Open access • Journal Article • DOI:10.1145/1185335.1185344

Gender, achievement, and persistence in an undergraduate computer science program — Source link

Sandra Katz, David Allbritton, John M. Aronis, Christine Wilson ...+1 more authors

Institutions: University of Pittsburgh, DePaul University, University of Virginia

Published on: 28 Nov 2006 - ACM Sigmis Database (ACM)

Gender, Achievement, and Persistence in an Undergraduate Computer Science Program

Sandra Katz
University of Pittsburgh

David Allbritton
DePaul University

John Aronis
University of Pittsburgh

Christine Wilson
University of Pittsburgh

Mary Lou Soffa
University of Virginia

Acknowledgements

This research was supported by a grant from the National Science Foundation (grant number EIA 0089963). The data presented and views expressed are not necessarily endorsed by this agency. The authors appreciate the helpful comments of the editors and anonymous reviewers.

Abstract

The proportion of computer science (CS) bachelor's degree recipients who are women has consistently been small and is declining. This study investigates factors that predict performance and persistence in an undergraduate CS program and explores why even high-achieving students leave the undergraduate "CS pipeline." The factors that predict achievement and retention sometimes interact in complex, unexpected ways. Male students who earned less than a B in an introductory CS course were more likely to take the next course in the curriculum than were women who earned less than a B. Achievement is a factor in even high-achieving students' decision to leave CS; loss of interest can accompany loss of confidence. Level of achievement was predicted by various background factors including Scholastic Aptitude Test scores, the number of Calculus courses taken before entering the CS program, amount of access to a computer at home, prior computing experience, and having a mentor or role model during high school. Most of these factors also predicted persistence beyond the first two courses required for a CS major. Curiously, women in the introductory CS course who reported having low exposure to specific programming skills outperformed women who reported having a high level of programming experience. The reverse was true of men. Further investigation provided evidence that women who develop programming skills while in high school might do so at the expense of developing other skills that strongly predict CS achievement, particularly math skills.

ACM Categories: K.3.2

Keywords: Gender and Computer Science Education, Achievement and Retention of Students in an Undergraduate Computer Science Program, Under-Representation of Women in Computer Science.

Introduction

According to the Bureau of Labor Statistics' *Occupational Outlook Handbook, 2004-05 Edition*, employment of software engineers, systems analysts, database administrators, and computer scientists is expected to grow much faster than the average growth rate of 10 to 20 percent through 2012. Employment of computer programmers is expected to grow more slowly, at about the average rate, with job prospects best for "college graduates with knowledge of a variety of programming languages and tools" (BLS, 2004). Hence, the demand for computer specialists in various areas is on the rise, as is the standard that computer programmers will have to meet in order to be competitive. Undergraduate computer science programs are challenged to turn out students who can meet this standard.

Unfortunately, while the demand for competent, versatile technology professionals is increasing, *the 2002-2003 Taulbee Survey* shows that undergraduate enrollment in computer science (CS) and computer engineering (CE) programs decreased significantly from 2002 to 2003 and is expected to decline further—most likely because the dot-com crash and move of more technology jobs overseas have made CS and CE less attractive to undergraduates (Zweben & Aspray, 2003). Although CS/CE enrollment levels are still considerably higher than they were before the dot-com era, there will nonetheless be a shortage of competent college graduates who will be qualified to fill the increasing number of information technology (IT) jobs predicted by the Bureau of Labor Statistics (2004).

It is frequently noted that the IT labor shortage could be resolved if more women chose to pursue a technology-related profession and persisted in post-secondary computer science programs. In addition, increasing the representation of women in CS would help to ensure that diverse perspectives are going into the development of technologies that can be used by a wide range of people (Roberts, 2003; cited in Beyer et al., 2003). Unfortunately, though, the dearth of women in IT, and disproportionate loss of women from undergraduate CS programs (relative to men) are well-documented phenomena (Camp, 1997; Camp et al., 1999; Crews & Butterfield, 2003; Freeman & Aspray, 1999; Gürer & Camp, 2002; Myers, 1999; NSF, 2000). Statistics available from the U.S. Department of Education show that, from 1993-1994 through 1997-1998, the percentage of women earning bachelor's degrees in computer science decreased to a low of 26.7% (NCES, 2001). This drop in female CS bachelor's degree recipients stands in contrast to the steady growth of women earning degrees in other sciences and in engineering. According to Camp, Miller, and Davies (1999), the "incredible shrinking pipeline" of women in CS is a trend that is likely to continue.

These nation-wide trends in decreased recruitment of women, and low retention of both men and women in undergraduate CS programs, are reflected in the Department of Computer Science at the University of Pittsburgh. In an on-going study of the factors that contribute to achievement in Computer Science (Katz et al., 2003), we have traced the progress of a cohort of 200 students throughout the sequence of courses that make up the programming core: *Introduction to Computer Science Using JAVA* (CS401)*, Data Structures* (CS445), and *Algorithm Implementation* (CS1501). Because students sometimes take another course required for the CS major (*Formal Methods*, CS1502) before taking CS1501, we also tracked student enrollment and performance in this course. These four courses are typically completed within the first two years of the program.

Figure 1 shows the sharp contrast between female and male entry into the program (CS401). Indeed, even after recruiting participants from CS401 across three academic semesters, we obtained a sample of only 35 women, compared with 165 men. These samples represent approximately 28% of the men in the population of students who were enrolled in CS401 during the recruitment period (165 men out of 590) and approximately 27% of the women in this population (35 women out of 130).[1] Furthermore, the rate of attrition from CS401 to CS445 is higher for women than for men—41% versus 57%, respectively, a marginally significant difference (χ^2 = 3.2; *p* = .07).

What is not directly apparent in Figure 1 is the fact that the Department of Computer Science is not only losing under-achieving students (i.e., students who earn less than C, which is the departmental standard for passing), but high-achieving students as well (i.e., students who earn B or above). This trend is shown in Figure 2. Reflecting the attrition trend for the cohort as a whole (Figure 1), the sharpest loss of high-achieving students takes place from CS401 to CS445. 23% of students who earned B or better in CS401 (25 out of 109) did not continue to CS445—that is, 22% of high-achieving men (20 out of 91) and 28% of high-achieving women (5 out of 18), a difference that was not statistically significant. This pattern continued after CS445. Among high-achieving students in CS445, 14% (7 out of 51) did not continue to CS1501 or CS1502—that is, 14% of high-achieving men (6 out of 43) and 13% of high-achieving women (1 out of 8).

The striking loss of high-achieving (B and above) students of both genders prompted us to investigate the relationship between achievement, student characteristics (including gender), and attrition in this program. Specifically, we asked: What factors predict achievement and do the same factors predict persistence?

[1] The population includes students who were not prospective CS majors, but who were enrolled in CS401 during the recruitment period. Frequencies for men and women in the population are approximate. They were derived from direct observation (a count of men and women who attended CS401 examinations), since we were unable to obtain demographic information for students who did not consent to participate in the study.

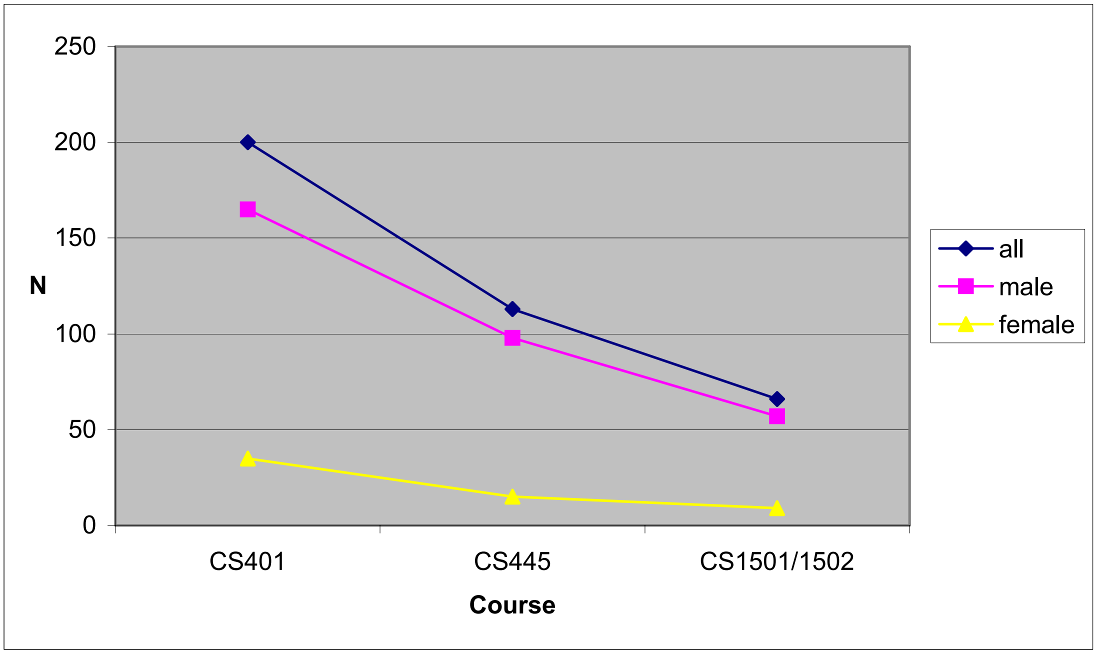

Figure 1. Enrollment in CS Courses Over Time

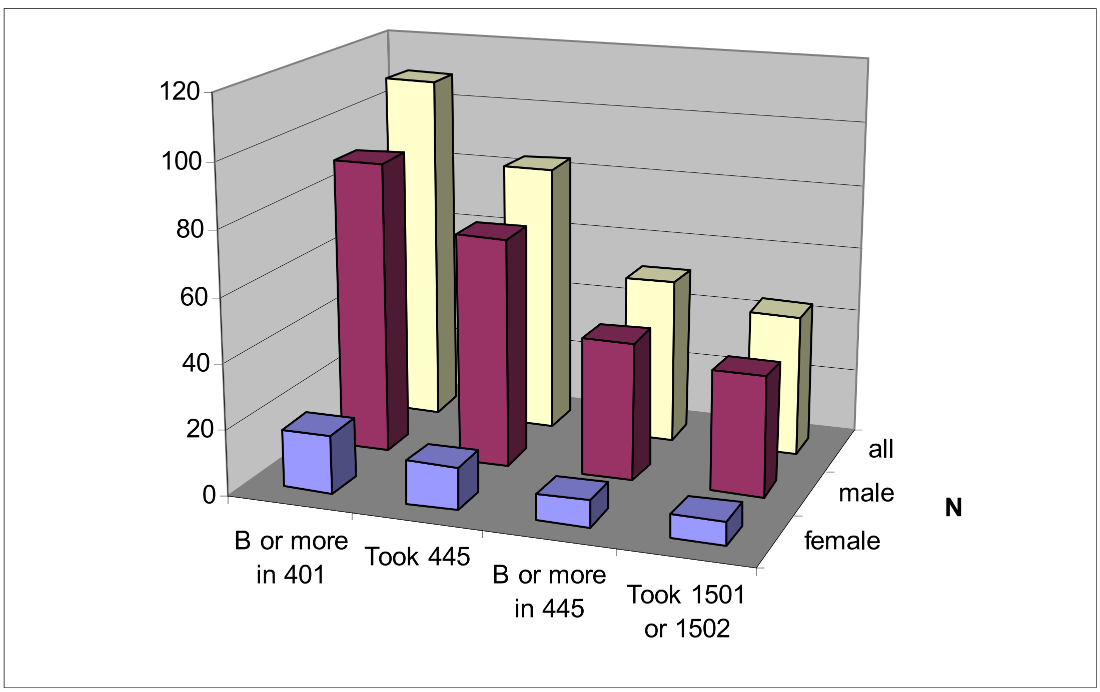

Figure 2. Loss of B-and-above Students After Two CS Courses

Do the factors that predict achievement persist through the programming core, or are they relevant to the entry-level course (CS401) only? Are there gender differences in the factors that predict achievement and persistence? What are the reasons that good students give for staying in or leaving the program, and, again, are there any gender differences in these reasons?

Our goals in investigating these questions are to build upon prior research dedicated to understanding the problem of attrition (especially female attrition) along the undergraduate segment of the CS "pipeline" and to contribute to the growing body of proposed solutions (e.g., Bunderson & Christensen, 1995; Cohoon, 2001; 2002; Margolis & Fisher, 2002; Pearl et al., 1990). Of course, as Figure 1 dramatically illustrates, the retention problem can not be solved independent of the recruitment problem. Much prior research has focused on understanding the complex forces that repel young girls and women from computers and computer science (see Davies et al., 2000 for a review), and many promising programs are in place to attract young girls to CS, such as SWIFT (Davies et al., 2000) and Eyes to the Future (Falk et al., 1999).

The problem of undergraduate attrition from CS is as complex as the problem of recruitment (Bunderson & Christensen, 1995; Crews & Butterfield, 2003; West & Ross, 2002). Since 1990, research has focused on several causes of female attrition, including (but not limited to) gender bias and stereotyping (Margolis et al., 1999; Spertus, 1991), attitudes towards computing (Charlton & Birkett, 1998; Shashaani, 1997), prior computing experience (Bunderson & Christensen, 1995; Fan et al., 1998; Taylor & Mounfield, 1994), motivation (Charlton & Birkett, 1999), confidence (Beyer et al., 2003), disciplinary characteristics (Mahony & Van Toen, 1990; Pearl et al., 2002), and departmental characteristics (Cohoon, 2001; 2002) including the "culture of computing" at particular institutions (Bunderson & Christensen, 1995; Margolis & Fisher, 2002). Because most studies have found that women perform equally as well or better than men in undergraduate CS programs, little attention has been paid to the relationship between achievement and persistence. In this exploratory study, we take a closer look at this relationship. Supporting research by Jagacinski, LeBold, and Salvendy (1988), we found evidence that achievement can play a role in even "successful" students' decision to leave CS. While there is considerable overlap between the factors that predict achievement and retention, these factors sometimes interact in complex and unexpected ways.

Method

Participants

Two hundred students at the University of Pittsburgh voluntarily participated in the study. Recruitment took place on the first day of the first course taken by all prospective computer science majors, *Introduction to Computer Science Using Java* (CS401). Students were paid a nominal amount for consenting to participate. In order to increase the number of women in our sample, we recruited from five sections of the course, spanning three academic semesters during 2001-2002. Despite this lengthy recruitment period, our sample consists of only 35 women (17.5% of participants in the cohort and 27% of women in the population), compared with 165 men (82.5% of participants in the cohort and 28% of men in the population).

Originally, 236 students signed up for the study. We eliminated thirty-six students because their planned major was not computer science and they were therefore not required to complete the programming core, they withdrew from CS401, or they did not have Scholastic Aptitude Test (SAT) scores. Included within the sample of 200 were 9 engineering students, 1 student from the Scientific Computing program, and 9 students who stated that they were "undecided" about majoring in CS. We included the engineering students because they are required to complete the programming core and several of these students planned a concentration in computer engineering.

As shown in Table 1, the majority of participants (78%) are Caucasian. Thirteen percent are African American, 5% Asian or Asian American, and 4% belong to another ethnic group. Because of the small sample of students who could be considered a "minority" in computer science (i.e., students who are not Caucasian, Asian, or Asian American), our analyses do not differentiate students based on race or ethnicity. We differentiate students only according to gender. The mean age of participants is 20 years.

Classification	Males	Females	Total
Caucasian	130 (79%)	26 (74%)	156 (78%)
African American	21 (13%)	5 (14%)	26 (13%)
Asian or Asian American	8 (5%)	2 (6%)	10 (5%)
Other (Hispanic American, American Indian, etc.)	6 (4%)	2 (6%)	8 (4%)

Table 1. Ethnic Classification of Participants

Instruments

In order to identify the factors that predict achievement and to determine if these factors also predict persistence, we developed and administered two surveys: a background questionnaire that was issued at the start of CS401 and an end-of-course questionnaire that was issued approximately a week before the final exam for each course. Students' responses to selected questions on these surveys were used in the analyses. Appendix A shows the questions used in the analysis, with their associated variables (factors) and values. These questions targeted factors shown in prior research to be predictive of achievement in undergraduate CS programs, such as math background (Cafolla, 1987; Clement et al., 1986; Jagacinski et al., 1988; Lai & Repman, 1996), prior computing experience (Byrne & Lyons, 2001; Taylor & Mounfield, 1989), home and school access to computers (Jones, 1987, cited in Bunderson & Christensen, 1995; Foorman et al., 1985, cited in Bunderson & Christensen, 1995), motivation (Charlton & Birkett, 1999), confidence in one's computer ability and the perception that one's confidence is backed by good performance (Charlton & Birkett, 1999; Jagacinski et al., 1988; Volet & Styles, 1992), and encouragement from mentors and role models (Falk et al., 2000). The background survey also included questions aimed at exploring the influence of parental encouragement on achievement and persistence. Since math ability has repeatedly been shown to correlate with CS achievement, we also obtained students' official SAT scores (math and verbal) from the university's Office of Admissions and Financial Aid.

As shown in Appendix B, the end-of-course surveys were primarily designed to identify students who no longer planned to major in computer science and students' self-reported reasons for this decision. Towards this end, it asked questions aimed at gauging students' confidence in their ability to succeed, interest in computer science, attitude towards the CS department, and perception that they had friends in the department.

The response rate for surveys is as follows:

- 156 out of 200 students completed the background survey at the start of CS401 (78%)
- 131 CS401 students completed the end-of-course survey (66%)
- 47 students completed the end-of-course survey for CS445, out of 110 students in the course (43%)

Data collection for the third course in the programming core (*Algorithm Implementation*, CS1501) is in progress, given that some students opt to take *Formal Methods* (CS1502) before this course. Hence, the end-of-course surveys for CS1501 have not been analyzed and will not be discussed here.

In addition to collecting the survey data, we acquired students' letter grades for each course from their instructors.

Results and Discussion

Letter grades are the measure of achievement in our analyses, on a scale from 1 to 13 (e.g., 1 = F, failure; 6 = C; 9 = B; 13 = A+). The number of courses that students completed, as evident by their having a grade for a given course, functions as a measure of persistence: 1, 2, or 3, where 3 means "3 or more courses" (i.e., students who took CS1501, CS1502, or both all received a persistence rating of 3). Table 2 presents means and standard deviations for the dependent variables (grades and persistence ratings) by gender. N (M,F) means number of men and women, respectively.

We assume that if a student did not have a grade for a course, the student did not take the course. We also assume that students should have completed the programming core (CS401, CS445, and CS1501)—or, alternatively, CS401, CS445, and CS1502—by the time we analyzed the data for this study. We believe that these are fair assumptions, given the amount of time that has elapsed since the last set of students signed up in the Spring of 2002 and the fact that most students take these courses in successive academic semesters, as recommended by the CS Department. There are probably some students who are progressing more slowly than the norm, but we do not expect there to be enough of these cases to significantly lessen the validity of our findings.

All findings reported in this section are statistically significant at the .05 level, unless otherwise stated.

Gender, Persistence and Achievement

Does achievement predict persistence? If so, what student characteristics predict achievement and are there gender differences among these factors?

35% of men versus 27% of women had a persistence rating of 3 (i.e., they completed CS401, CS445, and either CS1501, CS1502 or both). This difference is not statistically significant. Not surprisingly, students' persistence rating correlated strongly with their grades in CS401 and CS445 (r = .61, and r = .64), respectively. Among the 66 students who completed three or more courses, women outperformed men in all four courses, but not significantly so.

If we look at persistence over time—the transition from CS401 to CS445 and then from CS445 to CS1501—we find significant differences in persistence by gender. 60% of men who completed CS401 versus 43% of women continued to CS445, a marginally significant difference ($\chi^2 = 3.2$, $p < .07$). When we examined this difference further, we found that the nexus of the gender difference was at the B performance level. Comparing students who earned C or above with students who earned less than C in CS401, a Chi-square test revealed no significant gender differences in persistence to CS445. Indeed, very few students of either sex who earned less than a C in CS401 continued to CS445 (8% overall; 10% of men who earned C or less and 0% of women). However, when we performed the same test using B as the performance cut-off, we found that men who earned less than B were more likely to continue to CS445 than were women who earned less than B ($\chi^2 = 3.9$, $p = .05$).

The reasons for this gender difference are unclear and warrant further investigation. It could be that women tend to be more strongly affected by performing below the B level than men and this influences their decision to continue with a CS major. Although this explanation is speculative, it is consistent with prior research by Eccles-Parsons (1984; cited in Jagacinski et al., 1988), which noted the importance of students' performance expectations in their plans to enroll in advanced mathematics courses. Jagacinski et al. (1988) suggest a similar relation between performance expectations and persistence, but their work, like Eccles-Parsons' did not investigate this relation with respect to gender. Alternatively, or in addition, it could be that women are more cautious than men about pursuing a major that they might not be well-suited for; they leave CS to pursue a field that they are confident about being able to excel in. Whether or not this is a wise decision, and how long students should be encouraged to test their ability to succeed in CS before leaving the field (e.g., by repeating first-year courses that they do not do well in) are also important questions for further research.

Looking at the transition from CS445 to CS1501, Chi-square analysis did not reveal any significant gender differences in persistence, using either a letter grade of C or B as the performance cut-off. This is probably because the number of women who continued on to CS1501 or CS1502 (9) was too small for statistical analysis. Nonetheless, the same pattern that we observed at the first transition point (from CS401 to CS445) holds for this second transition point: more men who earned less than a B in CS445 (37%, 20 out of 54) continued to CS1501 and/or CS1502 than did women (33%, 2 out of 6).

Student Characteristics that Predict Achievement

Given that achievement predicts persistence as well as gender differences (above and below the B grade level), it is worthwhile to consider which of the background factors described in Appendix A significantly predict achievement.

		Men		Women	
	N (M,F)	Mean	SD	Mean	SD
CS401 letter grade	165, 35	8.2	3.9	7.6	4.7
CS445 letter grade	96, 14	6.9	4.1	7.9	4.3
CS1501 letter grade	47, 6	7.8	2.9	9.0	2.1
CS1502 letter grade*	45, 9	8.0	3.6	9.4	1.8
persistence	165, 35	1.9	.86	1.7	.86
* = marginally significant, $p = .08$					

Table 2. Gender Differences in Dependent Variables

Table 3 displays gender differences in mean scores (SD = standard deviation) for selected background and end-of term survey factors. As in Table 2, N in Table 3 represents the number of data points for each factor by gender (M = male, F = female), based on survey responses. Significant gender differences are flagged.

Consistent with prior research (e.g., Beyer et al., 2003), men reported more confidence in earning a B or above on all three surveys—$t(148) = 2.1$; $t(128) = 1.8$, $p < .08$, and $t(47) = 2.0$, for the background survey, CS401 end-of-course survey, and CS445 end-of-course survey, respectively. Also consistent with prior research (Taylor & Mounfield, 1989), men entered the program with a greater number of self-reported programming skills (exposure to specific constructs), $t(153) = 1.8$, $p < .07$. Women scored higher on self-reported home access to a computer during high school; $t(154) = 1.8$, $p < .07$.

Math and verbal SAT scores. Math Scholastic Aptitude Test (SAT) score predicted achievement in all four courses, overall (all students, irrespective of gender): $r = .38$ for CS401, $r = .23$ for CS445, $r = .34$ for CS1501, and $r = .37$ for CS1502. Number of Calculus courses taken predicted achievement in CS401 overall, $r = .16$. These correlations support much prior research on the importance of math skills in computer science (e.g., Lai & Repman, 1996).

Verbal SAT score predicted male performance in CS401 ($r = .20$), female performance in CS445 ($r = .78$), and CS401 performance overall ($r = .21$).

Factor		Men		Women	
	N (M,F)	Mean	SD	Mean	SD
Math SAT score	165, 35	618	81.0	609	83.5
Verbal SAT score	165, 35	595	83.5	604	107.7
Number of Calculus courses completed	128, 28	.79	.92	1.1	1.1
Number of CS courses that involved programming	128, 28	2.2	1.2	1.9	1.1
Access to computers at home during high school*	128, 28	1.9	.38	2.0	.19
Availability of computers at high school	128, 28	1.9	.31	1.9	.26
Informal computing activities	128, 28	2.0	1.1	1.6	1.1
Exposure to specific programming constructs*	127, 28	5.6	2.1	4.8	1.8
Goal of earning B or above	127, 28	1	.0	1	.0
Confidence in earning B or above at start of CS401**	122, 28	4.1	1.1	3.6	1.4
Had mentor or role model in high school	127, 28	.57	.50	.75	.44
Maternal encouragement	127, 28	3.5	1.3	3.5	1.2
Paternal encouragement	128, 28	3.8	1.2	.88	.33
Interest in CS at end of CS401	105, 26	4.1	1.0	3.7	1.2
Likes CS environment at end of CS401	105, 26	3.5	1.1	3.4	1.2
Has friends in department at end of CS401	105, 26	3.2	1.3	2.8	1.2
Goal of earning B or above at end of CS401	104, 26	.77	.42	.81	.40
Confidence in earning B or above at end of CS401*	105, 25	3.9	1.4	3.4	1.5
Interest in CS at end of CS445**	41, 8	4.2	.90	3.2	1.3
Likes CS environment at end of CS445	41, 8	3.0	1.2	2.8	1.5
Has friends in CS department at end of CS445	41, 8	3.2	1.3	3.1	1.5
Goal of earning B or above at end of CS445	41, 8	.63	.49	.38	.52
Confidence in earning B or above at end of CS445**	41, 8	3.5	1.4	2.4	1.5
* = marginally significant, $p < .10$ ** = $p < .05$					

Table 3. Gender Differences in Selected Factors

Computing experience and access to computers. Having a high level of access to computers in the home during high school predicted performance in CS401 for men (*rho* = .20) and overall (*rho* = .19), consistent with prior findings on the importance of home access to computers (Byrne & Lyons, 2001). Supporting prior research that demonstrated the importance of prior computing experience for CS performance at the post-secondary level (e.g., Byrne & Lyons, 2001), the number of informal computing activities that the student engaged in predicted male performance in CS401 (*rho* = .18). In addition, the number of specific programming constructs that the student had prior exposure to predicted male performance in CS401 (*rho* = .19) and in CS1501 (*rho* = .37). Curiously, an ANOVA revealed a significant cross-over interaction between gender and number of CS courses taken prior to CS401 as predictors of letter grade for this course, whereby women who took more programming courses before CS401 did more poorly than those who took fewer courses, while the reverse was true for men, $F(1, 152) = 3.4$, $p < .07$. (For this analysis, we redefined "number of CS courses that involved programming," as specified in Appendix A, into a categorical variable with 1-2 courses meaning "low" and 3 or more courses meaning "high.")

We hypothesized that women who had developed their programming skills while in high school, either by taking CS courses or on their own, might have done so at the expense of developing other skills that predict CS achievement, particularly math skills. As an initial test of this hypothesis, we performed an ANOVA of prior CS courses and the number of Calculus courses that women had taken before entering CS401 with the grade that women earned in this course as the dependent variable. (For this test, we redefined number of Calculus courses taken as a categorical variable, with 0 or 1 meaning "low" and 2-3+ meaning "high." We used the same specification of number of prior CS courses taken as described in the preceding paragraph—1-2 meaning "low" and 3+ meaning "high.") Although the number of prior CS courses taken marginally predicted female performance in CS401 ($F(1,24) = 3.6$, $p = .07$), there was no interaction between this variable and number of Calculus courses taken. However, when we ran the same test using number of Calculus courses taken and amount of exposure to specific programming constructs as predictor variables, we observed a significant cross-over interaction between the predictor variables ($F(1,23) = 4.2$, $p = .05$) as well as a positive relation between number of Calculus courses taken and CS401 grade ($F(1,23) = 6.5$, $p < .05$). Indeed, women who had low exposure to specific programming constructs but one or more Calculus courses performed nearly as well as women who had high exposure to specific programming constructs and several Calculus courses. The poorest performers were women who reported a high level of programming exposure and a low number of Calculus courses. The high correlation between Math SAT score and female performance in CS401 (.60) further supports the proposed importance of math preparation for CS achievement, particularly for women. These findings suggest that further research is needed to specify the relative importance of pre-college math and CS training for CS achievement at the undergraduate level and to determine if there are gender differences in what constitutes optimal preparation.

Encouragement. Having a mentor or role model, according to students' self-reports, predicted CS401 performance across genders, *tau-b* = .14.

Contrary to our expectations, parental encouragement *negatively* correlated with achievement in the second course, CS445. Specifically, maternal encouragement was negatively related to CS445 achievement overall (*rho* = -.33) and for each gender (*rho* = -.29 for men and *rho* = -.61 for women). Paternal encouragement was negatively related to CS445 achievement overall (*rho* = -.23). These findings suggest that at least some under-prepared students persist in a CS major in part because their parents encourage them to do so. The influence of parental encouragement on achievement and persistence is an interesting topic for further research.

Indicators of achievement. Several factors correlated with performance, though we suspect that they are more the result of students' perceptions that they are doing well rather than contributors to success. For example, a belief that CS is interesting at the end of CS401 correlated with male performance in this course (*rho* = .34), and with overall performance (*rho* = .22), while a belief that CS is interesting at the end of CS445 correlated with female performance (*rho* = .93) and overall (*rho* = .37).

Similarly, confidence at the end of courses may be as much a reflection of achievement as a contributor. Confidence in earning a B at the end of CS401 correlated with CS401 performance for men (*rho* = .71), women (*rho* = .41), and overall (*rho* = .65). Confidence in earning a B or above in CS445 correlated with female performance in this course (*rho* = .47), and overall (*rho* = .27). Apparently, students are fairly good judges of their performance status.

Another reflection of good performance is students' attitude towards the CS department at the university. This factor correlated with CS401 performance for males (*rho* = .31), and overall (*rho* = .19) and with male, female and overall performance in CS445 (*rho* = .41, *rho* = .87, and *rho* = .51, respectively).

Other Factors that Predict Persistence

Since, as we have shown, persistence is highly correlated with achievement, we can expect high overlap between the factors that predict achievement and the factors that predict persistence. Indeed, this appears to be the case. In the analysis reported below, we treated persistence rating as a categorical variable, with 1-2 meaning "low persistence" (students who left the program after CS401 or CS445) and 3 meaning "high persistence" (students who continued onto the upper-level courses, CS1501 and/or CS1502). This categorization was motivated by our observation that students who take three or more CS courses tend to continue with the major. As several students stated on the end-of-term surveys, they persisted with CS in part because they had already invested considerable time and effort.

If we look at students in the aggregate (without distinguishing by gender), without filtering by grade, we find several significant differences between students who persist beyond CS445 and students who do not. Students who took 3 or more CS courses had higher math SAT scores, $t(198) = 2.6$, and verbal SAT scores, $t(198) = 1.8$, $p < .08$. They reported taking more Calculus courses, $t(154) = 2.7$ and more CS courses, $t(154) = 2.1$. Consistent with a higher number of CS courses, higher-persisting students reported more prior exposure to specific programming constructs, $t(153) = 3.2$. They were more likely than students who left the program early on (after CS401 or CS445) to agree that CS is interesting at the end of CS401 and CS445—$t(129) = 2.9$ and $t(47) = 2.0$, $p < .06$, respectively; to agree that they liked the CS department at the end of the same courses—$t(129) = 2.2$ and $t(47) = 2.1$—and to state that they were confident about earning a B in CS401 and CS445 at the end of these courses—$t(128) = 4.3$ and $t(47) = 1.8$, $p < .08$, respectively. As we have seen, these factors (SAT scores, number of Calculus courses taken, prior computing experience, confidence, interest in CS, and a positive attitude towards the CS department) also predicted achievement.

What this analysis can not explain is why it is that a high number of promising students (earning B or above) leave the program, as shown in Figure 2. Are there specific factors that predict whether successful students stay in the program as opposed to leave? Given that the most significant loss of high-achieving students occurs after CS401—23% (5 women and 20 men out of 109 high-achieving CS401 students; Figure 1)—and that the number of high-achieving students who left the program after CS445 was too small for statistical analysis (1 woman and 6 men out of 51 high-achieving CS445 students), we focused the analysis that follows on attrition after CS401.

For both genders, B-or-above students who persisted to CS445 had more prior programming experience than B-or-above students who left the program, as evident in the number of programming courses taken before CS401, $t(94) = 2.8$, and amount of exposure to specific programming skills, $t(94) = 2.8$. High-achieving male students who persisted to CS445 had taken more programming courses than did high-achieving male students who left the program at this point, $t(78) = 2.8$, and reported exposure to more programming skills, $t(78) = 2.8$). High-achieving female students who persisted also reported more exposure to specific programming skills than did women who left, $t(14) = 2.2$. Not surprisingly, across both genders, high-achieving students who continued to CS445 were more likely to report that they found CS interesting at the end of CS401 than were high-achieving students who left, $t(78) = 2.8$.

What distinguishes women who stayed in the program from those who left after CS401, at the B-or-above level, is apparently the need for support and encouragement. Relative to high-achieving women who left, high-achieving women who persisted were more likely to claim that they had friends in the CS department, $t(13) = 2.3$, strong maternal encouragement to major in CS, $t(14) = 3.8$, as well as paternal encouragement, $t(14) = 2.5$ to do so. So, although parental encouragement to persist may be detrimental to weaker students (B- or lower), as suggested by the negative relation between parental encouragement and CS445 grades discussed previously, it seems to be an important factor in keeping strong female students in the program. We did not observe significant relationships between friendship, encouragement, and persistence among high-achieving male students.

Insights about Persistence Drawn from Students' Responses to Open-ended Questions

When we looked more closely at the survey responses of "good" students who left the program, we gained further insight into why it is that these students leave. Two of the main reasons were a loss of confidence that led to a loss of interest in the program, and a direct loss of interest in the program—that is, unmediated by a loss of confidence. The first type of relationship is discussed and illustrated extensively in Margolis and Fisher (2002).

Loss of confidence, possibly followed by a loss of interest, can stem from a gap between a student's performance expectations and actual performance (Volet & Styles, 1992). For example, one woman entered CS401 with the goal of earning a B or above. However, she got a 75 on the course's second exam, which might have shaken her confidence. On the end of term survey for this course, this woman showed that she was also among those in Margolis and Fisher's study who compared their required effort with that of their peers (mostly males), and concluded that CS was not right for them:

> *I don't think I have the skills to go on as a CS major. The programs that are assigned are incredibly difficult for me, while for others, they are very simple.*

Sometimes there was no indication of a loss of confidence, because the student earned high marks on each course exam administered prior to the end of term surveys. However, despite A-level performance, some students nonetheless left the program because they simply lost interest in the discipline. For example, one women with a strong A in CS401 gave the following reason for her consideration of leaving:

> *I found Java to be easy to learn as long as I paid attention and attended, yet not interesting at all. So, I'm not sure I have a lot of interest, but will probably stick with Computer Science.*

Apparently, she did not "stick with it." We heard similar comments about loss of interest from B-or-above male students who left, despite no apparent reasons for a loss of confidence, e.g.:

> *I feel confident in my programming ability and have an A in the class. However, I do not believe that I would be able to program all of my life in front of a computer. I switched to information science, which is still related to programming, just not as intense.*

This student's comment also represents another apparent reason for the loss of "good" students: disillusionment with what they perceive the discipline entails, whether or not their perceptions are entirely accurate. (For example, many computer scientists spend little time programming.) Another male student commented that he planned to leave the major because he did not want to take higher-level math courses.

These comments suggest the importance of sustaining students' interest in the program and of providing them with an accurate portrait of what the discipline entails and can offer with respect to career opportunities. The reasons that the handful of women who persisted through all three courses gave for staying underscore the importance of interest and an awareness of career opportunities. Four of these women talked about using their degree in computer science to complement or serve another career interest, such as biological research, computational physics, or neuroscience. This interest was acquired in high school or early in college, suggesting the importance of developing students' awareness of CS applications in high school and post-secondary CS courses. For example:

> *I am interested in technology in biological research and, while working in the research setting, I found a great need for computer knowledge and programming skills. Also, I enjoy programming and solving problems.*

Women tend to have an applied interest in computer science, while men tend to like computers and programming for their own sake (Margolis & Fisher, 2002). The only applied interest we found expressed among the B-and-above male students was video game programming.

Conclusions

This study investigated the factors that predict performance and persistence in an undergraduate CS program and explored why it is that even high-achieving students leave the undergraduate "CS pipeline." Not surprisingly, we found that several factors that predict achievement also predict persistence beyond the first two courses in an undergraduate CS major, such as math and verbal SAT score, the number of Calculus courses taken, and prior computing experience. Home access to a computer and having a mentor or role model during high school also predicted achievement.

Following related research (e.g., Beyer et al., 2003; Charlton & Birkett, 1999), this study demonstrates the value of investigating the interactions between variables that predict achievement and persistence in computer science. By doing so, we observed several complex and unexpected interactions between student characteristics (including gender), achievement, and persistence. Most notably, male students who earned less than a B in the introductory CS course (CS401) were more likely to take the next course in the curriculum for CS majors than were women who earned less than a B. Achievement is a factor in even high-achieving students' decision to leave CS; loss of interest can accompany loss of confidence. To take another example of interactions between background factors, women who reported more exposure to particular programming constructs before CS401 performed worse than women who entered with less programming experience. Upon further investigation, we found evidence to support our conjecture that substituting the development of math skill (a consistently strong predictor of achievement in CS programs) with programming skill development during high school may be detrimental to performance in an undergraduate CS program, especially for women.

Given the small sample size used in this study (especially of women) and the fact that it was conducted at one institution, there is a need for further research to test the validity and generalizability of our findings, and to uncover the reasons for the complex interactions and unexpected correlations we observed. For example, why are male students who perform below the B level more likely to persist to the next course in a CS program than women who perform at the same level? Is this a reflection of less confidence on women's part, better judgment, or both? At what point, and under what circumstances, should academic advisors suggest that students pursue a different major that they are more likely to excel in, as opposed to repeating CS courses? What is the optimal combination of pre-college math and CS training? What, if anything, can be done to sustain the confidence and interest of promising students whose grades do not rise up to their performance goals and expectations? And how can we sustain the interest of all students through the early stages of CS instruction, when a large body of rudimentary skills and concepts that are often not inherently interesting must be covered? We expect that investigation into these issues will lead to a deeper understanding of why many students of both sexes, even some of the most promising ones, leave the "CS pipeline" at the undergraduate level, and what can be done to reverse this trend.

References

Beyer, S., Rynes, K., and Haller, S. (2003). "What Deters Women from Taking Computer Science Courses?" *IEEE Technology & Society Magazine*.

Beyer, S., Rynes, K., Perrault, J., Hay, K., and Haller, S. (2003). Gender differences in computer science students. *Proceedings of the Thirty-fourth SIGSCE Technical Symposium on Computer Science Education*, pp. 49-53.

Bureau of Labor Statistics (BLS) (2004). *Occupational Outlook Handbook, 2004-2005 Edition*, U.S. Department of Education, Bureau of Labor Statistics, http://www.bls.gov.

Bunderson, E.D. and Christensen, M.E. (1995). "An Analysis of Retention Problems for Female Students in University Computer Science Programs," *Journal of Research on Computing in Education*, Vol. 28, No. 1.

Butcher, D.F. & Muth, W.A. (1985). Predicting Performance in an Introductory Computer Science Course," *Communications of the ACM,* Vol. 28, No. 3, pp. 263-268.

Byrne, P. and Lyons, G. (2001). "The Effect of Student Attributes on Success in Programming," *Proceedings of the 6th Annual Conference on Innovation and Technology in Computer* Science, pp. 49-52.

Cafolla, R. (1987). "Piagetian Formal Operations and Other Cognitive Correlates of Achievement in Computer Programming," *Journal of Educational Technology Systems*, Vol. 16, No. 1, pp. 45-55.

Camp, T. (1997). "The Incredible Shrinking Pipeline," *Communications of the ACM*, Vol. 40, No. 10, pp. 103-110.

Camp, T., Miller, K., and Davies, V. (1999). "The Incredible Shrinking Pipeline Unlikely to Reverse," http://www.mines.edu/fs_home/tcamp/new-study/new-study.html

Charlton, J.P. and Birkett, P.E. (1998). "Psychological Characteristics of Students Taking Programming-oriented and Applications-oriented Computing Courses," *Journal of Educational Computing Research*, Vol. 18, No. 2, pp. 163-182.

Charlton, J.P. and Birkett, P.E. (1999). "An Integrative Model of Factors Related to Computing Course Performance," *Journal of Educational Computing Research*, Vol. 20, No. 3, pp. 237-257.

Clement, C.A., Kurland, D.M., Mawby, R., and Pea, R.D. (1986). "Analogical Reasoning and Computer Programming," *Journal of Educational Computing Research,* Vol. 2, No. 4.

Cohoon, J.M. (2001). "Toward Improving Female Retention in the Computer Science Major," *Communications of the ACM,* Vol. 44, No. 5, pp. 108-114.

Cohoon, J.M. (2002). "Women in CS and Biology," *Proceedings of SIGSCE '02*, pp. 82-86.

Crews, T. and Butterfield, J. (2003). "Gender Differences in Beginning Programming: An Empirical Study on Improving Performance Parity," *Campus-wide Information Systems,* Vo.20, No. 5, pp. 186-192.

Davies, A.R., Klawe, M., Ng, M., Nyhus, C., and Sullivan, H. (2000). Gender Issues in Computer Science Education, http://www.wcer.wisc.edu/nise/News_Activities/Forums/Klawepaper.htm

Eccles-Parsons, J. (1984). "Sex Differences in Mathematics Participation," in Steinkamp, M.W., and Maehr, M.L. (Eds.), *Advances in Motivation and Achievement, Volume 2, Women in Science,* Greenwich, Connecticutt: JAI Press, Inc.

Falk, J., Drayton, B., Crawford, C., and Obuchowski, J. (1999). "Eyes to the Future: A Multi-age Mentoring Program for Girls," *Leadership and the New Technologies Perspectives*, http://www2.edc.org/LNT/news/Issue8/feature2.htm (accessed March/April 1999)

Fan, T., Li, Y., and Niess, M. (1998). "Predicting Academic Achievement of College Computer Science Majors," *Journal of Research on Computing in Education*, Vol. 31, No. 2, pp. 155-72.

Foorman, B., Yoshida, H., Swank, P., and Gatson, J. (1985). "The Effect of Visual and Verbal Strategies on Children's Solutions of Figural Matrices in Japan and the United States," Paper presented at the *Annual Meeting of the American Educational Research Association*, Chicago, IL.

Freeman, P. and Aspray, W. (1999). "The Supply of Information Technology Workers in the United

States," *Computing Research Association Report*, Washington, DC: Computing Research Association.

Gürer, D. and Camp, T. (2002). An ACM-W Literature Review on Women in Computing," *Inroads (SIGSCE Bulletin)*, Vol. 34, No. 2, pp. 121-127.

Jagacinski, C.M., LeBold, W.K., and Salvendy, G. (1988). "Gender Differences in Persistence in Computer- related Fields," *Journal of Educational Computing Research,* Vol. 4, No. 2.

Jones, P.K. (1987). "The Relative Effectiveness of Computer-assisted Remediation with Male and Female Students," *T.H.E. Journal*, Vol. 14. No. 7, pp. 61-63.

Katz, S., Aronis, J., Allbritton, D., Wilson, C., and Soffa, M.L. (2003). "Gender and Race in Predicting Achievement in Computer Science," *IEEE Technology and Society Magazine*, Vol. 22, No. 3, pp. 20-27.

Lai, S. and Repman, J. (1996). "The Effects of Analogies and Mathematics Ability on Students' Programming Learning Using Computer-based Learning," *International Journal of Instructional Media,* Vol. 23, No. 4, pp. 355-364.

Levin, T. and Gordon, C. (1989). "Effect of Gender and Computer Experience on Attitudes Towards Computers," *Journal of Educational Computing Research,* Vol. 5, No. 1, pp. 69-88.

Mahony, K. and Van Toen, B. (1990). "Mathematical Formalism as a Means of Occupational Closure in Computing—Why "Hard" Computing Tends to Exclude Women," *Gender and Education*, Vol. 2, No. 3, pp. 319-333.

Margolis, J. and Fisher, A. (2002). *Unlocking the Clubhouse: Women in Computing,* Cambridge, MA: M.I.T. Press.

Margolis, J., Fisher, A., and Miller, F. (1999). "Caring About Connections: Gender and Computing," *IEEE Technology and Society Magazine*, Vol. 18, No. 4, pp. 13-20.

Miura, I.T. (1987). "The Relationship of Computer Self-efficacy Expectations to Computer Interest and Course Enrollment in College," *Sex Roles*, Vol. 16, No. 5, p. 1985.

Myers, J. (1999). "Women in High Tech Fields in Science and Technology in British Columbia: Fact Sheet and Summary," Prepared for SCWIST/WISTTE Steering Committee, http://taz.cs.ubc.ca/wistte/exec summary.pdf

National Center for Educational Statistics (NCES) (2001). *Digest of Educational Statistics 2000*, US Department of Education, Office of Educational Research and Improvement, Washington, DC, NCES 2001-034.

National Science Foundation (NSF) (2000). *Women, Minorities and Persons with Disabilities in Science and Engineering,* National Science Foundation report, http://www.nsf.gov/cgi-bin/getpub?nsf00327

Pearl, A., Pollack, M.E., Riskin, E., Thomas, B., Wolf, E., and Wu, A. (1990). "Becoming a Computer Scientist: A Report by the ACM Committee on the Status of Women in Computing Science," *Communications of the ACM*, Vol. 33, No. 11, pp. 47-57.

Shashaani, L. (1997). "Gender Differences in Computer Attitudes and Use Among College Students," *Journal of Educational Computing Research*, Vol. 16, No. 1, pp. 37-51.

Spertus, E. (1991). "Why are There So Few Female Computer Scientists?" *AIT Technical Report 1315*, Cambridge: MIT Artificial Intelligence Laboratory.

Roberts, E. (2003). "Expanding the Audience for Computer Science." *Proceedings of the Thirty-fourth SIGSCE Technical Symposium on Computer Science Education*.

Taylor, H.G. and Mounfield, L.C. (1989). "The Effect of High School Computer Science, Gender, and Work on Success in College Computer Science," *Proceedings of the 20th SIGSCE Technical Symposium on Computer Science Education*, pp. 195-198.

Taylor, H.G. and Mounfield, L.C. (1994). "Exploring the Relationship Between Prior Computing Experience and Gender on Success in College Computer Science, " *Journal of Educational Computing Research*, Vol. 11, No. 4, pp. 291-306.

Volet, S.E. and Styles, I.M. (1992). "Predictors of Study Management and Performance on a First-year Computer Course: The Significance of Students' Study Goals and Perceptions," *Journal of Educational Computing Research*, Vol. 8, No. 4, pp. 423-449.

West, M. and Ross, S. (2002). "Retaining Females in Computer Science: A New Look at a Persistent Problem," *Journal of Computing in Small Colleges*, Vol. 17, No. 5.

Zweben, S. and Aspray, W. (2004). "2002-2003 Taulbee Survey," *Computing Research News,* Washington, DC: Computing Research Association, http://www.cra.org/statistics/survey/03/03.pdf

About the Authors

Sandra Katz received her Doctor of Arts degree in English from Carnegie Mellon University in 1985 and her M.S. in Information Science from the University of Pittsburgh in 1989. She is currently a Research Associate at the University of Pittsburgh's Learning Research and Development Center. Her published work includes research on instructional dialogue, the under-representation of women and minorities in computer science, and intelligent tutoring systems.

David Allbritton received his PhD in cognitive psychology from Yale University in 1992 and is currently Associate Professor of Psychology at DePaul University. His published work includes research on language comprehension, memory, instructional technology, tutoring, and intelligent tutoring systems.

John Aronis received his B.A. in Mathematics from the State College of New York at New Paltz in 1979 and his M.S. in Mathematics from Syracuse University in 1981. After working several years as a Research Programmer at Carnegie Mellon University, he entered the Intelligent Systems Program at the University of Pittsburgh and earned his Ph.D. in 1993. He was a Research Associate in the Department of Computer Science at the University of Pittsburgh until being appointed as a Lecturer in 1999.

Christine Wilson received her B.S. degree in computer science and M.S. degree in education from the University of Pittsburgh. She is presently employed as a computer programmer at the University of Pittsburgh's Learning Research and Development Center. She contributes to research on the under-representation of women and minorities in computer science and on natural-language tutorial dialogue systems.

Mary Lou Soffa is the Owen R. Cheatham Professor and Chair of the Department of Computer Science at the University of Virginia. She moved from the University of Pittsburgh to Virginia in September, 2004. She received her Ph.D. in computer science from the University of Pittsburgh in 1977. Her research interests include optimizing and parallelizing compilers, program analysis, and software tools for debugging and testing programs. She has been active for many years on improving the participation of women in computer science.

Appendix A: Background Characteristics Considered in the Analysis

Factor	Relevant Survey Question(s)	Values used in the Analysis
Math and Computing Background		
Calculus courses completed	Which Calculus courses have you completed? Check all that apply: • Currently taking Calculus • Calculus 1 • Calculus 2 • Calculus 3	0 = none (no selection) or currently taking Calc 1 = Calculus 1 only 2 = through Calculus 2 3 = through Calculus 3
Number of CS courses taken that involved programming	How many computer science courses did you take in high school? Please count only those courses that involved programming, as opposed to learning how to use software applications: • 0 • 1 • 2 • 3 How many computer science courses have you taken during college? (Same specification and options as the preceding question.)	Sum of programming-oriented CS courses taken before CS401, in high school and college. 3+ counted as "3"
Access to Computers		
Access to computers in home, while in high school	Did you or your family own a computer while you were in high school? • Yes • No … Skip next question Select the phrase that best describes your access to the computer in your home during high school: 1. I had to compete with other family members to use the computer; it was seldom available when I wanted to use it. 2. I shared the computer with other family members. It was sometimes unavailable. 3. I shared the computer with other family members, but it was usually available when I wanted to use it. 4. I was the sole computer user.	0 = "no" to question about home ownership of computers 1 = limited access; options 1 or 2 2 = adequate access; options 3 or 4
Availability of computers at high school	Select the phrase that best describes the availability of computers while you were in high school: 1. There were no computers for students to use at my high school. 2. I usually had to compete for access to a computer at school. 3. I sometimes had to compete for access to a computer at school. 4. A computer was usually free when I wanted to use one at school.	1 = limited availability; options 1 or 2 2 = adequately available; options 3 or 4
Informal activities that potentially develop programming skill	Which of the following informal computer activities did you participate in during high school or any time prior to taking CS401? Check all that apply: • Computer club • Computer camp • Email and chat rooms • Playing computer games • Programming for fun	Number of the following items selected: • Programming for fun • Building web pages • Building a computer or taking…

	• Building web pages • Browsing the Internet • Building a computer or taking a computer apart to explore how it works	
Exposure to specific programming constructs	Before taking CS401, which of the following could you write programs with? Check all that apply: • “if” statements • loops • arrays • linked lists • recursion • objects, classes, and inheritance • more than 100 lines of code • more than 300 lines of code	Number of items selected
Performance Goals and Expectations		
Goal of earning B or above	Select the phrase that best describes your performance goals for CS401: 1. Pass (earn a D or above) 2. Earn a C or above 3. Earn a B or above 4. Earn an A 5. Do more than is needed to earn an A	0 = no; selected items 1 or 2 1 = yes; selected items 3, 4, or 5
Confidence in earning B or above	How confident are you in your ability to earn a final grade of B or above?	Scalar selection; 1 = not confident at all; 5 = very confident
Motivation		
Reason for pursing CS	In one or two sentences, please describe what most influenced your decision to major in computer science or to consider majoring in CS.	Open-ended
Encouragement		
Mentor or role model	Did anyone serve as a mentor to you, with respect to computer science? (A mentor is someone who is knowledgeable and who takes an active interest in a person’s career development.) • Yes • No Did anyone serve as a role model for you, with respect to computer science? (A role model inspires a person, but does not directly help him or her to achieve a career goal.) • Yes • No	0 = no to both questions 1 = yes to either question
Maternal encouragement	Please rate the degree to which your mother (or step-mother) encouraged you in your decision to major in computer science, or to consider a CS major: 0 = Not applicable (NA) 1. Discouraged me 2. Neither discouraged nor encouraged me 3. Encouraged me somewhat 4. Encouraged me considerably 5. Encouraged me a great deal	Scalar selection.
Paternal encouragement	Same as preceding, for father (or step-father)	Scalar selection

Appendix B: End-of-course Survey Factors Considered in the Analysis

Factor	Survey Question	Values
Persistence and Motivation		
Plan to continue in CS	At the beginning of the semester, you stated that you are considering a major in CS. Do you still plan to major in CS? • Yes • No • Maybe	0 = no or maybe 1 = yes
Reasons for staying or leaving	Briefly explain your response to the preceding question. Why do you still plan to major in CS? Or, why are you considering a different major?	Open-ended
Interest in CS	Please rate the degree to which you agree with the following statement: I find CS interesting.	Scalar; 1 = strongly disagree; 5 = strongly agree
Attitude Towards the CS Environment at the University		
Likes CS environment	Please rate the degree to which you agree with the following statement: I like the CS environment at Pitt (i.e., lab facilities, faculty, other students, etc.).	Scalar; 1 = strongly disagree; 5 = strongly agree
Has friends in the department	Please rate the degree to which you agree with the following statement: I have several friends who are CS majors or prospective majors.	Scalar; 1 = strongly disagree; 5 = strongly agree
Performance Goals and Expectations		
Goal of earning B or above	Select the phrase that best describes your performance goals for CS401: 1. Pass (earn a D or above) 2. Earn a C or above 3. Earn a B or above 4. Earn an A 5. Do more than is needed to earn an A	0 = no; selected items 1 or 2 1 = yes; selected items 3, 4, or 5
Confidence in earning B or above	How confident are you in your ability to earn a final grade of B or above?	Scalar selection; 1 = not confident at all; 5 = very confident