

Understanding Sources of Student Struggle in Early Computer Science Courses

Adrian Salguero, Christine Alvarado, William G. Griswold, and Leo Porter
University of California San Diego
USA

ABSTRACT

Computer science students struggle in early computing courses as evinced by high failure rates and poor retention. As such, studies have attempted to characterize the root of student struggles from many perspectives, including cognitive, meta-cognitive, and social emotional. Typically, studies have limited their inquiry to a specific perspective or a single course. This paper reports the results of a broad student experience survey conducted across several computer science courses. Through a periodic survey, students rated various cognitive, socio-emotional, external, personal, and structural barriers in terms of how much each impacted their learning throughout the term. An exploratory factor analysis of these questions revealed four factors—personal obligations, lack of sense of belonging, in-class confusion, and lack of confidence—that capture a range of possible struggles students may face. We analyzed the prevalence of these factors across courses, performance quartiles, and demographic groups broken down by gender, race/ethnicity, and matriculation status. Students in lower performance quartiles report higher stress levels on multiple factors, with statistically significant differences found between all quartiles and courses, for most factors. Moreover, students from traditionally underrepresented groups report struggling more across all four factors, suggesting that they may be facing more challenges than classmates from represented populations. Overall, these findings indicate that student struggles are associated with stresses from many areas of their lives, suggesting that future interventions should target multiple areas of stress.

CCS CONCEPTS

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

KEYWORDS

computer science education, sense of belonging, student outcomes, diversity, underrepresentation

ACM Reference Format:

Adrian Salguero, Christine Alvarado, William G. Griswold, and Leo Porter. 2021. Understanding Sources of Student Struggle in Early Computer Science Courses. In *Proceedings of the 17th ACM Conference on International Computing Education Research (ICER 2021)*, August 16–19, 2021, Virtual Event, USA. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3446871.3469755>



This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivs International 4.0 License.

ICER 2021, August 16–19, 2021, Virtual Event, USA
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8326-4/21/08.
<https://doi.org/10.1145/3446871.3469755>

1 INTRODUCTION

Computing education is well known to suffer from poor retention of students interested in computing [8, 51] and students not learning what instructors expect [35]. As such, researchers have sought to understand why some students struggle in computing courses by examining a myriad of factors that might be connected to student success [6, 36, 65, 66]. In the United States (the context for our work), there also exists substantial divides by gender and race among those who major in computer science [17] and in those who have access to computing education before college [37, 38]. In addition, Black, LatinX, and Native American and women students who pursue computing in university are retained at lower rates than White and Asian men [70].

Given these challenges, prior studies in computer science education have investigated the impact of a variety of factors on students' outcomes. Some of these factors include prior experience [66], sense of belonging [31, 39, 40, 42, 43, 45, 52, 64], interest in the material [11], self efficacy [33], study skills [32], and alignment between student goals and the perceived goals of the field [31].

These previous studies typically study a single factor in isolation, making it difficult to gain a holistic view of multiple factors that might interfere with a student's learning. Some students might be affected by one factor and other students a different one. Yet other students may be affected by more than one factor at the same time, perhaps compounding their negative effects. Single-factor studies cannot see this larger picture. Similarly, many studies have focused on a single course [32, 45, 49, 68] (typically CS1), perhaps masking challenges that evolve over time, or are unique to a specific topic or timing in the curriculum. Finally, few of these studies have examined the role of students' personal lives (e.g., outside work commitments, family obligations) on their success; existing work in this area is not specific to CS [34, 57]. These limitations might partially explain why some interventions directed at mitigating a particular factor (e.g., meta-cognitive skills or growth mindset) have not had greater impact [10, 26, 58].

The present study investigates how multiple potential sources of student struggle relate to outcomes across multiple computer science courses at a research-intensive university in the United States. Specifically, we investigated how non-academic barriers (e.g., illness, family care) might interact with social emotional and social cognitive barriers (e.g., lack of belonging, lack of peer networks) or structural barriers (e.g., lack of peer support, inability to get help). Drawing on previous studies and our own experience as instructors, we developed a broad survey instrument that was compact enough to be administered regularly with high participation rates. Instructors administered it multiple times throughout the Fall 2019 term in four programming-intensive courses ranging from CS1 to Computer Organization. We then employed Exploratory Factor Analysis

to cluster questions into cohesive categories of factors. Four factors emerged: personal obligations, lack of sense of belonging, in-class confusion, and lack of confidence. We analyzed how these factors related to student performance on the final exam in each course. Finally, we examined differences by gender, race/ethnicity, and matriculation status,¹ again relating reported factors to performance on the final exam. Our results can be summarized as follows:

- When students struggle, they often struggle on multiple fronts. Over 70% of students in the lowest quartile of final exam performance report high levels of stress due to at least one of the four identified factors compared with less than half (30%) in the highest quartile. Over 50% of students in the lowest quartile report high levels of stress for two or more factors. By comparison, half as many students in the next higher quartile report struggling with multiple factors.
- For a given level of performance, women students; Black, LatinX, Native American, and Pacific Islander (BLN+) students; and transfer students report slightly more factors overall interfering with their performance, especially in the lowest quartile of final exam performance. However, which factor interferes the most depends on the demographic.
- Survey response rates for students in the lowest quartile of performance were significantly lower than for higher performing students. This may be a signal of struggle, rather than having given up, because these students still persisted to take the final exam. Because their voices weren't heard, our other results regarding students in the lowest quartile of performance are likely conservative. We believe this result is not unique to our work and exposes an under-acknowledged threat to this type of research (participation rates may vary within subgroups).
- Students in the CS1 course for students with no prior experience reported considerably more struggle than students in the CS1 course for students with prior experience, across all levels of exam performance. This result suggests that computing students with no prior experience need additional support.

2 PREVIOUS WORK

A variety of factors associated with student success have been studied by the community. Factors associated with student success include prior experience [66], sense of belonging [31, 39, 40, 42, 43, 45, 52, 64], interest in course content [11], achievement goals [68], self efficacy [33], study skills [32], and alignment between student goals and the perceived goals of the field [31]. Which students succeed may be impacted by the structure of CS courses or curricula, as the way courses are taught [46, 47, 51], the learning environment [3, 4, 18, 19, 53], and the competitive nature of the program [43] (among others). In this section we review the theory and prior work related to the factors most closely related to those examined in our current work.

¹In the United States, 2-year community colleges provide high school graduates with an opportunity to prepare themselves for acceptance to a 4-year university, for the latter half of their studies, at a lower net cost.

2.1 Social-Emotional Factors

Situated learning theory asserts that there exists an important relationship between the act of learning and the social context in which the learning occurs [30]. Legitimate Peripheral Participation says that learning occurs within a community and that "...mastery of knowledge and skill requires newcomers to move toward full participation in the sociocultural practices of a community" [30]. From this lens, if a student feels they do not belong in that community, it may be difficult for them to fully participate in the community, potentially limiting their ability to make connections with their peers and potentially impacting their performance in the course.

Studies in computing have shown how sense of belonging has been associated with various important student factors such as performance, retention, motivation, and persistence in computing, with race and gender being significant predictors of feelings of belonging [20, 25, 31, 39, 52, 64]. A study focusing on why students leave CS1 shows that sense of belonging from lack of social groups can impact the student experience [45]. Understanding how student belonging can be an influence might help in retaining more students. A recent study focused on how providing students networking, outreach, and mentoring opportunities positively influenced their sense of belonging in computing [42]. Apart from hurting a student's academic success, this lack of belonging has also been tied to feelings of depression [21].

Other studies have focused on self-efficacy in computer science and its association with anxiety, success, and interest in the field [16, 41, 66]. The concept of self-efficacy makes the claim that individuals with high self-efficacy are likely to face difficult challenges while those with low self-efficacy will try to avoid them [2]. Studies have shown that women in computing struggle with self-efficacy, which can impact their performance [7, 9].

Prior studies that have observed student outside obligations, such as work and family, have focused on how work-life balance is associated to well-being, anxiety and depression. Although we know these issues can have an impact on student performance, to our knowledge, no prior study has directly associated these outside obligations to student exam performance.

2.2 Student Goals or Behaviors Associated with Success

Goal congruity theory argues that student personal goals should be aligned with opportunities to achieve those goals in a field in order for students to enter and persist in the field. When these goals are not aligned, students are more likely to exit the field [13, 14]. In a recent study, Lewis et.al., demonstrated how high communal goals led to a lower sense of belonging among computing students [31]. Since communal goals have been shown to be heavily endorsed among women and BLN+ students, two groups that computing education struggles to retain [55, 61], a lack of alignment between goals and perceived opportunities in CS (and STEM) may be a source for the lack of retention of women and BLN+ students.

Bandura's Social Cognitive Theory asserts that learning occurs in a social context within a dynamic interaction between people, the environment, and behavior [29]. Focused on the behaviors of students, a recent study observing behaviours of high and low performers in a CS1 course [32] identified different help-seeking

behaviors among low and high performing students in terms of who they approached (friends, instructional staff) and the type of help they sought. The availability of help for students is a potential structural (environmental) barrier included in this study.

Achievement goals are students' goals for success in a particular setting. The goals that students' approach (mastery versus performance) have been shown to impact student success in computing courses [68, 69]. Although this area of goal types are relatively new in its application of CS, studies found that students who are focused more on mastering the material (mastery goals) are more likely to perform better in their CS courses and express interest in pursuing CS as a field than students who want to outperform their peers (performance goals).

2.3 Challenges due to Students' Personal Lives

A few studies have examined the role of students' personal lives on their success, but this work has generally not been specific to CS. A study of Scottish further education students found that full-time students in the study worked in a job between 16-20 hours per week and had different strategies for coping with the demands of job, school, family, and friends [34]. In a study of Midwestern college students, Sprung and Rogers found that students with a better work-life balance had lower levels of stress and depression [57]. Stratton et al. examined how different obligations impact post-secondary student enrollment and drop-out rates [59] and found that outside employment, marriage, work-study aid, and grants were all connected to the likelihood of staying in the program.

In computer science, previous studies have not examined the link between outside obligations and success explicitly, though a few studies have examined the effect of multiple different barriers to student retention and motivation. One study examined how multiple factors, such as personal values, teaching quality, satisfaction with their learning, and student motivations impacted retention in CS [44]. Another study in CS used a four-factor model to describe student motivations found that students' enjoyment of completing academic tasks and the utility of studying CS were most important to motivation, whereas factors such as the influence of friends and family, were less important [50].

3 STUDY DESIGN

3.1 Research Questions

As motivated in the Introduction, our study addresses two research questions:

- **RQ1**—What high-level social-emotional/social-cognitive, structural, or personal factors are related to student outcomes?
- **RQ2**—How do any identified factors and associated outcomes vary by demographic group?

3.2 Course Context

Our study was conducted in the Fall of 2019 at a large US research-intensive university operating on a quarter system with an 11 week term (10 weeks of instruction and one week for final exams). We studied four undergraduate lower-division computer science courses: A CS1 course requiring no prior programming experience

(CS1-NPE); a fast-paced CS1 for students with prior programming experience (CS1-PE); a CS2 course that includes Basic Data Structures and Object-Oriented Programming; and a Computer Organization course (CompOrg) focused on C, assembly, and basics of hardware design. CS1-NPE and CS1-PE have the same formal prerequisites, but students are advised to take one course or the other based on their prior experience. CS1-NPE is the first course in a pair of courses taken across two terms that together are equivalent to one term of CS1-PE. The second course in the pair was not included in our study because few students take this course in the fall quarter. CS1-PE, CS1-NPE, and CompOrg were taught using active learning techniques (Peer Instruction [12], worksheets, etc.) and CS1-NPE included an interactive textbook with auto-graded programming exercises. CS2 was taught using a more traditional lecture format. A breakdown of our study population can be found in Table 1.²

3.3 Survey Design

Our survey was designed not only for research, but also as a tool for instructors at our institution to gain insight into their students concerns, respond to student issues, and make course corrections throughout the term. This motivated designing a survey that collected actionable data throughout the term and maximized student response rate. As such, we collected data regarding a wide range of possible struggles drawn from previous literature, our own experience as educators and instructors, and student feedback. As researchers, we hoped a broad survey would allow new perspectives on these potential struggles to emerge during the analysis. Question (6) is the one question taken directly from an existing instrument, the sense of belonging instrument used in Smith et al. [56].

We trialed our survey in CS1-PE and CS2 in the summer of 2019, both to assess whether it could be completed within about 5 minutes and to solicit other barriers to learning that we might have missed. The final survey is shown below (questions used in this study have been bolded).

- (1) **In the last week, approximately how many hours did you spend outside of class time working on work for this course?** (*Open text box, numeric answer required*)
- (2) How challenging was the work for this class this week? (*Likert scale: 1–5 where 1 was labeled as "Not at all challenging" and 5 was labeled as "Extremely challenging"*)
- (3) In the past week, which of the following, if any, did you seek help from/work with for this course? (*Checkboxes, select one or more*)
 - (a) My peer(s) in this class (as pair programming/joint assignment submission)
 - (b) My peer(s) in this class (for help completing my own version of the assignment)
 - (c) My friends who are not in this class
 - (d) Piazza (I posted one or more questions)
 - (e) Piazza (I got help from reading responses to others' questions)
 - (f) A tutor
 - (g) A TA

²We acknowledge that gender extends beyond simply male and female. However, our data was collected as male/female so to remain consistent we decided to maintain this binary classification in our analysis.

Table 1: Summary of courses, with the number of participating students, breakdown by group, number of surveys administered, and quartile size (see Section 4.2). Numbers in parentheses give proportions between genders (Male vs. Female), race categories (Non-BLN+ vs. BLN+) and matriculation status categories (First-Year vs. Transfer).

Course	Total Students	Male	Female	Non-BLN+	BLN+	First-Year	Transfer	Surveys Given	Quartile Size
CS1-NPE	544	349 (0.64)	187 (0.34)	455 (0.84)	89 (0.16)	477 (0.88)	58 (0.11)	7	136
CS1-PE	436	340 (0.78)	94 (0.22)	410 (0.94)	26 (0.06)	390 (0.89)	44 (0.10)	8	109
CS2	268	187 (0.70)	79 (0.29)	241 (0.90)	27 (0.10)	198 (0.74)	67 (0.25)	10	67
CompOrg	428	318 (0.74)	109 (0.25)	380 (0.89)	48 (0.11)	365 (0.85)	61 (0.14)	4	107

- (h) The instructor
 (i) I did not seek help from anyone this week
 (j) Other (Please specify):
- (4) **In the past week, on a scale of 1 to 5, to what degree did each of the following interfere with your ability to learn and complete the work for this course? (Likert scale for each item: 1–5 where 1 was labeled as "Not at all" and 5 was labeled as "Significantly")**
- (a) Requirements for other classes
 - (b) Illness
 - (c) Family obligations
 - (d) Work obligations
 - (e) Social/personal life issues
 - (f) Confusion specifically about the assignment
 - (g) Confusion generally about the material
 - (h) Getting stuck on a bug
 - (i) Inability to get help
 - (j) Embarrassment/discomfort asking others
 - (k) Self-doubt/lack of confidence
 - (l) Lack of interest in the assignment or material
 - (m) Goofing off/procrastination
 - (n) Other (Please specify:)
- (5) **What is your current overall satisfaction with your performance in this class? (Likert scale: 1–5 where 1 was labeled as "Extremely dissatisfied" and 5 was labeled as "Extremely Satisfied")**
- (6) **Reflecting on your experiences over the past week, to what extent do you agree with the following statements. (Likert scale for each item: 1–5 where 1 was labeled as "Not at all" and 5 was labeled as "Completely")**
- (a) I feel accepted in this class
 - (b) I feel comfortable in this class
 - (c) I feel supported in this class
 - (d) I feel like I don't belong in this class
- (7) **At this time, approximately how many other students in this course would you be comfortable reaching out to study with? (Open text box, numeric answer required)**
- (8) **Reflecting on the last week, how stressed have you been overall? (Likert scale for each item: 1–5 where 1 was labeled as "Not stressed at all" and 5 was labeled as "Extremely stressed")**

- (9) **Optional: If you want, please enter any information to expand on or explain your answers to any of the questions on this survey. (Open text box.)**

For the study we present here, we focused on questions (1), (4), (5), (6), and (7), 18 questions in all, as question (4) is comprised of 14 questions. We chose to omit the remaining questions as they did not provide specific detail on why the student would be likely to struggle. Sentiments generated from question (2) and (8) may repeat in other questions such as questions on the assignments, material, or other personal issues. Question (9) is open-ended. Question (3) was not included because we were unclear on how to meaningfully combine this with the rest of our data.

3.4 Survey Administration and Data Cleaning

Students were given the survey by their instructor as part of the course's weekly or bi-weekly homework assignments, and they received a small amount of credit for completing it. Instructors maintained administrative control over the surveys in order to ease their use for responding to student issues and making course corrections on the fly. Survey responses were not anonymous to facilitate individual student support from the instructor, though all identifying information was removed before analysis according to our approved human subjects protocol, as described below.

Prior to analysis, all survey responses, as well as course performance data including overall course grade and final exam score, were sent to an external team that deidentified the data and removed any students who were under 18 years old or had opted out of the study. Using registrar data, this team also added demographic data about the students, notably race/ethnicity, gender, and matriculation status. Our university uses a binary gender classification scheme (Male/Female), which we acknowledge limits our analysis. For race/ethnicity, it uses several categories. However, for the privacy of students from groups that are represented in small numbers, the race/ethnicity data we received was aggregated into two groups: (1) White/Asian students and (2) Black, LatinX, Native American, and Pacific Islander (BLN+) students.

We performed a small amount of pre-processing on the data in order to maximize consistency. First, the answers for the sense of belonging questions (6a), (6b), and (6c) were reverse-coded, as the

1-5 range runs from negative to positive sentiments, whereas for all the other questions the 1-5 range runs from positive to negative responses. As a result, a higher number is negative for all these questions.

Second, in order to account for a different number of surveys offered per course, for each question we calculated each student's average response across all surveys they completed. Although using averages with Likert-scale data is controversial, we chose to average each student's responses instead of taking the student's median response for two reasons. First, students responded on a 5-point scale with labels on only the extremal values, making the difference between the levels appear linear. Second, median values would potentially drop important variations between a single student's responses across the quarter. For example, a student who responded with 1, 1, 1, 5, 5 on five surveys would be recorded with a response of 1 if we had used medians, completely losing the struggle the student experienced in the last part of the quarter. There were cases in which a student did not respond to any surveys during the term. These students were removed from any analysis, but were included when providing descriptive statistics of our courses, quartiles, and indirectly, response rates to surveys.

Finally, to associate survey responses with outcomes, we needed a metric to compare student performance across the four courses. We settled on using final exam score, as it is individual work, proctored, and designed to measure knowledge with no weight given to participation or effort. To address variations in difficulty and grading, we normalized the scores across the courses using z-scores.

3.5 Exploratory Factor Analysis

The breadth of our survey meant that it was not immediately clear which individual challenges might comprise higher-level themes. Thus, after cleaning the data, we applied Exploratory Factor Analysis (EFA) to discover underlying structure in the data [15]. EFA finds shared variance among the variables and combines them into factors, which may be non-observable. To ensure that EFA could be applied, we conducted both the Bartlett's Sphericity and Kaiser-Meyer-Olkin tests to our data. These tests determine if factor analysis can be applied to the data by testing the overall significance of the correlations within the correlation matrix and testing if relationships between variables was high respectively. Both tests passed, with Bartlett's Sphericity being significant ($\chi^2 = 27706$, $p \ll 0.01$) and KMO indicating appropriately strong relationships between variables (KMO = 0.94) [5]. In the EFA analysis itself, we assigned a question to a factor if it had a loading of 0.4 or greater [54]. We also applied an oblique rotation in our EFA process, as we assume that our questions are not independent of each other [5]. To determine the number of factors to which it would be appropriate to fit our data—balancing model parsimony and the captured variance—we included those factors that had an eigenvalue greater than 1. To validate the appropriateness of this cutoff, we examined both the last factor included and the first factor left out, including their eigenvalues and the amount of variance they explained. We also used a scree plot visualization as an additional check of sensibility. We describe the results of our factor analysis in the next section.

4 RESULTS

4.1 Factors Derived from Exploratory Factor Analysis

Using EFA as described in the previous section, we found that four factors were the right number to best describe our data set. Running EFA with one more factor explained only slightly more variance (2%), and generated a factor composed of one survey question. The survey questions that comprise our factors are shown in Table 2, along with their loadings onto their factors. We were able to identify a common theme among the questions that make up each factor and hence named them: Lack of Sense of Belonging (LSoB), In-Class Confusion (ICC), Personal Obligations (PO), and Lack of Confidence (LoC). We then ran Cronbach's alpha for each of our factors, all of which demonstrated good internal consistency [60] and are above the 0.7 threshold which is considered sufficient evidence for internal reliability [22].

Some questions did not load well onto any factors. Three questions are worth discussing. First, the EFA did not capture question (6d) in LSoB. Question 6d had a factor loading of 0.37, below our threshold of 0.4. It's possible that students were inconsistent in their responses to (6d) because the preceding questions (6a)-(6c) were worded for 1 being the worst and 5 being the best, while (6d) was coded the opposite way. Second, questions (1) and (7) were not captured by our model. We are not sure why these questions did not enter into our model, as we had expected them to be important.

The rest of our analysis will be driven by observing the frequency and associations of these factors to student performance in courses and among different demographics in our study population. When using the term "factor" in the remainder of this section, we are referring to one of the factors determined using EFA. When referring to students' reporting on a factor, we mean reporting on one or more questions that our EFA linked to that factor.

4.2 RQ1: Association between Factors and Outcomes

We explored the relationships of the factors to student performance using three sources of information: the distribution of reported factors among students at different performance levels and to highlight responses among struggling students, the combinations of reported factors at different performance levels, and the survey response rate at different performance levels.

To explore how responses differed among students across different performance levels, we partitioned students into four equal-sized quartiles based on their final exam scores, with Q1 being the top 25% and Q4 being the bottom 25%. (Quartile sizes are listed in Table 1).

To compute each student's score for each factor, we averaged the student's responses for the questions comprising the factor (cf. Table 2) [15]. For example, a student's score for the Personal Obligations factor is the sum of their (average) responses to questions 4a, 4b, 4c, 4d and 4e. As each question was a Likert score from 1–5, this score has a range from 4–20. For cross-factor comparison and visualization of factor scores, we normalized individual factor scores by dividing by the number of questions composing that factor, creating a shared range of 1–5.

Table 2: Survey question makeup of each EFA factor, along with factor loadings, variance explained, and Cronbach’s Alpha.

Factors	Component Survey Questions	Question Loading	Variance Explained	Cronbach’s Alpha
Lack of Sense of Belonging (LSoB)	5 (Performance Satisfaction)	0.54	16%	0.93
	6a (Feel accepted)	1.00		
	6b (Feel comfortable)	0.95		
	6c (Feel supported)	0.96		
In-Class Confusion (ICC)	4f (Confusion on assignment)	0.87	13%	0.89
	4g (Confusion on material)	0.83		
	4h (Getting stuck on a bug)	0.77		
Personal Obligations (PO)	4a (Other Course Requirements)	0.51	11%	0.78
	4b (Illness)	0.64		
	4c (Family obligations)	0.73		
	4d (Work obligations)	0.67		
	4e (Social/Personal life issues)	0.65		
Lack of Confidence (LoC)	4j (Embarrassment)	0.80	10%	0.86
	4k (Lack of Confidence)	0.81		
	4l (Lack of Interest)	0.59		

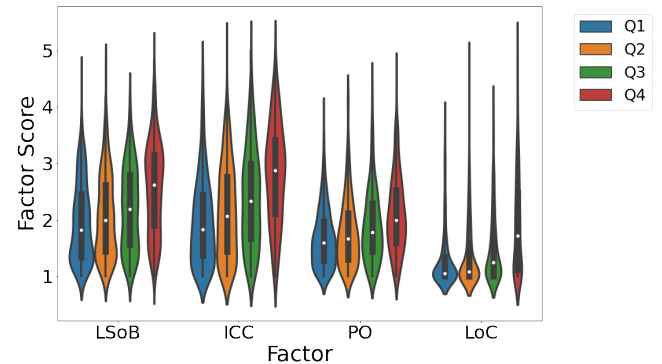
4.2.1 Distributions of Reported Factors. We first explore how responses for each factor vary for students at different performance levels. Figure 1 shows the distributions (as violin plots) of scores for each of the four factors, across all courses, separated by quartile. The small white dot in the center of each violin indicates the median of each (normalized) factor score, while the rest of the violin illustrates the distribution. Two trends are clear from Figure 1.

First, the distributions of responses across factors are different. Students report the highest scores on ICC and LSoB, slightly lower scores for PO, and lower scores for LoC. Only students in Q4 seem to really struggle with lack of confidence. Using Kruskal-Wallis tests, we determined that overall the differences between factors are statistically significant (see Table 3, top row). Post-hoc tests confirmed that all pairwise comparisons are also statistically significant. The large effect size for this difference between factors appears in the top row of Table 3 and was calculated with the eta-squared measure using the H-statistic generated by the Kruskal-Wallis test [62].

Second, for all factors, there is a statistically significant difference in distributions across quartiles (see Table 3, second row). In Figure 1, both the median factor score and the number of students at the high end of the distribution increases from Q1 to Q4, indicating that students report more struggle with all factors as the quartiles go from high performing to lower performing. Concerningly, some students in Q3 and Q4 still report low scores for ICC, perhaps evidence supporting the Dunning-Kruger effect [28] and showing that students may not be able to accurately self-evaluate their struggles. The effect sizes for “Quartiles” demonstrates the degree by which student struggle varies by performance quartile. For example, ICC has the largest effect size of the factors which is reasonable considering students who are lower performing should be more likely to report higher levels of confusion.

Figures 2, 3, 4, and 5 provide violin plots for each of the four factors, broken down by both course and quartile. Generally, we see the same trends that factor median scores increase and the distribution becomes more weighted toward the top as we move from Q1 to Q4, although the severity of the increase varies by

factor and by course. For LSoB (Figure 2), CS1-PE and CS2 do not completely fit the overall trend. For CS1-PE, Q1–3 are fairly similar with only Q4 showing a marked decrease in sense of belonging. CS2 reports the lowest sense of belonging with even students in Q1 reporting a sense of belonging comparable to Q3 or Q4 for the other courses. Its unclear why the CS2 course is different; potential sources may be the courses notoriously high difficulty and the more traditional lecture format used in that course (relative to the active learning methods used in the other courses).

**Figure 1: Distribution of factor scores across quartiles and factors. The distribution for each factor and quartile span across all courses.**

CS2 and CompOrg also seem to differ from CS1-NPE and CS1-PE in a couple of ways. First, students report higher rates of confusion in these courses overall, which may be an artifact of higher difficulty of these courses (Figure 3). Second, for LoC (Figure 5) CS2 and CompOrg have higher factor scores even among Q1–3, again perhaps due to the difficulty of these later courses in the curriculum. Finally, for PO (Figure 4), we see that CompOrg has higher numbers in general, even for students in Q1–Q2. This may be an artifact of students taking more courses, the addition of transfer

Table 3: Kruskal-Wallis score and effect size across factors, courses and quartiles. A * indicates statistical significance of $\alpha < 0.05$. Interpreting the eta-squared effect size for Kruskal Wallis, 0.01 through < 0.06 is small, 0.06 through < 0.14 is moderate, and anything greater than or equal to 0.14 is large [24].

Group	EFA Factor	Kruskal-Wallis Value	p-val	Effect Size
Factors	-	1163.04	$p < 0.01^*$	0.18
Quartile	LSB	129.28	$p < 0.01^*$	0.08
	ICC	196.63	$p < 0.01^*$	0.12
	PO	99.26	$p < 0.01^*$	0.06
	LoC	165.51	$p < 0.01^*$	0.10

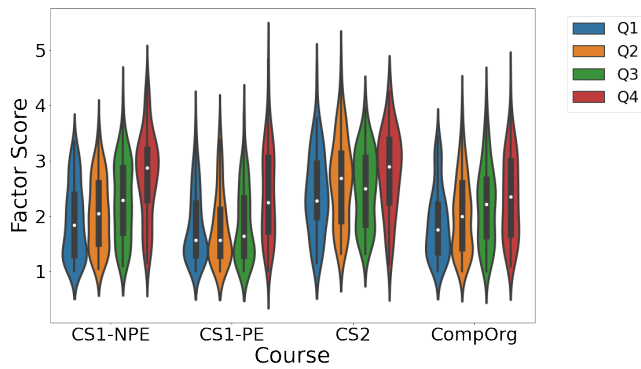


Figure 2: Lack of Sense of Belonging (LSb) factor score distribution by course and quartile.

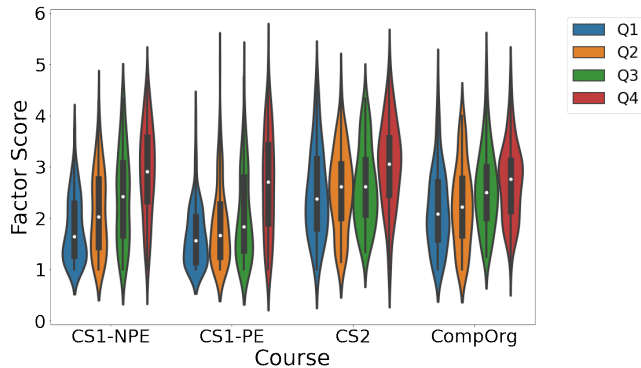


Figure 3: In-Class Confusion (ICC) factor score distribution by course and quartile.

students (who often start in CS2 or CompOrg) who tend to have more personal obligations as shown later this section, or that high performing students in CompOrg are often hired as instructional staff for the earlier courses.

4.2.2 Reporting of Simultaneous Factors. The previous section showed that higher factor scores are associated with lower performance. In this section, we examine how many students in each quartile are reporting high levels of of struggle with more than one factor.

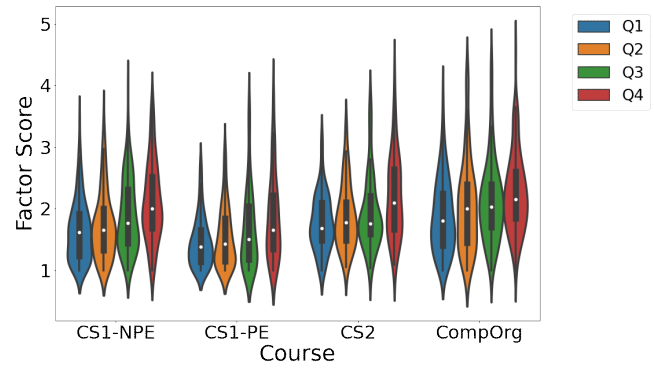


Figure 4: Personal Obligations (PO) factor score distribution by course and quartile.

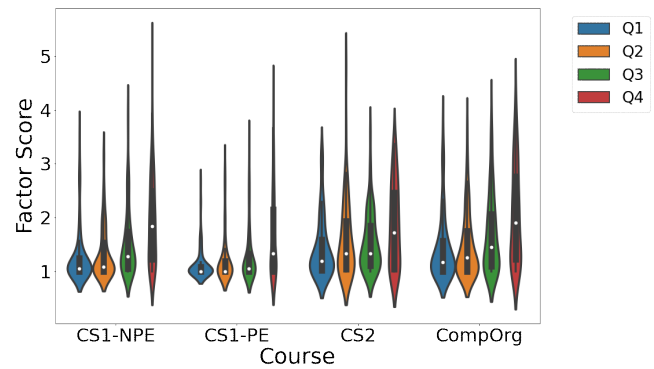


Figure 5: Lack of Confidence (LoC) factor score distribution by course and quartile

We defined a high level of struggle as having a factor score above the 75th percentile in comparison to the level reported by other students. (We note that we experimented with different thresholds and similar trends were present.)

Figure 6 shows the proportions of students from each quartile who report high levels of struggle for each factor. We find that students in the bottom quartile make up a far greater proportion (approaching half in some cases) of the students who report high levels of struggle on each factor compared to students in the other three quartiles. What is missing from this figure (and the previous section) is precisely how these factors might be combining for an individual student. As such, we sought to determine whether students are reporting interference with their performance from multiple factors. If students tend to struggle with just one factor, then perhaps targeted interventions aimed at the indicated factor would help. But if there is a group of students struggling with two or more factors, then more comprehensive interventions may be necessary.

Figure 7 shows the proportion of students who were above the 75th percentile in reporting high amounts of four, three, two, one, or zero of the factors. Recall that because we defined a student as experiencing a particular factor as a relative percentage to other students, each factor will be equally represented. We can clearly see that the proportion of students who report a high amount of

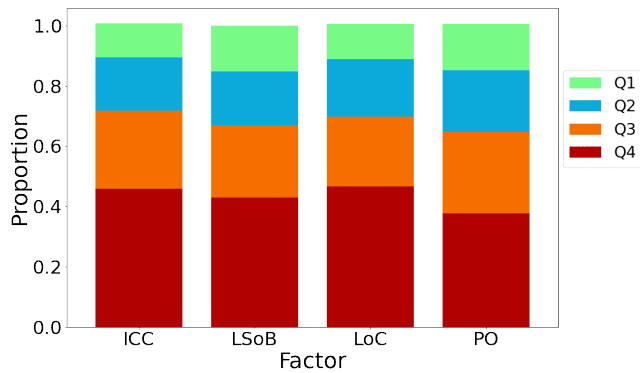


Figure 6: Proportion of students who report above the 75th percentile for each factor. The stacked bars represent quartiles 1 through 4.

struggle on zero factors decreases by quartile. Likewise, the number of students reporting high amounts of struggle on all four factors increases notably between Q1 and Q4. There are also consistent increases in the proportion of students reporting two or three factors across the quartiles. We conducted a Pearson’s correlation between the number of factors and normalized final exam score across all students and classes and found that number of factors reported as high struggle was negatively correlated with final exam score ($r = -0.35$, $p < 0.01$).

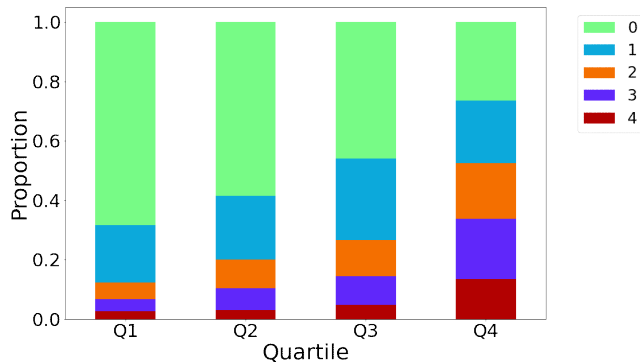


Figure 7: Proportion of factors students self-report above the 75th percentile per quartile. The data per quartile covers all courses.

Differences between CS1-NPE and CS1-PE. We next examined whether and how combinations of high degree of struggle with factors varied between courses. We found that for most courses the patterns were similar, except there is a notable difference in the results between CS1-NPE and CS1-PE. Figure 8 shows the factor breakdown by quartiles in CS1-NPE and CS1-PE, respectively. When examining the proportions of students reporting high degrees of struggle for a single course, recall that the 75th percentile threshold was set for all courses so one course may have more students above that threshold than another.

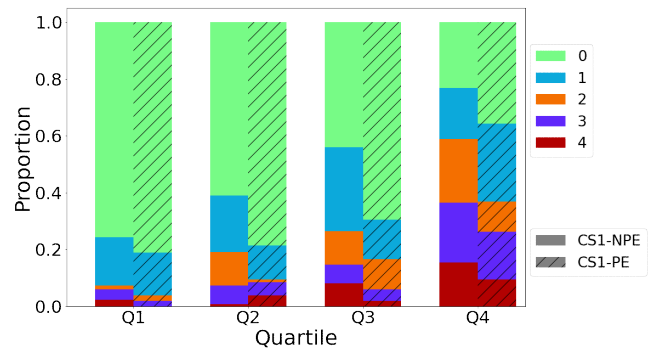


Figure 8: Proportion of factors students self-report above the 75th percentile, per quartile, in CS1-NPE and CS1-PE.

For all quartiles, CS1-NPE has a higher proportion of students with a high level of at least one factor. Surprisingly, students in CS1-NPE in Q1 and Q2 are reporting far more struggles than CS1-PE students in those same quartiles. These findings highlight the value of a CS1-NPE that offers a more “gentle” introduction to computing, as the CS1-NPE population faces more challenges in their computing course relative to those in CS1-PE. One reason for these differences might be due to CS1-NPE attracting more students who are not majoring in, nor intending to major in, computer science. The population of CS1-NPE can also skew toward students later in their studies so they may have more personal commitments relative to students earlier in their studies. It is also possible that these differences are due to (as we will see in the next section) the more diverse population of students who take CS1-NPE. In any event, the goal of the CS1-NPE course is to be sensitive to these issues and is perhaps not fully realizing that goal.

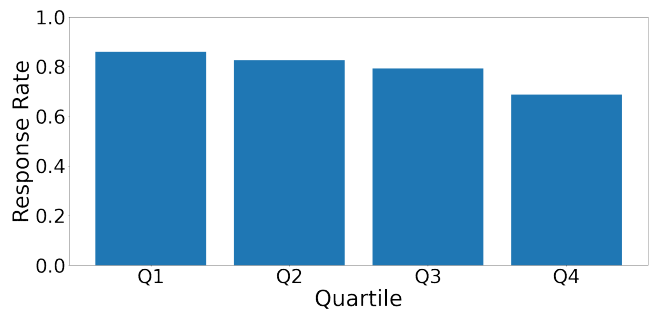


Figure 9: Survey response rates across all courses by quartile. Non-responders: Q1 = 3, Q2 = 2, Q3 = 8, Q4 = 27.

4.2.3 Response Rates. After examining some of student responses, we became concerned that survey response rates might not be uniform over quartiles, thus skewing the above results. Figure 9 presents the response rates, per quartile, across all courses. Consistently, lower performers completed fewer surveys, with students comprising the first quartile completing 86% of surveys and those in the fourth quartile completing 69%. Students who responded to zero surveys followed a similar, but somewhat stronger pattern across all courses: 3 in Q1, 2 in Q2, 8 in Q3, and 27 in Q4. The trends are the same at the course level. We note this both because:

(1) this pattern of response rates could itself be a signal of distress (e.g., lack of time) and (2) underreporting for students in the lower-performing quartiles suggests that the results presented above may be conservative, as students who do not feel they have enough time to complete a survey might report high levels of these factors interfering with their performance.

4.2.4 Take-Aways for RQ1: Association between Stress Factors and Outcomes. Overall, we see that students are struggling on many factors, at times even simultaneously, and these show a consistent pattern with respect to lower outcomes. Students in our CS1-NPE experience higher levels of outside stress than those in CS1-PE, supporting the need for such a course that can be tailored to this population. Interventions on a specific factor may help some students, but many of the students who need the most support, those in Q3 and Q4, may require interventions in multiple areas that span both inside and outside the classroom.

4.3 RQ2: Association between Stress Factors and Outcomes, by Demographic

As cited in Section 2, there is reason to believe that there are disparities in the challenges encountered by minority demographics. This section performs the same analysis as for RQ1, but factored by majority/minority group demographics. For demographic breakdowns across courses and quartiles, please refer to Figures 10, 14, and 18 in each group's respective section. Unfortunately, these figures show that for most courses women, BLN+, and transfer students are underrepresented in the higher quartiles and over-represented in the lower quartiles compared to men, non-BLN+ students and first-year students, respectively. The analysis in this section may explain some of these differences.

4.3.1 Gender. As shown in Figure 11, compared to men, women report a slightly higher incidence of most factors. Figure 12 reveals that women also consistently report experiencing interference from multiple factors, especially in Q2 and Q4. The differences are statistically significant as seen from our results in Table 4. Response rates for men and women, as shown in Figure 13, follow a similar pattern as seen for the whole cohort, with women overall responding at a slightly higher rate.

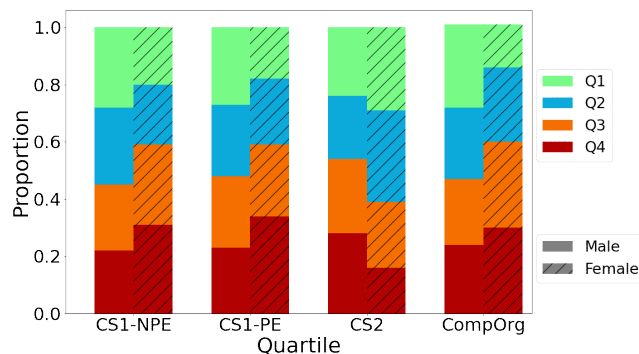


Figure 10: Proportion of male and female students in each quartile across courses.

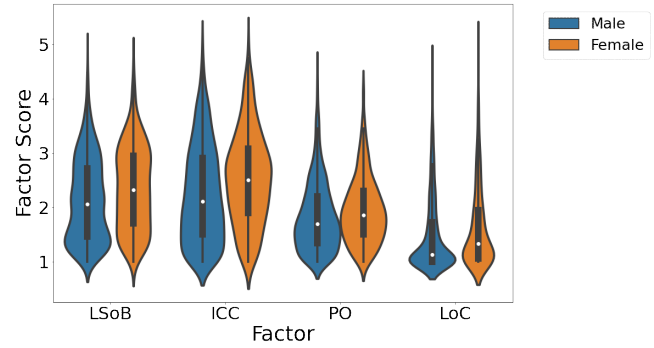


Figure 11: Distribution of factor scores across gender. The distributions for each factor and quartile span across all courses.

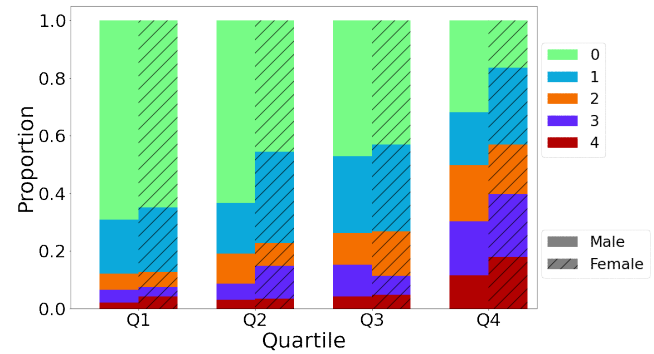


Figure 12: Proportion of number of factors students self-report above the 75th percentile, by gender across all courses.

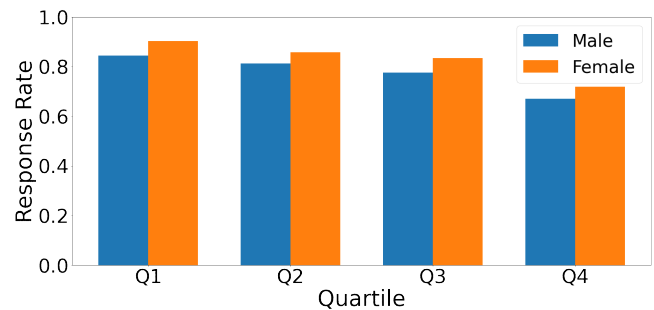


Figure 13: Survey response rates across all courses by gender. Male non-responders: Q1 = 3, Q2 = 2, Q3 = 5, Q4 = 20. Female non-responders: Q1 = 0, Q2 = 0, Q3 = 3, Q4 = 7.

4.3.2 Race/Ethnicity. Differences between BLN+ vs. non-BLN+ students for the four factors overall (Figure 15) are small and only significant for PO and ICC (Table 4). The stacked bar charts seen in Figure 16 reveal that BLN+ students report high degrees of struggle with the different factors in similar proportions to non-BLN+ students, with a few notable exceptions. First, over the first three quartiles BLN+ students report a lower incidence of multiple factors

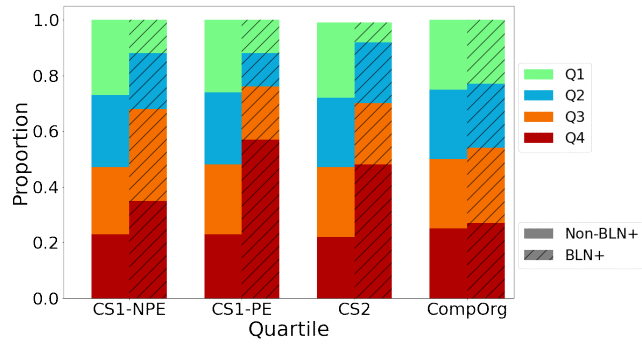


Figure 14: Proportion of non-BLN+ and BLN+ students in each quartile across courses.

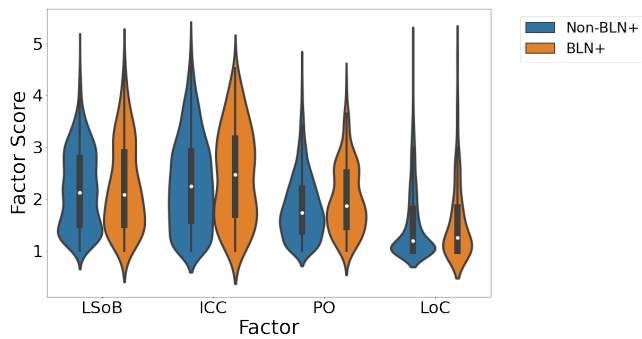


Figure 15: Distribution of factor scores across race. The distributions for each factor and quartile span across all courses.

than non-BLN+ students, but in Q4 double the proportion of BLN+ students report a high degree of interference from all four factors. This suggests that the educational structures at our institution may be doing a better job of helping non-BLN+ students cope when everything is going wrong, compared to BLN+ students, which could partially explain the high representation of BLN+ students in Q4 in most courses (Figure 14). Based on response rates in Figure 17, we see the same trend of lower quartiles responding at lower rates for both BLN+ and non-BLN+ students. Apart from Q4, however, the response rates between BLN+ and non-BLN+ students are very similar.

4.3.3 Transfer Status. A significant minority of students arrive at our university through the transfer pathway, with the majority of students matriculating directly out of high school. We wished to study transfer students as they are generally older and perhaps have a family to support, and so may struggle from different factors than an average student who enters our university directly from high-school. Figure 19 shows that transfer students experience higher levels of all factors than first-years. Higher levels of interference from personal obligations is not surprising given the aforementioned difference between transfer students and first-year students. In terms of ICC, more experienced students may be more capable of recognizing areas of confusion or they may not be as

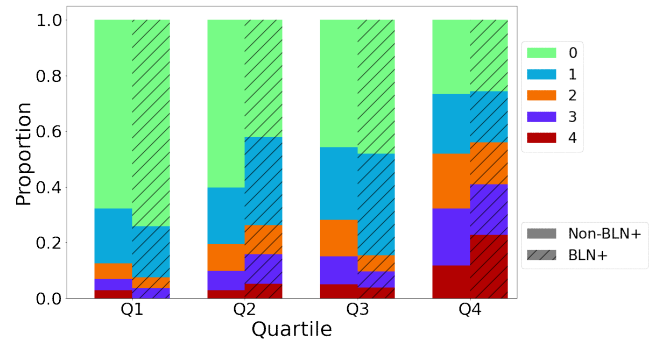


Figure 16: Proportion of number of factors students self-report above the 75th percentile, by race across all courses.

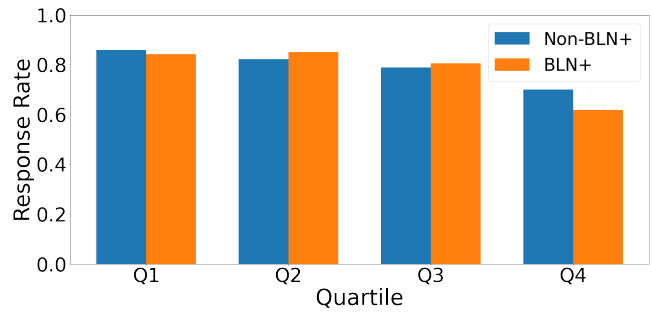


Figure 17: Survey response rates across all courses by students' race/ethnicity. Non-BLN+ non-responders: Q1 = 3, Q2 = 2, Q3 = 7, Q4 = 21. BLN+ non-responders: Q1 = 0, Q2 = 0, Q3 = 1, Q4 = 6.

well prepared for the class as their prior classes may not align perfectly with what is taught at our institution. For sense of belonging, they may feel the stigma of the age difference (e.g., being “behind”) compared to their majority first-year peers. From Figure 20 we can see that a larger proportion of transfer students report a high degree of interference for multiple factors than their first-year peers. In Q4, transfers report over 50% more multiple factors than their first-year counterparts, suggesting that the educational structures at our institution may be better serving their majority counterparts. In terms of response rates (Figure 21), we see a similar trend as with other groups.

4.3.4 Kruskal Wallis Tests. We ran Kruskal-Wallis tests and effect size across gender, race/ethnicity, and matriculation status. Results can be found in Table 4. For the majority of tests we found statistical significance between the groups based on their factor average. We see that comparing groups based on race/ethnicity was not significant for sense of belonging, which is interesting since prior work focuses heavily on how these groups suffer from lack of belonging in computing [52]. For those factors where the findings are significant, there are small effect sizes for gender, BLN+, and transfer students. These small effect sizes seem to indicate that the differences between the subgroups within each group is small for each individual factor. These results show us that struggles our

students experience may not be substantially increased or mitigated based on their gender, race/ethnicity, or matriculation status.

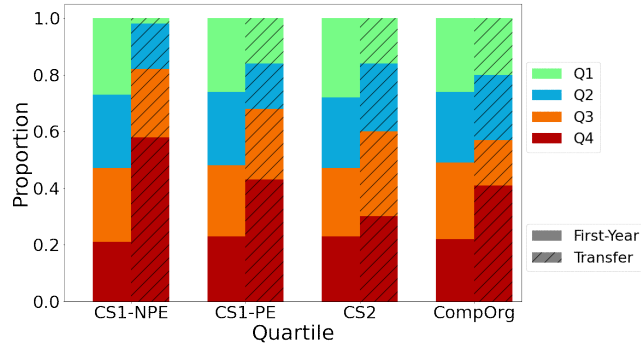


Figure 18: Proportion of first-year and transfer students in each quartile across courses.

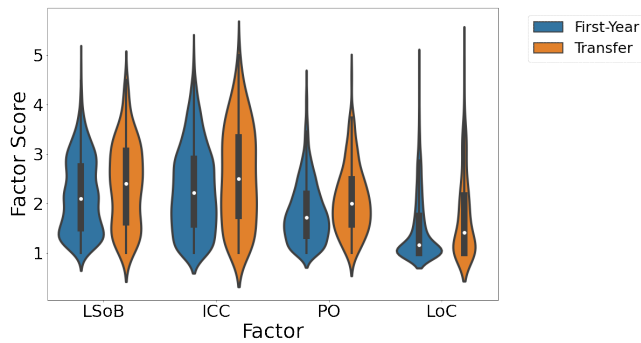


Figure 19: Distribution of factor scores across matriculation status. The distributions for each factor and quartile span across all courses.

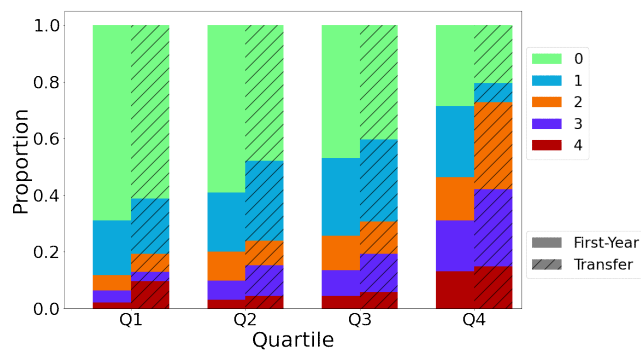


Figure 20: Proportion of number of factors students self-report above the 75th percentile, by matriculation status across all courses.

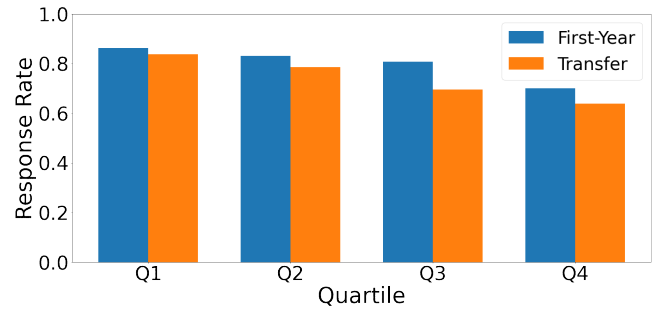


Figure 21: Survey response rates across all courses by whether a student is a first-year or transfer. First-year non-responders: Q1 = 3, Q2 = 2, Q3 = 3, Q4 = 16. Transfer non-responders: Q1 = 0, Q2 = 0, Q3 = 3, Q4 = 10.

Table 4: Kruskal Wallis test results for each factor across different groups. A * indicates statistical significance of $\alpha < 0.05$. Interpreting the eta-squared effect size for Kruskal Wallis, 0.01 through < 0.06 is small, 0.06 through < 0.14 is moderate, and anything greater than or equal to 0.14 is large [24].

Group	EFA Factor	Kruskal-Wallis Value	p-val	Effect Size
Gender	LSoB	26.84	$p < 0.01^*$	0.02
	ICC	36.16	$p < 0.01^*$	0.02
	PO	15.78	$p < 0.01^*$	0.01
	LoC	32.85	$p < 0.01^*$	0.02
Race/Ethnicity	LSB	0.29	0.59	0.00
	ICC	4.51	0.03*	0.00
	PO	10.09	$p < 0.01^*$	0.01
	LoC	0.93	0.34	0.00
Matriculation Status	LSB	13.86	$p < 0.01^*$	0.01
	ICC	13.94	$p < 0.01^*$	0.01
	PO	27.36	$p < 0.01^*$	0.02
	LoC	20.25	$p < 0.01^*$	0.01

5 LIMITATIONS TO GENERALIZABILITY AND THREATS TO VALIDITY

5.1 Limits due to Studying Four Courses, in a Single Term, at a Single Institution

The surveys analyzed for this study were given in four courses during one term at one research-intensive university. The reported results may not generalize to all institutions or other courses in the major. Still, the courses surveyed are critical in recruitment and retention, and for those who persist, play an important role in future student success as shown by recent work evaluating the role of prerequisites for courses similar to those in this study [27, 63]. Likewise, research-intensive universities graduate the majority of computing students in North America. Finally, the Fall term can be expected to present unique challenges for the many students matriculating in that term. However, that first experience is crucial for recruitment and persistence of students in computing [51].

5.2 Threats due to Student Self-Evaluation

Since students are required to self-evaluate to fill out the surveys, there is the potential that students may not be accurately reporting their level of stress on certain factors. For example, many students performing near the bottom of the course did not report serious issues for ICC. In terms of validity, the concern here is that students in different quartiles may exhibit different levels of accuracy.

6 DISCUSSION

Our central research question was to examine the social-emotional /social-cognitive, structural and/or personal factors that are related to student struggle, with the goal of gaining a more holistic view of pressures that are interfering with students' learning. We found that these pressures, and students' lives, are indeed complex. Not only do different students face different struggles, but that particularly lower-performing students are struggling across multiple factors. We find that gender, race, and matriculation status are related to these factors with students from underrepresented groups generally experiencing more sources of stress. Our study confirms results from previous studies as well as offers a new perspective on student struggle from a broad scale across several lower-division and mid-level courses. These results have implications both for researchers and for instructors and departments to build educational structures that allow all students to thrive.

6.1 Relationship to Previous Work

We found four factors that were centrally related to students' struggle across lower-division computer science courses; two of these factors (Lack of Confidence, and Lack of Sense of Belonging) were related to students' social-emotional experience. Consistent with previous work, a lack of sense of belonging was an important factor related to struggle. However, unlike Veilleux et al., who found that sense of belonging was related to students' perception of their grades but not their actual grades [64], we found that sense of belonging was indeed related to students' actual grades. Similar to Sax et al., we found that women report slightly lower sense of belonging than men. On the other hand, while Sax et al. found that BLN+ students reported a slightly higher sense of belonging than non-BLN+ students, we found that sense of belonging levels in BLN+ and non-BLN+ students were similar. We also found that transfer students—a population not often specifically studied in CS courses—have a lower sense of belonging than non-transfer students. Based on situated cognition, these students may have difficulty feeling like full members of the community which may limit their ability to master the knowledge and skill of that community.

The differences in confidence among low and high performing students and by gender are consistent with prior findings in computing that self-efficacy is associated with student success [33] and that women tend to have lower self-efficacy [7, 9]. Those students with lower self-efficacy may be struggling with higher levels of anxiety and have lower interest in the field. Our study adds to this previous literature by showing that lack of confidence persists beyond CS1, into mid-level CS courses. Perhaps most concerning are those students in the top quartiles of the class (particularly in CS2 and CompOrg) who express these low levels of confidence as this

may be an indication of poor metacognition or that these students are experiencing imposter phenomenon [48].

A third important factor that emerged in our study was Personal Obligations. As discussed in Section 2, we are not aware of any studies that relate outside commitments to success in CS courses, yet it is known that in general commitments outside schoolwork may impact dropout rates [59] and may require students to develop different strategies to succeed [34]. Their ability to manage these outside commitments may impact their mental health [57]. Our results show not only that CS students who struggle in class report higher instances of interference from outside sources such as work and other classes, but also that these struggles are often compounded or impacted by other struggles they face. Indeed, most lower-performing students who report struggling with personal obligations also report struggling with confusion, lack of confidence, lack of belonging, or more than one of these additional factors. As we find that BLN+ students disproportionately encounter more personal obligations (albeit with a small effect size) and one can expect these challenges to be more common for students with lower socioeconomic status [67], helping students address these challenges may be essential for improving the diversity of the field.

6.2 Future Research Directions

Understanding Interactions Between Factors: Our work has begun to reveal a complex web of struggle that students face in early CS courses. Our EFA found four separate factors, but the factors are all significantly correlated with one another: Spearman's rho varies between 0.36 and 0.67 for all of the pairwise factor comparisons with LSoB and PO having the lowest correlation and ICC and LoC having the highest. Better understanding the relationship between these factors, as well as considering additional sources of struggle, is critical to building structures where all students can succeed. For example, when a student reports confusion with the material, and also report a low sense of belonging and high personal obligations does this mean that the student's confusion causing them to have a lower sense of belonging? Or are the student's personal obligations keeping them from feeling that they belong, which in turn is making it harder for them to learn? These different scenarios could imply very different intervention techniques.

Closing Demographic Gaps: Although it was not the central focus of our study, our analysis revealed significant (and sometimes large) exam performance differences between men and women, non-BLN+ and BLN+ students, and transfer and first-year students. It is possible (even likely) that these differences are partially due to differences in prior experience; a recent study found that prior experience is correlated with higher grades even into the upper division [1]. Still, the magnitude and consistency of these differences is troubling. The results presented here on student struggle give us new insight into the broader barriers that women, BLN+ students, and transfer students are facing which might allow us to better address the uneven playing field for these students. Yet, more research is needed specifically to understand effective ways of closing these performance gaps.

Ensuring We Hear All Students: Our results on disproportionate response rates from students in different performance quartiles indicate that we are not hearing from precisely the the students

that we most want to understand. We believe that this pattern is not unique to our study, but is likely to be consistent for most survey-based research. Developing mechanisms that focus on response rates from the most affected groups is important to the validity of this approach. Furthermore, combining this work with deeper and more individualistic approaches such as critical theory approaches [23] will help ensure that the voices of those we most want to learn from are not left out.

6.3 Educators' Call to Action

Our results clearly demonstrate that students are struggling across multiple dimensions of their lives. As such, targeted interventions to address one factor, such as lack of sense of belonging or in-class confusion, may not be sufficient. Moreover, finding that a student is seriously struggling across multiple dimensions may be a good indicator that they will struggle to succeed in the course.

Furthermore, one of the main factors that emerged that may interfere with students' success, personal obligations, points to complex systemic issues that must be addressed beyond the design of a particular course. Departments, programs, and universities must consider whether the rules governing our programs are reasonable for all students, given the external pressures many students face. As just one example, students at our institution who are receiving financial aid are required to be enrolled in a minimum number of units—a full course load—which may be too much some quarters in view of their personal obligations and the way that our courses are currently designed. Given the presence of these constraints on students, there is a growing mandate to reflect on our current processes and make improvements to help our most at-risk students.

Finally, one of our notable findings was the substantial difference in response rates between high performing and low performing students. We have no way to recover the experience of students who did not respond to the surveys, but we can guess that the non-responders were experiencing even more barriers to learning than those who did respond. These surveys are meant to show us what students struggle with and if those most struggling are not responding, identifying how to help them becomes more difficult. This differential response rate is important to acknowledge in any survey-based study, and finding methods to ensure that these students are not excluded from this type of research is an important area of future work.

7 CONCLUSION

Through this study we have provided evidence that students in early computer science courses are struggling across different categories of stress: personal obligations, lack of belonging, lack of confidence, and in-class confusion. The lowest performing students are most likely to report struggling with the all four of these categories. The categories span from inside to outside the classroom, and suggest that interventions to help students better understand CS concepts or increase their sense of belonging must be paired with larger structural changes to address the barriers students face in their broader lives. Instructors can thus help by designing courses that will help support students who face a myriad of barriers. Lastly, our work adds support to the notion that it is time to start reflecting

on institutional shortcomings and start addressing compounded struggles rather than individual ones.

ACKNOWLEDGMENTS

This work was supported in part by NSF award 1712508, as well as a UCSD Sloan Scholar Fellowship, a UCSD STARS Fellowship, and a Gates Millennium Scholarship.

REFERENCES

- [1] Christine Alvarado, Gustavo Umbelino, and Mia Minnes. 2018. The Persistent Effect of Pre-College Computing Experience on College CS Course Grades. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*. 876–881.
- [2] Albert Bandura. 1982. Self-efficacy mechanism in human agency. *American psychologist* 37, 2 (1982), 122.
- [3] Lecia J Barker and Kathy Garvin-Doxas. 2004. Making visible the behaviors that influence learning environment: A qualitative exploration of computer science classrooms. *Computer Science Education* 14, 2 (2004), 119–145.
- [4] Lecia Jane Barker, Kathy Garvin-Doxas, and Michele Jackson. 2002. Defensive climate in the computer science classroom. In *Proceedings of the 33rd SIGCSE Technical Symposium on Computer Science Education*. 43–47.
- [5] Amy S Beavers, John W Lounsbury, Jennifer K Richards, Schuyler W Huck, Gary J Skolits, and Shelley L Esquivel. 2013. Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research, and Evaluation* 18, 1 (2013), 6.
- [6] Jens Bennedsen and Michael E Caspersen. 2007. Failure Rates in Introductory Programming. *ACM SIGCSE Bulletin* 39 (2007), 32–36.
- [7] Sylvia Beyer. 2014. Why are women underrepresented in Computer Science? Gender differences in stereotypes, self-efficacy, values, and interests and predictors of future CS course-taking and grades. *Computer Science Education* 24, 2-3 (2014), 153–192.
- [8] Maureen Biggers, Anne Brauer, and Tuba Yilmaz. 2008. Student Perceptions of Computer Science: A Retention Study Comparing Graduating Seniors with CS Leavers. In *Proceedings of the 39th SIGCSE Technical Symposium on Computer Science Education*. 402–406.
- [9] Jennifer M Blaney and Jane G Stout. 2017. Examining the relationship between introductory computing course experiences, self-efficacy, and belonging among first-generation college women. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*. 69–74.
- [10] Jeni Burnette, Crystal Hoyt, V Russell, Barry Lawson, Carol S. Dweck, and Eli Finkel. 2019. A Growth Mind-Set Intervention Improves Interest but Not Academic Performance in the Field of Computer Science. *Social Psychological and Personality Science* (2019), 194855061984163.
- [11] Lori Carter. 2006. Why students with an apparent aptitude for computer science don't choose to major in computer science. In *Proceedings of the 37th SIGCSE technical symposium on Computer Science Education*. 27–31.
- [12] Catherine H Crouch and Eric Mazur. 2001. Peer instruction: Ten years of experience and results. *American journal of physics* 69, 9 (2001), 970–977.
- [13] Amanda B Diekmann, Emily K Clark, Amanda M Johnston, Elizabeth R Brown, and Mia Steinberg. 2011. Malleability in communal goals and beliefs influences attraction to stem careers: evidence for a goal congruity perspective. *Journal of personality and social psychology* 101, 5 (2011), 902.
- [14] Amanda B Diekmann, Mia Steinberg, Elizabeth R Brown, Aimee L Belanger, and Emily K Clark. 2017. A goal congruity model of role entry, engagement, and exit: Understanding communal goal processes in STEM gender gaps. *Personality and Social Psychology Review* 21, 2 (2017), 142–175.
- [15] Christine DiStefano, Min Zhu, and Diana Mindrila. 2009. Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research, and Evaluation* 14, 1 (2009), 20.
- [16] Eileen Doyle, Ioanna Stamouli, and Meriel Huggard. 2005. Computer anxiety, self-efficacy, computer experience: An investigation throughout a computer science degree. In *Proceedings Frontiers in Education 35th Annual Conference*. S2H–3.
- [17] NSF (National Science Foundation). 2018. Science and Engineering Indicators. NSF (2018).
- [18] Kathy Garvin-Doxas and Lecia J Barker. 2004. Communication in computer science classrooms: Understanding defensive climates as a means of creating supportive behaviors. *Journal on Educational Resources in Computing (JERIC)* 4, 1 (2004), 2–es.
- [19] Michail N Giannakos, Ilias O Pappas, Letizia Jaccheri, and Demetrios G Sampson. 2017. Understanding student retention in computer science education: The role of environment, gains, barriers and usefulness. *Education and Information Technologies* 22, 5 (2017), 2365–2382.
- [20] Catherine Good, Aneeta Rattan, and Carol S Dweck. 2012. Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of*

- personality and social psychology 102, 4 (2012), 700.
- [21] Bonnie M Hagerty and Arthur Williams. 1999. The effects of sense of belonging, social support, conflict, and loneliness on depression. *Nursing research* 48, 4 (1999), 215–219.
 - [22] Geoffrey L Herman, Craig Zilles, and Michael C Loui. 2014. A psychometric evaluation of the digital logic concept inventory. *Computer Science Education* 24, 4 (2014), 277–303.
 - [23] Aleata Hubbard Cheuoua. 2021. *Confronting Inequities in Computer Science Education: A Case for Critical Theory*. Association for Computing Machinery, New York, NY, USA, 425–430. <https://doi.org/10.1145/3408877.3432453>
 - [24] Alboukadel Kassambara. 2021. Kruskal-Wallis Effect Size. https://rpkgs.datanova.com/rstatix/reference/kruskal_effsize.html. Accessed: 2021-03-23.
 - [25] Sophia Krause-Levy, William Griswold G., Leo Porter, and Christine Alvarado. 2021. The Relationship Between Sense of Belonging and Student Outcomes in CS1 and Beyond. In *Proceedings of the 2021 ACM Conference on International Computing Education Research*.
 - [26] Sophia Krause-Levy, Leo Porter, Beth Simon, and Christine Alvarado. 2020. Investigating the Impact of Employing Multiple Interventions in a CS1 Course. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. 1082–1088.
 - [27] Sophia Krause-Levy, Sander Valstar, Leo Porter, and William G. Griswold. 2020. Exploring the Link Between Prerequisites and Performance in Advanced Data Structures. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. 386–392.
 - [28] Justin Kruger and David Dunning. 1999. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology* 77, 6 (1999), 1121.
 - [29] Wayne W. LaMorte. 2019. The Social Cognitive Theory. <https://sphweb.bumc.bu.edu/otl/oph-modules/sb/behavioralchange/theories/behavioralchange/theories5.html>. Accessed: 2021-03-20.
 - [30] Jean Lave, Etienne Wenger, et al. 1991. *Situated learning: Legitimate peripheral participation*. Cambridge university press.
 - [31] Colleen Lewis, Paul Bruno, Jonathan Raygoza, and Julia Wang. 2019. Alignment of goals and perceptions of computing predicts students' sense of belonging in computing. In *Proceedings of the 2019 ACM Conference on International Computing Education Research*. 11–19.
 - [32] Soohyun Nam Liao, Sander Valstar, Kevin Thai, Christine Alvarado, Daniel Zingaro, William G Griswold, and Leo Porter. 2019. Behaviors of higher and lower performing students in CS1. In *Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education*. 196–202.
 - [33] Alex Lishinski, Aman Yadav, Jon Good, and Richard Enbody. 2016. Learning to program: Gender differences and interactive effects of students' motivation, goals, and self-efficacy on performance. In *Proceedings of the 2016 ACM Conference on International Computing Education Research*. 211–220.
 - [34] Janet Lowe and Vernon Gayle. 2007. Exploring the work/life/study balance: the experience of higher education students in a Scottish further education college. *Journal of further and Higher Education* 31, 3 (2007), 225–238.
 - [35] Andrew Luxton-Reilly. 2016. Learning to program is easy. In *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education*. 284–289.
 - [36] Andrew Luxton-Reilly, Vangel V Ajanovski, Eric Fouh, Christabel Gonsalvez, Juho Leinonen, Jack Parkinson, Matthew Poole, and Neena Thota. 2019. Pass rates in introductory programming and in other stem disciplines. In *Proceedings of the Working Group Reports on Innovation and Technology in Computer Science Education*. 53–71.
 - [37] Jane Margolis, Rachel Estrella, Joanna Goode, Jennifer Jellison Holme, and Kim Nao. 2017. *Stuck in the shallow end: Education, race, and computing*. MIT press.
 - [38] Jane Margolis and Allan Fisher. 2002. *Unlocking the clubhouse: Women in computing*. MIT press.
 - [39] Allison Master, Sapna Cheryan, and Andrew N Meltzoff. 2016. Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. *Journal of educational psychology* 108, 3 (2016), 424.
 - [40] Allison Master and Andrew N Meltzoff. 2020. Cultural stereotypes and sense of belonging contribute to gender gaps in STEM. *International Journal of Gender, Science and Technology* 12, 1 (2020), 152–198.
 - [41] Irene T Miura. 1987. The relationship of computer self-efficacy expectations to computer interest and course enrollment in college. *Sex roles* 16, 5-6 (1987), 303–311.
 - [42] Catherine Mooney, Anna Antoniadis, Ioannis Karvelas, Lána Salmon, and Brett A Becker. 2020. Exploring Sense of Belonging in Computer Science Students. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education*. 563–563.
 - [43] An Nguyen and Colleen M Lewis. 2020. Competitive enrollment policies in computing departments negatively predict first-year students' sense of belonging, self-efficacy, and perception of department. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. 685–691.
 - [44] Ilias O Pappas, Michail N Giannakos, Letizia Jaccheri, and Demetrios G Sampson. 2017. Assessing student behavior in computer science education with an fsQCA approach: The role of gains and barriers. *ACM Transactions on Computing Education (TOCE)* 17, 2 (2017), 1–23.
 - [45] Andrew Petersen, Michelle Craig, Jennifer Campbell, and Anya Tafilovich. 2016. Revisiting why students drop CS1. In *Proceedings of the 16th Koli Calling International Conference on Computing Education Research*. 71–80.
 - [46] Leo Porter, Cynthia Bailey Lee, and Beth Simon. 2013. Halving Fail Rates using Peer Instruction: A Study of Four Computer Science Courses. In *Proceedings of the 44th Special Interest Group on Computer Science Education Technical Symposium*.
 - [47] Leo Porter and Beth Simon. 2019. A Case Study of Peer Instruction: From University of California, San Diego to the Computer Science Community. In *The Cambridge Handbook of Computing Education Research*, Sally A Fincher and Anthony V Robins (Eds.). Cambridge University Press, Chapter 30, 861–874.
 - [48] Adam Rosenstein, Aishma Raghu, and Leo Porter. 2020. Identifying the prevalence of the impostor phenomenon among computer science students. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. 30–36.
 - [49] Nathan Rountree, Janet Rountree, and Anthony Robins. 2002. Predictors of success and failure in a CS1 course. *ACM SIGCSE Bulletin* 34, 4 (2002), 121–124.
 - [50] Merilin Säde, Reelika Suviste, Piret Luik, Eno Tõnisson, and Marina Lepp. 2019. Factors That Influence Students' Motivation and Perception of Studying Computer Science. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. 873–878.
 - [51] Adrian Salguero, Julian McAuley, Beth Simon, and Leo Porter. 2020. A Longitudinal Evaluation of a Best Practices CS1. In *Proceedings of the 2020 ACM Conference on International Computing Education Research*. 182–193.
 - [52] Linda J Sax, Jennifer M Blaney, Kathleen J Lehman, Sarah L Rodriguez, Kari L George, and Christina Zavala. 2018. Sense of belonging in computing: The role of introductory courses for women and underrepresented minority students. *Social Sciences* 7, 8 (2018), 122.
 - [53] Carsten Schulte and Maria Knobelsdorf. 2007. Attitudes towards computer science-computing experiences as a starting point and barrier to computer science. In *Proceedings of the third international workshop on Computing education research*. 27–38.
 - [54] Walter R Schuman. 1993. Applied multivariate statistics for the social sciences. *The American Statistician* 47, 2 (1993), 155–157.
 - [55] Jessi L Smith, Erin Cech, Anneke Metz, Meghan Huntoon, and Christina Moyer. 2014. Giving back or giving up: Native American student experiences in science and engineering. *Cultural Diversity and Ethnic Minority Psychology* 20, 3 (2014), 413.
 - [56] Tamara Floyd Smith, Denise Wilson, Diane Carlson Jones, Melani Plett, Rebecca A. Bates, and Nanette M Veilleux. 2012. Investigation of Belonging for Engineering and Science Undergraduates by Year in School. In *2012 ASEE Annual Conference & Exposition*.
 - [57] Justin M Sprung and Anna Rogers. 2020. Work-life balance as a predictor of college student anxiety and depression. *Journal of American college health* (2020), 1–8.
 - [58] Ben Stephenson, Michelle Craig, Daniel Zingaro, Diane Horton, Danny Heap, and Elaine Huynh. 2017. Exam Wrappers: Not a Silver Bullet. In *Proceedings of the 48th ACM Technical Symposium on Computer Science Education*. 573–578.
 - [59] Leslie S Stratton, Dennis M O'Toole, and James N Wetzel. 2008. A multinomial logit model of college stopout and dropout behavior. *Economics of Education Review* 27, 3 (2008), 319–331.
 - [60] Mohsen Tavakol and Reg Dennick. 2011. Making sense of Cronbach's alpha. *International Journal of Medical Education* 2 (2011), 53.
 - [61] Dustin B Thoman, Elizabeth R Brown, Andrew Z Mason, Allen G Harmsen, and Jessi L Smith. 2015. The role of altruistic values in motivating underrepresented minority students for biomedicine. *BioScience* 65, 2 (2015), 183–188.
 - [62] Maciej Tomczak and Ewa Tomczak. 2014. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in Sport Sciences* 21, 1 (2014).
 - [63] Sander Valstar, William G. Griswold, and Leo Porter. 2019. The Relationship between Prerequisite Proficiency and Student Performance in an Upper-Division Computing Course. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. 794–800.
 - [64] Nanette Veilleux, Rebecca Bates, Cheryl Allendoerfer, Diane Jones, Joyous Crawford, and Tamara Floyd Smith. 2013. The relationship between belonging and ability in computer science. In *Proceedings of the 44th ACM Technical Symposium on Computer Science Education*. 65–70.
 - [65] Christopher Watson and Frederick WB Li. 2014. Failure rates in introductory programming revisited. In *Proceedings of the 2014 Conference on Innovation & Technology in Computer Science Education*. 39–44.
 - [66] Brenda Cantwell Wilson and Sharon Shrock. 2001. Contributing to success in an introductory computer science course: a study of twelve factors. In *Proceedings of the 32nd SIGCSE Technical Symposium on Computer Science Education*. 184–188.
 - [67] Melissa R Witkow, Virginia Huynh, and Andrew J Fuligni. 2015. Understanding differences in college persistence: A longitudinal examination of financial circumstances, family obligations, and discrimination in an ethnically diverse sample. *Applied Developmental Science* 19, 1 (2015), 4–18.

- [68] Daniel Zingaro, Michelle Craig, Leo Porter, Brett A Becker, Yingjun Cao, Phill Conrad, Diana Cukierman, Arto Hellas, Dastyni Loksa, and Neena Thota. 2018. Achievement goals in CS1: Replication and extension. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*. 687–692.
- [69] Daniel Zingaro and Leo Porter. 2016. Impact of student achievement goals on CS1 outcomes. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*. 279–296.
- [70] Stuart Zweben. 2019. Enrollment and Retention in U.S. Computer Science Bachelor’s Programs in 2016–17. *ACM Inroads* 10, 4 (Nov. 2019), 47–59.