Student Motivations and Goals for CS1: Themes and Variations

David Liben-Nowell
Carleton College
Northfield, MN, USA
dln@carleton.edu

Anna N. Rafferty
Carleton College
Northfield, MN, USA
arafferty@carleton.edu

ABSTRACT

Students come to CS1 with a wide variety of motivations and goals, which may differ across subpopulations and be indicative of their future engagement with CS. While there is a rich literature relating success in CS1 to specific constructs, such as belonging, goal-orientation, or self-efficacy, less work has examined what motivations and goals students volunteer as most important for their enrollment in CS1. Here, we use qualitative coding to identify themes from students' open-ended descriptions of why they're taking CS1 and what they hope to get out of it, collected across fifteen years. Using quantitative analysis of these coded descriptions, and word-frequency analysis, we identify and name three clusters of students that encompass the majority of students taking CS1: Explorers, Planners, and Utilitarians. We also identify motivations and goals that are more common for particular populations, such as students who have not yet declared a major or students without prior programming experience, as well as factors predicting students' later engagement with CS. This work demonstrates the potential of qualitative coding and computational analyses to enable us to better understand a population of students based on their own words.

CCS CONCEPTS

• **Social and professional topics → Computing education**.

KEYWORDS

CS1; background surveys; subpopulations; goals; motivations

ACM Reference Format:
David Liben-Nowell and Anna N. Rafferty. 2022. Student Motivations and Goals for CS1: Themes and Variations. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2022), March 3–5, 2022, Providence, RI, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3478431.3499358

1 INTRODUCTION

Students come to CS1 with diverse prior experiences, varying in previous exposure to computing [e.g., 30], social support in engaging with computing [e.g., 34], and in their academic identities [e.g., 39]. A widely varying student population presents pedagogical challenges [e.g., 1]. Perhaps aware of these challenges, and the known benefits of close faculty–student interaction [e.g., 14], many instructors begin to get to know their students via early-course surveys. Surveying students' prior computing experiences can help instructors tailor a course to their students' common background. Surveys might also delve into students as individuals—e.g., seeking their motivations (why they're taking the course) or goals (what they'd like to gain from it), enabling instructors to use more relevant examples and to form individual relationships with students, which can positively influence student persistence [17, 18].

Background surveys provide an interesting lens into students, with the particular way they choose to respond lending insight into their feelings, fears, and desires [11]. (It is a filtered lens, though; students know they are writing for the course instructor.) While such personal information is most commonly used for class customization and student–faculty relationship building, when taken together these responses can also provide a "collective knowing" of students [35]; instructors can augment knowledge of individual students with an understanding of the population as a whole. Such an examination does not minimize the import of individual connections, but rather provides a complementary view, enabling instructors to adapt courses to better serve the full population of students who take their courses, and to recognize ways that students with particular motivations and course goals may not be well served by the class as designed. Yet examining student responses across classes and time is less common in the CS education literature.

In this paper, we focus on understanding students' motivations and goals for CS1 through a holistic examination of students' responses to background-survey questions—why they're taking CS1, and what they hope to get out of it—collected across fifteen years. These open-ended questions were not designed to probe particular concepts or academic orientations; students had full freedom to express a multitude of ideas, thereby allowing us to observe patterns of what students themselves saw as most important to communicate. We identify a diverse set of themes from 396 students' responses, illuminating the range of motivations and goals across a broad student population. Using qualitative coding, quantitative analysis of the coded responses, and word-frequency analysis, we address three types of questions about a population of CS1 students:

(A) Are there identifiable groups of students that differ in their motivations and goals, and how can we characterize such groups?

(B) How (if at all) do patterns of word usage, motivations, and goals vary systematically across demographic groups of students, such as those taking CS1 before declaring a major vs. taking it after?

(C) What patterns separate students whose post-CS1 trajectories of engagement with computer science in college differ, such as those who take a second course in CS vs. those who do not?

Over 70% of our CS1 population fell into three broad clusters that we identify and name—Explorers, Planners, and Utilitarians—who each enter the course with different goals. Students who had no pre-CS1 programming experience frequently cited social reasons

SIGCSE 2022, March 3–5, 2022, Providence, RI, USA.

ACM ISBN 978-1-4503-9070-5/22/03.
https://doi.org/10.1145/3478431.3499358

for taking CS1, echoing the importance of social support for learning [31]. Surprisingly, students who articulated a desire to learn (or improve) programming skills were somewhat *less* likely to continue in CS. While a great deal of variation is not captured by student narratives of their motivations and goals for CS1, our full results to (A–C) demonstrate systematic patterns in the population, affording us better knowledge of our students at a group level, and providing valuable context for better knowing future students as individuals.

2 RELATED WORK

CS1 student populations. The rising societal importance of computing has led to increased CS enrollments, with a wide variety of students choosing to take CS1 [26]. Much recent CS0/CS1 pedagogy has sought to make CS approachable and accessible to all students [e.g., 10]; some programs have created multiple CS1 "flavors" to meet the varying needs and desires of this diverse population (reviewed in, e.g., [39]; see also [9] for its success at one institution).

As described in a recent review [36], a number of computer science education publications report on student surveys, with foci including student demographics, motivations, perceptions of CS, and prior experience. For example, via closed-form questions probing specific possible motivations and perceptions, [25] found that students differed in whether they perceive CS as conferring social status and providing job security, and typically studied CS due to their perception of its value. Some of this work has identified particular subpopulations of students within CS1. For example, some students aim to be "conversational programmers," learning programming to enable them to communicate with, rather than become, programmers [7]. Identifying such subpopulations is key; current curricula may not serve these students well [e.g., 8, 33].

Predicting success based on student characteristics. While some work on student success in CS1 or persistence in CS uses interviews and open-ended responses [e.g., 22], a larger set of work aims to relate student demographics and characteristics as measured by closed-form scales to success/persistence [e.g., 24]. This work often uses theoretical frameworks for describing student attitudes that are relevant beyond computing, such as achievement orientation [29, 37, 38], growth mindset [29], self-efficacy [29], sense of belonging [20], and confidence [3]. Both student demographics and attitudes are predictive; male students and more confident students were likely to persist [3, 4] and mastery goals were associated with success [37].

In this paper, we also seek to identify differences across students, using a combination of demographic variables and factors based on students' open-ended responses; like [23], we use clustering as one way to identify subpopulations of students. The open-ended responses we analyze are answers to deliberately broad questions, providing us the opportunity to see what motivations and goals students see as salient enough to volunteer, rather than being limited to a fixed set of options. This approach is necessarily complementary to the literature focusing on well-established theoretical constructs: it has the drawback of using unvalidated survey questions, a common problem in CS education survey research [36], but allows for broader exploration of the space of possible student motivations.

Approaches to analyzing open-ended responses. We use two approaches for analyzing student responses: word frequencies, which

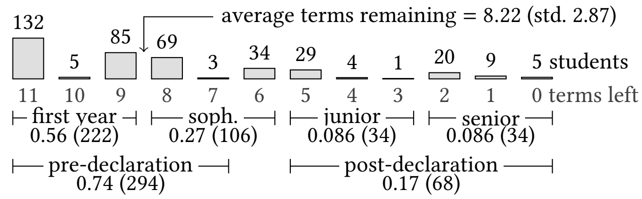

Figure 1: Distribution (counts and proportions) of class years, status of major declaration, and number of terms remaining at Carleton subsequent to a student's CS1 enrollment.

is a type of computational text analysis (see, e.g., [15] for an overview of computational social science) and qualitative coding of the features of student responses [e.g., 28, 32]. Classification analysis using qualitative codes and computational text analysis as input has been used to identify differences in subpopulations [e.g., 13, 19].

3 PARTICIPANTS AND PROCEDURES

To address our overarching questions—to understand the variegated landscape of student motivations and goals in taking CS1—we examine the responses to a beginning-of-term background survey of CS1 students at a particular institution, Carleton College.

Institutional context. Carleton is a highly selective undergraduate-only liberal arts college located in the midwestern United States, with ≈2000 total students. Carleton's academic calendar includes three 10-week terms. Students declare their major during their sixth term, and normally graduate after 12 terms.

CS1 "Introduction to Computer Science" at Carleton is a broad-spectrum course, intended for any and all students who want to take computer science, and in recent years, ≈60% of all students have taken it. The remainder of the CS curriculum is roughly in keeping with the ACM liberal arts CS model curriculum [16]. The number of CS majors at Carleton increased from ≈2% of the graduating class to ≈14% during the studied period (2006 to 2020).

The student population. We study $n = 396$ students who took CS1 over a 15-year period, from 2006 to 2020, with a particular instructor. This comprises >84% of all CS1 students taught by the instructor in this period. The IRB at Carleton approved this research (see Appendix A[1] for details and participation rates).

While most students were in their first year, all class years were represented (Figure 1; methods in Appendix B). According to Registrar's Office data,[2] 183 students (46%) were female and 213 (54%) were male. Most students were white (259, 65%); other reported race/ethnicity categories were Asian (36, 9.1%), Black/African American (9, 2.3%), Hispanic/Latino (20, 5.1%), and Two or more races (23, 5.8%). Six students (1.5%) had no reported race/ethnicity information, and 43 (11%) were reported only as international students.

The background surveys. On the first day of every CS1 offering, the instructor distributed an open-ended background survey, to

[1] OSF link to appendices: http://tiny.cc/StudentGoalsAppendices.

[2] The Registrar's database has notable limitations: it has capacity only for binary gender identification, and it stores identities as reported in the student's original application for admission to Carleton; thus in some cases the stored information fails to reflect the student's current (or contemporaneous) identity.

(Q9) Most students who take [CS1] have never done any computer programming before, but there are occasional exceptions. Do you have any computer programming experience? (How much? What languages?)

(Q10) Why are you taking this course? (If it's for a requirement, please be honest!)

(Q11) Why are you taking this course *now*?

(Q14) What do you hope to get out of this course?

(a) The key questions.

Student A *(Spring 2006, White, Female, 9 terms remaining):* (Q9) Very little, I did some calculator programming after the AP calc test. (Q10) I'm wondering if I might be interested in a CS major. (Q11) I'd like to know what I want to major in sooner than later. (Q14) A better sense of computer science, mostly. An idea of how much I like it.

Student B *(Fall 2011, Hispanic/Latino, Male, 11 terms remaining):* (Q9) Almost none, I can make basic lists and functions in Python but that's about it. (Q10) Because computer science would be important in whatever major I choose and friends who took the class liked it. (Q11) Because it works in my schedule and it was recommended recently. (Q14) An understanding of and basic ability in programming.

Student C *(Fall 2019, international student, Female, 11 terms remaining):* (Q9) No. (Q10) I think CS is very essential in our future life, so I'd better get to learn it. (Q11) I want to start programming right away and create cool stuff. (Q14) Become a skillful programmer.

(b) Some sample responses.

feature	*mean (count)*	**A**	**B**	**C**
CS1 suggested by someone else	0.088 (35)		B	
Classmate affinity [desire particular classmates]	0.025 (10)			
Others like CS1 or CS	0.126 (50)		B	
Instructor as reason for CS1	0.045 (18)			
Positive emotional response to CS	0.141 (56)			
Likes computers	0.083 (33)			
CS/CS1 is useful for getting a job	0.043 (17)			
CS is useful for understanding a particular topic	0.227 (90)			
Computing is useful or important	0.285 (113)		B	C
CS1 satisfies a general requirement	0.116 (46)			
Engage with something new	0.068 (27)			
Better understand computers	0.187 (74)			
Better understand CS	0.414 (164)	A		
Improve at logic/problem solving	0.167 (66)			
Better understand role of CS in society	0.015 (6)			
Better understand oneself or one's interests	0.374 (148)	A		
Develop computing skills/comfort	0.182 (72)			
Learn to program	0.505 (200)		B	C
Learn collaborative or communication skills	0.025 (10)			
Have fun in the course	0.086 (34)			
Build something cool	0.056 (22)			C

feature	*mean (std. dev)*	**A**	**B**	**C**
Intention to further study CS (0–10)	2.184 (2.748)	5	0	0
Amount of prior programming experience (0–3)	0.902 (1.138)	2	2	0

(c) Coded features, mean values, and values for Students A–C.

Figure 2: The background survey: the questions, example responses from three students, and the coding scheme.

start to get to know the students in the class. See Figure 2 (and Appendices C and D for the full survey and administration details).

4 UNCOVERING GOALS AND MOTIVATIONS

Each CS1 student arrives in the course with their own reasons for being there, their own desires for outcomes, and their own perceptions of CS and of themselves. To uncover some of this textured, nuanced variation among students from background-survey responses, we first (without preconceived categories) closely read all (anonymized) responses to Q10, Q11, and Q14, extracting common themes into a set of categories (see Appendix E for details).

We then coded each student's responses (see Figure 2c) for: (i) the presence or absence of the 21 possible motivations or goals for CS1 that we identified; and (ii) a 0-to-10 score [sum of 0-to-5 scores from two coders] of the student's intention to further study CS.

We also directly analyze the raw text (processed using NLTK [6]; see Appendix F) that students wrote in response to Q10, Q11, and Q14. The words used by the most students in describing their motivations and goals for the course are *CS* (used by 212 of 396 students, 54%); *computer* (185); *want* (169); *science* (149); *like* (146); *take* (143); *programming* (131); *learn* (123); *major* (118); and *hope* (114).

Clustering analysis: three large subpopulations of CS1 students. To cluster students based on their motivations and goals, we applied agglomerative clustering to their coded responses (implemented in sklearn 0.23 [21]). Because most features were binary, we used a Manhattan distance metric and an average linkage function, scaling intention to further study CS into $[0, 1]$ to match other features.

From the clustering results, we built a dendrogram of the students (Figure 3). Clustering is notoriously sensitive to small methodological perturbations [12]; in this case, we get broadly comparable clusters by using either of the two coders' 0-to-5 scores for intention to further study CS, or by summing them; because the cluster-identification methodology was slightly cleaner to describe in the former case, we use a single coder's ratings here. By applying a threshold distance cutoff to the dendrogram (< 4.5) and a cluster size threshold (> 15), we identify three large subpopulations taking CS1, each comprising $\approx$25% of the full population. We inspected these clusters for quantitative differences from the population at large, and labeled the clusters with subjectively determined names. Here are the three subpopulations that we identify and name:

(1) "The Planners" enter the course with significant intention to further study CS in their student career, and hope to gain more knowledge about themselves (including their self-perceived "fit" with the discipline) through the course;

(2) "The Explorers" seem to be taking the course hoping to have a fun experience, engage with a different way of thinking, and better understand what computer science actually is; and

(3) "The Utilitarians" enter the course with a more instrumental view of CS, seeking to gain skills which they perceive as useful for their future careers or for their study of some topic other than CS.

5 VARIATION ACROSS STUDENT GROUPS

As Figure 3 shows, background-survey responses provide a lens into meaningful differences among students who enroll in CS1.

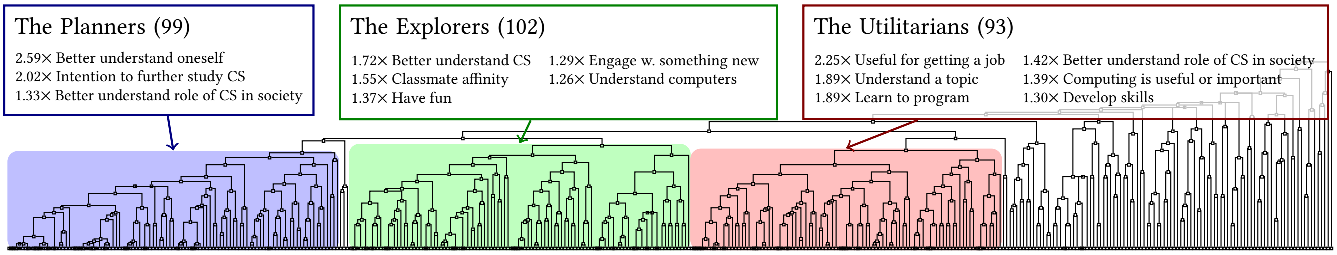

Figure 3: A dendrogram based on clustering of background surveys. Clusters are defined by a distance threshold and a minimum size; all features appearing at a ≥ 1.25 higher rate in the cluster than the baseline (the student population at large) are listed.

But, while clustering reveals groupings of students that emerge directly from the data, we can also examine differences between exogenously defined groups of students. We first focus on relating students' motivations and goals to characteristics of students "on their way in the door" to CS1, examining differences across students based on (a) gender, (b) prior programming experience, and (c) pre- vs. post-declaration of a major. We identify differences both quantitatively, by using regression to determine which features are predictive of the characteristic, and more qualitatively, by examining how student language differs across groups. While we find many similarities across groups, this more focused examination of specific subpopulations also shows consistent variation.

Our predictive models use regularized logistic regression with 10-fold cross validation ([27]; see Appendix G). We calculate (i) the mean explained variance score and, more importantly, (ii) the features that consistently appeared with non-zero coefficients (magnitude > 0.05 in all 10 trained classifiers). We add two further features: terms remaining (Figure 1) and (see, e.g., [4]) prior programming experience, based on coding responses to Q9 as reflecting no (0), minimal (1), some/ambiguous (2), or significant (3) experience.

Gender. (As a limitation, note that the institutional data on student gender was binary and sometimes inaccurate; see Section 3.)

Consistent with significant variation in student motivation and goals within each gender, only a small amount of the variance in unseen data was explainable by the logistic regression model (mean in training folds: 0.14; mean in test folds: 0.019). A number of features did occur consistently in the learned classifier, however (Figure 4a). Female students were more likely to cite a desire to engage with something new and satisfy a general requirement as reasons to take the course. They also tended to describe wanting to better understand CS or logic/problem solving, while males were slightly more likely to want to understand computers, as well as to like computers. These same trends can be seen in word-frequency analysis, too, with *new, way, problem,* and *solving* all occurring at least twice as often in responses from female students (Figure 5a).

All social reasons that were predictive of gender were more associated with female students (consistent with work examining social influences and career choices [2, 5]). Interestingly, female students tended to express positive emotions towards CS and used the words *love* and *fun* twice as often as males, although male students had greater intentions to study CS further. These relationships should be interpreted cautiously: expressing more positive emotions towards CS is consistent with multiple interpretations: e.g., at a group level, female students may need to exceed a higher threshold of excitement about CS (vs. lower for males) to consider taking CS1; or these language variations may reflect gendered expressive norms.

Prior experience. We compared students with *minimal* or *no* prior programming experience ("low") vs. *significant experience* ("high"), omitting intermediate *some/ambiguous experience* students. 17% of the remaining 326 students had high experience. Because programming experience was the outcome variable, it was omitted as a predictor. While still small, this analysis explained a larger amount of variance (mean for training folds: 0.21; mean for test folds: 0.13).

Students with low prior experience cited social reasons for taking CS more than those with high experience, as well as wanting to learn skills and understand CS (Figure 4b). These students were more likely to use words like *basic* and *knowledge* (e.g., *I would love to gain some basic programming knowledge* [Fall 2019, minimal experience]; Figure 5b). In contrast to low-experience students wanting to engage with something new and discover if they "might" have an interest, high-experience students came in with more positive emotions towards CS and a greater intention to further study CS.

Pre- vs. post-major declaration. Finally, we predict whether students took CS1 before declaring a major or after (i.e., with ≥ 7 vs. ≤ 5 post-CS1 terms remaining). Students with exactly 6 terms remaining were excluded, and terms remaining was omitted as a predictor. The variance explained was close to that for prior experience (mean for training folds: 0.21; mean for test folds: 0.10).

Most features consistent across folds were associated with being pre-declaration rather than post-declaration (Figure 4c), including a higher intention to further study CS and wanting to better understand oneself or one's interests. The latter motivation is typical of students exploring multiple possible majors, and the further association of a desire to engage with something new with pre-declaration students likely reflects the liberal arts ethos of exploration in the first two years. The language used by pre-declaration students (Figure 5c) also reflected this ethos, with more frequently occurring words including *new, try,* and *see,* such as the Fall 2014 student who said *I wanted to try something that I had no experience with . . . to see if it's something I might pursue further.*

The only consistent post-declaration–associated feature was studying CS because it was useful for understanding some specific topic. Such students often spoke about the usefulness of CS1 for their own major (e.g., *I think coding is an important skill for a math major . . . I'm hoping to really love it and continue with CS to do mathematical modeling (epidemiology, government, we'll see). I've heard*

(a) *feature*	*average coefficient (std.)*
coefficients > 0 ↔ female students; < 0 ↔ male	
Intercept	-1.01 (0.09)
Engage with something new	0.89 (0.18)
Likes computers	-0.73 (0.15)
Satisfies a general requirement	0.61 (0.14)
Better understand CS	0.59 (0.07)
Improve at logic/problem solving	0.55 (0.09)
Others like CS1 or CS	0.52 (0.08)
Computing is useful or important	0.40 (0.13)
Better understand oneself	0.37 (0.10)
CS1 suggested by someone else	0.31 (0.14)
Positive emotional response	0.27 (0.12)
Useful for understanding a topic	0.25 (0.07)
Intention to further study CS	-0.25 (0.04)
Better understand computers	-0.14 (0.04)

(b) *feature*	*average coefficient (std.)*
coefficients > 0 ↔ low experience; < 0 ↔ high	
CS1 suggested by someone else	1.87 (0.08)
Engage with something new	1.26 (0.11)
Intercept	1.15 (0.13)
Others like CS1 or CS	0.99 (0.21)
Develop skills	0.84 (0.17)
Positive emotional response	-0.80 (0.11)
Better understand computers	0.62 (0.19)
Better understand CS	0.46 (0.07)
Intention to further study CS	-0.42 (0.07)
Terms remaining	-0.37 (0.05)

(c) *feature*	*average coefficient (std.)*
coefficents > 0 ↔ pre-declaration; < 0 ↔ post-decl.	
Intercept	1.51 (0.11)
Engage with something new	1.08 (0.18)
Useful for understanding a topic	-1.07 (0.11)
Intention to further study CS	1.03 (0.06)
Likes computers	0.88 (0.23)
Improve at logic/problem solving	0.71 (0.15)
Better understand oneself	0.35 (0.12)
Better understand CS	0.31 (0.07)
Develop skills	0.27 (0.10)

Figure 4: Predicting (a) gender, (b) prior programming experience, and (c) being pre- vs. post-declaration of major, based on coded background surveys. All features occurring in all folds of the classifier are shown (all of which had a consistent sign).

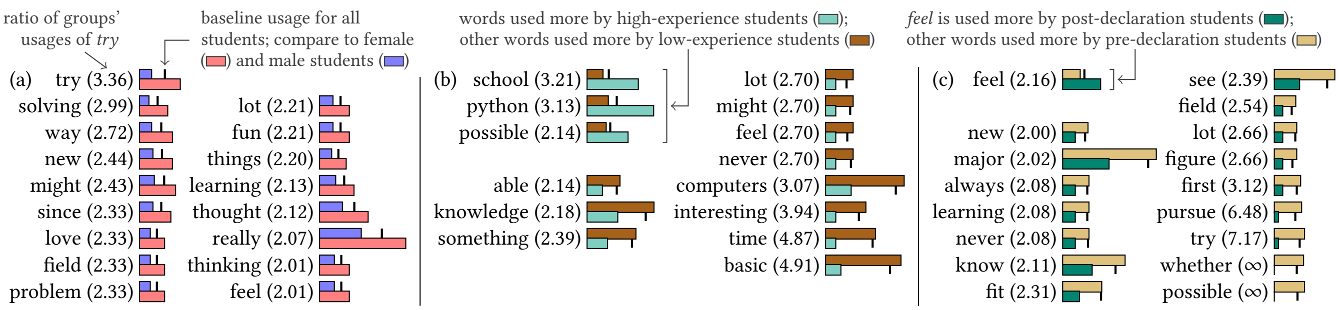

Figure 5: Words used ≥ 2× more by one group of students: by (a) gender, (b) programming experience, and (c) declaration status.

that if you have a math background and coding skills you can kinda do anything [Fall 2020] or *I was talking to a bio prof about bio major type stuff and he highly recommend that I take a CS course because it would probably be very relevant with all the new bioinformatics and sequencing being developed* [Fall 2011]). (Consistent with this application-oriented view, nine of the ten classifiers had useful for getting a job as post-declaration–associated.)

6 PREDICTING STUDENT TRAJECTORIES

Moving from how students differ when they arrive in CS1, we now turn to how students' motivations and desires relate to their differing paths *after* the course. We focus on students' further engagement with CS, using two measures: whether they take at least one more course in the department, and whether they major in CS.

We again expect that much of the variance in trajectories will be unexplained by a background survey completed at the outset of CS1. Presumably, students are changed by their experiences within the CS1 course, and further, many factors influence choice of major. Still, any differences by trajectory predicted by pre-CS1 student responses could help in better understanding different sets of students and perhaps identifying student motivations or goals that could be more directly targeted in CS1 or later courses.

Restricting attention to the 302 students who have completed their time at Carleton, 50.7% took one or more CS courses post-CS1 (methods in Appendix H). The distribution of the number of subsequent CS classes taken by these 153 students is bimodal, reflecting two large subpopulations of CS1 students: those who eventually declare a CS major, and those who do not.

Taking another CS course. As predicted, the classifiers explained a moderate amount of variance (mean for training folds: 0.25; mean for test folds: 0.18). Students who did not take further CS were more likely to be satisfying a general requirement. Interestingly, they were also more likely to view computing as useful or important in general or for understanding a particular topic (Figure 6). These "usefulness" features often corresponded to students who saw the importance of grounding in CS to their overall goals: e.g., *I . . . believe that no matter what I end up doing in life, a fundamental understanding of computers and possible computer software will be critical* [Fall 2014] for the former category, and *I took a computational chemistry class in the spring which made me feel like I might benefit from some exposure to computer syntax* [also Fall 2014] for the latter.

Students with more terms remaining and those with greater intention to further study CS were more likely to take another class, with words like *whether* and *possible* frequently present (Figure 7a), much like the pre-declaration students who took CS1 partially to explore possible majors (cf. Figure 5c).

Majoring in CS. Nearly all participants (382 of 396, 96%) had declared a major as of June 2021, including 81 who declared a CS major (21% of 382) and 301 who did not (79%).

feature	◆ take more CS? [all students]	● CS major? [all students]	■ CS major? [pre-declaration]
Satisfies a general requirement	-0.67 (0.10)	-1.34 (0.14)	-1.37 (0.19)
Useful for understanding a topic	-0.30 (0.10)	-1.28 (0.21)	-1.14 (0.18)
Terms remaining	0.43 (0.03)	1.02 (0.07)	0.33 (0.06)
Intercept	— (—)	-0.96 (0.07)	-0.35 (0.09)
Intention to further study CS	0.58 (0.04)	0.57 (0.06)	0.59 (0.05)
Better understand CS	— (—)	-0.56 (0.11)	-0.49 (0.05)
Others like CS1 or CS	— (—)	-0.46 (0.15)	-0.51 (0.24)
Learn to program	— (—)	-0.43 (0.10)	-0.31 (0.11)
Amount of prior programming experience	0.42 (0.03)	0.40 (0.02)	0.43 (0.06)
Computing is useful or important	-0.35 (0.07)	-0.35 (0.07)	— (—)
Build something cool	0.49 (0.19)	— (—)	— (—)

Features denoted with "—" did not occur in all 10 folds in the task.

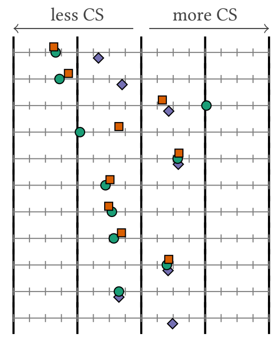

Figure 6: Predicting whether a student will go on to further study CS (either majoring in CS, or taking ≥1 further CS class).

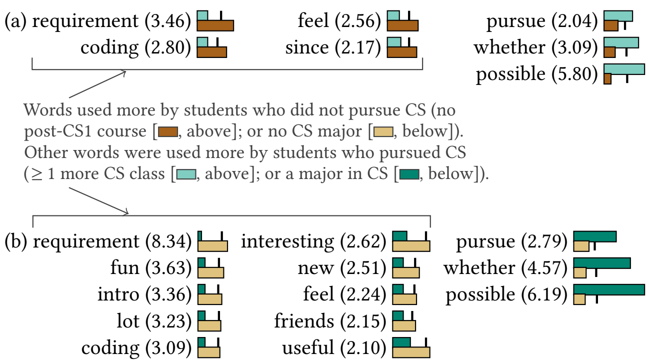

Figure 7: Terms used ≥ 2 times more or less frequently by students who take more CS vs. those who do not.

We first try to predict CS majors vs. non-majors among all 382 students with declared majors. This analysis explained more variance on average than the others (mean for training folds: 0.33; mean for test folds: 0.23), although many features were similar to those for whether a student took a second CS class (Figure 6). Intent to further study CS and prior programming experience were again associated with studying CS further (here, majoring in it); satisfying a general requirement or using CS to understand a topic was again associated with less future engagement. Students taking CS1 to learn to program were less likely to major, which could be due to factors like wanting to program for some specific purpose or to a misalignment between students' prior beliefs in what the discipline is about (e.g., the view CS = programming) and a CS curriculum emphasizing programming as a means to implement logic.

To examine whether the factors associated with majoring in CS differ for students who take CS1 before declaring a major, we repeated the analysis restricted to the 280 students who took CS1 pre-declaration and had subsequently declared a major. While a greater proportion of these students did in fact major in CS (29% versus 21% for all), the features and coefficients were similar for both groups, with a smaller weight on terms remaining, likely due to the fact that all pre-declaration students had sufficient time to complete the major. The regression on the pre-declaration students explained slightly less of the variance than for all students (mean for training folds: 0.27; mean for test folds: 0.16).

7 LIMITATIONS AND FUTURE WORK

The analyses we present in this paper reveal common themes in students' reasons for taking CS1, from the utilitarian (wanting a job, or seeking insight into another field) to the social, and a variety of outcomes desired by students—including some themes, like learning collaborative or communication skills, that echo existing literature [7]. Still, while our analyses uncovered some systematic patterns, there are several important limitations to the work.

Here, we have treated students' responses to background surveys as a window into their most strongly held motivations and goals. Importantly, though, this window does not provide an unfiltered view of their desires; a student's response may be influenced by their writing style and their audience (the instructor). The naturalistic context of the background survey can be seen as a strength, but interviews with students would help further illuminate their motivations and goals. The brevity of student responses also necessarily means that students did not write all goals, which could be due to these goals being of less importance or simply less salient.

We analyze CS1 students at one institution, with one instructor. Doing so eliminates any number of confounding factors, but this setting—like any one particular setting—is idiosyncratic.

Our analyses focus on linear predictive models. The sparse dataset necessitated simple models, but future work collecting additional data or considering lower-dimensional representations might be able to identify interesting interactions—e.g., are social influences more important for students who are uncertain about their desire to pursue CS in the future?

An unusual aspect of our dataset is its temporal breadth, spanning a 15-year period. While we treated this dataset monolithically, future work could examine changes in responses over time. Such changes might arise from evolving course pedagogy, mediated by course reputation, as well as broader shifts in societal perceptions of CS. Future work could also examine students' motivations and goals longitudinally, such as changes across CS1 and CS2. Such an exploration might lead to insights about how students' trajectories both influence and are influenced by these motivations and goals.

More broadly, this paper provides an example of both the richness and the potential explanatory power of students' own open-ended descriptions of their reasons and desires for enrollment in CS1, and a characterization of possible factors associated with student populations and trajectories that can be built upon in future work.

ACKNOWLEDGMENTS

We gratefully acknowledge the help and insights of Emy Farley and Ann May in the Carleton Registrar's Office.

REFERENCES

[1] S.O. Adodo and J.O. Agbayewa. Effect of homogenous and heterogeneous ability grouping class teaching on students' interest, attitude and achievement in integrated science. *International Journal of Psychology and Counselling*, 3(3):48–54, 2011.

[2] Amnah Alshahrani, Isla Ross, and Murray I. Wood. Using social cognitive career theory to understand why students choose to study computer science. In *Proceedings of the 2018 ACM Conference on International Computing Education Research*, pages 205–214, 2018.

[3] Christine Alvarado, Cynthia Bailey Lee, and Gary Gillespie. New CS1 pedagogies and curriculum, the same success factors? In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education*, pages 379–384, 2014.

[4] Lecia J. Barker, Charlie McDowell, and Kimberly Kalahar. Exploring factors that influence computer science introductory course students to persist in the major. In *Proceedings of the 40th ACM Technical Symposium on Computer Science Education*, pages 153–157, 2009.

[5] Sylvia Beyer. Gender differences and intra-gender differences amongst management information systems students. *Journal of Information Systems Education*, 19(3):301, 2008.

[6] Steven Bird, Ewan Klein, and Edward Loper. *Natural Language Processing with Python.* O'Reilly Media, Inc., 2009.

[7] Parmit K. Chilana, Celena Alcock, Shruti Dembla, Anson Ho, Ada Hurst, Brett Armstrong, and Philip J. Guo. Perceptions of non-CS majors in intro programming: The rise of the conversational programmer. In *2015 IEEE Symposium on Visual Languages and Human-Centric Computing*, pages 251–259, 2015.

[8] Kathryn Cunningham, Barbara J. Ericson, Rahul Agrawal Bejarano, and Mark Guzdial. Avoiding the Turing tarpit: Learning conversational programming by starting from code's purpose. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021.

[9] Andrea Forte and Mark Guzdial. Motivation and nonmajors in computer science: Identifying discrete audiences for introductory courses. *IEEE Transactions on Education*, 48(2):248–253, 2005.

[10] Dan Garcia, Brian Harvey, and Tiffany Barnes. The beauty and joy of computing. *ACM Inroads*, 6(4):71–79, 2015.

[11] Paul Grice. *Studies in the Way of Words.* Harvard University Press, 1989.

[12] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction.* Springer, 2nd edition, 2009.

[13] Jennifer E. John, Paul A. Nelson, Brittany Klenczar, and Rachael D. Robnett. Memories of math: Narrative predictors of math affect, math motivation, and future math plans. *Contemporary Educational Psychology*, 60:101838, 2020.

[14] Young K. Kim and Carol A. Lundberg. A structural model of the relationship between student–faculty interaction and cognitive skills development among college students. *Research in Higher Education*, 57(3):288–309, 2016.

[15] David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyn. Computational social science. *Science*, 323(5915):721–723, 2009.

[16] Liberal Arts Computer Science Consortium. A 2007 model curriculum for a liberal arts degree in computer science. *Journal on Educational Resources in Computing*, 7(2), June 2007.

[17] Ketevan Mamiseishvili and David M. Deggs. Factors affecting persistence and transfer of low-income students at public two-year institutions. *Journal of College Student Retention: Research, Theory & Practice*, 15(3):409–432, 2013.

[18] F. Marton, D. Hounsell, and N.J. Entwistle. *The Experience of Learning: Implications for Teaching and Studying in Higher Education.* Scottish Academic Press, 1997.

[19] Saif Mohammad and Tony Yang. Tracking sentiment in mail: How genders differ on emotional axes. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)*, pages 70–79, 2011.

[20] Sukanya Kannan Moudgalya, Chris Mayfield, Aman Yadav, Helen H. Hu, and Clif Kussmaul. Measuring students' sense of belonging in introductory CS courses. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, pages 445–451, 2021.

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[22] Andrew Petersen, Michelle Craig, Jennifer Campbell, and Anya Tafliovich. Revisiting why students drop CS1. In *Proceedings of the 16th Koli Calling International Conference on Computing Education Research*, pages 71–80, 2016.

[23] Beth A. Quinn, Wendy M. DuBow, and David Sul. Understanding who enrolls in introductory computing courses at community colleges. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, pages 49–55, 2019.

[24] Nathan Rountree, Janet Rountree, and Anthony Robins. Predictors of success and failure in a CS1 course. *ACM SIGCSE Bulletin*, 34:121–124, 2002.

[25] Merilin Säde, Reelika Suviste, Piret Luik, Eno Tõnisson, and Marina Lepp. Factors that influence students' motivation and perception of studying computer science. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, pages 873–878, 2019.

[26] Linda J. Sax, Jennifer M. Blaney, Christina Zavala, and Kaitlin N.S. Newhouse. Who takes intro computing? Examining the degree plans of introductory computing students in light of booming enrollments. In *2020 Research on Equity and Sustained Participation in Engineering, Computing, and Technology (RESPECT)*, pages 1–7. IEEE, 2020.

[27] Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with Python. In *9th Python in Science Conference*, 2010.

[28] Heather B. Shapiro, Clara H. Lee, Noelle E. Wyman Roth, Kun Li, Mine Çetinkaya-Rundel, and Dorian A. Canelas. Understanding the massive open online course (MOOC) student experience: An examination of attitudes, motivations, and barriers. *Computers & Education*, 110:35–50, 2017.

[29] Duane F. Shell, Leen-Kiat Soh, Abraham E. Flanigan, and Markeya S. Peteranetz. Students' initial course motivation and their achievement and retention in college CS1 courses. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, pages 639–644, 2016.

[30] Anya Tafliovich, Jennifer Campbell, and Andrew Petersen. A student perspective on prior experience in CS1. In *Proceeding of the 44th ACM Technical Symposium on Computer Science Education*, pages 239–244, 2013.

[31] Vincent Tinto. Classrooms as communities: Exploring the educational character of student persistence. *Journal of Higher Education*, 68(6):599–623, 1997.

[32] Janet G. Walter and Janelle Hart. Understanding the complexities of student motivations in mathematics learning. *Journal of Mathematical Behavior*, 28(2-3):162–170, 2009.

[33] April Y. Wang, Ryan Mitts, Philip J. Guo, and Parmit K. Chilana. Mismatch of expectations: How modern learning resources fail conversational programmers. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018.

[34] Brenda Cantwell Wilson and Sharon Shrock. Contributing to success in an introductory computer science course: a study of twelve factors. *ACM SIGCSE Bulletin*, 33(1):184–188, 2001.

[35] Laurie Woollacott, Shirley Booth, and Ann Cameron. Knowing your students in large diverse classes: A phenomenographic case study. *Higher Education*, 67(6):747–760, 2014.

[36] Angela Zavaleta Bernuy and Brian Harrington. What are we asking our students? A literature map of student surveys in computer science education. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education*, pages 418–424, 2020.

[37] Daniel Zingaro, Michelle Craig, Leo Porter, Brett A. Becker, Yingjun Cao, Phill Conrad, Diana Cukierman, Arto Hellas, Dastyni Loksa, and Neena Thota. Achievement goals in CS1: Replication and extension. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, pages 687–692, 2018.

[38] Daniel Zingaro and Leo Porter. Impact of student achievement goals on CS1 outcomes. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, pages 279–284, 2016.

[39] Madeline Zug, Hanna Hoffman, Forest Kobayashi, Miles President, and Zachary Dodds. CS for all academic identities. *Journal of Computing Sciences in Colleges*, 33(4):130–137, 2018.