

Predicting Women's Persistence in Computer Science- and Technology-Related Majors from High School to College

TIMOTHY J. WESTON University of Colorado, Boulder

WENDY M. DUBOW University of Colorado, Boulder

ALEXIS KAMINSKY, Kaminsky Consulting, LLC.

While demand for computer science and information technology skills grows, the proportion of women entering computer science (CS) fields has declined. One critical juncture is the transition from high school to college. In our study, we examined factors predicting college persistence in computer science and technology-related majors from data collected from female high school students. We fielded a survey that asked about students' interest and confidence in computing as well as their intentions to learn programming, game design, or invent new technology. The survey also asked about perceived social support from friends and family for pursuing computing as well as experiences with computing, including the CS Advanced Placement (AP) exam, out-of-school time activities such as clubs, and internships. Multinomial regression was used to predict persistence in computing and tech majors in college. Programming during high school, taking the CS Advanced Placement exam and participation in the Aspirations awards program were the best predictors of persistence three years after the high school survey in both CS and other technology-related majors. Participation in tech-related work, internships or after school programs was negatively associated with persistence, and involvement with computing sub-domains of game design and inventing new applications were not associated with persistence. Our results suggest that efforts to broaden participation in computing should emphasize education in computer programming.

CCS Descriptors:

- Social and professional topics ~Computing education

Additional Key Words and Phrases: survey research, diversity, broadening participation, women and computing.

ACM Reference format:

1. INTRODUCTION

According to the US Bureau of Labor Statistics, computer science and information technology jobs are likely to grow 13% between 2016 and 2026 [1]. While the increasing demand for such workers has been met by greater enrollment of students into computer science majors, women as a proportion to all students in the US receiving computer science (CS) related bachelor's degrees has (with some fluctuations) declined over the past three decades from 37% in the 1984-85 academic year to 19% in 2017 [2]. A critical step in pursuing a technical occupation is college preparation in related fields. Women in the US receive bachelor's degrees in engineering and computer science at lower rates than men, largely due to women entering these undergraduate programs in much lower numbers rather than because of differences in attrition by gender from degree programs [3 - 4]. Learning more about whether or not students transition from a high school interest in computing to actually majoring in computer science and other tech-related disciplines in college provides insight into what types of experiences in high school support women pursuing these majors.

2. LITERATURE ON CS PERSISTENCE

2.1 General factors associated with underrepresentation of women in CS

In the last decade, significant resources have been applied to understanding and addressing problems with the recruitment and retention of young women in computer science in the US. Researchers have identified a host of contextual factors and perception-based variables that seem to influence young women's persistence in computer-related fields [5 – 7]. Contextual obstacles to women's participation identified in the research include: limited access to high quality computing experiences; lack of early exposure, including opportunities at home and out of school; implicit and explicit stereotyping by teachers and guidance counselors; and media representations about who "belongs" in computing [7-9]. The roots of underrepresentation are social and cultural; in some countries outside the US women are not a minority in computer-related work [10], and women have seen greater representation in both CS education and the CS workplace in the past than is present today [11].

Along with other internal and external constraints, some US high school girls may have negative perceptions of the personality traits associated with computing that could steer them away from CS and other tech-related fields. These perceptions may act as an initial barrier which then prevents them from gaining experience with coding in high school. Cheryan, Master, and Meltzoff [12], Rommes, Oerbeek, Scholte, Engles and Decamp [13], and Miller [14] all described the negative stereotype of some STEM fields as work pursued by socially awkward males. Although some high school girls embrace (often ironically) the "geek" stereotype, others shy away from associating with this image. One related study found that students often choose gender-stereotyped careers -- interaction with people (for females) and work with "things, machines and tools" for males [15]. Another study by Eccles and Wang [16] showed that those high school students who were working in STEM careers ten years later tended to "value things over people" (p.104), a trait strongly associated in the study with being male. Moreover, the same interest predicted STEM careers in computer science, mathematics, engineering and physics, while personality traits linked to altruism and working with people predicted STEM careers in health-care and life sciences. The latter had a strong association in the study with being female. The authors discuss the role of stereotypes linking gendered traits to specific career paths as culturally determined and, thus, alterable through educational practices and interventions.

While less is known about the availability of adequate preparatory K-12 computing experiences in the US, a nationwide survey conducted by Gallup and commissioned by Google that was published in 2016 reported that 76% of K-12 principals reported having CS-related classes or clubs at their schools and 60% indicated having at least one CS class [17]. Although these statistics suggest that many young people may have access to some sort of computing experiences, it begs the question of whether all computing experiences are equally valuable. CS-related classes and clubs can, of course, have widely varying degrees of rigor when it comes to computer science content. The study also does not shed light on which students avail themselves of non-mandatory computing classes and clubs. Typically, extra-curricular or elective K-12 classes in computer sciences and related topics tend to be populated by male students [18].

2.2 Predictive and concurrent studies on persistence

Research on high school and early college persistence in computer science and technical majors often involves studies which either ask for retrospective assessments of high school experiences, or track persistence over a short period of time at the university. Fewer studies are truly longitudinal and track students from high school to college.

Sax et al. [11] conducted the largest study to date examining determinants of students intending to major in computer science, and how determinants changed over time. The researchers used over eight million records from the Cooperative Institutional Research Program (CIRP) Freshman Survey which examines the majors students planned to enter as of their first year, and asks for retrospective information about experiences in high school. This version of the survey did not track actual longitudinal persistence. Determinants for women across years included strong positive associations with self-ratings of math ability and aspirations to make scientific contributions to the computer science field; planning to major in CS was associated negatively with social activism and with leadership. These factors were also present for men with some differences in the magnitude of predictors. Trends across years (1976 – 2011) pointed to smaller differences in self-ratings of mathematical ability between genders, while both men and women with orientations toward leadership and

raising families were less inclined to pursue CS degrees over the years. The results of this study point to a link between self-reported mathematical ability in high school and becoming a CS major.

Wang, Hong, Ravitz and Ivory [19] surveyed Juniors and Senior in high school, and recent college graduates. Researchers conducted factor analysis to learn which factors influence students' early decisions to pursue computer science. Prominent factors influencing decisions included social encouragement from family, friends, and peers (accounting for 28% of variability), with an emphasis on participation in extracurricular activities around computing and the influence of peers and friends. Almost as influential were career perceptions (27.5%), with both males and females pursuing CS careers less likely to show an interest in careers with social impact. Academic exposure was associated with high school girls pursuing a CS career with girls who took, or had access to an AP course in CS in high school more likely to consider a CS degree. The effect was more pronounced among girls than boys. Confidence in one's math ability was also associated with wanting to pursue a CS degree for both boys and girls in the study, accounting for 17% of variance. While both high school students and recent college graduates were surveyed, the study did not track any of the students over time.

In a recent study, Chen, Jeckel, Sonnert and Sadler [20] surveyed a large stratified random sample of students at US colleges and universities asking them (retrospectively) about their pre-college computing experience. Hierarchical regression models predicted grades in a CS course, attitudes toward computing and if students dropped out of the CS course; all models included demographic and other covariate information. The researchers were specifically interested in "cowhand" programming, or students who learned programming on their own in high school, or those who learned on their own and in conjunction with a formal class. The analysis found that those students who did learn programming on their own (both by themselves or in conjunction with a class) had both higher grades and a more positive attitude toward computing. Additionally, those who learned strictly by themselves dropped out of CS classes at lower rates. The research suggests that self-directed learning in programming is an indicator of greater success and persistence in CS.

In an early study conducted in 1994, Taylor and Mounfield [18] examined the relationship between prior computing experience and academic success in college-level computing courses among non-computer science majors at one US university. They found that prior computing experience—having a computer at home, taking a high school programming course, having any computing experience, or taking a general computing course—positively and significantly correlated with academic success in the college computer course for females. Only high school programming courses and owning a computer were significantly associated with grades for males. The authors conclude that pre-college computing experiences *of any type* position females to be successful in college computing courses. Another early study conducted by the College Board found that students who took an advanced placement exam in a discipline were more likely to major in that field in college than those who did not [22]. However, they did not specifically study the effect of the computer science AP exam, and the study was conducted prior to the CS AP exam's radical transformation as a result of the CS Principles movement.

Barker, McDowell and Kalahar [21] used the Input-Environment-Output model to investigate factors that influenced persistence in computer science majors at one university. They surveyed 113 students in an introductory course in computer science. Using regression analysis, they found that prior programming experience, perceptions of the speed and workload of the course, and being male all predicted persistence, while issues with the students' current course including bad grades, relationships between students, relationships with faculty, and certain pedagogical practices all discouraged persistence. Similar to Taylor and Mounfield [18], prior programming experience in high school emerged as critical to persisting in a computer science major. These studies also suggest that prior experience affects perceptions of pace and workload and that students without previous experience were more likely to drop out of computing.

In a 2011 study investigating the motivation to study computing in college, Trevisan, McKlin, and Guzdial [22] asked 1434 students at Georgia colleges and universities taking Introduction to Computer Sciences classes if there was one experience that led them to computer sciences and if so, what. Slightly less than 18% responded affirmatively. Responses clustered into five categories listed in order of frequency: (1) social influence such as parents, friends, and early access (31%); (2) special event such as a specific teacher or class, a summer program or work experience or programming language (31%); (3) game play (18%); (4) outcome expectations around earning potential and prestige (14%); and (5) innate curiosity (6%). Although informative, this study did not investigate the relative impact of different types of computing experiences, nor does it shed light on the ways that multiple experiences inform students' choices to be in computer science.

Katz et al. [23] tracked students at one institution through three required gateway Computer Science courses. Persistence through all three courses was predicted from expected variables such as grades, standardized math scores and previous math credits. The study found that men with the same grades as women

had greater rates of persistence. For women, prior experience with programming was unexpectedly associated with lower grades in the CS courses, while the opposite was true for men. To better understand these counter-intuitive findings, the authors performed an ANOVA to determine the influence of prior computing courses and calculus courses on achievement in the first of the three required CS courses. They found that women who had taken more calculus courses did better in the CS course than those who had reported having more computing exposure prior to the college courses. The authors conclude that women who report having more programming exposure may gain this experience at the expense of developing other important skills, calculus in particular. That said, the authors acknowledge the small sample size (women comprised only 17.5% of the 200 person sample) and the need for further research to better understand the complex interactions and unexpected associations they found.

Other studies have longitudinally tracked student persistence into engineering or STEM majors (versus CS). Lent et al. [24] followed students for two years, comparing survey responses from engineering students during the first year of college with persistence at the end of the third year. Using Social Cognitive Career Theory, Lent concluded that satisfaction with major, social support and self-efficacy all significantly predicted persistence, while SAT mathematics scores, outcome expectations and interests were linked to persistence through other intervening variables. Maltese and Tai [25] used national survey data to learn which high school factors predicted who received bachelor's degrees in STEM disciplines. These researchers found that students made the choice to pursue careers in science and engineering if they made their career choice early and were interested in math and science during high school. Whether or not a student took an advanced placement course did not seem to be associated with greater persistence. Ellis, Fosdick and Rasmussen [26] studied students through the calculus sequence of courses fundamental to persistence in STEM majors and found that after controlling for academic preparedness and career intentions, women were 1.5 times more likely to leave the sequence than men. The authors concluded that a measure indicating lower mathematical confidence for women was the main predictor of lower rates of persistence for otherwise higher performing women. Larson, et. al [27] also found that self-efficacy in mathematics predicted four- and eight-year graduation status in STEM disciplines after controlling for prior performance and math aptitude.

3. RATIONALE FOR STUDY

The predictive studies in the literature point to a wide range of factors that can predict persistence. Experience with computers and exposure to programming and other technical instruction are seen in a number of studies as important predicates to entering and persisting in CS in both college and the workplace. Higher traditional academic performance in mathematics and computing in high school also play an important part, as do the social support and encouragement of teachers, family and peers. Persistence for women in science and engineering is also tied to self-perceptions of efficacy and confidence which appear to be separate from actual ability.

While important research has been conducted about students transitioning from high school to college in CS and in other STEM disciplines, few studies have tracked student interests from high school to college major, and none have done so with a focus on gender. We believe that a longitudinal study following girls who are interested in computing in high school through to their college choices fills an important gap in the literature. Learning about women students' interests during high school and how these translate into college outcomes can help educators in both secondary and post-secondary contexts gauge which interventions and curricular choices may be the most effective to broaden participation in computing. While much attention is being given to providing exposure to various aspects of computer sciences for groups historically underrepresented in computing, these efforts vary widely and there is no clear consensus among practitioners of interventions which intervention characteristics are necessary to actually produce computer scientists from diverse backgrounds in the future.

In our study, we have followed a group of women students across three years to help us understand the following question:

RQ. Which variables related to high school experience with computing predict high school women's persistence in computer science and technology-related majors three years later?

4. METHODS

The findings we report on here are part of a larger research project that has a goal of understanding the formative experiences that contribute to young women pursuing and persisting in computing despite the many obstacles they face. To this end, our study includes matched comparison groups and gathers data via surveys, interviews, focus groups, social media, and other artifacts. This paper focuses on the survey component of the study as well as one other measure, applications submitted by a sub-sample in high school (“application data”).

4.1 Survey Respondents

The full sample for this study was drawn from a database of young women who had won the Aspirations in Computing Award in 2007-2008, or after the program went online, who had registered on the award website between 2009-2013. The dataset includes award winners and runner-ups, girls that applied but did not win, and girls who registered on the award’s website but did not submit an application. All students are from the United States, including the US territories of Puerto Rico and Guam. It should also be emphasized that the makeup of this sample fundamentally limits the kind of students to which we can generalize our findings to, specifically girls who are initially interested in some aspect of computing during high school. This sample was intentionally chosen so we could learn more about the population of girls who are most likely to pursue computing after high school, and thus are theoretically the most reachable population for expanding the number of women pursuing computer science related fields. The women in our sample differ from the general population of high school students in the US at least by CS Advanced Placement exam participation. While only a small fraction of female high school students in the United States take the Computer Science Advanced Placement exam, 42% of students in our sample have done so. The survey sample is therefore open to bias from self-selection and is not representative of high school girls in general, but rather of college-bound girls who have some interest in computing. Although the makeup of this sample limits the kind of students to which we can generalize our findings to (specifically girls who are initially interested in some aspect of computing during high school), it was chosen intentionally to learn more about the population of girls who are most likely to pursue computing after high school, and thus are theoretically the most reachable population for expanding the number of women pursuing computer science related fields. We also know that our group is higher achieving than girls in general given that they were college bound.

Survey respondents included 494 females who identified themselves as high school students in the first survey administered in spring 2013. All 494 students took both 2013 and 2016 surveys. By the second survey in spring 2016, most had graduated from high school and entered college. Although all respondents were female, they were diverse in race and ethnicity. Because we know that there are differentiated experiences for students of color in secondary and post-secondary education in the US, and especially women of color, we wanted to make sure we captured any differences in outcomes in our analysis. To do so we created a variable called Under-represented Minority in Computing (URMC) status that grouped students by race/ethnicity. URMC indicated persons from groups historically under-represented in computing--African-American, Hispanic, or Native American. White, Asian and students of two or more races were coded as “Majority” in this variable. Unfortunately, further disaggregation by specific race/ethnicity was not possible due to low numbers. Thus, even though the numbers in the respondent pool were not high enough to disaggregate by specific race/ethnicity, we could still identify trends by over-representation and under-representation.

4.2 Survey development

The structural validity of the survey used in the study was validated using confirmatory factor analyses after identifying underlying factors using Exploratory Factor Analysis (EFA) in the first survey. Using EFA we found that items from the survey that referenced the same activities and skills formed better factors than grouping related to interest, confidence or other affective domains. This method also tests the commonality of terms found in different items such as “computer science,” “software,” and “programming.” As a rule of thumb, if items load on the same factor this suggests that respondents find common meaning between items. We used the software AMOS 24 to conduct Confirmatory Factor Analysis (CFA). CFA is a statistical method that tests how well different models fit empirical survey data. Fit indices (the statistics that assess how well the model fits the data) included Chi-square (χ^2), RMSEA, CFI, and SRMA indices, and followed best practices for standards of model fit [27]. We used Maximum Likelihood estimation procedures to calculate our factor analysis. We

checked the adequacy and assumptions of the data for modeling and found 1) an acceptably high ratio of persons to parameter for the survey (14:1), 2) univariate and multivariate normality were within acceptable ranges for skewness and kurtosis for individual variables and multivariate normality and, 3) the reliability of composites were adequate to good, ranging from .77 (gaming) to .85 (programming). These tests indicate that the assumptions for using factor analysis were achieved. The survey met or nearly met accepted model fit standards with $\chi^2 = 242$, $df = 84$, $CMIN/df = 2.9$, $CFI = .95$, $RMSEA = .061$ and $SRMR = .067$. Individual items showed high factor loadings for all variables greater than .5 except for a slightly lower path for one of the *Social Support* items, "*My friends think it is cool that I learn about technology*" ($FL = .47$). Correlations between factors were relatively low ($r = .3$ to $.4$), bolstering the divergent validity of this model, meaning that separate factors most likely stand for different and separate constructs. Table 1 shows the component items of each composite variable, the means and standard deviations of each scale, the factor loadings of each item, and the composite variable's internal reliability.

Table 1. Items, mean/SD, factor loading (FL) and reliability for survey (α).

<i>Question</i>	<i>Mean/SD</i>	<i>FL</i>	<i>α</i>
Programming			
How interested are you in solving software problems?	2.8 (.8)	.63	.85
How interested are you in programming computers or other technologies (in other words writing code)?	3.2 (.9)	.82	
How much do you want to learn how to program computers or other technologies?	3.3 (.8)	.89	
How much do you want to learn computer science?	3.35 (.9)	.83	
Game Design			
How interested are you in playing computer games?	2.7 (1)	.55	.78
How interested are you in designing computer games?	2.6 (.9)	.89	
How much do you want to learn about computer game design?	2.85 (1)	.79	
How confident are you in your ability to design computer games?	1.8 (.85)	(.52)	
Inventing Applications			
How confident are you in your ability to think of new technology inventions?	2.55 (.9)	.83	.76
How confident are you in your ability to actually create new technology inventions?	1.9 (.9)	.66	
How interested are you in actually creating new technology inventions?	3.1 (.9)	.59	
How much do you want to learn how to create new technology inventions?	3.3 (.8)	(.75)	
Social Support			
Important people in my life think it is good for me to learn about technology.	3.5 (.6)	.86	.76
I believe people like me can do well learning technology.	3.6 (.5)	.58	
My family likes me to learn about technology.	3.5 (.6)	.84	
Other students generally think it is cool that I learn about technology.	3.1 (.75)	.47	

We used responses to two items from the 2016 survey (also present in the 2013 version) to create the dependent variable in our analysis, CS Persistence and Tech-related Persistence. Three open-ended questions in the 2016 survey asked for college major and minor, or if graduated, the type of undergraduate degree (e.g., BS in computer science). This variable was then subdivided into three distinct and non-overlapping groups based on college major/minor and degree: (1) *CS-Persister* - those who were pursuing, or who had graduated, at the time of the second survey with a Computer Science or Computer Engineering degree, (2) *Tech-Persister* - anyone who was pursuing an information technology-related major (other than CS or Computer Engineering) in college, or who graduated with a degree in these fields or minored in one of these fields, and (3) *Non-*

Persisters - all remaining respondents. The Tech-persister major included a range of non-computer science majors including informatics, information science, electronic media, cybersecurity, data science, and mechanical engineering. Figure 1 shows the three groups in the outcome variable and numbers and a short description for each.

Fig. 1. Three (nonoverlapping) groups in the dependent (outcome) variable.

CS-persister	Tech-persister	Non-persister
N = 177	N = 137	N = 180
Computer Science and Computer Engineering majors.	Pursuing (non-CS) information technology-related majors or other engineering major.	Students not in CS, engineering or information technology-related majors.

Other variables included in the final predictive regression model were drawn from the application form for the Aspirations award (see below for fuller description), specifically items corresponding to how much high school girls reported engaging in different types of computing experiences. One-hundred and thirty-two (132) students in our sample did not complete this application, so we had application data for the majority of our sample. We compared non-applicants (e.g., those students who did not complete the application versus those who did) to those who did on our main survey variables and found no significant differences between the demographic characteristics of the groups.

4.3 Measures and data sources

We administered surveys online in 2013 and 2016 using the online survey software SurveyMonkey. Response rates were 59% (2013) and 53% (2016). The survey was developed by researchers at the National Center for Women and Information Technology (NCWIT) and was initially based upon the Social Cognitive Career Theory (SCCT) model that describes the relationships between five affective and contextual constructs related to educational and career decision making. After testing the structural validity of the survey (described in the previous section), we found that survey items better fit empirical response patterns associated with skill domains such as programming than any specific affective domain such as interest or confidence. Therefore we used composite variables based on the better fitting model. We also attempted to use the SCCT variables in a structural equation model adhering to previous literature using these variables in a predictive model [24], but the resulting model showed poor fit with the data, so this analysis approach was abandoned.

The survey included items about designing computer games, trying new computer software, solving software problems, fixing or building computers or other technologies out of parts, programming computers, thinking of, or creating new technology inventions (inventing applications), and finding technological solutions to world problems. The *Social Supports* construct contained items measuring perceived peer and family support for computing. The survey also included a question asking if students had taken the CS Advanced Placement exam while in high school along with a series of skip logic questions asking about their application to the Aspirations award program. Aside from some formatting and ordering changes, the 2013 and 2016 surveys were the same.

Other measures included in this study were responses from an application form to the Aspirations program that a portion of the sample completed in high school. The application form included 20 survey-like questions about how much computer experience they had in the following areas: programming, networks, multi-media, and computing-related community involvement (such as volunteer tech support, etc.). Response options utilized a 4-point scale (Not at all, Only a little, Pretty much, and A lot). The items on the application were developed by an advisory board of faculty from CS and IT academic departments from across the US.

4.4 Analytic methods

We used multinomial logistic regression (MLR) to predict persistence with independent variables derived from the survey and from the application data, as well as independent variables for *URM Status*, *Awardee Status*, *Took CS AP exam*, and *Time in college (Early/Late)*. The *Time in college* variable is split between first-year and sophomore students (coded 0) and Juniors and higher (coded 1). We included *Time in college* and *Awardee Status* as covariates in our various regression analyses in order to control for their potential confounding impact on our findings. We also wanted to learn any substantive predictive effect for students receiving the award. As with logistic regression analyses, MLR analyses finds the probability of group membership conditioned on the values of the dependent variables.

The MLR simultaneously compared 1) CS-persisters to Non-persisters, 2) Tech-persisters to Non-persisters, and 3) CS-persisters to Tech-persisters. The odds ratios calculated in the models are identical to those used in logistic regression; odds ratios are the ratio of the odds of group membership given an increase of one unit in the dependent variable compared to the odds of one unit less. Odds ratios less than one indicate that increases in the independent variable are correlated with a smaller probability of group membership. We tested two models, the first included only students that took our first and second surveys (n = 494); the second model included students who took both first and second surveys but also had matching application survey data (n = 362). Table 2 shows all variables used in multinomial logistic regression, the source of the data, the mean and standard deviation, and the scale of each variable.

Table 2. Variables in multinomial logistic regression.

<i>Variable</i>	<i>Source</i>	<i>Mean/SD, proportion</i>	<i>Scale</i>
<i>Programming</i>	Survey-1		
<i>Game design</i>	Survey-1		
<i>Inventing applications</i>	Survey-1		
<i>Social Support</i>	Survey-1	0, 1 for all standardized factor variables	-3 to +3 for all standardized variables
<i>Tech Work & Community</i>	Application		
<i>Multimedia</i>	Application		
<i>Computer Networks</i>	Application		
<i>Award Status</i>	Survey-2	61% Awardee	1 = Awardee 0 = Non- awardee/Non- applicant
<i>Took CS AP exam</i>	Survey-2	42% took exam	1= Took exam 0 = Did not take exam
<i>URMC Status</i>	Survey-2	18% URMC	1 = URMC 0 = Non-URMC
<i>Time in College</i>	Survey-2	50% Early	1 = Early 0 = Late

We created composite variables for multinomial regression analysis given the mix of observed and latent variables we used in our predictive model. Composites were created using the “regression” method in SPSS

for the four-factor solution. For our predictive models we used multinomial regression to (simultaneously) predict the Tech-persister and CS-persister dependent variables. Comparisons for multinomial regression were between Tech-persister and Non-persister on one hand, and CS-persister and Non-persister on the other. Attempted models comparing CS- and Tech- persister groups did not show any meaningful differences, so we did not include it in this analysis.

Data from the application included a list of computer activities rated 1- 4 for how much they had engaged in each activity. We created four factor variables from the 20 items using Maximum Likelihood Extraction and Orthogonal rotation procedures; the first four factors accounted for 43% of total variance. We found four factors representing involvement in *Programming*, *Technology Work and Community*, *Multimedia*, and *Network*. Composite variables made up of individual items in the application were also created with the “regression” method in SPSS. Individual survey items in composites included: experience with basic programming, web development languages, advanced data structures, presentation software and desktop publishing (*Multimedia*), experiences with technology related jobs, internships or summer camps (*Technology Work and Community*) and network design, development and maintenance (*Computer Network*), plus the Programming variable from the application. We found the programming variable based on the application data to be collinear (redundant) with the programming variable from the survey data and removed it from our analyses.

5. RESULTS

We wanted to know if composite variables from our survey as well as other variables predicted the dependent variable for Tech-persister and CS-persister. As well as composites, we included other substantive and covariate measures in our models. We created a multinomial regression model that compared simultaneously 1) CS-persisters to Non-persisters, and 2) Tech-persisters to Non-persisters.

In the first MLS model, CS-persister is compared with the Non-persister group. This comparison included the 494 students who took both Survey-1 and Survey-2. In this comparison, *Programming* significantly predicted CS persistence, with one unit increase corresponding to approximately four times more likely odds of persistence (odds ratio = 4.12). Other significant predictors included *Award status* (odd ratio = 2.24), and *Took the AP CS exam* (odds ratio = 2.75). These two latter variables may also reflect student participation in technically-oriented courses related to programming in high school. Also, because AP courses in computer science are less accessible in rural, poorer or smaller high schools this variable may also be

Table 3 presents all indices for the part of the multinomial model that predicts CS-persister status using only variables from the survey but does not incorporate application data.

Table 3. Multinomial logistic regression model predicts outcome variable CS-persister (n = 494). Uses only variables from survey; does not use variables from application data.

	<i>B</i>	<i>Std. Error</i>	<i>Wald</i>	<i>df</i>	<i>Sig.</i>	<i>Exp(B)</i>
<i>Programming</i>	1.416	.197	51.841	1	<.001**	4.12
<i>Game design</i>	-.013	.153	.007	1	.933	.98
<i>Inventing applications</i>	-.080	.159	.250	1	.617	.923
<i>Social Support</i>	.209	.146	2.063	1	.151	1.23
<i>Award Status</i> [^]	.806	.267	9.095	1	.003**	2.24
<i>Took CS AP exam</i>	1.012	.256	15.591	1	.001**	2.75
<i>URMC</i>	-.238	.328	.528	1	.468	.79
<i>Time in College</i> [^]	.100	.250	.161	1	.689	1.12
<i>Intercept</i>	-1.147	.278	17.060	1	.001**	

[^] Covariate * $\alpha < .05$, ** $\alpha < .01$

Programming involvement also significantly predicted persistence for Tech-persisters, although with a smaller magnitude than it predicted CS persistence, with an odds ratio of 1.61 (versus 4.12). Those receiving the Aspirations Award also had a higher chance of persisting (2.17), but taking the AP exam in Computer Science did not significantly predict persistence for those persisting in non-CS technology majors. Table 4 presents all indices for the part of the multinomial model that predicts Tech-persister status using only variables from the survey but not application data.

Table 4. Multinomial logistic regression model predicts Tech-persister, (n = 494). Does not use variables from application data.

	<i>B</i>	<i>Std. Error</i>	<i>Wald</i>	<i>df</i>	<i>Sig.</i>	<i>Exp(B)</i>
<i>Programming</i>	.480	.144	11.187	1	.001**	1.61
<i>Game design</i>	-.161	.147	1.205	1	.272	.85
<i>Inventing applications</i>	-.075	.151	.244	1	.621	.93
<i>Social Support</i>	.130	.135	.934	1	.334	1.14
<i>Award Status</i> [^]	.775	.252	9.443	1	.002**	2.17
<i>Took CS AP exam</i>	.250	.256	.956	1	.328	1.28
<i>URMC</i>	.057	.298	.037	1	.847	1.06
<i>Time in College</i> [^]	-.093	.239	.150	1	.699	.91
<i>Intercept</i>	-.657	.248	7.03	1	.008**	

[^] Covariate * $\alpha < .05$, ** $\alpha < .01$

We also made the same comparisons for students who took both Survey-1 and Survey-2 surveys *and* had completed the items from the application (n=362). For the CS-persister comparison, *Programming*, *Took CS AP*, and *Awardee Status* all strongly predicted persistence (odds ratio = 3.71, 2.64, 3.85, respectively). The variable for *Tech Work and Community* predicted non-persistence for CS-persister with an odds ratio of 0.68, indicating that those who expressed interest in these areas were less likely to persist in CS. Table 5 presents all indices for the part of the multinomial model that predicts CS-persister status using both variables from the survey and application data.

Table 5. Multinomial logistic regression predicts CS-persister (n = 362). This model uses variables from application data.

	<i>B</i>	<i>Std. Error</i>	<i>Wald</i>	<i>df</i>	<i>Sig.</i>	<i>Exp(B)</i>
<i>Programming</i>	1.313	0.225	33.94	1	.001**	3.71
<i>Game design</i>	0.043	0.179	0.057	1	0.811	1.04
<i>Inventing applications</i>	-0.185	0.196	0.887	1	0.346	0.83
<i>Social Support</i>	0.141	0.174	0.656	1	0.418	1.15
<i>Tech Work & Community</i>	-0.388	0.186	4.361	1	0.037*	0.68
<i>Multimedia</i>	-0.02	0.182	0.013	1	0.911	0.98
<i>Network</i>	0.256	0.167	2.345	1	0.126	1.29
<i>Award Status</i> [^]	0.969	0.345	7.885	1	0.005**	2.64
<i>Took CS AP exam</i>	1.347	0.321	17.59	1	.0001***	3.85
<i>URMC</i>	0.206	0.407	0.256	1	0.613	1.23
<i>Time in College</i> [^]	0.406	0.301	1.818	1	0.178	1.50
<i>Intercept</i>	-1.54	0.357	18.645	1	.0001***	

[^]Covariate * $\alpha < .05$, ** $\alpha < .01$, *** $\alpha < .0001$

Finally, we examined which variables predicted Tech-persister using the additional application data. Again, *Programming* (1.64) and *Aspirations Award Status* (2.29) were significant predictors, while the variable *Network* also predicted tech persistence (1.50). *Tech Work and Community* again predicted non-persistence (0.63). Table 6 presents all indices for the part of the multinomial model that predicts CS-persister status using both variables from the survey and application data.

Table 6. Multinomial logistic regression predicts Tech-persister, (n = 362). This model uses variables from application data.

	<i>B</i>	<i>Std. Error</i>	<i>Wald</i>	<i>df</i>	<i>Sig.</i>	<i>Exp(B)</i>
<i>Programming</i>	0.493	0.173	8.112	1	0.004**	1.64
<i>Game design</i>	-0.149	0.17	0.774	1	0.379	0.86
<i>Inventing applications</i>	-0.141	0.187	0.565	1	0.452	0.87
<i>Social Support</i>	0.064	0.161	0.159	1	0.69	1.07
<i>Tech Work & Community</i>	-0.459	0.181	6.426	1	0.011*	0.63
<i>Multimedia</i>	-0.043	0.168	0.064	1	0.801	0.96
<i>Network</i>	0.407	0.158	6.657	1	0.01*	1.50
<i>Award Status</i>	0.828	0.318	6.804	1	0.009**	2.29
<i>Took CS AP exam</i>	0.565	0.322	3.087	1	0.079	1.76
<i>URMC</i>	0.438	0.372	1.383	1	0.24	1.55
<i>Time in College[^]</i>	0.047	0.286	0.027	1	0.87	1.05
<i>Intercept</i>	-0.841	0.309	7.398	1	0.007**	

[^] Covariate * $\alpha < .05$, ** $\alpha < .01$

Variables that did not predict persistence in any model included *Social Supports*, *Game Design*, *Inventing Applications*, *Multi-media* and *URMC status*. Each model we used accounted for a third of the variability in both dependent variables with R-square values of .31 for the model using the larger sample (n = 494) and $R^2 = .33$ for model with the smaller sample (n = 362).

6. DISCUSSION

As noted earlier, the longitudinal nature of our project, spanning from high school to college and beyond, gives us a unique opportunity to examine actual persistence and not just self-reported intent to persist in the field. Additionally, the sheer number of young women who have participated in this project from all over the US, many of whom have had a variety of computing experiences, is unusual to find in any single group. We found that in this college-bound all-female sample, involvement in programming, and a stated intent in high school to pursue learning computer science and other technical fields were the best predictors of both persistence in computer science and other tech-related fields. This is consonant with previous research [20] suggesting that programming in high school is associated with college success in computer science.

Involvement in programming in high school was the best predictor of persistence in college for both CS and the larger, non-overlapping category of technology persistence. Other researchers [19, 29] have found that students who took Advanced Placement courses in CS also tended to have an interest in computer science in college, supporting our finding that students taking a CS Advanced Placement exam tended to persist in CS. Only one of the other computer fields in the second model (work with computer networks) also predicted Tech-persistence, albeit weakly, but did not predict becoming a CS major. In fact, while many of the non-CS Tech-persister students pursued alternative branches of computing and engineering in college such as information science, electronic media, cybersecurity or data science, early involvement in alternative areas of computing (other than networks) was *not* significantly associated with continuing in these areas, while involvement in programming did predict later persistence. This also applies to non-CS disciplines such as aerospace, civil or biological engineering where (presumably) some proficiency in coding supplemented other skills for the major. Although it can be argued that a computer science major is not necessary for doing intensive, innovative computing work [20, 30], increasing the diversity of those who complete computer science majors is one direct way to address the national imperative to broaden participation in computing [31].

While we see some evidence for students' involvement in computing diverging and stratifying after high

school, it seems that involvement in general tech-related fields -*other than programming*- in high school does not transfer to entering and persisting in computer science in college for the girls in our sample. Understanding the centrality of programming is important to the field's push to broaden participation in computing. While a number of efforts have been made to make coding more accessible and engaging and have more "real life" application for K-12 students [32], often, alternative uses of computers such as web design are framed as a way to attract students into an interest in computer science [33]. However, students adept at using technology in the service of other projects (e.g., designing a web page to illustrate a history paper), may not be interested in technology in-and-of-itself as a field of study. So while these interventions may expose students who would not otherwise have been exposed to computing, research suggests a clear distinction between students who pursue more technical aspects of computing (usually in Advanced Placement courses in computer science) and students who use computers as a "means to an end" to other activities [33].

In sum, our results suggest that involvement in less technical aspects of computing does not lead girls to later involvement in computer science or alternative technical majors. While other research shows that younger girls may not gain the experience and exposure to programming that could lead to choosing a CS major in college, our study suggests that not all exposure to computer science concepts is equally as influential in girls' persistence. Efforts (including the Aspiration Awards) attempt to support girls learning technical aspects of computing may have more power over long term retention of these girls in computing fields. As other research suggests, these women will persist in CS in spite of social pressures to leave the field. The current study supports the idea that encouraging girls to engage in more technical aspects of computing, and supporting them as they do so, can be an important pathway into college computing majors. While we recognize that these findings may seem discouraging to those who see alternative computing disciplines as a gateway to students pursuing CS in college, it is also important to note that multiple exposures and experiences influence what students choose to do in high school [35 -36]. Our findings in this study do not rule out the idea that exposure to non-technical aspects of computing may contribute to girls ultimately choosing to engage in more technical aspects of computing.

7. LIMITATIONS AND FUTURE RESEARCH

Generalizing the findings of this study is limited for several reasons. The individuals in this sample all identified as female and so no comparison to boys or young men in college is possible with this dataset, although this can also be seen as beneficial given the oversized role of the influence of male gender in predictive studies. A range of gender differences in students aspiring to become CS majors is described in Lehman, Sax and Zimmerman [37] with females as a group more diverse than males, rating themselves lower than men in computer and math skills, and less likely to aspire to becoming a computer programmer than men. A more consequential limitation of this study is that the high school girls included were at least somewhat interested in computing at the time of the first survey, as evidenced by their seeking out an award in computing and registering on the Aspirations website. Given that 73% of our sample ultimately completed an application for the Aspirations award and were much more likely to take the CS AP exam, our sample seems to have had at least some pre-existing interest in or affinity to computing. Thus, our findings are likely not generalizable to all individuals who identify as women but rather to women who also have interest in computing as girls. Finally, all of the young women included in our survey were living in the US during this study. Further research including individuals who identify as female but express a variable interest in computing as well as research looking at individuals in other nations is needed to know if the findings generalize outside the US.

8. ACKNOWLEDGEMENTS

This research is supported by the National Science Foundation under CNS 1441071 and CNS 0813956. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

- [1] Bureau of Labor Statistics, www.bls.gov/ooh/computer-and-information-technology/home.htm. Accessed 24 May, 2018.
- [2] DuBow, W. & Pruitt, A.S. (2018) *NCWIT Scorecard: The Status of Women in Technology*. Boulder, CO: NCWIT.
- [3] Cohen, Clemencia Cosentino de, and Nicole Deterding. "Widening the net: National estimates of gender disparities in engineering." *Journal of Engineering Education* 98.3 (2009): 211-226.
- [4] Dee, T., and S. Gershenson. "Unconscious Bias in the Classroom: Evidence and Opportunities, 2017." *Stanford Center for Education Policy Analysis* (2017)
- [5] Google (Firm). (2014). Women who choose computer science: what really matters: the critical role of encouragement and exposure.
- [6] Kanny, M. A., Sax, L. J., & Riggers-Piehl, T. A. (2014). Investigating forty years of STEM research: How explanations for the gender gap have evolved over time. *Journal of Women and Minorities in Science and Engineering*, 20(2).
- [7] Fisher, A., & Margolis, J. (2002). Unlocking the clubhouse: the Carnegie Mellon experience. *ACM SIGCSE Bulletin*, 34(2), 79-83.
- [8] Crenshaw, Tanya L., et al. "Ten Years toward Equity: Preliminary Results from a Follow-Up Case Study of Academic Computing Culture." *Frontiers in psychology* 8 (2017): 816.
- [9] Clarke, Valerie A., and Susan M. Chambers. "Gender-based factors in computing enrollments and achievement: Evidence from a study of tertiary students." *Journal of Educational Computing Research* 5.4 (1989): 409-429. Powell, R. M. (2008, March). Improving the persistence of first-year undergraduate women in computer science. In *ACM SIGCSE Bulletin* (Vol. 40, No. 1, pp. 518-522). ACM.
- [10] Michell, D., Szorenyi, A., Falkner, K., & Szabo, C. (2017). Broadening participation not border protection: how universities can support women in computer science. *Journal of Higher Education Policy and Management*, 39(4), 406-422.
- [11] 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.
- [12] Cheryan, Sapna, Allison Master, and Andrew N. Meltzoff. "Cultural stereotypes as gatekeepers: Increasing girls' interest in computer science and engineering by diversifying stereotypes." *Frontiers in psychology* 6 (2015): 49
- [13] Rommes, E., Overbeek, G., Scholte, R., Engels, R., & De Kemp, R. (2007). 'I'm not Interested in Computers': Gender-based occupational choices of adolescents. *Information, Community and Society*, 10(3), 299-319.
- [14] Miller, P. H., Slawinski Blessing, J., & Schwartz, S. (2006). Gender differences in high-school students' views about science. *International journal of science education*, 28(4), 363-381.
- [15] Katz, S., Allbritton, D., Aronis, J., Wilson, C., & Soffa, M. L. (2006). Gender, achievement, and persistence in an undergraduate computer science program. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 37(4), 42-57.
- [16] Eccles, J. S., & Wang, M. T. (2016). What motivates females and males to pursue careers in mathematics and science?. *International Journal of Behavioral Development*, 40(2), 100-106.
- [17] Google Inc. & Gallup Inc. (2016). *Trends in the State of Computer Science in U.S. K-12 Schools*. Retrieved from goo.gl/j291E0. Accessed 14 May, 2018.
- [18] Taylor, H. G., & Mounfield, L. C. (1994). Exploration of the relationship between prior computing experience and gender on success in college computer science. *Journal of educational computing research*, 11(4), 291-306.
- [19] Wang, J., Hong, H., Ravitz, J., & Ivory, M. (2015, June). Gender differences in factors influencing pursuit of computer science and related fields. In *Proceedings of the 2015 ACM Conference on Innovation and Technology in Computer Science Education* (pp. 117-122). ACM.
- [20] Chen, C., Jeckel, S., Sonnert, G., & Sadler, P. M. (2019). "Cowboy" and "Cowgirl" Programming and Success in College Computer Science. *International Journal of Computer Science Education in Schools*, 2(4), n4.
- [21] Barker, Lecia J., Charlie McDowell, and Kimberly Kalahar. "Exploring factors that influence computer science introductory course students to persist in the major." *ACM SIGCSE Bulletin*. Vol. 41. No. 1. ACM, 2009.
- [22] Trevisan, B., McKlin, T. & Guzdial, M. 2011. "Factors influencing CS Participation: Introductory computer science students describe what led them to computing." GaComputes! Technical Report, Atlanta: The Findings Group, LLC.
- [23] Katz, Sandra, et al. "Gender, achievement, and persistence in an undergraduate computer science program." *ACM SIGMIS Database: the DATABASE for Advances in Information Systems* 37.4 (2006): 42-57.
- [24] Lent, R. W., Miller, M. J., Smith, P. E., Watford, B. A., Hui, K., & Lim, R. H. (2015). Social cognitive model of adjustment to engineering majors: Longitudinal test across gender and race/ethnicity. *Journal of Vocational Behavior*, 86, 77-85.
- [25] Maltese, Adam V., and Robert H. Tai. "Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among US students." *Science Education* 95.5 (2011): 877-907.
- [26] Ellis, Jessica, Bailey K. Fosdick, and Chris Rasmussen. "Women 1.5 times more likely to leave STEM pipeline after calculus compared to men: Lack of mathematical confidence a potential culprit." *PloS one* 11.7 (2016): e0157447.
- [27] Larson, Lisa M., et al. "Predicting graduation: The role of mathematics/science self-efficacy." *Journal of Career Assessment* 23.3 (2015): 399-409.
- [28] Mueller, R. O., & Hancock, G. R. (2008). Best practices in structural equation modeling. *Best practices in quantitative methods*, 488508.
- [29] Ashcraft, C. & Blithe, S. (2010). *Girls in IT: The Facts*. National Center for Women & IT. Retrieved from www.ncwit.org/thefactsgirls. Accessed 1 June, 2018.
- [30] Weidler-Lewis, J. & DuBow, W. (2017). Defining a Discipline or Shaping a Community: Constraints on Broadening Participation in Computing." ACM Technical Symposium on Computer Science Education (SIGCSE '17) March 8-11, 2017, Seattle, WA, USA. DOI: <http://dx.doi.org/10.1145/3017680.3017776>
- [31] National Science Foundation. 2012. CISE Strategic Plan for Broadening Participation. National Science Foundation, Alexandria, VA. Retrieved from https://www.nsf.gov/cise/oad/cise_bp.jsp.
- [32] Kafai, Y. B., Lee, E., Searle, K., Fields, D., Kaplan, E., & Lui, D. (2014). A crafts-oriented approach to computing in high school: Introducing computational concepts, practices, and perspectives with electronic textiles. *ACM Transactions on Computing Education (TOCE)*, 14(1), 1.

Predicting Persistence in Computer Science and Technology Related Majors from High School to College

- [33] Huang, C. Y., Liou, J. C., & Huang, G. (2015, January). Pathway from Learning Multimedia Software to Computer Science Education. In *Proceedings of the International Conference on Frontiers in Education: Computer Science and Computer Engineering (FECS)* (p. 181). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).
- [34] Dimond, J., & Guzdial, M. (2008). More than paradoxes to offer: Exploring motivations to attract women to computing. *Georgia Institute of Technology Technical Report*.
- [35] Robin, B. R. (2015). The effective uses of digital storytelling as a teaching and learning tool. *Handbook of research on teaching literacy through the communicative and visual arts*, 2, 429-440.
- [36] DuBow, Wendy, Alexis Kaminsky, and Joanna Weidler-Lewis. "Multiple Factors Converge to Influence Women's Persistence in Computing: A Qualitative Analysis." *Computing in Science & Engineering* 19.3 (2017): 30-39.
- [37] Lehman, Kathleen J., Linda J. Sax, and Hilary B. Zimmerman. "Women planning to major in computer science: Who are they and what makes them unique?." *Computer Science Education* 26.4 (2016): 277-298.