Attrition in Introductory Computer Science at the University of California, Berkeley

Colleen Lewis

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2010-132
http://www.eecs.berkeley.edu/Pubs/TechRpts/2010/EECS-2010-132.html

October 17, 2010

Acknowledgement

I would like to acknowledge Marcia Linn and Michael Clancy for their helpful feedback on earlier drafts.

Attrition in Introductory Computer Science at the University of California, Berkeley

Colleen M. Lewis

Abstract

At the University of California, Berkeley, the introductory programming course, Computer Science 3: Introduction to Symbolic Programming, serves students with no programming experience and functions as a gatekeeper to further study of computer science or electrical engineering. This study documents patterns of attrition in this introductory computer science class from fourteen semesters, from the fall of 2002 to the spring of 2009. The purpose of the study is to document attrition and does not identify causal mechanisms for the patterns observed.

Introduction

Introductory computer science courses at the University level serve as gate-keepers to the STEM fields. While studies have addressed underrepresentation of females in the STEM fields as a whole, this study examines patterns of attrition within the introductory computer science course at the University of California, Berkeley.

The introductory computer science course offered at UC Berkeley is Computer Science 3: Introduction to Symbolic Programming (CS3). The enrollment of female students is typically much higher in CS3 than in computer science programs nationally. The increased enrollment of female students may be in part because the course is designed for students that will not necessarily focus on computer science. However, accompanying the increased representation of female students in CS3, the instructors for the course have observed that there appear to be higher rates of attrition amongst female students in CS3, which brought to light the research questions addressed:

- Controlling for what semester in the students' college career they took the course, students' major and the semester and year the course was taken, do female students have a higher odds of dropping Computer Science 3: Introduction to Symbolic Programming?

This study used the CS3 online curriculum database to analyze the attrition patterns for 14 semesters, from the fall of 2002 to the spring of 2009.

Background

National Context

Nationally, there has been attention to the underrepresentation of women in computer science. Although 2008 experienced the first increase in the percentage of female college-bound SAT-takers who intend to major in a Computer and Information Sciences major, the percentage of female students has decreased from a high in 1995 of 28% to only 13% in 2008 (College Board, 2008).

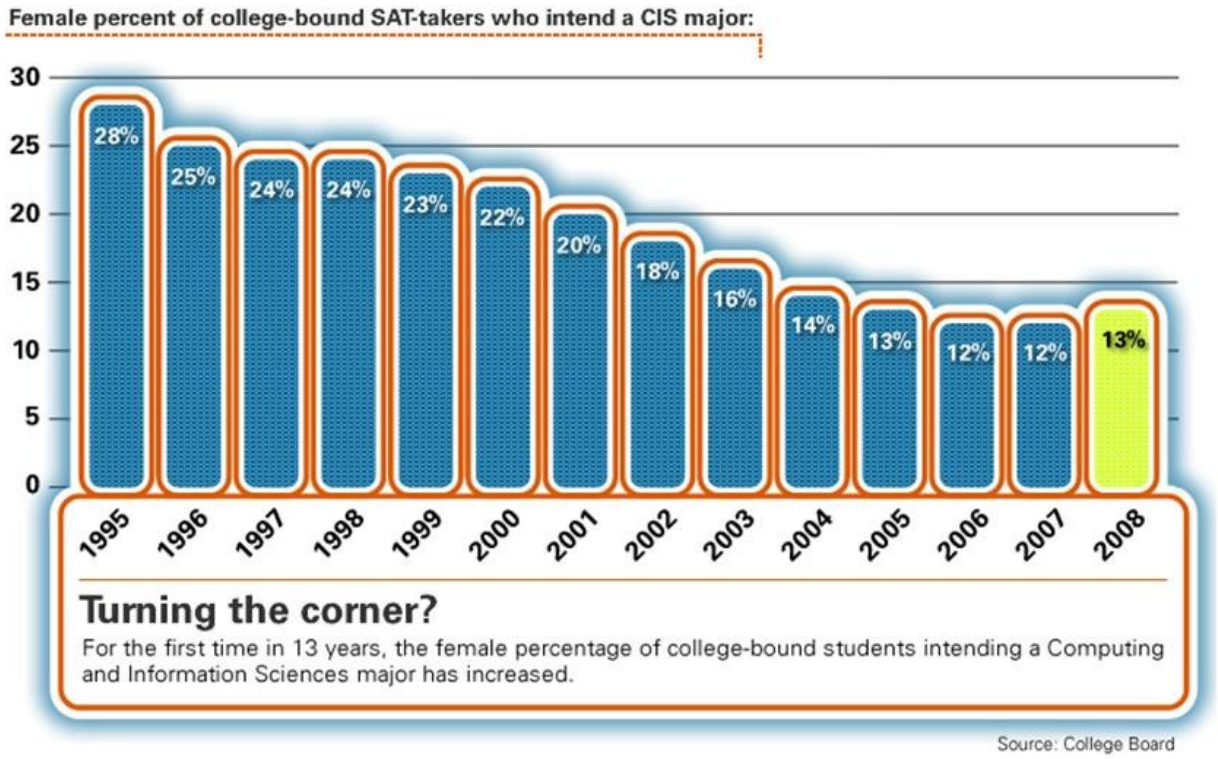


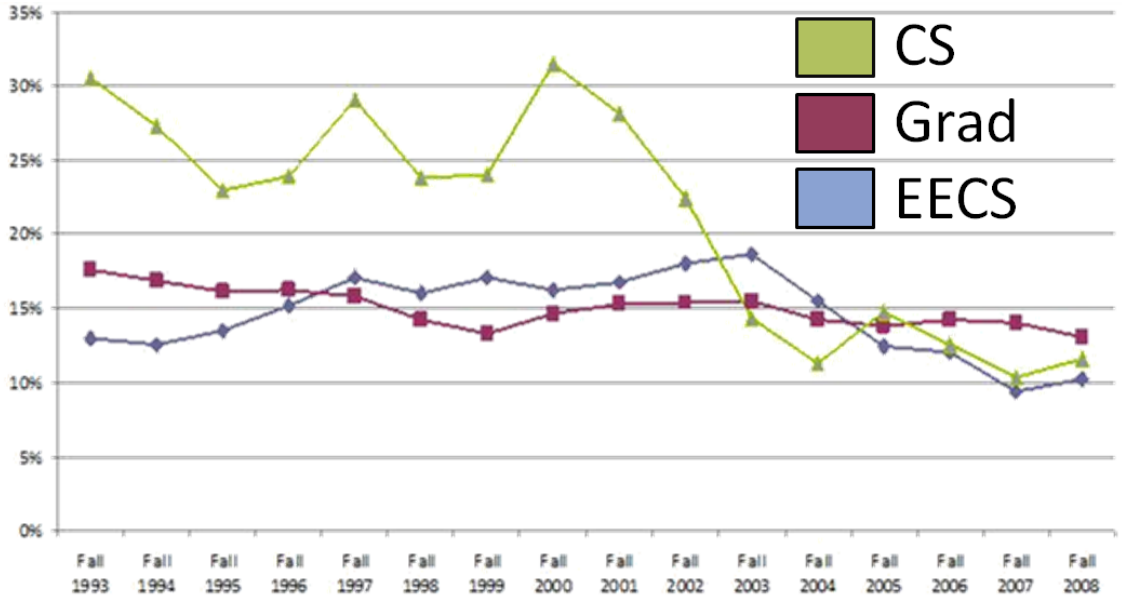
Figure 1: National data of students' reported interest in Computing and Information Science majors.

In addition to a decrease in student reported interest, the proportion of undergraduate computer science degrees awarded to women has decreased as well. In 1985, 37% of undergraduate CS degrees in the U.S. were earned by women, while in 2005 that number dropped to only 22% (Klawe, Whitney & Simard, 2009).

Context at UC Berkeley

Female representation at UC Berkeley in the graduate program and undergraduate majors of EECS and CS is lower than the national average of college-bound SAT-takers who intend to pursue a related major and UC Berkeley has experienced a corresponding decrease in female representation (Judson, 2008). Figure 2 shows the proportion of the respective majors who were female. The graph shows two undergraduate majors, computer science (CS) and electrical engineering and computer science (EECS). The graph also shows the representation within the graduate programs of computer science and electrical engineering.

Figure 2: Representation of female students at UC Berkeley in the undergraduate electrical engineering and computer science major (EECS), the undergraduate computer science major (CS), and the graduate EECS program (Grad).



CS3 at UC Berkeley

The computer science "61-series", CS61a, CS61b, and CS61c, is required for both the College of Engineering Electrical Engineering and Computer Science major (EECS) and College of Letters and Science Computer Science major (CS). In addition, students in the Cognitive Science major (CogSci) are required to take the first of these courses, *CS61a – Structure and Interpretation of Computer Programs*. This course CS61a requires a level of proficiency with computer programming, as might be gained from the opportunity to enroll in Advanced Placement (AP) computer science in high school. While in the past there has been an entrance exam, students may now choose to begin in CS61a regardless of having taking the AP computer science exam.

The computer science department has had a longstanding goal to provide alternative paths to prepare students for CS61a. The traditional path to CS61a, of taking the AP computer science test, suffers from little participation by populations that are typically under-represented in computer science. In 2007, only 17% of AP computer science test-takers were female. In the same year, black, Hispanic, and Latino students comprised only 10.2% of all AP computer science test-takers (College Board, 2008). To broaden access to the EECS, CS and CogSci majors, UC Berkeley provides an alternative to AP computer science as preparation for CS61a.

The course CS3 was created to prepare students for CS61a who have no background in programming. The course is designed in particular to help these students succeed in the subsequent CS61a by providing an introduction to the programming language used in CS61a as well as a foundation in some of the topics covered. In addition to supporting the population that intends to take CS61a, the course can be used to fulfill a technical requirement in the college of natural resources. In the past it has also fulfilled a requirement for business administration students.

The participation of female students in CS3 has been higher than the College Board data would predict for an introductory computer science course, possibly because CS is accessible to non-majors. Between the fall of 1997 and the fall of 2008, CS3 averaged 31.4% enrollment by female students. Figure 3 shows the representation of female students by semester, including summer school courses. The representation within the EECS and CS majors of less than 15% appears to highlight a missed opportunity in terms of attracting female students in CS3 to the major.

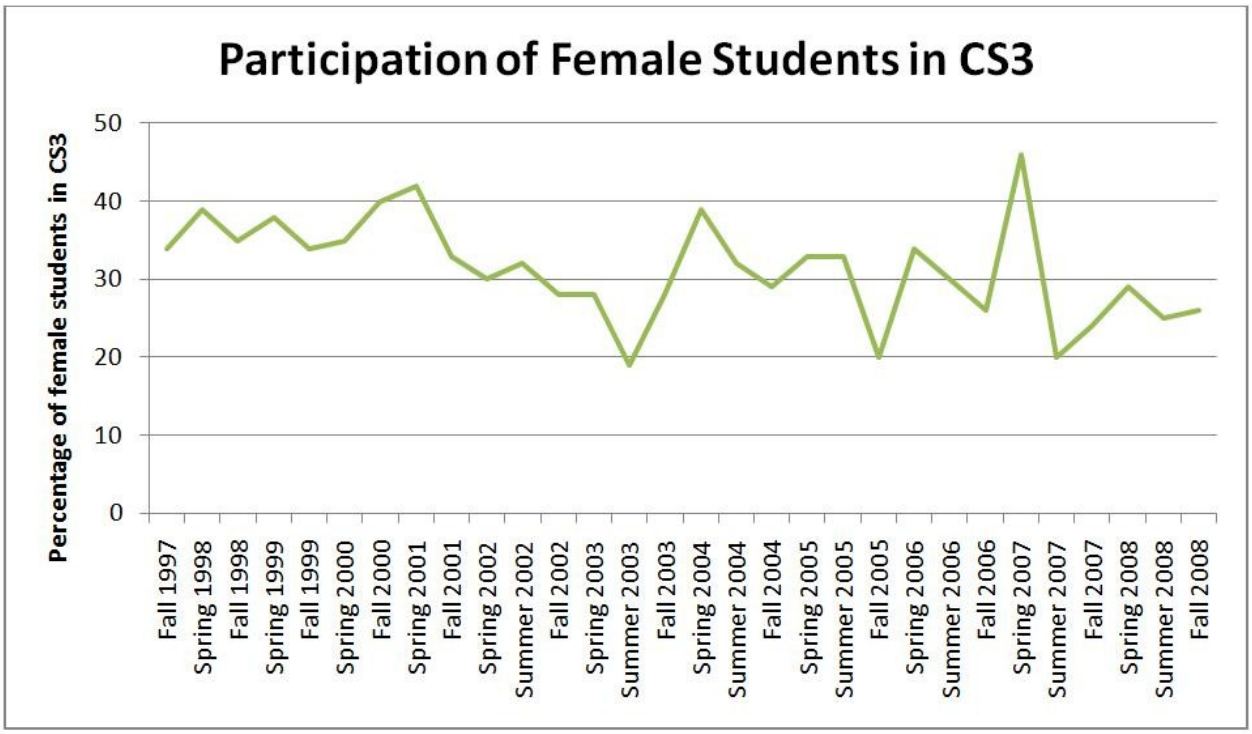


Figure 3: Representation of female students in CS3 from the fall of 1997 to the fall of 2008

Enrollment in CS3 echoes the national trend of decreased participation of female students. Enrollment of female students in CS3 for a letter grade has decreased from the fall of 1997 to the spring of 2008, significant at the 1% level (p=0.008). However, each course offering accounts for only a 0.35% decrease in representation of female students.

Research Question

Controlling for what semester in the students' college career they took the course (1-10), students' major and the semester and year the course was taken, do female students have a higher odds of dropping *Computer Science 3: Introduction to Symbolic Programming*?

Methods

The data come from 14 offerings of the course CS3, from the fall of 2002 to the spring of 2009. Data were gathered from the online-curriculum database rather than official University registration information. All students who attend a CS3 laboratory section create an account, regardless of whether they are ever

officially enrolled in the course. The number of students who have created an account is assumed to be different than the official enrollment, however attending a laboratory section for the course is used as an indicator for interest in the course regardless of official enrollment. These students with an account are considered to have dropped if there is no score recorded for the final.[1] [2]

Upon creating an online account, students were asked to indicate their gender, year in school and major. Dummy variables were created for these three categorical variables (`gender`, `year_in_school`, and `major`). When students indicated their gender they had two options, coded as `male` and `female`. When asked to indicate their level in school, students had 5 options, coded as `freshman`, `sophomore`, `junior`, `senior`, and `grad`. Students' response to their level in school is converted to a continuous variable, `semester_num`, representing what semester in college they were in when they took the course, ranging from 1 to 10. A student that is coded as a freshman would have either a 1 or 2 for `semester_num`, depending upon whether they took it in the fall or spring. During the fall semester the variable `semester_num` can take the values 1, 3, 5, 7 or 9. During the spring semester the variable `semester_num` can take the values 2, 4, 6, 8, or 10.

Students were given the option of choosing 6 classifications of their major: computer science (`lscs_major`), electrical engineering and computer science (`eecs_major`), cognitive science (`cogsci_major`), business (`business_major`), other (`other_major`) or decline to state (`undeclared_major`).[3] Students have been dropped from analysis if they declined to state their gender or year in school. Students who declined to state their major are assumed to have no major and are classified as "undeclared" (`undeclared_major`). Table 1 shows a summary of all categorical variables.

Table 1: Description of categorical variables

Question	Options	Dummy variable
Gender	Female	`female`
Year in School	Freshman	`freshman`
	Sophomore	`sophomore`
	Junior	`Junior`
	Senior	`Senior`
	Graduate Student	`Grad`
Major	Electrical Engineering and Computer Science (EECS)	`eecs_major`
	Computer Science	`lscs_major`
	Cognitive Science	`cogsci_major`

[1] At the time of data collection the spring 2009 semester was in process. For this semester, a student is considered to have dropped the course if they have no grade for the last 3 homework assignments.

[2] It is possible, however assumed to be rare, that students dropping the course decide instead to take the next course in the sequence, Computer Science 61A. These students that they are prepared for CS61A are, at this point, unavoidably included in the analysis.

[3] These majors are used as the options because of the typical demographics of the course. The course CS3 is not required for students of any major. It is recommended for students who are required to take CS61a, which includes students majoring in electrical engineering and computer science (EECS), computer science and cognitive science. Students who are majoring in business are not required to take CS61a, but can take CS3 as one of a few options to complete a requirement.

	Business	`business_major`
	Other	`other_major`
	Not given (assumed to be undeclared)	`undeclared_major`
Semester	Fall 2002 – Spring 2009	`offering1` … `offering14`

To investigate the research question, a logistic regression was run for whether students dropped the course on the variables for gender, semester in the students' college career that they took the course, major and the semester and year. The reference group is from the fall of 2002, male freshman students, majoring in electrical engineering and computer science. All other dummy variables were included in the logistic regression (`female`, `lscs_major`, `cogsci_major`, `business_major`, `other_major`, `undeclared_major`, `offering2` … `offering14`).

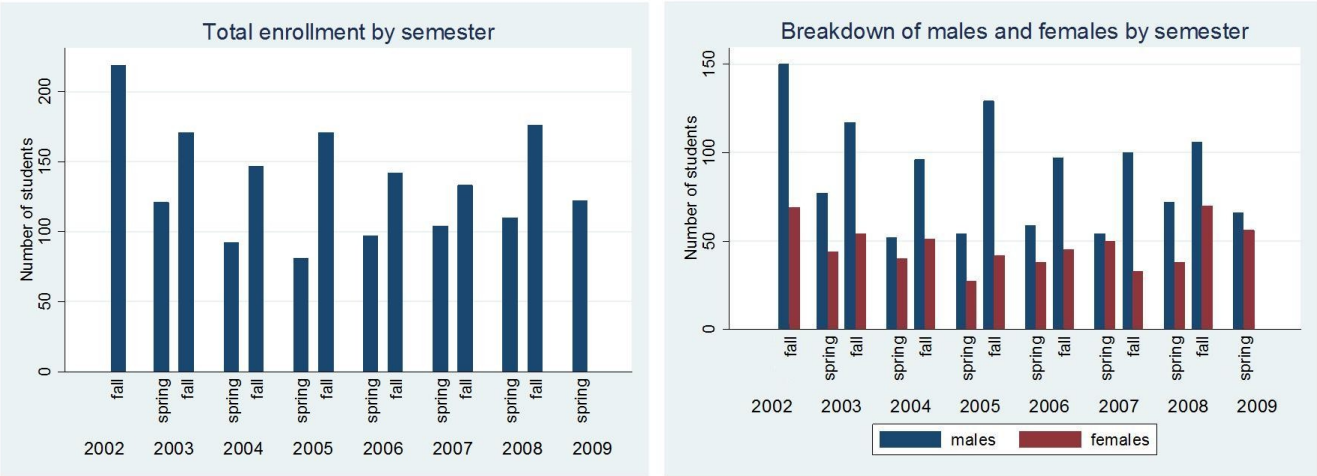
Results

Over the fourteen semesters that the course has been offered using the online curriculum, there were a total of 2209 students. Students were dropped from analysis if they did not report both their gender and year in school. These qualifications caused 323 students to be dropped from analysis, with the final data accounting for 85.4% of the original population. The data considered in this analysis includes 1886 students.

Enrollment by semester

As expected, enrollment varies by semester. Figure 4 shows how the number of students varies by semester and further how the number of male and female students varies by semester.

Figure 4: Total enrollment per semester and enrollment of male and females by semester



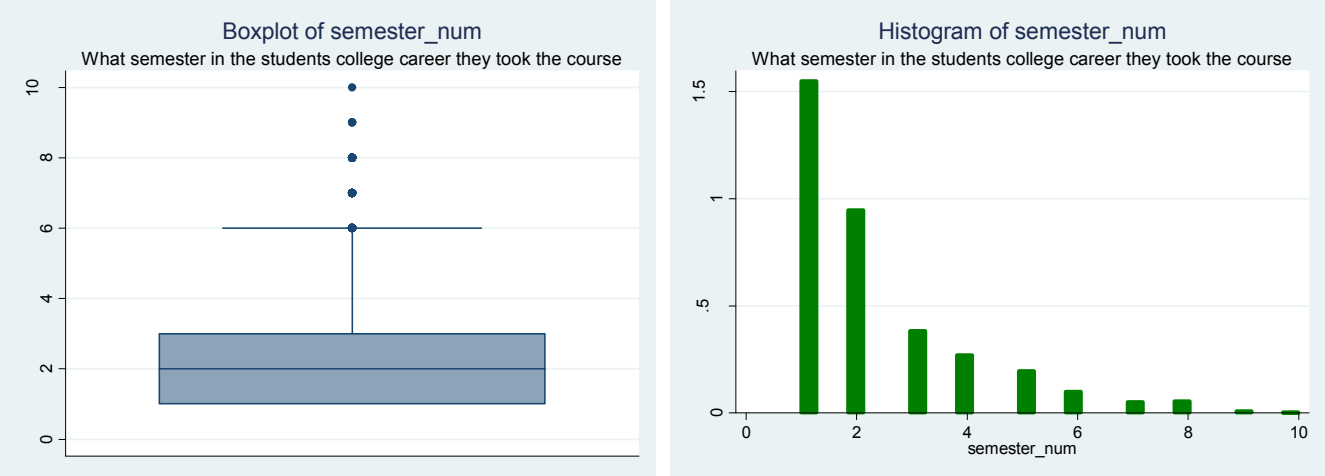
Composition of course: Fall versus Spring

The composition of the class differs greatly between the fall and spring semesters. The highest enrollment is during the fall semester, as students often take the course during their first semester at UC Berkeley, see Figure 5 and Table 2 below.

Table 2: Number of students who enrolled in the course for each value of semester_num

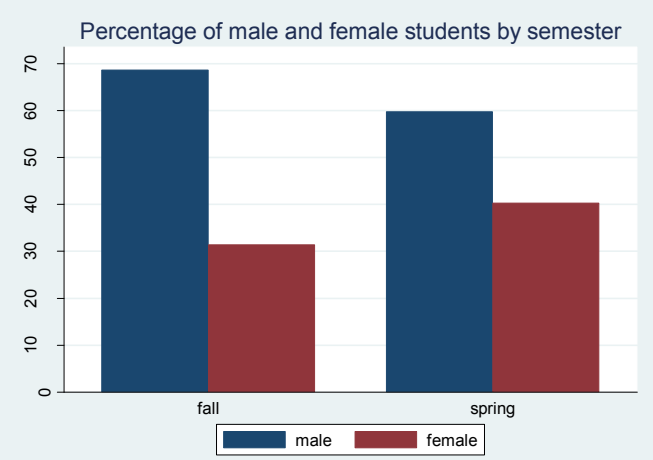
	Freshman		Sophomore		Junior		Senior		Graduate	
	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring
semester_num	1	2	3	4	5	6	7	8	9	10
N	821	502	203	143	104	52	27	19	4	1
%	43.5%	26.6%	10.8%	7.6%	5.5%	2.8%	1.4%	1.5%	0.2%	0.1%

Figure 5: Distribution of in which semester of the students' college career they take CS3



There is a difference in the percentage of the population that female students represent during the fall and spring semesters. Over the 7 semesters in the spring and 7 semesters in the fall, the mean percentage of female students in the spring was 40.3%, compared to only 31.4% during the fall.

Figure 6: Distribution of gender and student level in fall versus spring



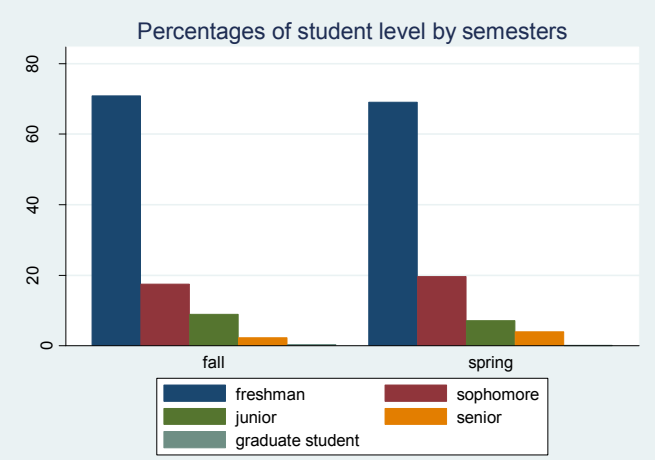
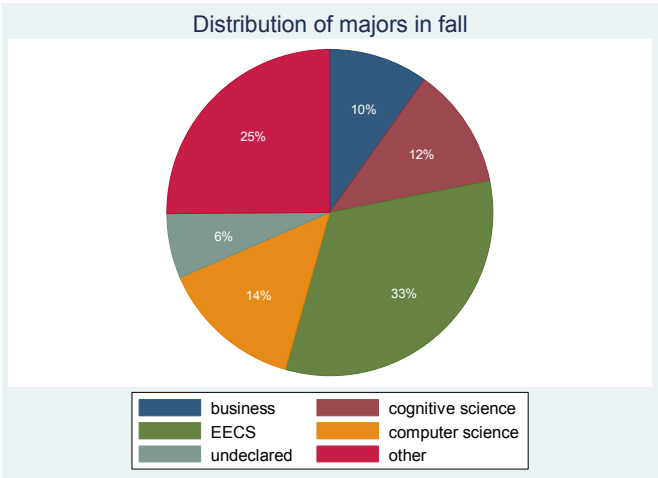
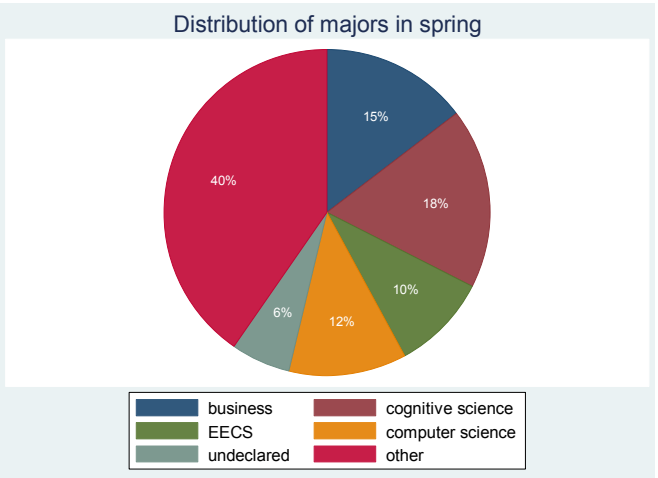


Figure 6 shows this difference in gender composition of the course between fall and spring and shows that there is almost no difference between the distribution of level in school between the fall and spring. However, as shown in Figure 7, there is a difference in the distribution of majors between the semesters in the fall and spring. Together, the two main technical majors, computer science and EECS (Electrical Engineering and Computer Science), make up 46.7% of the course participants in the fall and only 21.3% in the spring.

Figure 7: Distribution of major in fall versus spring





Dropout Rate by Gender

We find that a higher percentage of the female students drop the course, 27.5% of females drop the course in comparison with 20.2% of males. This difference of 7.5% is rather small when compared to differences in attrition within individual semesters. For example in the fall of 2006, 46.7% of the female students dropped the course and only 17.5 % of the males dropped the course. Unlike the spring of 2006 that had universally high attrition rates (42.4% of males and 47.4% of females), it appears that whatever caused the

high rates of female students dropping the course in the fall of 2006 did not have a uniform effect. These semester variations substantiate the observations of the instructors that at times a much greater proportion of female students are dropping the course.

	Female		Male	
	N	%	N	%
Completed	476	72.5%	981	79.8%
Dropped	181	27.5%	248	20.2%
Total	657		1229	

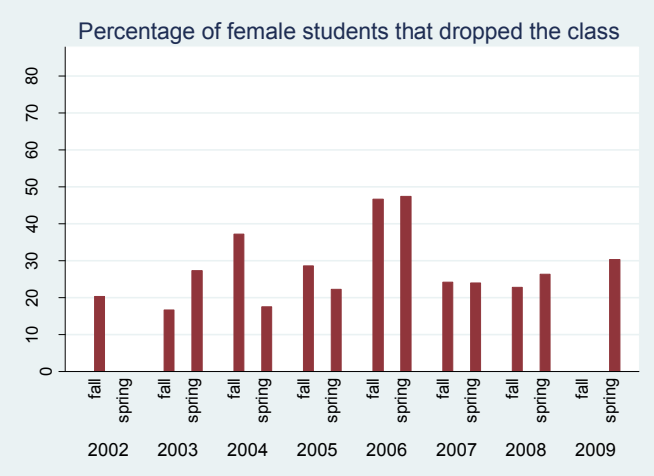
Table 3: Frequency table for the entire sample

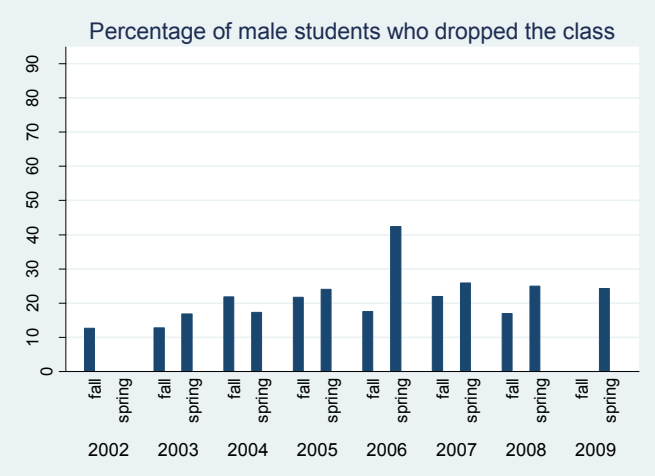
	Female		Male	
	N	%	N	%
Completed	24	53.3%	80	82.5%
Dropped	21	46.7%	17	17.5%
Total	45		97	

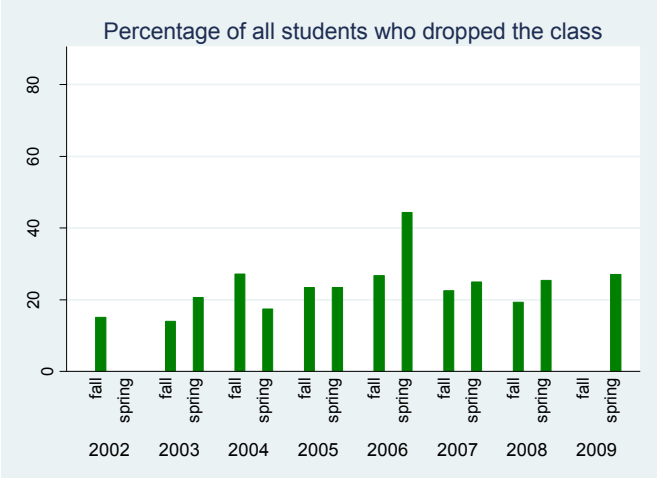
Table 4: Frequency table for the fall of 2006

As can be seen in Figure 8, there were variable percentages of the male students and female students who dropped the course each semester. The fall of 2004 and fall of 2006 seem to have particularly drastic differences in the percentage of male and female students who dropped the course. Table 4 shows the frequencies during the fall of 2006. The fall of 2004 is less drastic, however much greater than the average 7.5% difference, with 37.3% of the female students dropping the course and 21.9% of the male students dropping the course.

Figure 8: Percentages of male and female students who dropped the class







Dropout Rate by Class Standing and Major

What is even more startling than the differences in percentages of males and females dropping the course are the percentages of more senior students that are dropping the course. Figure 9 and Table 5 show the increasing level of drop rates with increasing school level. While 35.2 % of non-freshman dropped the course, only 17.5% of freshman dropped the course. Figure 9 and Table 6 show the differences in percentages of each major that drop the course. The drop rates range from the lowest rate of 10.5%, for electrical engineering and computer science (EECS) majors, to the highest rate of 36.4%, for business majors.

Figure 9: Percentages of each major and level in school that dropped CS3

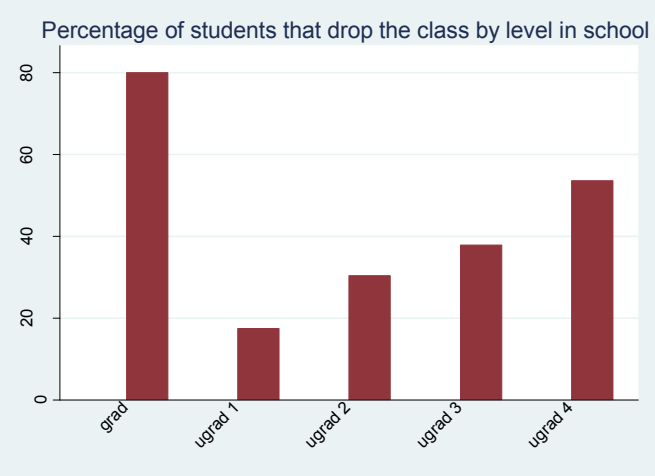


Table 5: Frequency table for students dropping based upon level in school.

	Freshman		Sophomore		Junior		Senior		Graduate Student	
	N	%	N	%	N	%	N	%	N	%
Completed	1092	82.5%	241	69.65%	97	62.2%	26	46.4%	1	20.0%
Dropped	231	17.5%	105	30.35%	59	37.8%	30	53.6%	4	80.0%
Total	1323		346		156		56		5	

Figure 9: Percentages of each major and level in school that dropped CS3

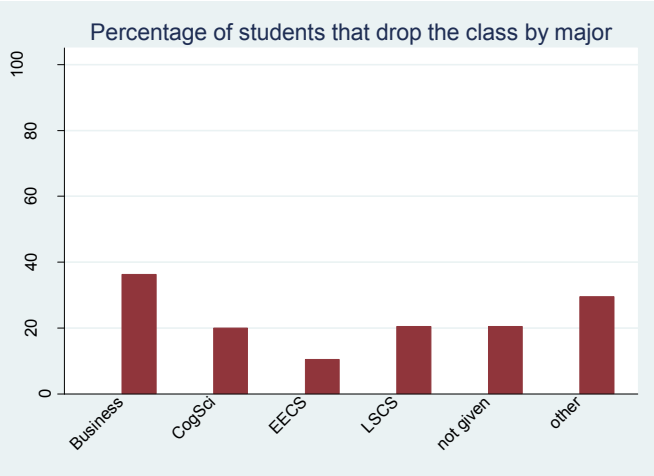


Table 6: Frequency table for students dropping based upon major.

	EECS	Computer	Cognitive	Business	Other	Undeclared

			Science		Science							
	N	%	N	%	N	%	N	%	N	%	N	%
Completed	401	89.5%	197	79.4%	215	79.9%	140	63.6%	411	70.4%	93	79.5%
Dropped	47	10.5%	51	20.6%	54	20.1%	80	36.4%	173	29.6%	51	20.6%
Total	448		248		269		220		584		144	

A logistic regression was performed, as described, to address the research question. When controlling for level in school, major and the semester and year the course was taken, the odds of a female student dropping the course is 32.0% higher than for a male student; this was significant at the 5% level (p=0.024, z=2.26). Table 7 shows the predicted probabilities for males and females. All results can be found in Table 8.

Table 7: Predicted probabilities for males and females with 95% confidence interval shown in brackets [ ]

	Male	Female
All	**0.179** [0.158, 0.203]	**0.255** [0.221, 0.291]

Table 8: Logistic regression results for the model including gender, semester_num, major and semester

		Odds Ratio (Std Err)	Log Odds (Std Err)	Log odds z-value	Log odds p-value	Log Odds: 95% Confidence Interval	
Gender	Female	1.320 (0.162)	0.277 (0.123)	2.26	0.024	0.371	0.518
Student level	Semester number	1.303 (0.46)	0.265 (0.035)	7.51	<0.001	0.196	0.334
Major	Computer Science	2.191 (0.494)	0.784 (0.226)	3.48	0.001	0.342	1.226
	Cognitive Science	1.297 (0.304)	0.260 (0.234)	1.11	0.267	-0.199	0.719
	Business	3.483 (0.767)	1.248 (0.220)	5.67	<0.001	0.816	1.679
	Undeclared	1.877 (0.537)	0.630 (0.286)	2.20	0.028	0.069	1.191
	Other	2.418 (0.468)	0.883 (0.194)	4.56	<0.001	0.503	1.263
Semester	Spring 2003	1.045 (0.317)	0.044 (0.303)	0.14	0.886	-0.550	0.638
	Fall 2003	0.886 (0.265)	-0.121 (0.300)	-0.41	0.685	-0.709	0.466
	Spring 2004	0.695 (0.241)	-0.363 (0.346)	-1.05	0.294	-1.042	0.315
	Fall 2004	2.049 (0.564)	0.717 (0.275)	2.61	0.009	0.178	1.257
	Spring 2005	1.086 (0.366)	0.082 (0.338)	0.24	0.808	-0.579	0.744
	Fall 2005	1.675 (0.458)	0.516 (0.274)	1.89	0.059	-0.020	1.052
	Spring 2006	2.790 (0.814)	1.026 (0.292)	3.52	<0.001	0.454	1.598
	Fall 2006	1.895 (0.430)	0.639 (0.279)	2.29	0.022	0.092	1.187
	Spring 2007	1.081 (0.335)	0.078 (0.310)	0.25	0.802	-0.529	0.684
	Fall 2007	1.519 (0.443)	0.418 (0.292)	1.43	0.152	-0.153	0.989
	Spring 2008	1.140 (0.346)	0.131 (0.304)	0.43	0.665	-0.464	0.726
	Fall 2008	1.263 (0.353)	0.233 (0.280)	0.83	0.405	-0.315	0.781
	Spring 2009	1.277 (0.376)	0.245 (0.295)	0.83	0.406	-0.333	0.822

The odds of a cognitive science major dropping the course is increased by 29.7%, however this result is not significant at the 5% level (p=0.267, z=1.11). The odds for dropping the course for each of the other major classifications were increased, significant at the 5% level. The odds for a computer science major were 2.19

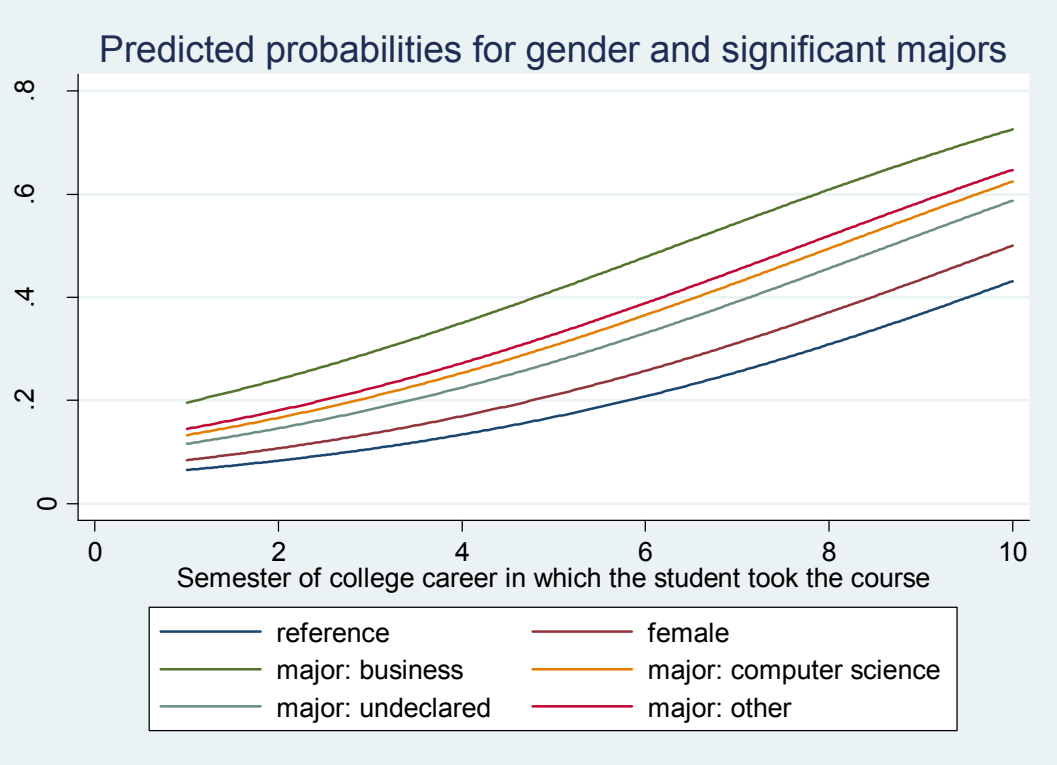
times greater (p=0.001, z=3.48), the odds for a business major were 3.48 times greater (p<0.001, z=5.67), the odds for someone who selected the classification of "other" were 2.42 times greater (p<0.001, z=4.56), and the odds for someone who choose to not select a major and are assumed to be undeclared were 87.7% greater (p<0.028, z=2.20). A Wald test showed that the categorical variable `major` was significant as a category in the logistic regression (p <0.0001). Table 9 shows the predicted probabilities for the categories of `major`.

Table 9: Table of predicted probabilities for each major with 95% confidence interval shown in brackets [ ]

	Business	Cognitive Science	EECS	Computer Science	Undeclared	Other
Male	**0.332** [0.269, 0.402]	**0.170** [0.128, 0.224]	**0.094** [0.071, 0.125]	**0.186** [0.142, 0.240]	**0.173** [0.115, 0.252]	**0.250** [0.213, 0.291]
Female	**0.390** [0.318, 0.468]	**0.206** [0.158, 0.263]	**0.118** [0.085, 0.161]	**0.228** [0.171, 0.296]	**0.221** [0.150, 0.314]	**0.348** [0.296, 0.404]
All	**0.356** [0.293, 0.423]	**0.190** [0.147, 0.241]	**0.099** [0.074, 0.130]	**0.197** [0.152, 0.252]	**0.195** [0.133, 0.277]	**0.281** [0.245, 0.321]

Students who are further along in their college career when they take CS3 are more likely to drop the course. For an additional semester delay in taking the course, the odds of the student dropping the course are increased by 30.3% (p<0.001, z=7.51). Figure 10 shows this relationship expressed as predicted probabilities. In addition to showing the predicted probability by `semester_num`, the semester of the students' college career in which they took CS3, this graph of predicted probabilities shows the predicted probabilities for other significant terms in the logistic regression: `female`, `business_major`, `undeclared_major`, `lscs_major` (computer science), and `other_major`.

Figure 10: Predicted probabilities for gender and majors that were found to be statistically significant



Conclusion

We found that in fact female students do have a higher rate of dropping the class when controlling for what semester in the students' college career they took the course (1-10), the students' major and the semester and year the course was taken. We found different patterns of attrition based upon major, level in school, and offering of the course. The increased odds of students of certain majors and levels in school of dropping the class were greater in magnitude than the increased odds of a female student dropping the course.

Exploring the reasons behind these patterns of behavior was not within the scope of the current study. Future work will examine causal factors behind the observed patterns of attrition.

References

College Board. (2008). AP *Report to the Nation*. Princeton, NJ: Author. Retrieved March 28, 2009, from http://professionals.collegeboard.com/data-reports-research/ap/nation/2008.

Judson, Eugene. (2008). University of California, Berkeley, Baseline Report for Undergraduate Computer Science and Engineering Programs. National Center for Women & IT.

Klawe, M., Whitney, T., & Simard, C. (2009). Women in Computing – Take 2: Inspiring, recruiting, and retaining women for a career in computing remains a challenge. *Communications of the ACM.* 52 (2), 68-76.

Office of Planning & Analysis, University of California, Berkeley. (2008) CS3-CS61a Fall Update. Berkeley, CA: Author.

Acknowledgements

I would like to acknowledge Marcia Linn and Michael Clancy for their helpful feedback on earlier drafts.

Appendix

Frequency table of students dropping for offering1 through offering14

	Fall 2002 (offering1)		**Spring 2003** (offering2)		**Fall 2003** (offering3)		**Spring 2004** (offering4)		**Fall 2004** (offering5)	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Completed	186	84.9%	96	79.3%	147	86.0%	76	82.6%	107	72.8%
Dropped	33	15.1%	25	20.7%	24	14.0%	16	17.4%	40	27.2%
Total	219		121		171		92		147	
	Spring 2005 (offering6)		**Fall 2006** (offering7)		**Spring 2006** (offering8)		**Fall 2007** (offering9)		**Spring 2007** (offering10)	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Completed	62	76.5%	131	76.6%	54	55.7%	104	73.2%	78	75.0%
Dropped	19	23.5%	40	23.4%	43	44.3%	38	26.8%	26	25%
Total	81		171		97		142		104	
	Fall 2007 (offering11)		**Spring 2008** (offering12)		**Fall 2008** (offering13)		**Spring 2009** (offering14)			
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**		
Completed	103	77.4%	82	74.55%	142	80.7%	89	73.0%		
Dropped	30	22.6%	28	25.45%	24	19.3%	33	27.1%		
Total	133		110		176		122			