Student-Teacher Demographics in Code.org Computer Science Principles Classrooms

Mary Barnes and Ben Brooks
Code.org
April 16, 2018

Abstract

Prior research has shown that stereotypes of computer scientists as white and Asian men has discouraged many female and students from underrepresented minority (URM) backgrounds from enrolling in high school computer science electives. This study presents evidence that female and URM teachers of Code.org's CS Principles course have higher proportions of female and URM students in their classes than male and white teachers, respectively. The findings suggest that recruiting female and URM teachers to teach computer science alone could be a key strategy for improving the gender and racial makeup of computer science classrooms. This study correlated demographic data from teacher participants in Code.org's Professional Learning Program for AP Computer Science Principles with the gender and race composition of the students in their classrooms. Using a chi squared test for independence, we found that female teachers tend to have higher proportions of female students than male teachers. A multivariate regression indicated that URM teachers in our sample tend to have higher proportions of URM students in their classes than white teachers, regardless of the proportion of URM students attending the school. However, the URM regression analysis was underpowered and the finding was not statistically significant. We suggest future research with a larger sample size.

Background

One of Code.org's core goals is to ensure that students of all backgrounds have the opportunity to learn computer science. We know that the stereotype that computer scientists are white and Asian men has discouraged many female and URM[1] students from enrolling in high school computer science electives.[2] In general, we believe that any teacher can help students break

---

[1] For an explanation of which races are considered underrepresented minorities (URM) in technology fields, see the section "Racial Categories" on pages 3 and 4.

[2] Master, A., Cheryan, S., & Meltzoff, A. N. (2016). Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. Journal of Educational Psychology, 108(3), 424.; Pajares, F. (1996). Self-efficacy beliefs in academic settings. Review of Educational Research, 66(4), 543-578.; Unfried, A., Faber, M., & Wiebe, E. N. (2014, April). Gender and Student Attitudes toward Science, Technology, Engineering, and Mathematics. Presented at the AERA Annual Meeting, Philadelphia, PA.

stereotypes about who can succeed as a computer scientist through what and how they teach. Our professional learning workshops focus on issues of equity and emphasize equitable teaching practices along with strategies to recruit and support female and URM students.[3] [4] That said, we are always seeking to understand other factors leading to higher female and URM student participation in computer science classes. Compelling research shows that female and URM students are more likely to achieve when they have same-gender and same-race teachers.[5][6] This led us to wonder whether the presence of a female or URM computer science teacher might correlate with more female and URM students enrolling in our CS Principles course, a high school elective that introduces students to the foundational concepts of computer science designed to meet the requirements of *AP Computer Science Principles*.

Purpose

The purpose of this study is to determine if there are correlations between the race and gender of teachers in our CS Principles Professional Learning Program (PLP) and the race and gender of their students. In order to determine if such correlations exist, we compared student demographic data across a variety of teacher groups. We had two hypotheses: 1) female teachers would have a higher proportion of female students in their classes than male teachers and, similarly, 2) teachers who identify their race as a URM race would have a higher proportion of URM students in their classes than white teachers, regardless of the URM makeup of their schools. Finally, we wanted to determine how the intersection of a teacher's race and gender correlated with the composition of their classrooms.

Data

Teacher Data – Code.org PLP Teachers

This study is based on demographic data collected from 392 teachers who met the following criteria:

---

[3] Any teacher is welcome to use Code.org's curriculum, but we try to prioritize space in our PLP for teachers who will make the most impact in recruiting female and URM students to their classrooms. One of the primary goals of our professional development is to increase teachers' confidence and willingness to teach computer science.
[4] We must point out that offering computer science coursework alone is not enough. Teachers also need quality curriculum and professional support in order to help students achieve equity in computer science. Goode, J., Margolis, J., & Chapman, G. (2014). Curriculum is not enough: The educational theory and research foundation of the exploring computer science professional development model. Proceedings of the 45th ACM Technical Symposium on Computer Science Education, 493-498.
[5] Dee, T. (2004). Teachers, Race, and Student Achievement in a Randomized Experiment. Review of Economics and Statistics, 86(1), 195-210.
[6] Bottia, Stearns, Mickelson, Moller, & Valentino. (2015). Growing the roots of STEM majors: Female math and science high school faculty and the participation of students in STEM. Economics of Education Review, 45, 14-27.

- joined the Code.org CS Principles PLP in summer 2017[7] (549 teachers),
- taught at least five students with the CS Principles curriculum through our platform in the 2017-2018 school year[8] (418 teachers),
- and voluntarily self-reported information about their race and gender (392 teachers).

We refer to this group as "Code.org PLP teachers" throughout this paper.

Student Data - Students in Classrooms with a Code.org PLP Teacher

Code.org collects self-reported gender data from all students and self-reported race data from students in the United States who are at least 13 years old[9]. Of the 12,045 students who started the CS Principles curriculum with one of the 392 teachers in this analysis, 9,799 provided their race and 6,553 provided their gender[10]. These numbers are shown in the table below disaggregated by teacher race and gender[11].

Table 1: 2017 Code.org CS Principles Professional Learning Program Teachers

		# Teachers	Students per bin	Students who provided gender	Students who provided race	Median Students per Teacher
All Teachers Total		**392**	**12,045**	**9,799**	**6,553**	**23**
Asian Teachers	Combined	**12**	**373**	**325**	**210**	**30**
	Female Teachers	4	108	100	65	25
	Male Teachers	8	265	225	145	32
URM Teachers	Combined	**46**	**1,629**	**1,220**	**760**	**25**
	Female Teachers	32	1,155	771	479	25
	Male Teachers	14	474	449	281	29
White Teachers	Combined	**334**	**10,054**	**8,265**	**5,590**	**23**
	Female Teachers	166	5,037	4,110	2,859	23
	Male Teachers	168	5,088	4,200	2,772	23

---

[7] Currently, Code.org doesn't ask teachers on our platform to provide demographic data and 2017 was the first year we collected demographic information from teachers attending our professional learning workshops.

[8] Since this analysis was conducted in early 2018, we counted students who started the CS Principles curriculum. We did not filter out students who did not complete the curriculum.

[9] Teachers can submit gender information on behalf of students but cannot submit student race information.

[10] The discrepancy in the number of students who provide their race vs provide their gender is largely accounted for by the fact that teachers have the option to provide the student's gender, but not their race.

[11] Please note that the Code.org platform allows teachers to co-teach, thereby 'sharing' students. Only 734 students belong to multiple teachers, but, for this reason, the totals are not equal to the summation of the numbers in the more granular cells.

Racial Categories

At Code.org, students and teachers are considered underrepresented minorities in computer science if they listed their race as including one or more of the following: "American Indian/Alaska Native", "Black or African American", "Hispanic or Latino", or "Native Hawaiian or other Pacific Islander." Students are considered non-URM if they listed their race as “White,” “Asian,” or both “White” and “Asian.”.

In the following analysis we categorize students as URM or non-URM, whereas we categorize teachers as URM, white, or Asian. We made the decision to categorize students and teachers differently for two reasons. First, Asian Americans are underrepresented in the teaching profession and we wanted to see if any patterns emerged[12]. Second, by separating Asian teachers, we were able to discern discrepancies between white and URM teachers more directly.

School Data

In analyzing whether a teacher’s race correlates with the proportion of URM students in their classroom, we had to control for the URM proportion of the school. School-level data was obtained from the National Center for Education Statistics (NCES)[13], which provides the most comprehensive and trusted dataset available on student demographic data.

Code.org’s race categories do not match perfectly with the NCES race categories. In the NCES dataset, if a student identifies as more than one race, they select the category “two or more races” rather than choosing the races individually. For these students, it is not possible to identify whether one or more of their races would qualify them as a URM student in the Code.org framework. Only 3% of students fall into this category in the NCES data. We decided to count them as part of the URM population of their school, understanding that a small number of them may be Asian and white. If anything, this inflates the number of URM students in the NCES data by a small amount.[14]

The parts of our analysis that include school data are based on a subset of 291 teachers (33 URM teachers and 258 white teachers) who met the following criteria:

- taught at a public or charter school included in the NCES school statistics data,
- taught at least five students who reported their race,
- were white or URM (we filtered out the Asian teachers to highlight the difference between URM and white teachers)

[12] There are only 12 Asian teachers included in this analysis, as shown in Table 1.
[13] U.S. Department of Education, National Center for Education Statistics, Common Core of Data (CCD), “Public Elementary/Secondary School Universe Survey,” 2015–16.
[14] For comparison, 15% of the Code.org students in this survey report that they are of two or more races, with only 20% of those selecting a combination of “Asian”, “White”, or “Other”.

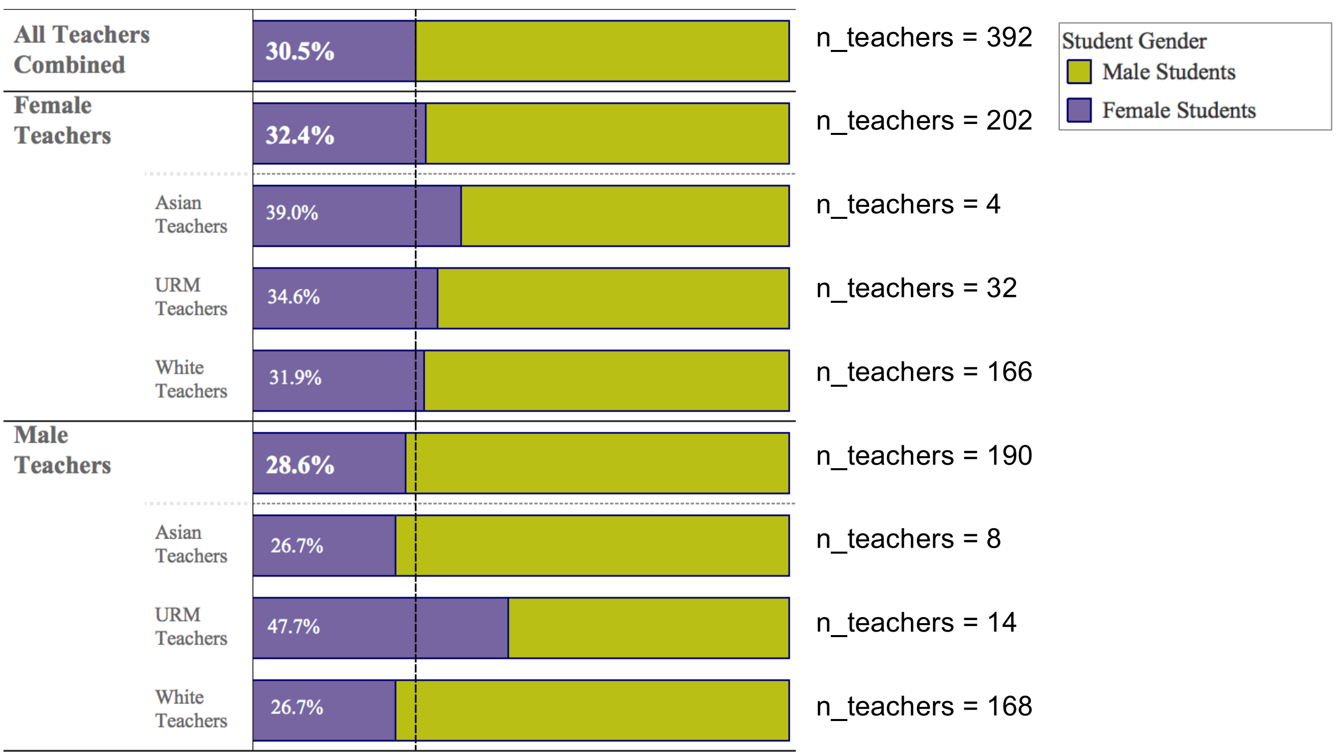
Finally, we also filtered out 6 teachers who had students with self-reported race data that was "impossible." These teachers had more URM or non-URM students in their class than in the entire school according to the NCES data[15].

Findings

Student-Teacher Gender Correlation

Chart 1 shows that, in our group of CS Principles teachers, female teachers had a higher proportion of female students than male teachers did. Of the students reporting their gender, 32.4% were female in female-led classrooms while only 28.6% were female in male-led classrooms. By constructing a confidence interval for the difference of two population proportions we find that we can be 95% confident that the true proportion of female to male students was between 2.0 and 5.7 percentage points higher in female-led Code.org PLP classrooms than in male-led Code.org PLP classrooms.

Chart 1: Percentage Female Students by Code.org PLP Teacher Gender and Race



[15] We suspect these are teachers who switched schools after signing up for our professional learning program. National statistics show that around 8% of teachers switch schools each year. We suspect that this accounts for the schools where we see more URM students in Code.org classrooms than there are URM students in the school (and vice versa for non-URM students). Of course, our filter does not catch all cases; a small percentage of the remaining teachers will have switched to lower URM schools while others will have switched to higher URM schools.

Interestingly, when we disaggregate male teachers by race, we find that URM male teachers had a higher proportion of female students in their classrooms than female teachers. White and Asian male teachers had fewer female students in their classrooms than female teachers.

Chart 2: Percentage Point Difference in Female Enrollment by Code.org PLP Teacher Gender and Race

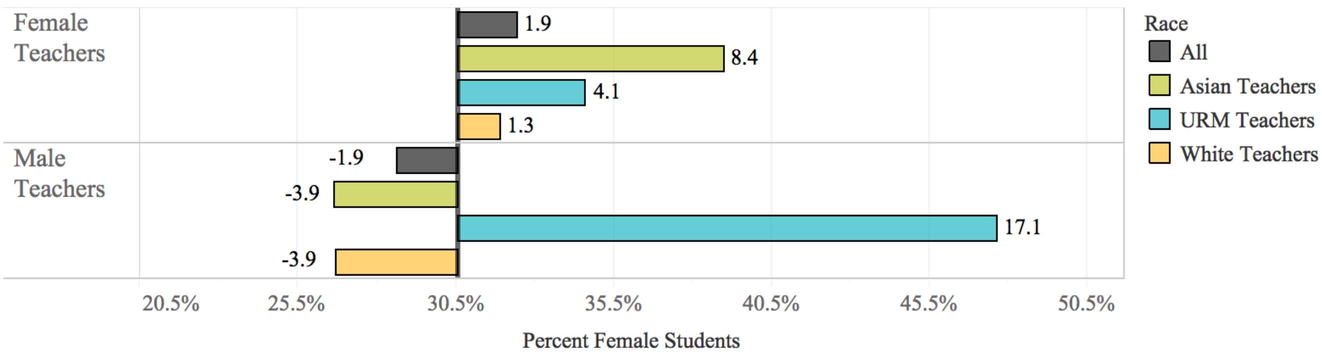


Chart 2 shows the difference between female enrollment and the overall female enrollment of 30.5% split by teacher category. For example, female enrollment in classes taught by male URM teachers was 17.1 percentage points higher than the overall female enrollment (i.e. 47.6% female versus the combined proportion of 30.5% female).

Table 2: Proportions of female students (chi-squared expected vs actual values)

	Actual Female Students	Actual Male Students	Expected Female Students	Expected Male Students	Difference between Actual and Expected	Number of Teachers	Avg. Difference Actual and Expected
Asian Female Teachers	39	61	30.6	69.4	8.4	4	2.11
URM Female Teachers	267	504	235.6	535.4	31.4	32	0.98
White Female Teachers	1311	2799	1256.1	2853.9	54.9	166	0.33
Asian Male Teachers	60	165	68.8	156.2	-8.8	8	-1.10
URM Male Teachers	214	235	137.2	311.8	76.8	14	5.48
White Male Teachers	1121	3079	1283.7	2916.3	-162.7	168	-0.97`

The p-value for this chi-squared test is < .001

Table 2 provides the results of a chi-square test of independence of variables. The “expected female students” column contains the number of female students that each group of teachers would contain if there were actually no difference in the proportion of female students between the groups. The final column presents the average difference between the actual and expected number of female students. For example, on average, white male teachers have about one fewer female student than would be expected if all groups had the same proportion of female students. Only white and Asian male teachers have fewer female students than expected.

Student-Teacher Race Correlation

In our group of CS Principles teachers, even when controlling for their school's URM population, URM teachers had more URM students than white teachers.

URM teachers are more likely to teach at a school with a high proportion of URM students than are white teachers. For this reason, we controlled for the URM proportion at the school when we evaluated whether there is a correlation between teacher race and student race in our sample of professional learning program teachers.

Chart 3: School URM vs. Code.org CS Principles PLP Teachers' Classroom URM

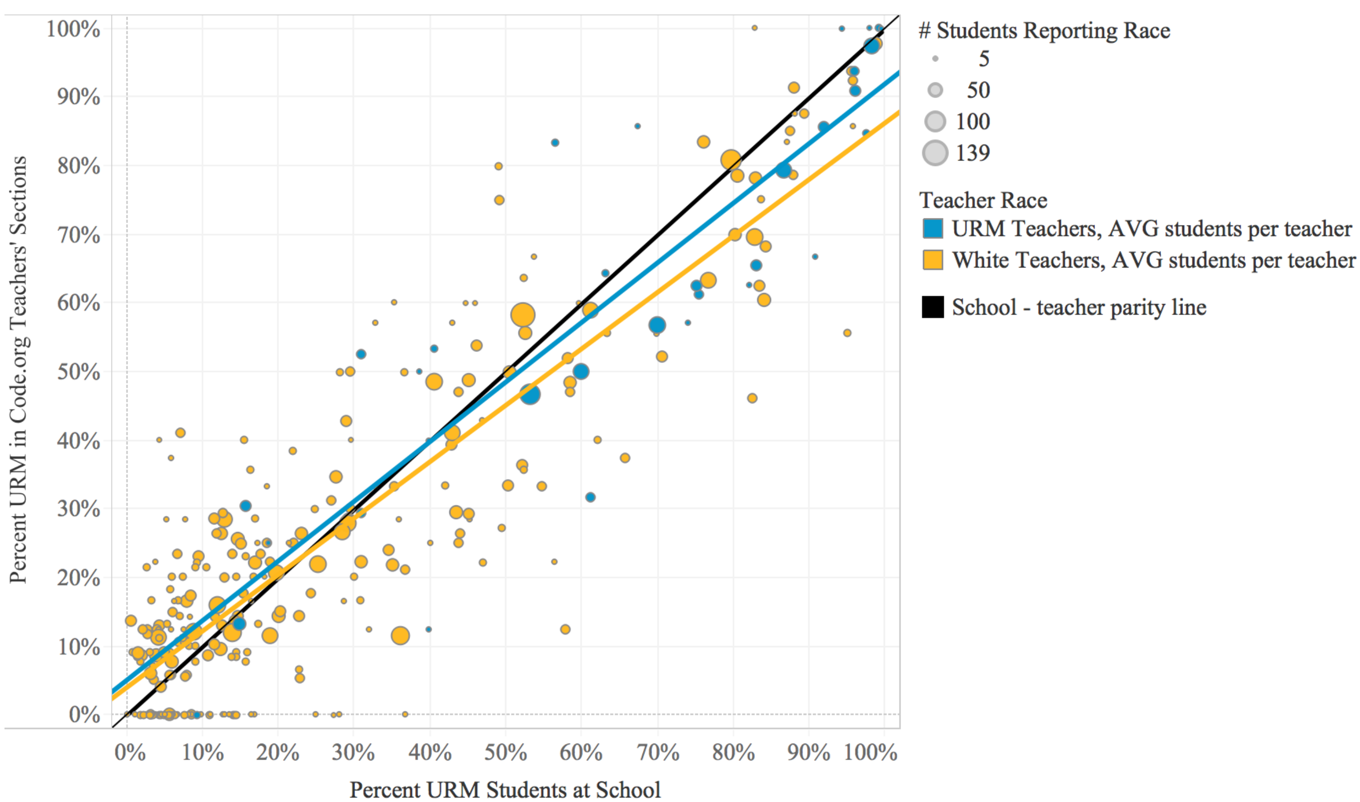


Chart 3 depicts URM teachers clustering in high URM schools and white teachers clustering in low URM schools, confirming that URM teachers are indeed more likely to teach in high URM schools. It also confirms that the proportion of URM students at the school explains most of the variation in percentages of URM students in Code.org PLP CS Principles classrooms. The combined regression has an R^2 value of 0.799, indicating that 79.9% of the variation in percentage of URM students in PLP Code.org classrooms is explained by the percentage of URM students at the school.

The regression lines show that the URM teachers in our sample tended to have a higher percentage of URM students in their classrooms than white teachers did, regardless of the percentage of URM students at the school. The difference increases as the percentage of URM students in the school increases -- in the lowest URM schools, URM teachers are expected to have classrooms with about the same percentage of URM students as white teachers but are expected to have 4.5 percentage points more URM students in the highest URM schools.

Table 3: School URM vs Classroom URM in Code.org Professional Learning Program Teachers' Classrooms (regression values)

Teacher Race	Term	Value	StdErr	t-value	p-value
White Teachers	School URM Percent	0.823	0.029	28.11	< 0.0001
	Intercept	0.039	0.011	3.68	0.00028
URM Teachers	School URM Percent	0.869	0.082	10.62	< 0.0001
	Intercept	0.050	0.058	0.86	0.397

Table 3 provides the parameters used to generate the regression lines shown in Chart 3 and their related statistics.

At schools with lower URM proportions, the proportions of URM students in the Code.org teachers' classrooms tend to be higher than that of the school. However, in schools with higher URM proportions, the proportions of URM students in Code.org teachers' classrooms tend to be below that of the school's. It is also worth noting that according the NCES data, 50% of all schools in the US have less than 25% URM.

Table 4: Code.org PLP Teachers' Classroom URM proportion (multivariate regression)

Term	Value	StdErr	t-value	p-value	95% C.I.
(Intercept)	0.038	0.010	3.609	0.0003	[0.017, 0.058]
School URM Percent	0.828	0.028	30.023	<0.00002	[0.774, 0.882]
Teacher is URM	0.039	0.025	1.550	0.121	[-0.010, 0.088]

In considering whether these differences can be generalized to all teachers in the CS Principles Professional Learning Program, we fit a multivariate linear regression model to predict the URM percentage of teachers' classrooms with two explanatory variables: a binary 'teacher is URM' variable as well as the URM percentage of the school. The results are presented in Table 4. The coefficient on the 'teacher is URM' term was 0.039, indicating that URM teachers are expected to have 3.9 percentage points more URM students in their classrooms than white teachers. However, the p-value on this coefficient is 0.121, meaning that there is a 12.1% chance that we would see these results if there were actually no difference between URM and white teachers. While not statistically significant at the 0.05 level, this finding is encouraging and suggests that further research with a larger sample size may uncover a significant relationship between teacher URM status and the URM proportion of their classrooms[16].

Limitations

One limitation on this study is that the race and gender information we collect is self-reported and voluntary. Our analyses assume that students of all races and genders report their demographic information at the same rate.

We only have demographic data from teachers who joined the Professional Learning Program in 2017, meaning that we are unable to assess the impact of teaching CS Principles over multiple years. We hope that teachers who succeed in making their computer science classrooms welcoming and supportive of female and URM students will see the number of those students who enroll in their course increase year over year. Furthermore, strategies for recruiting diverse sets of students to classes are discussed during Code.org professional learning workshops and teachers will have a chance to actively recruit underrepresented students before their second year of teaching.

Finally, the findings in this paper cannot be applied to high school computer science teachers more generally. The teachers in this study are sampled from the population of CS Principles teachers who have joined the Code.org Professional Learning Program and are therefore not representative of the larger population of computer science teachers. Teachers either apply for Code.org professional learning on their own, or as part of a regional cohort, but are given preferential admission if they teach at a high needs or high URM school. Due to this selection

[16] For example, in order to achieve 80% power for the 'teacher is URM' coefficient at a significance level of 0.05 if we would need to increase the number of observed teachers from 329 to 1567. This number was calculated with MBESS package in R, where Rho2.Y_X is the R^2 of the model including the predictor in question and Rho2.Y_X.without.j is the R^2 of the model when excluding the predictor in question: ss.power.reg.coef(Rho2.Y_X = 0.8007, Rho2.Y_X.without.j = 0.7997, p = 2, desired.power = 0.80, alpha.level = 0.05)

process, the teachers for whom we have data might prioritize classroom diversity more than the general population of teachers.

Conclusion and implications

Code.org strives to create equity in computer science education by reaching groups of students who are underrepresented in computer science, including girls and students from underrepresented minority backgrounds. In analyzing demographic information from our PLP CS Principles teachers and their students, we found that teachers' gender and race correlate with the gender and race composition of the students in their computer science classes. For gender, this finding was statistically significant at the .05 level but for race it was not. Regardless, these correlations imply that one way to increase the number of female and URM students studying computer science might be to recruit and train more female and URM teachers to teach computer science.

These findings suggest many avenues for future research. For example, this analysis focuses on classroom enrollment rather than student attitudinal data, achievement, or progress. Future work might consider whether teachers' gender and race correlate with these additional metrics.

Want to do research with Code.org data? We'd love to partner with you. If you would like to partner with us on research or access de-identified datasets for your own research, please email research@code.org.

Many thanks to Baker Franke, Matt Drury, Poorva Singal, Alice Steinglass, and Marina Taylor for invaluable support in structuring this paper.