The Effect of Computer Science Classes on Education Outcomes, Education Returns, and Job Market Returns

Daniel Cheung

I. Introduction

Chicago Public Schools, CPS, had added a high school graduation requirement for the class of 2020 and beyond that required a computer science course according to Amina Elahi's "CPS to roll out computer science requirement" (2016). Elahi's article also stated that 23.295% of high schools were able to provide computer science courses at an acceptable level with every high school being able to comply in the 2018-2019 school year. So, all CPS students in the class of 2020 had the opportunity to take computer science in their junior and senior years of high school. All cohorts before the class of 2020 were not required to take a computer science course to graduate high school but will have an increased opportunity to do so in schools that increased their computer equipment and possibly staff to offer the course. Elahi mentioned that CPS is the third-largest school district in the United States and CPS proclaimed that the computer science graduation requirement was the first of its kind in the United States for a school district (2016). Research on education outcomes, education returns, and job market returns related to the graduation requirement is necessary to see if this additional requirement will benefit students in the long run and whether other school districts should follow. The additional teachers and computer equipment would make it difficult for some school districts to follow suit, which is why it is important to show whether a computer science graduation requirement will have benefits that outweigh the costs.

Computer skills are essential in college and the labor market. As more everyday tasks become digital, computer skills are seen as more of a necessity. Therefore, students need to have basic computer knowledge. CPS' computer science classes will expose students to online and computer resources to assist them in learning computer science concepts. According to CPS' blog post "CPS Computer Science Graduation Requirement Fact Sheet", Code.org is a website that CPS encourages computer science teachers to use in class. Many other resources vary depending on the teacher of the course and the school.

The effect of CPS' high school graduation requirement on education outcomes, education returns, and job market returns can be seen with more data on individual students' future earnings through a social security matching approach, standardized tests, GPA, class rank, class schedule, computer science teacher, and college admission. More data is needed to run the regressions to see standardized test scores, GPA, college acceptance rate, and earnings as outcomes.

II. Literature Review

Mark Hoekstra's *The Effect of Attending the Flagship State University on Earnings: A Discontinuuity-Based Approach* (2009) used attending a flagship public state college as a treatment on earnings to see the impact it would have (717). This paper used a sample of 28 to 33-year-olds so the subjects had enough time to enter and sort themselves into the labor market. Hoekstra also stated that the earnings of a 28 to 33-year-old are a good predictor for career earnings. For the computer science graduation requirement model, a similar approach is needed to have somewhat accurate and relevant results for earnings. This also means that a lot of time needs to pass since the class of 2020 is around 20 years old in 2022 (718).

Lisa Barrow's, Silvana Freire's, and Marisa de la Torre's *Trends in Computer Science Education* (2020) found that there was selection before the computer science graduation requirement based on race and gender. There were overall increases in the takeup of computer science between all races and genders. The class offering of computer science was the main contributor to disparities in computer science takeup. Lastly, there was confirmation that CPS had leaped the number of high schools that offered computer science with over 90% of high schools offering computer science in CPS.

III. Contribution

This proposal can contribute to education literature by showing the effect of technology and computer science courses on education outcomes, education returns, and job market returns. This will be done through different stages of education and the job market. First, standardized test scores will be the outcome with GPA as the second outcome, college acceptance as the third, and earnings as the fourth. Each stage will be connected to the prior in the form of a control variable, which will to the presence of the treatment in each regression. This also acts as a timeline of events to show the overall effect of CPS' computer science graduation requirement.

IV. Theory

An important assumption about the computer science graduation requirement would be that almost every high school student in the graduating class of 2020 and beyond will have had a computer science class. In these classes, students will be exposed to and learn various skills and concepts like binary, computer parts, coding, picture manipulation, and search engine utilization using services like Code.org, GIMP, Google, Notepad++, Scratch, and W3Schools. The coding includes JavaScript and HTML which is a markup language. Students would also be exposed to resources that can assist them in academics and standardized test scores through information in-class activities and lessons like learning about paid online educational resources. Students learn about resources like CPS provided resources, Kahn Academy, and College Board practice tests. The additional search engine skills will allow students to better find the resources they are looking for. Navigating CPS' website will allow students to see other in-person programs and resources that would have not been utilized otherwise like tutoring and standardized test preparation. Additionally, students will be able to use these resources and search engine skills to apply to colleges and financial aid. Students would be more likely to use and apply for FAFSA and college applications if it is easier for them to access Common Application and FAFSA on their own. This could be potentially seen with higher college acceptance rates, enrollment rates, and an easier transition to college as many resources are now online.

There is no assumption that all the students who take or meet the computer science graduation requirement will have been exposed to a set number of concepts or skills. The variation in computer knowledge and skills learned depends on the student, the teacher, and the resources available to the student. Students who have greater access to technology at home and school are more likely to practice and apply the computer knowledge and skills they have learned in class. There is additional variation among students who have had prior computer knowledge and technology access in elementary school.

There is an assumption that the conversion between SAT and ACT is correct and on average close to the true conversion. This is crucial as before 2017, ACT was the primary standardized test score for CPS students which would restrict the data if ACT scores could not be compared with SAT scores between cohorts (CPS).

It is assumed that recent cohorts before the class of 2020 are not underlyingly different from the class of 2020 and beyond with no reason to believe that the types of students after the class of 2019 are any different.

Lastly, a potential assumption that an additional graduation requirement would decrease graduation rates based on another criterion to meet does not qualify in this case as students can swap an elective with the computer science course. Also, the potential skills learned in a computer science course can translate to better grade outcomes in other high school classes which will lead to a greater probability of graduating. An example of this would be utilizing search engines where a student would browse the internet more effectively which can assist them in understanding class material faster with external explanations and examples found online.

V. Identification Strategy

The implementation of CPS's computer science graduation requirement will be the treatment. This is an exogenous treatment on standardized test scores because this policy was not implemented based on prior standardized test scores results or students' GPA. The counterfactual would be suburban Cook County school districts that are pretrending similarly to CPS. City, state, or country averages can be used to compare estimates to see if there are drastic differences that could be alarming.

VI. Data

The data used in this study will be from CPS' Assessment Report and Chicago's Open Data Portal. Both of these data sources are publicly available and have over a decade's worth of test scores. The data is for school averages that have SAT, PSAT, and ACT scores. There are no data on the dates of the exact year of computer science compliance. This data could be found through school websites, with links or posts that explicitly state, show, or indicate that computer science is offered as a course for students to take. Although, the problem is that students might offer these courses without having an "acceptable" curriculum or materials that could make the offering invalid by CPS before the expected full high school compliance of 2018-2019.

Currently, the standardized test score data is available through CPS' Assessment Report page on their website and Chicago Open Data Portal includes school graduation rates and college enrollment data (CPS, 2021 and Chicago Open Data Portal). The data still needed are class schedules to determine what semester students take computer science, the CS teacher who taught the course, the GPA for each semester, individual college acceptance rates, individual standardized test scores, class ranks, students' number of extracurricular activities, work experience, education attained, earnings, industry worked in, and city of residence. Some of this data can be found publically on individual CPS schools' websites. Some of the individual data needs to be done through a Conduct Primary Research request on CPS' District Data page on their website (CPS).

The amount of counterfactual data needed depends on how much data can be collected for CPS. CPS' Conduct Primary Research request page on their website seems to be against individual-level data that can be used to determine earnings as this is a big privacy risk (CPS). So, it will take a formal request with all the information that is needed and a list of the steps that will be conducted to see if it is possible to find earnings data. If all the necessary CPS data can be collected, then the same counterfactual data needs to be found in a school district that is near CPS like a suburban school district in Cook County.

Estimate comparison groups like large school districts, city, state, and country-level averages can be found on education departments' and government websites. This data will be used to compare CPS estimates to non-local averages to see if there are any alarming discrepancies.

Two standardized test score approaches could be used to see the effect of the computer science requirement by looking at PSAT and SAT scores. A majority of the weight will fall on SAT scores with PSAT as a substitute to make up for the lack of SAT data caused by COVID-19. Using PSAT as a lesser weighted second approach increases the number of observations that can be seen and the number of cohorts that can be looked at. Since there are at most 4 cohorts that can be seen with SAT data.

The issue with PSAT data is that there is selection with who tries their best since there is very little to gain from putting effort into the test. The two main benefits are practicing for the SAT and getting the opportunity to win a scholarship (The Princeton Review). Due to the diminishing value of the SAT, it would be even more difficult to get students to take the PSAT seriously. This will lead to bias in the PSAT estimate with varying degrees of effort that students put into the test, which will make the data very noisy as students have very little incentive to do well.

Mark Hoekstra's *The Effect of Attending the Flagship State University on Earnings: A Discontinuuity-Based Approach* (2009) used university admission and unemployment insurance and connected the observations using social security numbers (718). A similar matching approach using social security numbers can be used to find earnings for CPS and other school district cohorts. In Hoekstra's paper, there were different amounts of earning data collected that can cause noise if a similar approach is used with CPS' and other school districts' student earnings (Hoekstra 2009, 718).

VII. Empirical Strategy

A difference-in-difference research design will be used with standardized test scores, GPA, college acceptance, and earnings as outcomes. There are both school average and individual outcomes looked at based on the potential of enough data appearing to regress individual outcomes. The threshold for treatment is the implementation of the computer science class graduation requirement for the class of 2020 and beyond. All classes before the class of 2020 will not be considered treated. The counterfactual that can be used is suburban school districts in Cook County that are trending the same way in the years before the introduction of CPS' policy. As long as both school districts are trending the same way before the introduction

of the policy, then the difference-in-difference model will be following the common trends assumption.

Other counterfactuals could be used to check estimates in the difference-in-difference model like other large districts. Large districts can include the public school system of New York, Los Angeles, and Houston. City, state, and country averages can only be used as counterfactuals for comparing estimates and not as counterfactuals in the model based on the sheer size of observations at the city, state, and country levels. There would be data issues with some districts at any of the levels not having all the data necessary or unwilling to share the data to run the regressions. Other large school districts could be used as counterfactuals in the model because of the similar size and atmosphere of being in a city.

(1) $Standardized\ Test\ Score_{sy} = \beta CSGradReq + School_{s} + Year_{y} + \epsilon_{sy}$

(2) $Standardized\ Test\ Score_{isy} = \beta CSGradReq + School_{s} + CSTeacher_{sty} + Year_{y} + \epsilon_{isy}$

Equation 1's and 2's Variables and Comments:

Standardized Test Score is standardized test scores as an outcome. Equation 1 has a school's average standardized test score as an outcome while equation 2 has an individual's standardized test score. CSGradReq is a dummy variable for whether there was a computer science graduation requirement for that class or student. In equation 1, CSGradReq is 1 if a majority of the students in a school who are receiving standardized test scores are in the class of 2020 or beyond that are receiving or have received the treatment. In equation 2, CSGradReq is 1 if the student who is receiving their standardized test score is in a class that is receiving or has received the treatment. CSTeacher, Individual, School, and Year are the fixed effects for the computer science teacher that taught the student's computer science course, the student, the school the student attended, and the year. Each fixed effect as a subscript is represented by the first letter in its name. ϵ is the error term.

(3) $GPA_{sy} = \beta_0 + \beta_1 CSCourse_{sy} + \beta_2 PriorGPA_{sy} + School_{s} + Year_{y} + \epsilon_{sy}$

(4) $GPA_{isy} = \beta_0 + \beta_1 CSCourse_{isy} + \beta_2 PriorGPA_{isy} + School_{s} + Year_{y} + \epsilon_{isy}$

Equation 3's and 4's Variables and Comments:

GPA is grade point average as an outcome. Equation 3 has a school's average GPA as the outcome while equation 4 has an individual's GPA as the outcome. PriorGPA is the school average GPA the year before for equation 3 and the student's GPA the year before for equation 4. CSCourse is a dummy variable for if the student had taken a required computer science course in a high school. The fixed effects and error term are the same as the first two equations.

(5) $College\ Acceptance\ Rate_{sy} = \beta_0 + \beta_1 Standardized\ Test\ Score_{sy} + \beta_2 GPA_{sy} + \beta_3 ClassRank_{sy} + \beta_4 NumberofExtracurricularActivities_{sy} + \beta_5 CSCourse_{sy} + \epsilon_{sy}$

(6) $College\ Acceptance\ Probability_{isy} = \beta_0 + \beta_1 Standardized\ Test\ Score_{isy} + \beta_2 GPA_{isy} + \beta_3 ClassRank_{isy} + \beta_4 NumberofExtracurricularActivities_{isy} + \beta_5 CSCourse_{isy} + \epsilon_{isy}$

Equation 5's and 6's Variables and Comments:

College Acceptance Rate/Probability is the probability of getting accepted to college, and this can be used for any college or 4-year college. The probability of getting accepted to a 4-year college would be used as an outcome to exclude colleges that have a very high probability of getting accepted like community colleges that have basic requirements like having a high school education or an equivalent to a high school education. Equation 5 has a school's college acceptance rate as the outcome while equation 6 has the probability for an individual student to receive acceptance from a college. GPA and ClassRank are the grade point average and the class rank for the student respectively. NumberofExtracurricularActivities is the number of total extracurricular activities the students have done in high school in terms of semesters. So, participating in an extracurricular activity for a majority of a semester would count as 0.5 and 1 for a majority of both semesters. CSCourse is a dummy variable for if the student had taken a required computer science course in a high school. The fixed effects and error term are the same as the first two equations.

(7) $Earnings_{sy} = \beta_0 + \beta_1 CSCourse_{sy} + \beta_2 Education_{sy} + \beta_3 WorkExperience_{sy} Year_y + \epsilon_{sy}$

(8) $Earnings_{isy} = \beta_0 + \beta_1 CSCourse_{isy} + \beta_2 College_{isy} + \beta_3 WorkExperience_{isy} + \beta_4 Industry_{isy} + \beta_5 Education_{isy} + City_c + Year_y + \epsilon_{cisy}$

Equation 7's and 8's Variables and Comments:
Earnings are the outcome that is adjusted to 2017 dollars. Equation 7 has a school's average earnings as an outcome while equation 8 has a student's individual earnings as the outcome. CSCourse is the same dummy variable in equations 3-6. College is the college attended by the student. Education is the amount of education achieved either by the school average in equation 7 or by the individual student in equation 8. Work experience is the number of years worked with the school average in equation 7 and the individual student's total in equation 8. Industry is the industry the student is currently working in. Equation 7 has a Year fixed effect and school and year as the subscripts. Equation 8 has City and Year fixed effects with city, individual student, school, and year as subscripts.

Equation 3 and 4 use GPA as a proxy for graduation rate because GPA almost perfectly correlates with graduation rates and provides greater detail on changes in grades that can be caused by the introduction of a required computer science course. GPA does not perfectly predict graduation rates because other factors contribute to graduating like not completing certain class requirements, not having enough service learning projects, and not having post-secondary plans.

Each regression equation includes the computer science requirement treatment in some form. Equation 1 includes the treatment in the form of whether the majority of those receiving standardized test scores are in the class of 2020 or beyond. Equation 2 includes the treatment of whether the individual who is receiving their standardized test score is in the class of 2020 or beyond. In equations 3-8, the treatment is in the form of whether the school average or the individual had taken a required computer science course. This links together the treatment with

the subsequent stages of education and job market outcomes in addition to the previous stages. An example of this is the coefficient of CSCourse in College Acceptance where College Acceptance includes GPA that also has a coefficient for CSCourse. This allows for a less noisy measure of the computer science requirement on outcomes since it measures prior educational stages that are also impacted by the computer science requirement. Both coefficients can be examined and weighed to get a more accurate measurement of CPS' policy.

VIII. Mechanisms

The mechanisms studied in this paper are how standardized test scores, GPA, college acceptance, and earnings are affected by the introduction of CPS' computer science graduation requirement. The difference-in-difference model uses between variation between CPS and a suburban Cook County school district which is the primary counterfactual. Since almost all school districts are trying to increase standardized test scores, GPA, college acceptance, and job market outcomes, it is fair to interpret that on average the school and district efforts to increase these metrics will not cause bias. Almost every school district tries to increase the metrics mentioned since they can be a determining factor in funding, prestige, and enrollment. These school districts try to increase standardized test scores, GPA, college acceptance, and job market outcomes by providing information about external resources, advising teachers to spend time on education and standardized test preparation, after-school tutoring and mentorship programs, and providing resources themselves.

VIIII. Validity

The primary issues with validity are changes with standardized testing, COVID-19, and some CPS high schools not being able to offer computer science courses to students with computer science as a graduation requirement in their freshman and sophomore years.

Changes in the 2017-2018 school year saw a shift from the ACT standardized test to the SAT standardized test that was seen in the standardized test score data from CPS' Assessment Reports page on their website (CPS). On average, this should not make a difference as there are ways to convert ACT scores to SAT scores. There should be no validity concerns with colleges weighing standardized tests differently as there are time-fixed effects that would take this into account. There are other crucial changes made to the SAT due to the presence of technology and COVID-19 like online SAT testing and the SAT removing the essay test (College Board, 2021 and Morning Edition, 2022). A complete change in the standardized test format will make it hard to compare prior test scores with new test scores due to variability in how the test compares to itself. Different changes can affect individual students' test scores differently based on their strengths, weaknesses, and how they think. National Public Radio's Morning Edition article on the SAT changes that are soon to be enacted in 2024 shows that the changes in the SAT are meant to be test-altering to adjust to the current landscape of schooling. There are changes to the structure of the reading test and the rules of the math test which would create bias in a straight comparison between the current SAT and the future SAT (2022). This means that any new score adjustments might not capture the true conversion rate and might have an upward or downward bias on the estimates that are converted scores. As these scores could not be the value of the true

converted scores. The SAT removing certain test also has this same effect as there is one less metric that can be used to compare individual students. SAT's essay test did have an issue of selection due to its optional nature of it, so eliminating it would still have difficulties comparing before and after the removal. This also affects college acceptance rates as non-test-optional schools might have valued those who were selected into taking the essay test before its removal. Therefore, this is another systematic difference in the test over time. Although the time fixed effect will still capture this.

The drastic change among college to test-optional applications has decreased the value of doing well on the SAT, especially during COVID-19. With less of an incentive to do well on the SAT, there will be more students who choose to not take it seriously or prepare for the test. This will see very low and noisy estimates for the coefficient of the college acceptance rate. The time-fixed effects would have captured the changes in the SAT over time and would be large. Although, the estimates would not be externally valid since the rapid digitalization of education and COVID-19 together have pushed for drastic and permanent changes to be made.

COVID-19 does cause bias in students who took the SAT and the time they took the SAT. From March-June 2020, a majority of activities in the United States were closed which means that almost all students who planned to take the SAT had to postpone their test. The lack of students taking the SAT during 2020 and the beginning of 2021 means that there will be very low observation amounts. Even with time-fixed effects, there will not be enough observations to get meaningful results compared to prior years. In addition to mostly a majority of SAT test-takers following COVID-19 protocols like wearing a mask and social distancing, test-taking will be different for students and will reflect in their scores. Time-fixed effects can capture this but would have already captured a lot of other dirty variation and it would be hard to tell if what is left is clean. In combination with the low observation amounts, there might not be enough variation left to have an unbiased estimate.

The projected date of the beginning of the 2018-2019 school year for all high schools to be compliant and offer acceptable computer science courses might not give enough time to have a common pretrend with the counterfactual. The fact that many CPS high schools are actively changing to become compliant with the new computer science graduation requirement will cause there to be other changes too. This will introduce bias and could be captured by time fixed effect but more precise data is needed on the dates that each high school in CPS became compliant with the new graduation requirement. Additionally, students in the class of 2020 or beyond who take the required computer science course in their senior year of high school would not have their standardized test score impacted unless they take and retake the standardized test in their senior year and have taken the computer science course long enough to learn about the educational resources and apply them. This would also be controlled for with the CSGradReq variable that assigns the computer science graduation requirement treatment to those who are receiving or have received the treatment.

Using PSAT scores to increase the number of cohorts and observations introduces selection bias in the form of individuals choosing whether they should try their best on the PSAT

or not. The PSAT offers minimal benefits to a majority of students who take it with a percentage of those who do well being offered a scholarship (The Princeton Review). This will create a downward bias on the estimates as it will students intentionally not try their best. The lack of effort will be a problem due to the varying magnitude at which this can occur. An individual fixed effect can be used, but will not do very well to find the clean variation to get close to a true estimate.

X. Conclusion

This research proposal seeks to find the effect of CPS' computer science graduation requirement on students' standardized test scores, GPA, college acceptance, and earnings. The research design for this proposal is a difference-in-difference model that uses the implementation of CPS' computer science graduation requirement as the treatment. The primary counterfactual is a nearby school district like a suburban school district in Cook County that meets the common trends assumption. There are data concerns because there is not enough data to run all the regression in this proposal due to CPS' policy being somewhat recent. COVID-19 also affected the number of observations and the reliability of observations in 2020 and 2021 due to a different type of testing environment. Overall, with more reliable data and the cooperation of CPS and another school district, this research can be done to examine the effects of CPS' policy.

XI. References

Barrow, L., Freire, S., & de la Torre, M. (2020). Trends in Computer Science Education. *UChicago Consortium on School Research*. https://consortium.uchicago.edu/publications/trends-in-computer-science-education

Chicago Public Schools. (n.d.-a). *Assessment Reports*. Chicago Public Schools; Chicago Public Schools. Retrieved April 27, 2022, from https://www.cps.edu/about/district-data/metrics/assessment-reports/

Chicago Public Schools. (n.d.-b). *Chicago Public Schools—School Progress Report*. Chicago Open Data Portal; Chicago Public Schools. Retrieved April 28, 2022, from https://data.cityofchicago.org/browse?q=chicago+public+schools+-+school+progress+report&sortBy=relevance&page=1

Chicago Public Schools. (n.d.-c). *Conduct Primary Research*. Chicago Public Schools; Chicago Public Schools. Retrieved April 27, 2022, from https://www.cps.edu/about/district-data/conduct-primary-research/

Chicago Public Schools. (n.d.-d). *CPS Computer Science Graduation Requirement Fact Sheet*. Retrieved April 27, 2022, from https://blog.cps.edu/wp-content/uploads/2016/02/CS_Grad_Fact_Sheet.pdf

Chicago Public Schools. (2019). *SAT Suite of Assessments*. Chicago Public Schools; Chicago Public Schools. https://www.cps.edu/globalassets/cps-pages/academics/student-assessments/sat_suiteofassessments_fact_sheet.pdf

Chicago Public Schools. (2022, April 27). *Chicago Public Schools—School Progress Reports SY2122*. Chicago Open Data Portal; Chicago Public Schools. https://data.cityofchicago.org/Education/Chicago-Public-Schools-School-Progress-Reports-SY2/ngix-dc87/data

Code.org. (n.d.). *Learn computer science. Change the world.* CODE. Retrieved April 27, 2022, from https://code.org/

College Board. (2021, January 19). College Board Will No Longer Offer SAT Subject Tests or SAT with Essay [Blog]. *College Board Blog*. https://blog.collegeboard.org/January-2021-sat-subject-test-and-essay-faq

Elahi, A. (2016, March 1). CPS to roll out computer science requirement. *Chicago Tribune*. https://www.chicagotribune.com/business/blue-sky/ct-computer-science-graduation-cps-bsi-20160225-story.html

Hoekstra, M. (2009). The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach. *The Review of Economics and Statistics*, *91*(4), 717–724. https://content.blackboardcdn.com/

Inskeep, S., & Martin, R. (n.d.). *U.S. students will take the SAT entirely online beginning in*

2024. Retrieved April 27, 2022, from https://www.npr.org/2022/01/28/1076306310/u-s-students-will-take-the-sat-entirely-online-beginning-in-2024

The Princeton Review. (n.d.). *National Merit Scholarships and the PSAT*. The Princeton Review; The Princeton Review. Retrieved April 28, 2022, from https://www.princetonreview.com/college/psat-nmsqt