |
| Black | 7% | 4% |
| Hispanic | 20% | 20% |
| Asian | 24% | 40% |

We see from Tables 5 and 6 that the UTeach CSP schools had a different blend of racial and ethnic groups than CS Principles schools overall, and a smaller percentage of students eligible for Free and Reduced-Price Lunch than U.S. schools overall. (These estimates could be improved if we had College Board data by school on the racial/ethnic composition of participating students.) Returning to Equation (3) with these adjusted distributions, one can calculate the expected results for UTeach CSP using the probabilities in Tables 5 and 6. The result is the adjusted expectations in column 3 of Table 1.

Students in UTeach CSP schools are economically disadvantaged by roughly 10 fewer percentage points than students of the same racial/ethnic group in U.S. schools. This results in a 1% to 4% increase in those expected to pass the AP CS Principles exam. However, the actual performance of students in the UTeach CS Principles course was significantly better than this.

Thus, in addition to effects from economic status, race, and ethnicity of the students, the data suggest that the favorable exam results for students taking UTeach CS Principles were due to the quality of teacher professional development, support, and curriculum.

**Acknowledgment**

These computations were done in response to questions from Hadi Partovi of Code.org about the significance of UTeach CSP results in light of the difference between the demographics of UTeach CSP students, those of the national CSP sample, and those of Code.org.