Computer Science Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ncse20

In-school and/or out-of-school computer science learning influence on CS career interests, mediated by having role-models

Chen Chen, Jonathan Rothwell & Pedrito Maynard-Zhang

To cite this article: Chen Chen, Jonathan Rothwell & Pedrito Maynard-Zhang (26 Dec 2023): In-school and/or out-of-school computer science learning influence on CS career interests, mediated by having role-models, Computer Science Education, DOI: 10.1080/08993408.2023.2290435

To link to this article: https://doi.org/10.1080/08993408.2023.2290435

Published online: 26 Dec 2023.

COMPUTER SCIENCE EDUCATION
https://doi.org/10.1080/08993408.2023.2290435

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

In-school and/or out-of-school computer science learning influence on CS career interests, mediated by having role-models

Chen Chen [a,b], Jonathan Rothwell [c,d] and Pedrito Maynard-Zhang [e]

[a]Faculty of Education, University of Hong Kong, Hong Kong; [b]Science Education Department, Center for Astrophysics, Harvard | Smithsonian, Cambridge, MA, USA; [c]Gallup, Washington, DC; [d]Metropolitan Policy Program, Brookings Institution, Washington, DC; [e]Amazon Future Engineer, Amazon in the Community, Seattle, WA

ABSTRACT

Background and Context: Both in- and out-of-school computer science (CS) learning opportunities are expanding, but their influences on CS career interests are unclear.
Method: To investigate, we applied multinomial propensity score weighting analysis on a 2021 U.S. nationally representative sample of 4,116 5th-to-12th-grade students.
Findings: The odds of expressing CS career interest increase by 171%, 94%, and 40%, respectively, when students pursue CS learning both in and out-of-school, out-of-school only, and in-school only. These effects were similar across race/ethnicity but stronger for girls. Out-of-school learning was the strongest predictor of having CS role-models, though each experience was positive. One third of the effect of both and about half of the effect of each separately were mediated by having CS role models.
Implications: Our findings suggest that domain-focused learning experiences are generally effective in shaping career interests, and out-of-school learning, specifically, may enhance exposure to role models.

ARTICLE HISTORY
Received 15 February 2023
Accepted 29 November 2023

KEYWORDS
Computer science education; career interest; out-of-school time learning; computer science in school; role model

Introduction

Bolstering the supply of adults with computer science proficiency and a willingness and capacity to work in computer-related professions has long been recognized as a national priority for the United States and other developed countries. As an academic field, computer science (CS) has unusual characteristics in that it aligns with rapidly growing employer demand for computer science skills, a shortage of workers, and as a relatively new field, a shortage of qualified teachers, particularly in elementary and secondary school settings. It is also, arguably, one of the few STEM subjects that large numbers of students can master by learning independently out-of-school, given the widespread

CONTACT Chen Chen chen.chen@cfa.harvard.edu Faculty of Education, University of Hong Kong
Supplemental data for this article can be accessed online at https://doi.org/10.1080/08993408.2023.2290435.

availability of online training resources and an increasing number of community-based learning centers.

To inspire the future generation to pursue the unfilled, high-paying jobs in computing, and to fulfill a nation's interest in innovation and technology, various groups of stakeholders have called for both the expansion of in-school offerings of CS courses and the investment in out-of-school CS experience opportunities. These two solutions largely summarized the approaches adopted by the U.S. federal initiative "CS for All" (e.g. DeLyser, 2018). This approach raises two important questions. First, to what extent do childhood CS learning experiences affect the probability of interest in working in a computer science career (e.g., as an information scientist, software developer, or network architect)? Secondly, do in-class experiences have different effects than out-of-class experiences? This paper addresses both of these questions and considers how the development of computer science role models, either in or outside of class, may mediate career interest.

Academic literature about the impact of in-school and out-of-school experience, and the comparison between the two, has been limited. A review of the literature published between 2012 and 2018 showed that most research on the topic has been limited by sample sizes, sampling biases, or underreported demographic information, which constrained the generalizability of these studies (Decker & McGill, 2019; Upadhyaya et al., 2020). In this study, we use a nationally representative sample of 4,116 U.S. students in grades 5 to 12 to investigate the effect of in-school, out-of-school or both CS learning experiences on students' CS career interests. Adopting an expectancy-value theory framework, we study how role models mediate the effects of formal learning experiences, and whether or not this dynamic may be different across gender or race/ethnicity.

Literature

Theoretical framework

Expectancy-value theory (EVT) (Ball et al., 2017; Eccles & Wigfield, 2020; Wigfield, 1994) is widely adopted in explaining students' learning and career interests. This theory has three key components: 1) the expected chance of success (e.g., Can I get a job in an IT company?), 2) the task value for personal goals (e.g., Will a computing career enable me to pursue my financial or social goals?), and 3) the expected cost to achieve success (e.g., How much stress – emotionally or financially – will I have to endure?). In a nutshell, a person tends to have high motivation to pursue and persist in a task when the anticipated benefits of the task are aligned with personal values, and when the value and chances of success outweigh the estimated cost. These variables – personal value system, the anticipation of benefits afforded by a task, estimation of success or cost – are not fixed or pre-determined; rather, they adjust dynamically with one's experience. In the context of computing career interest, influential experiences may come from within or outside of schools.

Learning CS in-school

The in-school offerings of CS education are expanding (Upadhyaya et al., 2020). According to Code.org (2022), 53% of high schools in the U.S. offer CS courses in 2022, a rapid increase from 35% in 2018. In the U.S., more than 50 million dollars from state level budget were allocated to K-12 computer science education in year 2022 (Code.org, 2022). AP Computer Science Principles, launched in 2016, focusing on the big ideas in computing, helped recruit students from historically underrepresented groups (The College Board, 2020).

Some scholars had concerns that high school CS courses, downplaying the technical and mathematical nature of CS, might give students an over-simplified impression of a CS career, and students would be disappointed or overwhelmed when they encountered advanced levels of computing (in EVT terms, an overestimation of chance for success contradicted by the realization of emotional-cognitive cost) (Beaubouef & Mason, 2005). Nevertheless, large volumes of evidence have shown that students who enroll in high school CS coursework are more likely to major in CS in college (Armoni & Gal-Ezer, 2014; College Board, 2011, 2020). From the EVT perspective, high school CS – although surveying the surface – offers a roadmap to the field, a preview of the endgame, and many chances for self-assessment. Such offerings not only provide information to assess one's expected chance for success and expected obstacles, but also pave a gentle introduction that boosts one's confidence in achieving success (Authors, 2019a). Moreover, the foundational courses in CS, such as AP CS Principles, connect CS with society and students' lives, which may help students to decide if CS careers are aligned with their personal values and goals. In addition, the more technical high school CS courses, such as AP CS A or Carnegie Mellon University (CMU) CS Academy's CS2 (Stehlik et al., 2020), offer students in-depth knowledge in programming and transition to higher education. The authors (2020) further show that students who enroll in AP CS have similar performance in college-level introduction to CS courses than students who enroll in AP Calculus. It is noteworthy that, despite the national push to broaden access to CS education in schools, only 5.6% of high school students in the U.S. (sampled in 36 states) enrolled in any foundational computer science courses (Code.org, 2022). The rest of the students either do not have any foundational CS learning experience or learn CS out-of-school.

Learning CS out-of-school

The CS profession is a one-of-a-kind STEM occupation in which a large number of the professionals are trained out-of-school. According to a computer programming developer survey conducted by Stackoverflow.com (2021), only 53.59% of developers learned coding in school. The younger the developers, the more likely they learned programming out-of-school such as from online forums/platforms or from bootcamps. In fact, this pattern applies to K-12 CS teachers as well – as of 2021, only 30% of them held a degree in a CS or IT-related field, only 46% held CS teaching credentials, and 62% were out-of-field teachers who were also teaching other subjects (Code.org, 2021). The lack of formal education in CS doesn't prevent them from growing into the CS teaching positions, thanks to alternative paths in out-of-school professional development (Ni et al., 2023).

Scholars once worried that students learning coding out-of-school in a "cowboy/cowgirl" hacking style (informal learning) would end up developing many bad habits in programming, and that correcting these bad habits might be more difficult compared to formal learning against a "blank sheet" (Kölling, 1999, p. 4). However, this argument only focused on code tidiness as taught in schools but neglected the fact that the CS profession is full of hackery and messiness. Students who have had the experience of programming out-of-school may have more chances to "geek out and mess around" (Ito et al., 2009; Liggett, 2014; McKenna & Bergie, 2016), which may be discouraged in schools or in exams but to a certain degree reflects the workspace of the CS profession (at least partially, in the initial stages of problem-solving).

From the perspective of EVT, learning CS out-of-school may situate learners in the context of real-world challenges. In this way, learners can more easily decide if the task is manageable (in terms of difficulty) and relevant (in terms of personal goals). In addition, "messing around" may reduce the perceived cost of errors or dead ends. This argument is strongly supported by a national sample study of U.S. college students in introductory CS courses that showed that students who learned programming out-of-school performed significantly better in these courses and had significantly more positive attitudes towards CS than those who enrolled in high school CS courses but never learned or practiced it out-of-school (Authors, 2019). Interestingly, half of the effect of out-of-school learning was mediated by the boost in attitude towards CS. This cited study, however, only sampled introductory CS students in colleges and did not sample the students who didn't go to college or students who might be so advanced in CS (due to their pre-college learning experience) that they skipped the introductory level courses in college. There is also a dearth of research that studies the influence of out-of-school learning experiences on students' CS career interests.

Role models

Role models can have a major influence on students' career interests. They may closely communicate with the students about an occupation, and/or they are often identified by the students to be similar to themselves in certain aspects (Betz & O'Connell, 1992). They influence learners by showing the possibility of career paths, setting up behavior norms, and promoting personal values (Morgenroth et al., 2015). In the EVT framework, the presence of role models helps students to assess if "people like me" can succeed in the field, and if "people who share my values" can actualize their goals in their professions (Gartzia et al., 2021). In addition, by providing behavior demonstrations in everyday life, they normalize the task by making it seem less daunting, nerdy or boring (Woods-Townsend, 2016). Authors' (2021) previous study showed that 73% of grades 5–12 students who strongly agree they have computer science role models say they plan to have a job in CS, compared to only 7% who strongly disagree. Research has shown that role models are particularly important for underserved and underrepresented students to aspire and persist in CS (Kearney & Levine, 2020; Mani & Riley, 2019).

Classroom teachers can usually serve both of these two functions: they give students career advice (Jacobs & Bleeker, 2004; Sonnert, 2009), and students see the resemblance in them because they consider teachers as in-group persons (Lockwood, 2006; Nixon & Robinson, 1999; Authors, 2019b). The first teacher in a subject is usually the role model that exercises lasting influence (Bettinger & Long, 2005; Authors, 2019b). For this reason, high school CS teachers are often anticipated to become role models. The question remains to what extent high school teachers can fulfill the role of CS career role models by sharing CS career insights if a large proportion of them are out-of-field teachers and do not have a CS degree or CS professional experiences. Marken and Crabtree (2021) showed that 5th-12th-grade students in big cities are more likely to report having CS role models than students in rural areas. They interpret this difference as a result of schools in big cities being more likely to offer CS courses than rural schools, which further suggests that most of the students' CS role models were in their schools. Nevertheless, big cities may have more IT companies and professionals as well as out-of-school CS learning opportunities and infrastructure that contribute to a larger pool of mentors and role models (Clevey, 2018).

School teachers don't have to assume the burden of role models. As once suggested by Bryan Twarek (2018), the director of education at CSTA (Computer Science Teachers Association), teachers can expose students to more CS role models by inviting guest speakers, arranging field trips (in-person or virtual) to explore careers, or watching movies (or reading biographies) about the great figures in the history of computing – see examples from CS Journeys (Code.org, 2022) or Class Chats (Amazon Future Engineer, 2021). Arguably, all these approaches borrow some forms of out-of-school learning.

Underrepresented groups in CS

The opportunities to learn CS in-school or out-of-school, or to identify with role models, are neither equal nor random for all students (Kearney & Levine, 2020). Women and racial/ethnic minorities (except for Asian) are underrepresented in high school CS course enrollment and in AP computer science exams (Code.org, 2022). Ashcraft et al. (2012) attributed it to the abstract curriculum offered in schools that focuses on independent work/tests and "innate" talent but downplays communal goals and collaborations. The "innate" talent mindset leads underrepresented students to the self-defeating expectations of their failure, as opposed to the "growth" mindset that anticipates the task difficulty to decrease in the long run as one persists and grows (Dweck, 2007; Wang & Degol, 2013). Women students are more likely to value community belonging and mutual help and may consider CS professions inconsistent with their personal goals (Diekman et al., 2017; Tellhed et al., 2018). At the university level, programs that have successfully promoted STEM completion for Black students feature community-building and summer-bridge programs (Domingo et al., 2019). Recently, numerous initiatives have tried to amend this shortcoming by developing school curriculums that are more culturally responsive (Codding et al., 2021; Eglash et al., 2013). In the meantime, out-of-school programs, such as Girls Who Code (Stern et al., 2015) or Black Girls Code (Braswell & Rankin, 2023), harness the power of community that may foster identity, give exposure to potential role models and reveal the diversity in the CS workforce.

Gender and race/ethnicity are only two of many dimensions where we may observe gaps in the CS learning opportunities. Students who go to schools that offer CS courses may live in resourceful school districts, and students who attend out-of-school programs may have strong parental support. Seldom has any research thus far teased out these entangled confounding factors when considering the influence of learning experiences on CS career interests.

Research questions

Here, three questions are at stake: 1) Does CS learning in-school or out-of-school influence CS career interests even if we curb the self-selection bias due to the under-sampling of disadvantaged and underrepresented students? 2) To what extent do the effects of in-school or out-of-school CS learning (if the effects are present and robust) differ by gender or race/ethnicity? And 3) what role does having role models play in the above relationships? To answer these questions, we collected a nationally representative sample of pre-college students and used statistical methods such as propensity score weighting/matching (randomization is impossible) to balance students' background information in order to yield a more robust estimation than typical regression models. We also adopted mediation models (causal mediation in the case of propensity score weighting/matching) to estimate the proportion of the total effect that has been mediated.

Formally, we ask:

RQ1: 1) What are the effects of *in-school*, *out-of-school*, or *both* CS learning experiences on grades 5–12 students' CS career interests? 2) Do these effects vary by student's gender or race/ethnicity?

RQ2: 1) What are the effects of in-school, out-of-school, or both CS learning experiences on grades 5–12 students' likelihood of reporting having a CS role model(s)? 2) Do these effects vary by student's gender or race/ethnicity?

RQ3: What proportion of the effects of *in-school*, *out-of-school*, or *both* CS learning experiences on grades 5–12 students' CS career interests are mediated by having a role model(s)?

Methods

Sample

Our survey data are from the Amazon Future Engineer/Gallup Student Survey. The data were collected by Gallup through a web survey conducted between June 2 and June 20, 2021, with a sample of 4,116 U.S. public and private school students in grades 5–12. Respondents were contacted via the Gallup Panel, a probability-based panel of U.S. adults, and an opt-in sample from Dynata.

Gallup obtained parental consent before requesting the participation of a child. Parents with more than one child in grades 5 through 12 were instructed to select the child with the most recent birthday to complete the questionnaire.

Gallup statisticians weighted the obtained samples to correct for nonresponse. This was done by adjusting the sample to match the national demographics of gender, grade, race/ethnicity, and school type. Demographic weighting targets were based on the U.S.

Census Bureau's American Community Survey for 2019. In addition to sampling error, question-wording and practical difficulties in conducting surveys can introduce error or bias into the findings of surveys.

Measurement

The questionnaire was developed in a partnership between Gallup and Amazon Future Engineer. Two of the current authors (Pedrito Maynard-Zhang and Jonathan Rothwell) contributed to the development of the survey along with Gallup methodologists, who performed cognitive testing with several of the proposed items on earlier drafts to check for clarity and respondent understanding.

Independent variables

To study computer science learning among children and young adults in our sample, computer science was defined in the survey using the following language: *Now, we would like to ask you some questions about computer science.*

Computer science includes learning about or doing things like:

- *Developing hardware and software for computers*
- *Programming a computer*
- *Writing and running computer code*
- *Robotics*
- *Machine learning and artificial intelligence*

Computer science does not include things like:

- *Simply using a computer, tablet, or smartphone*
- *Using the internet to search for things or viewing social media*
- *Creating documents or presentations on the computer*

Participants were then asked "Have you ever learned any computer science?" If they answered yes, they were asked to select where they learned it from a list that includes "in a class at school during school hours" and eight other categories, including in an after-school class, a club, or "on my own, by teaching myself".

Students were classified into one of four categories based on their responses to these items: 1) Have not studied computer science, 2) Studied only in-school during school hours, 3) Studied only out-of-school or outside school hours and 4) Studied both in-school during school hours and elsewhere. In our full sample, 35.1% of students have not studied computer science, 23.4% have studied in school during school hours, 16.9% have studied outside of school or school hours, and 25.1% have studied both in and outside of school. Table 1 show the summary statistics for all variables grouped by the above four categories.

Dependent variables

Our primary outcome of interest is whether or not the student plans to have a career in computer science. This was collected by asking, "Which of the following fields do

Table 1. Summary statistics for all variables, grouped by CS learning experiences.

	Full sample			Did not study			Outside of school			In school			Both in and outside school		
Descriptive variable name	N	Mean/Percent	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Plans to work in computers and technology field someday	4,107	0.276	0.447	1,446	0.141	0.348	710	0.364	0.482	930	0.240	0.427	1,021	0.441	0.497
Age	4,116	13.680	2.405	1,448	13.806	2.393	713	13.240	2.426	933	13.985	2.343	1,022	13.506	2.412
Female	4,116	0.459	0.498	1,448	0.529	0.499	713	0.362	0.481	933	0.503	0.500	1,022	0.385	0.487
American Indian or Alaskan Native	4,110	0.007	0.081	1,447	0.008	0.090	712	0.010	0.098	931	0.005	0.074	1,020	0.003	0.057
Asian	4,110	0.046	0.210	1,447	0.043	0.203	712	0.050	0.218	931	0.047	0.212	1,020	0.047	0.212
Black	4,110	0.117	0.322	1,447	0.142	0.349	712	0.109	0.311	931	0.094	0.292	1,020	0.110	0.313
Hispanic	4,110	0.199	0.399	1,447	0.219	0.414	712	0.212	0.409	931	0.190	0.392	1,020	0.171	0.377
Multi-racial in any combination	4,110	0.036	0.187	1,447	0.044	0.205	712	0.021	0.144	931	0.033	0.178	1,020	0.038	0.192
Native Hawaiian or Pacific Islander	4,110	0.002	0.049	1,447	0.002	0.046	712	0.006	0.077	931	0.003	0.055	1,020	0.000	0.000
Other race	4,110	0.025	0.156	1,447	0.031	0.174	712	0.012	0.107	931	0.029	0.168	1,020	0.021	0.144
White	4,110	0.567	0.496	1,447	0.510	0.500	712	0.581	0.494	931	0.599	0.490	1,020	0.608	0.488
Parental IT ability (reverse coded)	4,107	1.593	0.712	1,446	1.706	0.771	708	1.562	0.645	933	1.592	0.716	1,020	1.455	0.635
Internet reliability	4,111	3.505	0.662	1,448	3.398	0.682	710	3.543	0.686	932	3.522	0.671	1,021	3.616	0.585
Number of children in home	4,116	1.880	0.994	1,448	1.784	0.888	713	2.048	1.154	933	1.853	0.954	1,022	1.928	1.040
Has computer or laptop in house	4,116	0.922	0.269	1,448	0.917	0.276	713	0.866	0.341	933	0.915	0.278	1,022	0.970	0.170
Has tablet at home	4,116	0.564	0.496	1,448	0.560	0.497	713	0.531	0.499	933	0.559	0.497	1,022	0.595	0.491
Has smartphone at home	4,116	0.742	0.437	1,448	0.777	0.416	713	0.628	0.484	933	0.745	0.436	1,022	0.765	0.424
No device in home	4,116	0.006	0.077	1,448	0.007	0.083	713	0.015	0.121	933	0.005	0.068	1,022	0.000	0.015
Child owns a device	4,116	0.998	0.039	1,448	1.000	0.019	713	0.994	0.080	933	0.998	0.040	1,022	1.000	0.000
Geography is missing	4,116	0.006	0.079	1,448	0.005	0.073	713	0.006	0.080	933	0.006	0.079	1,022	0.008	0.087
Lives in large city	4,116	0.294	0.456	1,448	0.194	0.396	713	0.407	0.492	933	0.280	0.449	1,022	0.372	0.484
Lives in suburb	4,116	0.375	0.484	1,448	0.398	0.490	713	0.321	0.467	933	0.374	0.484	1,022	0.377	0.485
Lives in town	4,116	0.202	0.402	1,448	0.244	0.430	713	0.184	0.387	933	0.204	0.403	1,022	0.155	0.362
Lives in rural area	4,116	0.123	0.328	1,448	0.158	0.364	713	0.082	0.274	933	0.136	0.343	1,022	0.089	0.284
Household income category (1–6)	4,062	4.566	1.517	1,424	4.273	1.579	704	4.638	1.517	920	4.606	1.499	1,014	4.884	1.370
Parent has high school diploma or less education	4,093	0.110	0.312	1,437	0.143	0.351	707	0.093	0.290	929	0.124	0.330	1,020	0.060	0.238
Parent has some postsecondary education but less than Bachelor's	4,093	0.326	0.469	1,437	0.366	0.482	707	0.313	0.464	929	0.331	0.471	1,020	0.274	0.446
Parent has a Bachelor's degree but not a graduate degree	4,093	0.318	0.466	1,437	0.272	0.445	707	0.339	0.474	929	0.320	0.467	1,020	0.367	0.482
Parent has graduate degree	4,093	0.246	0.431	1,437	0.218	0.413	707	0.255	0.436	929	0.224	0.417	1,020	0.299	0.458
School performance	4,103	3.366	0.743	1,445	3.172	0.804	707	3.495	0.622	930	3.378	0.751	1,021	3.540	0.652
Missing data on school sector	4,116	0.003	0.057	1,448	0.004	0.060	713	0.010	0.102	933	0.001	0.033	1,022	0.000	0.000
Public	4,116	0.771	0.420	1,448	0.817	0.387	713	0.655	0.476	933	0.813	0.390	1,022	0.743	0.437
Charter school	4,116	0.083	0.276	1,448	0.058	0.233	713	0.131	0.337	933	0.076	0.265	1,022	0.094	0.293
Private school	4,116	0.100	0.299	1,448	0.061	0.239	713	0.121	0.326	933	0.103	0.304	1,022	0.137	0.344

(*Continued*)

Table 1. (Continued).

	Full sample			Did not study			Outside of school			In school			Both in and outside school		
Descriptive variable name	N	Mean/Percent	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Home school	4,116	0.037	0.189	1,448	0.052	0.222	713	0.073	0.260	933	0.006	0.075	1,022	0.022	0.148
Other school, type	4,116	0.006	0.077	1,448	0.009	0.093	713	0.011	0.105	933	0.002	0.044	1,022	0.003	0.052
Have role models in computer science: strongly disagree	4,096	0.258	0.438	1,442	0.440	0.497	710	0.111	0.314	926	0.256	0.437	1,018	0.102	0.303
Have role models in computer science: somewhat disagree	4,096	0.213	0.410	1,442	0.250	0.433	710	0.130	0.337	926	0.254	0.436	1,018	0.178	0.382
Have role models in computer science: somewhat agree	4,096	0.271	0.445	1,442	0.223	0.417	710	0.373	0.484	926	0.254	0.435	1,018	0.288	0.453
Have role models in computer science: strongly agree	4,096	0.258	0.437	1,442	0.087	0.282	710	0.386	0.487	926	0.236	0.425	1,018	0.433	0.496

you plan or hope to have a job in someday?" One of the available choices was "computers and technology", and a positive response was deemed to indicate career interest, even if other careers were selected. There were 18 other career choices and a 19th option for "unsure". The options roughly correspond to the major occupational categories from the Standard Occupational Classification system used by the U.S. Bureau of Labor Statistics.

In our full sample, 28.2% of students selected computers and technology, and that was the modal choice. Other common choices included "Arts or Entertainment" (20.0%), "Engineering" (18.8%), "Health or Medicine" (17.2%) and "Business or Finance" (13.9%). The median student chose two careers. Other career choices were weakly correlated with career choices in computers. The strongest correlation was with engineering. Those who selected engineering were somewhat more likely than those who did not to select computers, with a bivariate correlation of 0.21.

As discussed, we hypothesize that learning computer science has a direct effect on career interests by increasing knowledge and interest in computers, but also generates additional motivation by exposing students to role models. To study role models as potentially mediating factors between studying computer science in childhood and career aspirations, we examine whether the students have **role models** in computer science. The survey asked students for agreement with the following statement: "I have role models in computer science". Response options were 1) strongly disagree, 2) somewhat disagree 3) somewhat agree and 4) strongly agree.

Control variables

The goal of this paper is to make a robust estimation to the effects of computer science learning experiences on career interest. A principal threat to identification is selection bias into computer science learning experiences. For example, parents who are software engineers may be more likely than parents with other careers to provide instruction to their children at home, sign them up for extracurricular computer science classes, or strategically enroll them in schools that offer computer science. If so, then career interest in computer science may appear to be the outcome of these experiences, when the fundamental cause is identification with or emulation of a parental figure. Regardless of parental occupation, this form of bias could cloud interpretation of our results if children from higher socio-economic status households are more likely to be exposed to learning experiences deemed valuable in the future labor market. More generally, these confounding factors could be described as familial predisposition toward computer science careers.

There may also be child-specific confounds. Students with higher cognitive ability, particularly in math, may be more likely to study computer science regardless of childhood experiences. Their participation in computer science in middle and high-school may be driven by their cognitive performance in math and related areas, which makes computer science more engaging and interesting. This could more generally be considered a child-specific bias, grounded in characteristics present before they studied computer science that are the fundamental cause of both study choice and career interest. We take several steps to address these concerns, as described below.

General demographics

First, we control for demographic characteristics of the parent and child. Specifically, we have binary (dummy) variables for each major racial group and Hispanic ethnicity as provided by the parent. We code children who are two or more races as multi-racial, all Hispanic children as Hispanic, and all single-race children as belonging to that group. Racial demographics, after weighting, are largely aligned with the 2020 American Community Survey, though our sample is somewhat over-represented by non-Hispanic White students (56.7% in the survey vs 48.7% in the Census) and under-represents Hispanic students (19.9% vs 25.6%). For this analysis, census data were restricted to currently enrolled students in grades 5 to 12, using data from Ruggles et al. (2022).

We also control for parental education and household income. Household income is entered linearly, using the ordinal category number, where a value of "1" is assigned if household income is < $24k, "2" if $24 to $48k, "3" if $48 to 60k, "4" if $60 to 90k, "5" if $90 to 120k, and "6" if > $120k. Some are wider in terms of actual monetary value and wider in terms of the share of respondents. The largest group is the highest (6), representing 34.4% of the sample. The smallest group is group 1 (7.7%). To confirm the validity of these categories, we merged respondent zip codes to IRS Individual Income Tax ZIP Code Data for 2020 (the latest available) and calculated average household income per zip code, as adjusted gross income per tax return, a proxy for household income at the neighbourhood level. When compared against our household income variable, the log of mean household income is correlated at 0.35, and the rank order is preserved. Respondents who did not select a household income group on our survey had the lowest median values (using IRS data), compared to the 6 other groups.

We create binary variables for parents with a 1) high school diploma or less, 2) some postsecondary experience or associate's degree, 3) a bachelor's degree or some post-graduate education and 4) a graduate degree.

Parents were also asked to provide the age of their child, which ranged from 9 to 19 (mean is 13.7).

Potential confounds related to computers and technology

Our survey has several items that capture access to technology and parental expertise in using it. To measure parental expertise in computer science, the survey asks parents: "How would you rate your ability with computers and computer-related technology?" Answer choices were strong, good, fair, weak. 52% responded "strong" and 37% responded "good", with only 9% and less than 2% replying fair and weak, respectively. There were two ways to quantify this variable: either treating it as a numerical Likert scale variable, or categorizing it into smaller bins. We tried both approaches, intending to examine if the two yield the same estimations. We first treated this variable as a numerical Likert scale variable and entered it into a series of models. We separately categorized it as a binary variable (1 = strong; 0 = not strong) to examine the difference it brought to model estimations.

We also asked about the strength of the home internet connection, which speaks to both socio-economic status, geographic characteristics, and technology savvy: "How reliable is your internet access at home?" Answers were very reliable

(58%), reliable (35%), not too reliable (5.4%), not at all reliable (0.6%), and do not have internet access at home (0.4%). Similar to parental IT skills, we first entered this variable to our models as a numerical Likert scale variable, and separately categorized it as a binary variable (1 = reliable; 0 = not reliable). The survey asks parents, "What type of devices, if any, does your child have access to **at home**?" Parents could select 1) a computer or laptop, 2) a tablet, 3) a smartphone or 4) none. A follow-up question asked if any of these devices were the child's own. We created binary variables for each type of device (including none) and whether the child had ownership of one.

Academic variables

To account for the baseline cognitive or scholastic ability of children, we asked "Which of the following best describes the grades you typically get at school?" Answers were 3 = excellent (49.6%), 2 = good (36.6%), 1= average (11.8%), 0 = poor (2.1%). This variable is entered as a categorical variable in our model. We separately categorized it as binary (1 = good or excellent, 0 = average or poor) to examine if the model results would change. Our model also controls for the type of school the child attends, including public (77%), public charter (8%), private (10%), home school (4%). These choices were coded as binary variables in our model. Since our survey was conducted during the pandemic, we also asked whether students were mostly in-person or learning from home or both. We controlled for whether the student was mostly learning from home (32%). We control for grade level linearly, since we assume knowledge about and access to computer science learning will increase incrementally with level of study. Indeed, this relationship was established in Gallup's preliminary report on the survey (Gallup-Amazon, 2021). Finally, the survey asks students about whether there were computer science learning opportunities in various settings:

- Are there classes that teach computer science **at your school** that you could take if you wanted to?
- Are there any groups or clubs at your school where students learn computer science that you could participate in if you wanted to?
- Are you aware of any **online** computer science classes **not** taught by your school that you could participate in if you wanted to?
- Are you aware of any opportunities **in your community** to learn computer science **outside of school** that you could participate in if you wanted to?
- Are you aware of any **camps** where you could go to learn computer science if you wanted to?

These variables were recoded as "1" if the student has participated in computer science learning in these locations.

Analysis approach

A critical challenge in investigating the influence of learning experience is the self-selection bias. People don't have equal or random chances of enrollment. Exogenous variables, such as parental education and residency area, may confound the potential effect of learning experiences. Traditional regression analysis can hardly tease out these confounding effects, but its estimation may be heavily influenced by extreme values, particularly if (in our case) sample sizes and the distribution of covariates vary greatly between groups (Ho et al., 2007). For this reason, we adopt the multinomial propensity score weighting technique that reduces the self-selection bias by balancing the pre-treatment variables between multiple groups. In our analysis, the above-mentioned pre-treatment covariates were successfully balanced, as shown in the supplementary material.

In the meantime, a traditional regression analysis may be useful as a reference as this reflects what teachers may observe in their classroom without doing the propensity score weighting in their head (a student learned CS out-of-school, do I notice him/her to be more interested in a CS career than others?). Therefore, we also provide this estimation as a means to examine to what extent the raw observation may be validated by a more robust estimation.

To answer our question regarding the mediation effect (learning experience → role-model → CS career interest), we need to adopt a causal mediation modeling approach. Role-model was a post-treatment variable because we conceptualized and modeled it as an outcome of the treatment variable and other covariates. Including role-model as the mediator may reintroduce bias to our estimation of the treatment effect of learning experiences, so we avoid using this procedure to estimate the treatment effect of learning experiences (treatment effects have been robustly estimated above). Rather, our purpose in including the role-model in the equation was to estimate the proportions of the treatment effects that were mediated by role-model. This proportion can be estimated using causal mediation analysis, following the procedure introduced by Imai, et al. (2010a) and Jo et al. (2011).

The procedure is as follows: 1) model the mediator (role-model) as a function of treatment variables (learning experience) and covariates, with propensity score weighting; 2) model the outcome variable (CS career interest) as a function of treatment variable, mediator and covariates, with propensity score weighting; 3) estimate the mediation effect using the joint estimation algorithm proposed by Imai et al. (2010b), p. 4) conduct a sensitivity analysis to examine the robustness of the estimation – to what extent can the estimation withhold the violation of ignorability assumption (treatment is assumed to be ignorable given the pre-treatment variables, and the mediator is assumed to be ignorable given the treatment and pre-treatment variables). This approach avoids violating the assumptions needed to apply propensity score weighting (see explanation in detail from Jo et al. (2011)).

Results

CS career interests

Table 2 shows the coefficients of the logistic regression that predict CS career interests as a function of CS learning experience. In M1.1, no weighting is applied. In this model, we

Table 2. Logistic regression models predicting *CS career interest* as a function of *CS learning experience* (and, for M1.4, *gender*). M1.1 did not apply any weighting; M1.2–1.4 applied multinomial propensity score weighting.

	M1.1				M1.2				M1.3				M1.4			
	β	se	OR		β	se	OR		β	se	OR		β	se	OR	
Intercept	−1.578	0.068	0.206	***	−1.334	0.040	0.264	***	−0.154	0.257	0.858		−0.048	0.258	0.953	
Learning experience (reference group = Did-not-learn)																
Out-of-school	1.049	0.106	2.855	***	0.682	0.052	1.977	***	0.663	0.055	1.940	***	0.541	0.067	1.7184	***
In-school	0.511	0.099	1.668	***	0.326	0.054	1.385	***	0.339	0.056	1.403	***	0.149	0.069	1.160	*
Both	1.323	0.095	3.756	***	0.976	0.052	2.654	***	0.997	0.054	2.710	***	0.897	0.066	2.4512	***
Interaction effect																
Out-of-school × Female													0.379	0.119	1.4608	**
In-school × Female													0.561	0.120	1.753	***
Both × Female													0.321	0.116	1.3788	**
Covariates:																
Female (vs. Male)									−0.761	0.040	0.467	***	−1.096	0.092	0.3343	***
Age									−0.066	0.008	0.936	***	−0.067	0.008	0.9355	***
Black (vs. White)									−0.071	0.067	0.931		−0.076	0.067	0.9265	
Hispanic (vs. White)									0.047	0.052	1.048		0.043	0.052	1.0436	
Asian (vs. White)									0.546	0.113	1.726	***	0.534	0.113	1.7049	***
Other race (vs. White)									−0.138	0.080	0.871		−0.140	0.080	0.8693	
parents have weak IT skills									−0.247	0.031	0.781	***	−0.249	0.031	0.7796	***
parent has BA degree									0.022	0.042	1.022		0.027	0.042	1.0271	
household income									−0.023	0.013	0.977		−0.023	0.013	0.9768	
Has computer or laptop in house									0.258	0.077	1.294	***	0.255	0.077	1.2909	***
Has tablet at home									−0.024	0.039	0.976		−0.025	0.039	0.9756	
Has smartphone at home									0.035	0.045	1.036		0.031	0.045	1.0316	
Lives in large city									0.282	0.051	1.325	***	0.280	0.051	1.3228	***
Lives in suburb									0.041	0.047	1.042		0.040	0.047	1.0406	
public school (vs. private)									−0.317	0.052	0.729	***	−0.317	0.052	0.7283	***
charter school (vs. private)									−0.359	0.086	0.698	***	−0.359	0.086	0.6981	***
other school (vs. private)									−0.003	0.092	0.997		0.000	0.092	1.000	
school grade-average (vs. poor)									−0.274	0.172	0.760		−0.257	0.173	0.7733	
school grade-good (vs. poor)									0.306	0.164	1.357		0.319	0.164	1.3761	
school grade-excellent (vs. poor)									0.325	0.163	1.385	*	0.339	0.164	1.4041	*
internet reliability									0.082	0.033	1.086	*	0.083	0.033	1.0864	*
number of children									−0.113	0.022	0.893	***	−0.113	0.022	0.8934	***
AIC.		4672				18418				17146				17131		

Notes: School grade is students' self-report of their typical grades at school.

found that all three learning experiences had significantly positive effects, compared to the *none* group as reference, on CS career interests. Converting the logit coefficients to odd ratios (by exponentiating the coefficients), we found that students in the *out-of-school* group had 2.85 times the odds of students in the *none* group to report CS career interests ($\beta = 1.05$, se = 0.10, $p < 0.001$), an 185% increase in odds. Students in the *in-school* group had 1.67 times the odds of the students in the *none* group ($\beta = 0.51$, se = 0.10, $p <$ 0.001; 67% increase in odds), and students in the *both* group had 3.76 times the odds of the students in the none group ($\beta = 1.32$, se = 0.09, $p < 0.001$; 276% increase in odds) to report CS career interests.

M1.2 applied multinomial propensity score weighting. As noted above, we also tried a model in which Likert scale variables – parental IT skills, internet reliability and academic performance – were treated as binary variables. This yielded nearly identical estimations to the main effects ($\beta_{\text{out-of-school}}$ changed from 0.68 to 0.71, $\beta_{\text{in-school}}$ changed from 0.32 to 0.31, β_{both} remained to 0.97). Therefore, we kept treating the Likert scale variables as numerical.

M1.3 applied the weighting while controlling for the covariates. Expectedly, the M1.2 and M1.3 yielded similar estimates of the effects of learning experiences because the covariates had been balanced by weighting, and the covariates should minimally confound the treatment variables. Thus, we only report M1.3 here. The null model (a model without predictors) using propensity score weighting yielded an AIC of 18,853. In comparison, M1.3 was a significant improvement.

The estimated effects of learning experiences in M1.3 were positive and significant, although smaller than the estimated effects from M1.1. Compared to *none* group, out-of-school group's odds in reporting CS career interest increased by 94% ($\beta = 0.66$, se = 0.05, $p <$ 0.001, OR = 1.94), in-school group's odds increased by 40% ($\beta = 0.34$, se = 0.05, $p < 0.001$, OR = 1.40), and the *both* group's odds increased by 171% ($\beta = 0.99$, se = 0.05, $p < 0.001$, OR = 2.71). The *out-of-school* group's odds was 1.34 times of the odds of the *in-school* group (1.94/1.40 = 1.34, $p < 0.001$ as tested after switching the reference group). Interestingly, the effect of *both* was nearly equal to the sum of *out-of-school* and *in-school* (on the logit scale), indicating that the effects of *out-of-school* and *in-school* were additive to each other.

M1.3 also showed a significant gender difference: female students' odds of reporting CS career interest was 0.467 times that of the male students' odds (53% decrease in odds, $\beta = -0.76$, se = 0.04, $p < 0.001$, OR = 0.47). Further, in M1.4 where we include gender in the treatment variables, we found significant and positive interaction effects between learning experiences and female (vs. male), meaning the effects of learning experiences were stronger for female than for male students. For male students, *out-of-school* experience increased the odds of CS career interests by 72% compared to *none* ($\beta = 0.54$, se = 0.06, $p < 0.001$, OR = 1.72), *in-school* experience increased the odds by 16% ($\beta = 0.15$, se = 0.07, $p < 0.05$, OR = 1.16), and *both* experiences increased the odds by 145% ($\beta = 0.89$, se = 0.06, $p < 0.001$, OR = 2.45); post-hoc testing showed that all of these three effects were significantly different from each other at $p < 0.001$. For female students, *out-of-school* experience increased the odds of CS career interests by 151% ($\exp(0.54 + 0.38) = 2.51$), *in-school* experience increased the odds by 103% ($\exp(0.15 + 0.56) = 2.03$), and *both* experiences increased the odds by 238% ($\exp(0.89 + 0.32) = 3.38$); post-hoc testing showed that all of the three effects were significantly different

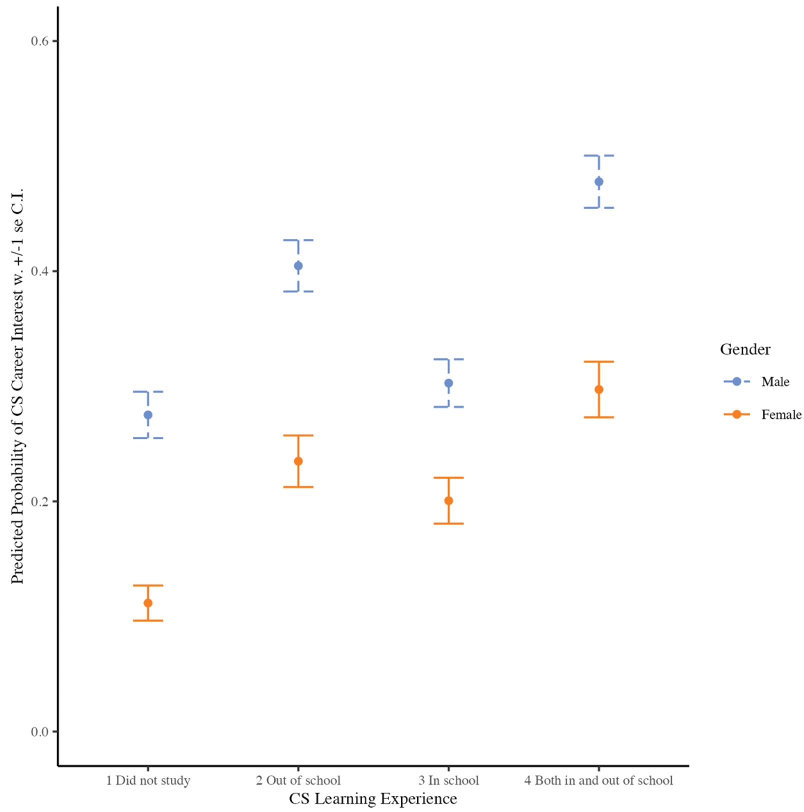

Figure 1. Model 1.4 predicted probability of CS career interests (with ± 1 se confidence interval) as a function of learning experiences × gender, after fixing all other covariates at their sample means.

from each other ($p = 0.014$ between *in-school* and *out-of-school*, and $p < 0.001$ between *both* and *in-school* or *out-of-school*). We show the interaction effects (with 95% confidence interval) in Figure 1 after calculating the predicted probability (converting from logit to the probability scale in M1.4) of reporting CS career interests by learning experiences × gender, after fixing other covariates at their sample means (race was fixed at White because White was the largest category). Female students with no learning experience had the lowest probability in CS career interest (~16%); their male counterparts had nearly twice the probability (~30%). Female students who experienced both out-of-school and in-school CS learning were expected to have the same probability of CS career interests as male students who had in-school experience (~35%). We tested for but did not find any significant interaction effects between the learning experience and race/ethnicity.

Given the striking difference in career interest effects between classroom experiences and out-of-class experiences, we also test whether students rate these classes differently. For each type of CS learning experience, the survey questionnaire asked whether the experience "was fun", whether the student "learned a lot about computer science" and whether "It made me want to learn more about computer science". Responses were collected on a 1–4 scale, where 4 means "strongly agree". We calculated the percentage who strongly agree on each item across the learning experiences.

On each measure, the in-class experience was less likely to be rated high than every other type. For example, 47% of students strongly agreed that the CS class at school was fun. Among the 7 other experiences measured, learning "on your own" or "from a family member or friend" were the next lowest with 55% strongly agreeing. The most fun were "from a tutor or mentor" (64%) and camp (65%). When asked about whether the experience made them want to learn more about CS, in-school performed 12 to 21 ppt lower than other types of experiences. Learning from a tutor or mentor performed the best (Supplemental Table 1).

CS role models

Turning to the role-modeling outcome, Table 3 shows the results of the logistic regressions that predict whether students report having a CS role model as a function of their CS learning experiences. M2.1 does not apply weighting. It shows that *out-of-school* experience increases the odds of having CS role models by 702% ($\beta = 2.08$, se = 0.110, $p < 0.001$, OR = 8.02), and *in-school* increases the odds by 134% ($\beta = 0.85$, se = 0.084, $p < 0.001$, OR = 2.34), and experience *both* increases the odds by 592% ($\beta = 1.93$, se = 0.095, $p < 0.001$, OR = 6.92), compared to students who had *none* of the experiences. We also tried a model that treated parental IT skills, internet reliability, and academic performance as binary variables. This yielded nearly identical estimations to the main effects ($\beta_{out\text{-}of\text{-}school}$ changed from 1.65 to 1.68, $\beta_{in\text{-}school}$ changed from 0.76 to 0.74, β_{both} changed from 0.51 to 0.50). Therefore, we kept treating the Likert scale variables as numerical.

M2.2 and M2.3 applied multinomial propensity score weighting. The null model using propensity score weighting yielded an AIC of 20,890. Thus, M2.2 and M2.3 significantly improved the model quality from the null model. Focusing on M2.3, we find that *out-of-school* experiences increase the odds of having CS role models by 462% ($\beta = 1.73$, se = 0.054, $p < 0.001$, OR = 5.62), *in-school* increases the odds by 131% ($\beta = 0.84$, se = 0.051, $p < 0.001$, OR = 2.31), and experiencing *both* increases the odds by 410% ($\beta = 1.63$, se = 0.05, $p < 0.001$, OR = 5.10), compared to students who had *none* of the experiences. *Out-of-school* group had 2.43 times the odds of the *in-school* group (5.62/2.31 = 2.43, $p < 0.001$). *Out-of-school* group also had similar odds to the *both* group (no statistically significant difference), in having CS role models. Thus, unlike the additive effect of *both* on CS career interest reported above, having both in-school and out-of-school experiences didn't appear to be additive for CS role models. We propose possible explanations in the Discussion.

We do not find significant interaction effects between learning experiences and gender or race.

CS career interest and role model

Finally, we include CS role model in the causal mediation model that uses learning experiences to predict CS career interests. In the baseline model with no weighting, we find that having CS role models is positively associated with CS career

Table 3. Logistic regression models predicting *CS role model* as a function of *CS learning experience*. M2.1 did not apply any weighting; M2.2 and M2.3 applied multinomial propensity score weighting.

	M2.1				M2.2				M2.3			
	β	se	OR		β	se	OR		β	se	OR	
Intercept	−0.850	0.056	0.427	***	−0.681	0.034	0.506	***	0.882	0.244	2.415	***
Learning experience (reference group = Did-not-learn)												
Out-of-school	2.082	0.110	8.023		1.653	0.050	5.221	***	1.726	0.054	5.621	***
In-school	0.850	0.084	2.341	***	0.763	0.047	2.145	***	0.839	0.051	2.314	***
Both	1.935	0.095	6.921	***	1.514	0.049	4.546	***	1.630	0.053	5.103	***
Covariates:												
Female (vs. Male)									−0.108	0.038	0.898	**
Age									−0.094	0.008	0.911	***
Black (vs. White)									−0.164	0.063	0.849	**
Hispanic (vs. White)									0.028	0.052	1.029	
Asian (vs. White)									0.271	0.119	1.311	*
Other race (vs. White)									−0.602	0.074	0.548	***
parents have weak IT skills									−0.223	0.028	0.800	***
parent has BA degree									0.095	0.042	1.100	*
household income									−0.059	0.013	0.943	***
Has computer or laptop in house									0.009	0.073	1.009	
Has tablet at home									−0.232	0.039	0.793	***
Has smartphone at home									−0.269	0.046	0.764	***
Lives in large city									0.721	0.052	2.057	***
Lives in suburb									0.215	0.044	1.240	***
public school (vs. private)									−0.420	0.056	0.657	***
charter school (vs. private)									−0.237	0.087	0.789	**
other school (vs. private)									−0.195	0.095	0.823	*
school grade-average (vs. poor)									−0.417	0.156	0.659	**
school grade-good (vs. poor)									0.079	0.150	1.082	
school grade-excellent (vs. poor)									0.567	0.150	1.763	***
internet reliability									0.061	0.031	1.063	
number of children									0.135	0.022	1.144	***
AIC		4998				19304				15975		

interests ($\beta = 0.96$, $se = 0.08$, $p < 0.001$, $OR = 2.62$). After applying weighting, the effect of CS role model remains similar ($\beta = 0.95$, $se = 0.04$, $p < 0.001$, $OR = 2.58$). Applying causal mediation analysis on the weighted models, we find that, on average, 45% (95% C.I. [28%-80%]) of the total effect of *out-of-school* learning, 56% (95% C.I. [30%-100%]) of the total effect of *in-school*, and 32% (95% C.I. [21%-44%]) of the total effect of both, on CS career interests were mediated by having role models.

Discussion

This analysis makes several contributions to the literature.

First, we estimated that in-school CS learning increases grade 5–12 students' CS career interests by 40% compared to students who do not learn any CS. Out-of-school learning increases students' career interests by 94%. If students learn CS both in and outside of their schools, their CS career interests increase by 171%.

Secondly, we find that the effects of these learning experiences are larger for girls. The in-school effect is only 16% for boys but 103% for girls. The out-of-school effect is 72% for boys versus 151% for girls. Both experiences increase boys' interest by 145% compared to 238% for girls. Given the lower mean level of CS career interest for girls (20% versus 36% for boys), these results suggest formal learning experiences both in and out of school may help close the gender gap.

We do not detect similar interaction effects by race and ethnicity, though race and ethnicity predict participation in learning activities. 43% of Black students and 39% of Hispanic students said they had not studied CS either in school or outside of school. These were significantly higher rates than we found for Asian and White students, which were 33% and 32%, respectively. Taken together, these findings suggest that participation in CS learning opportunities may raise the interests of students equally across racial and ethnic groups, and closing gaps in career participation will require making these opportunities more equitable.

Third, we show that these learning experiences sharply raise the probability of students identifying with CS role-models. For in-school learning, the effect is a 131% increase, for out-of-school, it is a 462% increase, and for both it is a 410% increase in the probability of having a role model. CS role-models, meanwhile, explain approximately 30% to 50% of the overall effect of learning experiences on career interest. We did not detect a significant interaction effect between CS learning experience and gender or race/ethnicity on having CS role models.

Out-of-school vs. in-school

The relative importance of learning outside of school is a noteworthy finding in its own right. We have a few informed speculations as to what may explain this. Self-learning or bootcamp sessions may be oriented around projects or inquiries where creativity, overlap with personal interests, hacking, and encouragement of trial-and-error are more frequent. In comparison, school education is likely to be more formally structured, less individually tailored, and more stressful. This may explain why students were much less likely to rate in-school CS learning as fun, as we show in the supplement, and why in-school learning also performed relatively poorly on whether the students learned a lot of CS or whether it made students want to learn more about CS compared to the learning opportunities outside of school.

Likewise, in-school classes might discourage students from copy-and-paste solutions, despite their real-world utility. Looking up a solution on Stack Overflow may inform learners what professional developers are struggling with, but listening to a lecture about how a function works in the classroom can hardly connect to the endgame (a glimpse at the profession). Out-of-class experiences may more readily preview the end-game and provide more opportunities for authentic coding and learning experiences that connect to the profession, boosting opportunities for self-assessment and confidence-building.

One may worry that learning CS both in and out of school may overwhelm a student's busy schedule, lead to boredom due to oversaturation, or otherwise hamper students' interest in CS careers. However, our data showing the additive effect of the *both* condition on CS career interest helps allay this concern. This additive effect might be due to in-school and out-of-school learning providing non-overlapping curricula and pathways to introduce CS skills and career prospects (as we discussed above). Interestingly, however, this additive effect didn't apply to CS role models – indeed, the odds of having a role model were lower for *both* than for *out-of-school* alone. A possible explanation is that the available role models are limited and overlapping. It may be that the individuals encountered by students involved in both in-school and out-of-school experiences may often have a different inspirational impact than those who students only involved in out-of-school experiences are typically exposed to, e.g., due to perceived differences in career experiences. More research is needed to better understand who are the role models students are citing.

It is noteworthy that "out-of-school" learning entails a wide range of experiences – we don't know to what degree the effects of these experiences may vary, or to what extent our interpretation above applies to all of them. In any case, the lower subjective ratings of in-school CS experiences and the lower effect sizes on building career interest should prompt discussion about how to make the CS in-school experience as inviting as the out-of-school tutoring or camp experiences.

Gender interaction effects

The literature documents a strong gender disparity in CS interests, enrollment, and persistence (Margolis et al., 2017; Sax et al., 2017). In a national sample study of college students in the U.S., authors (2020) showed that only 27% of enrollees in college level Introduction to CS (e.g. CS50) were women. In the same study, authors (2020) showed that high school CS course enrollment had positive impact on college CS grades, but didn't interact with gender. Consistent with the literature, in this study, we also find women students to report less interest in CS and fewer CS role models. However, different from our previous findings, here we detect a gender interaction effect with all three types of pre-college CS learning experiences, meaning the positive influence of these experience on CS careers interests are more pronounced for women than for men. The previous study couldn't detect this interaction effect perhaps due to its "survivor bias" by only sampling the college students in CS50s and collecting their high school information retrospectively. In this study, when we ask grade 5–12 students to envision their future occupations, we do see pre-college CS experience giving an extra boost to women students compared to men, and the boosted amount coming highest from in-school learning experiences. Research shows that men and women tend to develop a narrow and stereotypical view of STEM and who can succeed in STEM by the time they enter college (Kahle & Lakes, 1983; Potvin & Hasni, 2014; Tytler, 2014). Women who did choose a career in STEM have often fostered such an interest pre-college (Hazari et al., 2010; Ivie & Guo, 2006; Ivie et al., 2002). Our results suggest that learning CS before college may remediate some of the gender stereotypes in computing that pose particular threats to women's CS identity. Women benefited even more from in-school learning, possibly due to a more gender balanced environment in school as compared to the male-dominated environment in out-of-school CS learning facilities (Author, 2019; Bettinger & Long, 2005; Miller et al., 2015).

Role-model as a meditator

About half of the effects of either in-school or out-of-school were mediated by CS role models. From an EVT perspective, this shows that role models may provide important information about what success may look like in the CS profession. In particular, if the students see the resemblance between themselves and their role models, they may develop affirmative affinities to the CS career. There might be numerous routes for in-school or out-of-school learning to have impact on students' career interests, such as content knowledge, career knowledge, real-life relevance, or fun factors. Yet, role models explained half of all effects, meaning it is likely to be the most influential route toward CS career affinities.

When students learn CS out-of-school, such as online or in a boot camp, they may have lots of opportunities to interact with and learn from CS professionals. Many boot camps are organized or supported by IT companies; most respondents to coding questions posted online are advanced programmers. Students may indirectly yet frequently communicate with computer scientists even if students are learning CS completely by themselves – almost every package that a learner loads for free was developed by an altruistic CS expert. As students have more exposure to the professionals in the industry (this can be done in-school but typically requires teachers to reach out to out-of-school resources), they may identify role models who can share insider perspectives.

In our literature review, we anticipated that in-school learning experience may also connect the classroom with the role models from the outside world (such as via CS Journeys from Code.org, or Class Chats from Amazon Future Engineer). However, our results show that in-school learning predicts less likelihood in having CS role models than out-of-school learning. This suggests that students' exposure to role models in schools, compared to out-of-school, may be more confined to the figures in the textbooks or their classroom teachers, and less likely to be expanded to real-world professionals. We don't know to what extent the role models from the two channels overlap.

Interestingly, the proportion that role models explain the effect of the *both* condition was smaller than that of either in-class or out-of-class. One possible explanation is that the role models from the two channels may overlap. Repeating the story of the same (so similar) role models cannot add up the influence of role modeling. Another possible explanation is that when students are only learning through one channel, they rely strongly on the support and encouragement from external role model(s) – perhaps some influential figures who inspire them to persevere. In comparison, students who have learned through both channels may have learned CS so intensively and been exposed to role models so frequently that they have internalized the messages from the role models and don't need to rely on a specific person of inspiration.

Much of our abovementioned interpretations are based on speculation. More research is needed to investigate *how* role models mediate the effects of CS experiences, the characteristics of the role models, and the natures of the interactions with specific learning experiences.

Limitations

This research is confronted by several limitations. Access to computer science learning opportunities in and out of school is not random, and while our matching method explicitly mitigates selection bias, it is possible that unobservable factors – such as parental occupation – bias our analysis, nonetheless. Additionally, selection bias may be more pronounced for out-of-school learning opportunities, which are likely to be more self- or parent-initiated than in-class learning. We do not have pre-class measures of career interest, which would be needed to see how these learning experiences affected interest above the baseline. We also do not know who the role models are in the students' minds. Are they teachers, family members, computer scientists in history or modern times, or entrepreneurs in tech companies? Future studies should collect detailed information about role models, as such information may help us anticipate students' modeling behavior. Some variables were difficult to measure precisely, such as household income that has a wide range in each bin, or students' self-reported school grades that might be subjective. These imprecisions might misalign participants when we apply the weighting algorithms to balance between groups. Finally, our dependent variable is limited to career interest. It is not established how strongly expressions of interest in secondary or elementary school predict whether someone actually majors in computer science or pursues a career in the field. A longitudinal research design featuring random assignment of exposure to computer classes starting in childhood and following students into early adulthood would be needed to address these concerns directly.

Conclusion

In this study, we provide estimates of the effect of learning opportunities for U.S. students in grades 5 through 12 on interest in a computer science career. Our method explicitly considers that family socio-economic status, student academic performance, demographics, parental computer use, and other factors affect the probability of gaining access to computer science learning opportunities. These results are encouraging for those who would like to bolster the CS workforce through both school and non-school opportunities in that these experiences predict higher interest in a computer science career independently (40% higher interest for in-school and 94% for out-of-school) and even more strongly in combination (a 171% increase), with a significant portion of the effect mediated by providing the students with CS role models for the students.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by Amazon Future Engineer, Amazon's global philanthropic computer science education program.

ORCID

Chen Chen http://orcid.org/0000-0002-6065-8889

Data availability statement

Data is available upon request.

References

Amazon Future Engineer. (2021). Amazon employees share their stories to spark student curiosity in tech. [Online]. Available: https://www.aboutamazon.com/news/community/amazon-employees-share-their-stories-to-spark-student-curiosity-in-tech

Armoni, M., & Gal-Ezer, J. (2014). High school computer science education paves the way for higher education: The Israeli case. *Computer Science Education*, *24*(2–3), 101–122. https://doi.org/10.1080/08993408.2014.936655

Ashcraft, C., Eger, E., & Friend, M. (2012). *Girls in IT: The facts*. National Center for Women & IT. Boulder, CO.

Ball, C., Huang, K. T., Cotten, S. R., & Rikard, R. V. (2017). Pressurizing the STEM pipeline: An expectancy-value theory analysis of youths' STEM attitudes. *Journal of Science Education and Technology*, *26*(4), 372–382. https://doi.org/10.1007/s10956-017-9685-1

Beaubouef, T., & Mason, J. (2005). Why the high attrition rate for computer science students: Some thoughts and observations. *ACM SIGCSE Bulletin*, *37*(2), 103–106. https://doi.org/10.1145/1083431.1083474

Bettinger, E. P., & Long, B. T. (2005). Do faculty serve as role models? The impact of instructor gender on female students. *American Economic Review*, *95*(2), 152–157. https://doi.org/10.1257/000282805774670149

Betz, M., & O'Connell, L. (1992). The role of inside and same-sex influencers in the choice of nontraditional occupations. *Sociological inquiry*, *62*(1), 98–106.

Braswell, K., & Rankin, Y. A. (2023). Intech: Designing intersectional learning experiences for black girls. *Interactions*, *30*(1), 66–69. https://doi.org/10.1145/3575868

Clevey, L. (2018). *Comparing the impacts of mentors vs. role models across domains: A meta-analysis*. Eastern Michigan University.

Codding, D., Alkhateeb, B., Mouza, C., & Pollock, L. (2021). From professional development to pedagogy: Examining how computer science teachers conceptualize and apply culturally responsive pedagogy. *Journal of Technology & Teacher Education*, *29*(4), 497–532.

Code.org. (2022). Teachers guide: Engaging your students in CS Journeys. [Online]. Available: https://code.org/files/csjourneys/csjourneys-teacherguide.pdf

Decker, A., & McGill, M. M. (2019). A systematic review exploring the differences in reported data for pre-college educational activities for computer science, engineering, and other STEM disciplines. *Education Sciences*, *9*(2), 69.

DeLyser, L. A. (2018, July). A community model of csforall: Analysis of community commitments for cs education. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*, Larnaca Cyprus, (pp. 99–104).

Diekman, A. B., Steinberg, M., Brown, E. R., Belanger, A. L., & Clark, E. K. (2017). A goal congruity model of role entry, engagement, and exit: Understanding communal goal processes in STEM gender gaps. *Personality and Social Psychology Review*, *21*(2), 142–175. https://doi.org/10.1177/1088868316642141

Dweck, C. S. (2007). Is math a gift? Beliefs that put females at risk. In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science?: Top researchers debate the evidence* (pp. 47–55). American Psychological Association. https://doi.org/10.1037/11546-004

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, *61*, 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

Eglash, R., Gilbert, J. E., & Foster, E. (2013). Toward culturally responsive computing education. *Communications of the ACM*, *56*(7), 33–36. https://doi.org/10.1145/2483852.2483864

Gallup-Amazon. (2021). Developing careers of the Future: A study of student access to, and interest in, computer science, https://www.gallup.com/analytics/354740/study-of-student-interest-in-computer-science.aspx

Gartzia, L., Morgenroth, T., Ryan, M. K., & Peters, K. (2021). Testing the motivational effects of attainable role models: Field and experimental evidence. *Journal of Theoretical Social Psychology*, *5*(4), 591–602. https://doi.org/10.1002/jts5.121

Hazari, Z., Sonnert, G., Sadler, P. M., & Shanahan, M. C. (2010). Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study. *Journal of Research in Science Teaching*, *47*(8), 978–1003. https://doi.org/10.1002/tea.20363

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*(3), 199–236. https://doi.org/10.1093/pan/mpl013

Imai, K., Keele, L., & Tingley, D. (2010b). A general approach to causal mediation analysis. *Psychological Methods*, *15*(4), 309–334. https://doi.org/10.1037/a0020761

Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2010a). Causal mediation analysis using R. In H. D. Vinod (Ed.), *Advances in social science research using R* (pp. 129–154). Springer.

Ito, M., Baumer, S., Bittanti, M., Cody, R., Stephenson, B. H., Horst, H. A., & Perkel, D. (2009). *Hanging out, messing around, and geeking out: Kids living and learning with new media*. MIT Press.

Ivie, R., Czujko, R., & Stowe, K. (2002). Women physicists speak: The 2001 international study of women in physics. AIP Conference Proceedings, Paris (France), 628(1), 49–70. AIP.

Ivie, R., & Guo, S. (2006). Women physicists speak again. American Institute of Physics. Retrieved from http://www.aip.org/statistics/trend/report/iupap05.pdf

Jo, B., Stuart, E. A., MacKinnon, D. P., & Vinokur, A. D. (2011). The use of propensity scores in mediation analysis. *Multivariate Behavioral Research*, *46*(3), 425–452. https://doi.org/10.1080/00273171.2011.576624

Kahle, J. B., & Lakes, M. K. (1983). The myth of equality in science classrooms. *Journal of Research in Science Teaching*, *20*(2), 131–140. https://doi.org/10.1002/tea.3660200205

Kearney, M. S., & Levine, P. B. (2020). Role models, mentors, and media Influences. *The Future of Children*, *30*(1), 83–106. https://doi.org/10.1353/foc.2020.0006

Kölling, M. (1999). The problem of teaching object-oriented programming. *Journal of Object Oriented Programming*, *11*(8), 8–15.

Liggett, J. B. (2014). *Geek as a constructed identity and a crucial component of STEM persistence*. [Master of Science thesis]. University of North Texas.

Lockwood, P. (2006). Someone like me can be successful": Do college students need same-gender role models? *Psychology of Women Quarterly*, *30*(1), 36–46. https://doi.org/10.1111/j.1471-6402.2006.00260.x

Mani, A., & Riley, E. (2019). Social networks, role models, peer effects, and aspirations, WIDER Working Paper, No. 2019/120, ISBN 978-92-9256-756-9, The United Nations University World Institute for Development Economics Research (UNU-WIDER), Helsinki. https://doi.org/10.35188/UNU-WIDER/2019/756-9

Margolis, J., Estrella, R., Goode, J., Holme, J. J., & Nao, K. (2017). *Stuck in the shallow end: Education, race, and computing*. The MIT Press.

Marken, S., Crabtree, S. (2021). Role models spark students' interest in computer science. Retrieved from https://news.gallup.com/poll/355070/role-models-spark-students-interest-computer-science.aspx

McKenna, B. W., & Bergie, L. (2016). Creating the next generation of innovators. *Publications & Research* Paper 6. http://digitalcommons.imsa.edu/stratinnov_pr/6.

Miller, D. I., Eagly, A. H., & Linn, M. C. (2015). Women's representation in science predicts national gender-science stereotypes: Evidence from 66 nations. *Journal of Educational Psychology*, *107*(3), 631–644. https://doi.org/10.1037/edu0000005

Morgenroth, T., Ryan, M. K., & Peters, K. (2015). The motivational theory of role modeling: How role models influence role aspirants' goals. *Review of General Psychology*, *19*(4), 465–483. https://doi.org/10.1037/gpr0000059

Ni, L., Tian, Y., McKlin, T., & Baskin, J. (2023). *Who is teaching computer science? Understanding professional identity of American computer science teachers through a national survey*. Computer Science Education.

Potvin, P., & Hasni, A. (2014). Interest, motivation and attitude towards science and technology at K-12 levels: A systematic review of 12 years of educational research. *Studies in Science Education*, *50*(1), 85–129. https://doi.org/10.1080/03057267.2014.881626

Ruggles, S., Flood, S., Goeken, R., Schouweiler, M., & Sobek, M. (2022). *IPUMS USA: Version 12.0 2020 American community survey*. IPUMS. https://doi.org/10.18128/D010.V12.0

Sax, L. J., Lehman, K. J., Jacobs, J. A., Kanny, M. A., Lim, G., Monje-Paulson, L., & Zimmerman, H. B. (2017). Anatomy of an enduring gender gap: The evolution of women's participation in computer science. *The Journal of Higher Education*, *88*(2), 258–293. https://doi.org/10.1080/00221546.2016.1257306

Stack Overflow. (2021) Stack Overflow developers survey 2021. [Online]. Available: https://insights.stackoverflow.com/survey/2021

Stehlik, M., Cawley, E., & Kosbie, D. (2020, February). Cmu cs academy: A browser-based, text-based introduction to programming through graphics and animations in python. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, USA, (pp. 1420–1420).

Stern, J., Reid, E., & Bancroft, K. (2015, February). Teaching introductory computer science for a diverse student body: Girls who code style. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*, Kansas, Missouri, USA, (pp. 705–705).

Sto. Domingo, M. R., Sharp, S., Freeman, A., Freeman, Jr, T., Harmon, K., Wiggs, M., Sathy, V., Panter, A.T., Oseguera, L., Sun, S., & Williams, M.E. (2019). Replicating Meyerhoff for inclusive excellence in STEM. *Science*, *364*(6438), 335–337.

Tellhed, U., Bäckström, M., & Björklund, F. (2018). The role of ability beliefs and agentic vs. communal career goals in adolescents' first educational choice. What explains the degree of gender-balance ? *Journal of Vocational Behavior*, *104*(1), 1–13. https://doi.org/10.1016/j.jvb.2017.09.008

Twarek, B. (2018). A call to celebrate diversity in computer science. Retrived from https://advocate.csteachers.org/tag/role-models/

Tytler, R. (2014). Attitudes, identity, and aspirations toward science. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of research on science education volume II* (pp. 82–102). Taylor & Francis.

Upadhyaya, B., McGill, M. M., & Decker, A. (2020, February). A longitudinal analysis of k-12 computing education research in the United States: Implications and recommendations for change. In *Proceedings of the 51st ACM technical symposium on computer science education*, USA, (pp. 605–611).

Wang, M. T., & Degol, J. (2013). Motivational pathways to STEM career choices: Using expectancy–value perspective to understand individual and gender differences in STEM fields. *Developmental Review*, *33*(4), 304–340. https://doi.org/10.1016/j.dr.2013.08.001

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review*, *6*(1), 49–78. https://doi.org/10.1007/BF02209024