INVESTIGATING STUDENTS' PERCEPTIONS AND EXPERIENCES TO BROADEN PARTICIPATION IN COMPUTING

BY

VIDUSHI OJHA

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Computer Science
in the Graduate College of the
University of Illinois Urbana-Champaign, 2024

Urbana, Illinois

Doctoral Committee:

Associate Professor Colleen Lewis, Chair
Professor Craig Zilles
Assistant Professor Kathryn Cunningham
Dr. Christopher Hovey, University of Colorado Boulder

ABSTRACT

My work seeks to investigate students' perceptions of and experiences in higher education computer science (CS) courses in order to inform efforts to broaden participation in computing (BPC). In the U.S., people who identify as women, Black/African American, Hispanic, Latina/o/x/*, Native American, Native Alaskan, Native Hawaiian, Pacific Islanders, and/or disabled are considered historically underrepresented in computing. I refer to these groups as historically underrepresented groups, or HUGs. Individuals from HUGs are not proportionally represented in the CS educational system or in the computing workforce; BPC efforts aim to recruit and retain students from HUGs in order to address this inequity. Prior work has identified many factors that may affect students' recruitment and retention in CS. Of particular interest to this dissertation are the influences of students' prior perceptions of computing and of their experiences once they are in the field. This dissertation consists of four studies investigating aspects of students' perceptions and experiences in computing and how they relate to important outcomes for recruitment and retention in the field.

The first study investigated students' perceptions of specializations within the field of CS, specifically artificial intelligence (AI) and cybersecurity, before they have personal experience in these specializations. This qualitative study examined students' expressed beliefs about AI and cybersecurity that may affect their interest in pursuing these fields. Understanding these perceptions may inform efforts to bolster students' interest in AI and cybersecurity towards the goal of increasing participation of students from HUGs in these specializations. The second study in this dissertation addressed the experiences students from HUGs may have regarding understanding the expectations of a doctoral CS program. Through interviews with PhD students, we found that students' initial expectations of graduate school were often incomplete or inaccurate, and that policies such as formal mentorship systems may positively impact PhD students' experiences and ability to succeed, thereby increasing their retention in CS. The third study examined undergraduate students' computing self-efficacy, i.e., students' beliefs that they can achieve desired outcomes in their computing courses. We found that identifying as Asian, Black, Native, Hispanic, non-binary, and/or a woman were statistically significantly associated with lower computing self-efficacy, even when controlling for prior CS experience. This work further used an intersectional approach to show that identifying as Asian *and* non-binary correlates with lower computing self-efficacy. The fourth study investigated whether self-efficacy and sense of belonging correlate with instructional transparency, which aims to make course's learning goals, evaluation criteria, and path to

success clear and accessible to students. Our findings show group differences in students' perception of transparency, as students who identify as women, first-generation college students, and/or disabled reported perceiving less transparency. We also demonstrate that perception of instructional transparency has a positive correlation with students' self-efficacy and sense of belonging in computing while controlling for important confounding variables, such as prior CS experience.

As a whole, my work aims to investigate students' perceptions of and experiences in CS education in order to inform and support BPC efforts. Collectively, this research provides insights into the kinds of institution-level policies and classroom-level practices that may contribute to diversity, inclusion, and equity efforts. Findings from each of the studies suggest particular policies and practices, such as a formal mentorship system for PhD students and using transparent teaching in CS classrooms, as well as avenues for future research that may shed further light on how to recruit and retain students from HUGs in computing.

To Austin – you were right, it is "our" PhD!

ACKNOWLEDGMENTS

I am immensely grateful to so many people for so much support over the course of my PhD. Those named here are really only scratching the surface; a big, blanket thanks to all the friends, acquaintances, and coworkers who have helped me get to this point.

First and foremost, *thank you* to Colleen Lewis, whose influence on me and my career probably cannot be overstated. Since my first research experience with you ten (!) years ago, I feel incredibly lucky to have been able to learn so much about how to be a teacher, researcher, and mentor from you. I would also like to thank Craig Zilles, Katie Cunningham, and Chris Hovey. In your roles as members of my dissertation committee and as research mentors more generally, you have all taught me a lot about academia and academic life, and my work is so much stronger for all of your feedback. Speaking of feedback, many other members of the CS Education area at Illinois have provided me with so much helpful feedback on endless research ideas, paper drafts, and practice talks over the years. Thank you to Luther Tychonievich, Yael Gertner, Geoffrey Herman, Paul Bruno, and all other faculty and students who have supported my growth as a teacher and researcher during my PhD.

To all members of my lab: teamwork makes the dream work, and I have no idea where I would be without you all. Many, *many* thanks go to Kari George; your expertise in topics from higher ed to quantitative research methods, as well as your support over the years, have been beyond valuable. To Mariam Saffar Perez, Kathleen Isenegger, Chris Perdriau, and Andrea Watkins, I'm grateful for all the ways you have all influenced me. I am *so* lucky to know you, to be inspired and informed by your work, and, most importantly, to be your friend. The same goes for our original CS Ed PhD crew: Morgan Fong, Max Fowler, and Seth Poulsen. You're all the best colleagues and friends I could have hoped for in grad school. I'm truly thankful to have been able to work and spend time with all of you these past few years, and I have no intention of stopping that anytime soon.

Shout out to the #QUAD for so much PhD support over the years. Max Hlavacek, Annisa Dea, and especially Cherie Ho, it's bizarre that all four of us ended up in PhDs, but I couldn't be more grateful to have had you all to swap stories with and vent to this whole time. Thanks also to all Mudd friends, who have continued to be such an important part of my life years after graduation.

As has always been the case in my life, I can't thank my family enough for all their love, encouragement, and support. My parents, Alok and Sumedha Ojha, have never doubted my ability to pull off even the most harebrained of ideas, whether it was moving across the world to a new country for college, or moving across *that* country for grad school. Thanks to my brother Vivaswat Ojha for being possibly even more enthusiastic about my becoming Dr. Ojha than I am. I'm grateful also to my in-laws, Kwang and Keum Rhan Shin, who, in the past few years, have made me feel like I've always been a part of the family. Finally, to Austin Shin. Thanks for all the nights you made dinner because I was too stressed; for all the convoluted research methods you let me talk through; for all the commas you've tried to edit in my writing; for all the things I can't even remember right now because you took care of them all. I genuinely couldn't have done it without you.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION . . . 1
1.1 Motivation . . . 1
1.2 Summary of Work . . . 2
1.3 Significance & Broader Impacts . . . 5

CHAPTER 2 PERCEPTIONS OF AI & CYBERSECURITY . . . 6
2.1 Introduction . . . 6
2.2 Previous Research . . . 7
2.3 Methods . . . 10
2.4 Results . . . 11
2.5 Discussion . . . 16
2.6 Implications for Practice . . . 17
2.7 Limitations and Future Work . . . 18
2.8 Conclusion . . . 19

CHAPTER 3 PHD EXPECTATIONS . . . 20
3.1 Introduction . . . 20
3.2 Previous Work . . . 21
3.3 Research Questions . . . 23
3.4 Methods . . . 23
3.5 Results . . . 25
3.6 Discussion . . . 28
3.7 Limitations & Future Work . . . 30
3.8 Conclusion . . . 31

CHAPTER 4 SELF-EFFICACY . . . 33
4.1 Introduction . . . 33
4.2 Previous Research . . . 34
4.3 Research Question and Hypotheses . . . 36
4.4 Data . . . 37
4.5 Methods . . . 40
4.6 Results . . . 43
4.7 Discussion and Future Work . . . 44
4.8 Conclusion . . . 46

CHAPTER 5 INSTRUCTIONAL TRANSPARENCY 47
5.1 Introduction 47
5.2 Background: Transparent Teaching Practices 49
5.3 Previous Research 50
5.4 RQ0: Theoretical Framework 53
5.5 Research Questions and Hypotheses 57
5.6 Data 58
5.7 Measures 59
5.8 Methods 62
5.9 Limitations 65
5.10 Results 65
5.11 Discussion & Future Work 69
5.12 Conclusion 73

CHAPTER 6 CONCLUSION 74
6.1 Intellectual Merit 74
6.2 Broader Impacts 75

REFERENCES 77

APPENDIX A PERCEPTIONS OF AI & CYBERSECURITY INTERVIEW QUESTIONS 90
A.1 Math 90
A.2 Narratives within CS 90
A.3 Perceptions of AI 90
A.4 Perceptions of Cybersecurity 91

APPENDIX B PERCEPTIONS OF AI & CYBERSECURITY CODING SCHEME 92
B.1 Category: Perception of AI 92
B.2 Category: Perception of Security 93
B.3 Category: The Hardest Subfield 94
B.4 Category: The Prestigious Subfield 94

APPENDIX C PHD EXPECTATIONS INTERVIEW QUESTIONS 95
C.1 Introduction 95
C.2 Advisor Relationship 95
C.3 Expectations 95
C.4 Learning/Acquiring Skills 96
C.5 Conclusion 96

APPENDIX D PHD EXPECTATIONS CODING SCHEME 97
D.1 Category: Positive Advisor Actions 97
D.2 Category: Negative Advisor Actions 97
D.3 Category: Expectations 97
D.4 Category: Pre-Grad School Experience 98

D.5 Category: Challenges 98
D.6 Category: Measuring Progress 98
D.7 Category: Impact of Identity 99
D.8 Category: Advisor Relationship 99
D.9 Category: Strategies 99

APPENDIX E CATEGORIZATION OF DATA BUDDIES SURVEY ITEMS 100
E.1 Categorization of Prior Computing Experience 100
E.2 Categorization of Institutional CS Requirement 100
E.3 Student Disability Options 101
E.4 Categorization of Students' Majors 101

APPENDIX F SELF-EFFICACY CORRELATIONS 103
F.1 Correlations Between Self-Efficacy and Nominal Variables 103
F.2 Overlap Between Nominal Variable Categories 104

APPENDIX G INSTRUCTIONAL TRANSPARENCY FACTORS 106

APPENDIX H INSTRUCTIONAL TRANSPARENCY CORRELATIONS 108
H.1 Correlations Between Continuous Variables 108
H.2 Correlations Between Continuous and Nominal Variables 109
H.3 Overlap Between Nominal Variable Categories 109

CHAPTER 1: INTRODUCTION

1.1 MOTIVATION

The focus of my PhD research has been on broadening participation in computing (BPC), with the goal of fostering greater representation, equity, and inclusion for students from historically underrepresented groups (HUGs) in computing. In the US, HUGs in computing include people who identify as women, Black/African American, Hispanic, Latina/o/x/*, Native American, Native Alaskan, Native Hawaiian, Pacific Islanders, and/or disabled [1]. Individuals from HUGs are not proportionally represented in the computer science (CS) educational system [2, 3, 4, 5] or in the computing workforce [6, 7]. As recently as 2021-2022, women earned only 21.3% of the bachelor's degrees in computer and information sciences [8], despite representing 50.4% of the U.S. population [9]. In terms of racial and ethnic identities, of the recipients of bachelor's degrees in computer and information sciences in 2022, 7.2% were Black students, 13.1% were Hispanic students, and 0.34% were Native students [8], despite constituting 13.6%, 19.1%, and 1.6% of the U.S. population, respectively [9]. Prior work has documented the need to broaden the participation of individuals from HUGs and has identified many factors that may affect their recruitment and retention into the field of CS, including their socio-cultural identities, background, and goals (e.g., [2, 10, 11, 12, 13, 14]).

The lack of representation has several consequences for technological growth, the field of computing, and individuals from HUGs. First, the above patterns of underrepresentation lead to a loss of the potential contributions that individuals from HUGs would make in computing. As a result, the problems technological leaders choose to address, and the solutions they create, may not align with the needs and values of communities not represented by those individuals. It is particularly important for individuals from HUGs to have a say in the direction of technological growth as computing becomes increasingly ubiquitous and impactful in daily life [15]. Second, within computing higher education, institutions are failing to produce enough computing graduates to fulfill predicted market demand in the workforce [4] and in academia [16, 17]. The dearth of qualified graduates is exacerbated by the lack of representation from underrepresented groups [4, 5]. Broadening participation from currently underrepresented groups is crucial for increasing the number of qualified professionals in the field. Thirdly, at an individual level, computing is a source of high-paying jobs [18] and therefore a key opportunity for social mobility. For these reasons, it is incumbent upon the field of computing to broaden the participation of people from HUGs.

The motivation behind my work is to inform efforts to broaden participation in computing by investigating students' perceptions of and experiences in CS.

1.2 SUMMARY OF WORK

A growing body of work has sought to understand the unique experiences of students from HUGs in CS higher education. From this work, it is clear that students' recruitment and retention in CS is impacted by their prior perceptions of the field *and* by their experiences once they are in the field (e.g., [11, 19, 20, 21, 22, 23, 24, 25]). My dissertation work, presented in the following four chapters, consists of four studies that investigate different aspects of students' perceptions and experiences in computing. Each paper is presented with minor edits from its final published (or currently submitted) form.

1.2.1 Chapter 2: Perceptions of AI & Cybersecurity

The first study investigated students' pre-existing perceptions about specializations within the field of CS, specifically artificial intelligence (AI) and cybersecurity [26]. AI and cybersecurity are in-demand skills, but little is known about what factors influence CS undergraduate students' decisions on whether to specialize in AI or cybersecurity and how these factors may differ between populations. In this study, we interviewed undergraduate CS majors about their perceptions of AI and cybersecurity. Qualitative analyses of these interviews showed that students have narrow beliefs about what kind of work AI and cybersecurity entail, the kinds of people who work in these fields, and the potential societal impact AI and cybersecurity may have. Specifically, students tended to believe that all work in AI requires math and training models, while cybersecurity consists of low-level programming; that innately smart people work in both fields; that working in AI comes with ethical concerns; and that cybersecurity skills are important in contemporary society. Some of these perceptions reinforce existing stereotypes about computing, such as beliefs that work in computer science requires innate brilliance [20, 27], is only for men [13, 28], and is not conducive to benefiting society [11, 28, 29]. These beliefs may disproportionately deter the participation of students from groups historically underrepresented in computing. Our key contribution is identifying beliefs that students expressed about AI and cybersecurity that may affect their interest in pursuing the two fields and may, therefore, inform efforts to expand students' views of AI and cybersecurity. Expanding student perceptions of AI and cybersecurity may help correct misconceptions and challenge narrow definitions, which in turn can encourage participation in these fields from all students.

1.2.2 Chapter 3: PhD Expectations

The second study in this dissertation addresses the experiences students from HUGs may have in trying to understand the expectations of a doctoral CS program [30]. Previous research shows that one key factor affecting the rate of degree completion among doctoral students is a clear understanding of what is expected of them in order to complete the program [22, 31, 32]. These expectations may concern requirements from the department or from a student's advisor, who has an important say in awarding a doctoral degree. In this paper, we investigated the different sources that students use to form their expectations, and how each of these sources provides different kinds of guidance. Our research team conducted a survey and semi-structured interviews with doctoral students from HUGs at a large institution with high research productivity. Our analysis found that participants had three primary sources from which they deduce what they are expected to do, and how to do it: research experience prior to beginning their program, their PhD advisor, and their peers. Many students began the program anticipating more hands-on support from their advisor, but instead found themselves relying more on their labmates and peers. We also found that students' initial expectations of what graduate school would entail did not always provide a complete picture, often giving them incomplete or inaccurate expectations. To address these issues, we suggest developing lab-based mentorship systems and encouraging clear communication between advisors and students not only on high-level goals, but on the daily tasks needed to accomplish them. Policies of this kind may positively impact PhD students' experiences and ability to succeed, thereby increasing their retention in CS.

1.2.3 Chapter 4: Computing Self-Efficacy

The third study pertains to undergraduate students' computing self-efficacy [33]. Self-efficacy refers to an individual's belief that they can act in ways that will lead to a desired outcome [34, 35] and has been shown to contribute to student retention in computing [36, 37, 38, 39]. Therefore, investigating computing self-efficacy may help to improve the persistence of students from HUGs in computing. Previous research has shown that computing self-efficacy is positively correlated with prior computing experience [40, 41, 42, 43], but negatively correlated with some demographic identities (e.g., identifying as a woman) [38, 39, 44, 45, 46]. However, existing research has not demonstrated these patterns on a large scale while controlling for confounding variables and institutional context. In addition, there is a need to study the experiences of students with multiple marginalized identities through the lens of intersectionality [24, 47]. Our goal was to investigate the relationship between students' computing self-efficacy and their prior experience in computing, demographic identities, and institutional policies. We conducted this investigation using a large, recent, and multi-institutional dataset with survey responses from 31,425 students. Our findings confirmed that more computing experience positively predicts computing self-efficacy. However, identifying as Asian, Black, Native, Hispanic, non-binary, and/or a woman were statistically significantly associated with lower computing self-efficacy. The results of our work point to several future avenues for self-efficacy research in computing.

1.2.4 Chapter 5: Instructional Transparency

The fourth study included in this dissertation investigates the relationship between students' perceptions of instructional transparency in their CS courses and their self-efficacy and sense of belonging in computing [48]. Instructional transparency makes a course's learning goals, evaluation criteria, and path to success clear to students, with the goal of improving equity in higher education [49, 50]. Increased transparency may improve equity by bolstering students' self-efficacy and sense of belonging in computing [49, 51], both of which are correlated with persistence in the field [38, 39, 52, 53, 54]. We aim to understand whether there are group differences in how students perceive and benefit from instructional transparency. We are additionally interested in understanding whether perceiving instructional transparency can positively influence students' self-efficacy and sense of belonging and, therefore, contribute to the persistence of students from historically underrepresented groups in computing. To investigate these relationships, we used linear regressions to analyze survey responses from 11,046 undergraduate students from 203 institutions. We found that there are group differences in students' perception of transparency in their CS courses: students who identify as women, first-generation college students, and/or disabled reported perceiving less instructional transparency than their peers. We also found that perceiving more transparency has a positive correlation with students' self-efficacy and sense of belonging in computing while controlling for important confounding variables, such as prior CS experience. We further demonstrated that this relationship is different for certain groups of students: first, for Black students and first-generation college students, perceiving transparency has a *larger* positive impact on their self-efficacy, and second, for Hispanic students, perceiving transparency has a *smaller* positive impact on their sense of belonging. Our work constitutes one of the first empirical, multi-institutional investigations of the perceptions and benefits of transparency in CS classrooms that focuses on group differences. Our work also includes a theoretical articulation of the mechanisms through which transparent teaching practices may influence students' self-efficacy and sense of belonging in computing. Taken together, our empirical findings and theoretical argument provide important evidence for the benefits of instructional transparency in CS courses, particularly as it relates to improving equity in computing.

1.3 SIGNIFICANCE & BROADER IMPACTS

As a whole, my work aims to investigate students' perceptions of and experiences in CS education in order to inform and support BPC efforts. In particular, my research provides insights into the kinds of institution-level policies and classroom-level practices that may contribute to diversity, inclusion, and equity efforts. My investigation of students' perceptions of computing reaffirms that many students have incomplete and inaccurate beliefs about the field (and its subfields) and that these beliefs deter the participation of students from HUGs (see Chapter 2). Concretely, these findings underscore the need for practices that demonstrate to students the breadth of opportunities available in computing, such as by inviting speakers whose work benefits society. Further, my work investigated a number of different experiences students may have in CS spaces, namely the process of navigating expectations during a doctoral program (Chapter 3), how their identities and backgrounds may relate to their computing self-efficacy (Chapter 4), and the impact of transparent teaching practices in CS classrooms (Chapter 5). Findings from each of these studies suggest particular policies and practices, such as a lab-based mentorship system and using transparent teaching, as well as avenues for future research that may shed further light on how to support students from HUGs in computing.

Understanding students' perceptions and experiences, and assessing resultant policies and practices, may lead us to suggested best practices that bolster the recruitment and retention of students in computing, particularly those from HUGs. In future work, I intend to further investigate these and other policies and practices in order to further the goal of inclusive excellence in computing.

CHAPTER 2: PERCEPTIONS OF AI & CYBERSECURITY

2.1 INTRODUCTION

Artificial intelligence (AI) and cybersecurity are growing fields: individuals with expertise in AI, cybersecurity, or both are increasingly in-demand for roles in industry and government [55, 56, 57]. Current research suggests that there are insufficient numbers of individuals with appropriate expertise to fulfill this increasing demand, particularly in cybersecurity [56] and in applying AI to cybersecurity [58]. A significant number of occupations in the technology industry require AI or cybersecurity skills, a number that is likely to increase in the near future [55]. Jobs in cybersecurity already exceed the number of qualified professionals to fill them: according to Cyberseek, a website providing data on the cybersecurity job market, "[o]n average, cybersecurity roles take 21% longer to fill than other IT [information technology] jobs" because employers cannot find workers with appropriate qualifications [56].

A factor further exacerbating the gap between qualified professionals and available positions is that the workforce in computing in general, and cybersecurity in particular, lacks proportional representation of people from different races, ethnicities, and gender identities [6, 7]. Specifically, in the U.S., people who identify as women, Black/African American, Hispanic, Latina/o/x, Native American, Native Alaskan, Native Hawaiian, Pacific Islanders, and/or disabled belong to historically underrepresented groups (HUGs) in computing [1]. These patterns of underrepresentation are also apparent in the educational system [2, 3, 4, 5], suggesting that in order to broaden participation, it is important to understand how students decide whether to pursue AI and cybersecurity, as these students go on to become qualified professionals in the workforce.

Previous work suggests that, in general, computer science (CS) students make decisions regarding specializations based on limited information [59], but little is known about specific factors that may either increase their interest in AI or cybersecurity, or discourage them from pursuing these fields entirely. A greater understanding of how students perceive these two fields may lead to interventions that correct misconceptions and highlight factors likely to draw students' interest, especially students from HUGs. To this end, we address the following research question:

RQ: How do undergraduate computing students perceive AI and cybersecurity?

We interviewed 17 students majoring or minoring in CS at a large, public university about their experiences and perceptions of AI[1] and cybersecurity. Using qualitative analysis methods, we identified patterns in students' beliefs regarding the nature of the work in AI and cybersecurity, the kind of people who work in the fields, and the societal impact such work may have. Our results suggest that participants have narrow perceptions of AI and cybersecurity. In particular, students tended to believe that all work in AI requires math and training models, while cybersecurity consists of low-level programming; that innately smart people work in both fields; that working in AI comes with ethical concerns; and that cybersecurity skills are important in contemporary society.

Our key contribution is identifying beliefs that students expressed about AI and cybersecurity that may affect their interest in pursuing the two fields and may, therefore, inform efforts to expand students' views of AI and cybersecurity. Many of our participants' beliefs mirror existing stereotypes about the work and the people in computing at large. Some of these stereotypes have been shown to impact participation of people from HUGs in computing [11, 27], suggesting that they may play a similar role in AI and cybersecurity. In addition, some of the beliefs expressed by students are not supported by experts in the field. For example, participants suggested that all occupations in cybersecurity require low-level programming, but there exist a wide variety of cybersecurity occupations in management, analysis, policy, ethics, and software engineering [60]. These findings may provide partial explanations for why students do not choose to pursue AI or cybersecurity and, as such, can be used to create interventions to expand student perceptions of what these fields are and who works in them.

2.2 PREVIOUS RESEARCH

Previous research has investigated how stereotypes about computing might affect student decisions to pursue it, how students specialize within computing, and efforts to recruit students into AI and cybersecurity courses.

2.2.1 Computing stereotypes

Understanding stereotypes about AI and cybersecurity is important as they may affect student decisions to pursue the fields, especially since individuals working in these fields are often represented in popular media [61]. Although we are not aware of work on stereotypes specific to AI or cybersecurity, previous research has documented stereotypes that students

[1] As most of our participants equated AI and machine learning (ML), we combine the two and refer only to AI for the remainder of this paper, excepting participant quotes.

tend to have about the people and the work in science, technology, and mathematics (STEM) fields generally [11, 13, 27, 62] and in computing specifically [20, 28, 29, 63].

Like in other STEM fields, people in computing are stereotyped as requiring innate brilliance. For instance, in a study by Lewis et al. [20], undergraduate computing students seemed to believe that computing was an ability you were "just born with". In a survey study by Leslie et al. [27], respondents from STEM fields, including those in CS, were more likely to agree with statements such as "Being a top scholar of [their discipline] requires a special aptitude that just can't be taught". Endorsement of such beliefs was inversely correlated with the percentage of PhDs in the field awarded to women [27], suggesting that the prevalence of this stereotype may be particularly detrimental to broadening the participation of women in computing.

Some students also stereotype people in computing as not working with or for the benefit of others. In a survey of undergraduate students by Diekman et al. [11], respondents did not believe that STEM fields, including computing, would allow them to positively impact society. The perceived asocial nature of CS was also a theme identified by a Lewis et al. [28] study, where they found that many students viewed having to work alone as a requirement of the field. Beliefs about the ability of computing to benefit society have been shown to affect the likelihood of an individual pursuing CS [11, 29].

Another stereotype that many students have about people in CS is that they are typically men. When Lewis et al. [28] identified a list of factors that students used to assess their fit in computing, they found that students included the perceived masculinity of the field. This perception likely contributes to the relative lack of women pursuing computing. Indeed, Cheryan et al. [13] suggest that the "masculine culture" of some STEM fields, such as CS, contributes to the gender gap in these fields.

Our focus on AI and cybersecurity allows us to investigate the open question of whether students have stereotypes particular to these fields and whether such beliefs impact their decision to specialize in AI or cybersecurity.

2.2.2 How students pick CS specializations

Prior work has shown that students typically do not have clear goals when choosing a specialization within computing. To understand how CS majors make these decisions, Hewner interviewed computing students and advisors about how students choose courses [59]. Hewner's resulting model suggests that computing majors did not begin their academic careers with concrete goals and simply followed the curriculum until they encountered a course experience that was markedly more or less enjoyable than other courses [59]. This experience then helped the students create more specific goals and pursue a newly-defined interest [59]. This model suggests that students do not necessarily have clear preferences or goals of specialization within the major at the outset of their degree program.

However, many students have likely encountered, for example, "hackers" or AI robots in popular culture, which prior research has shown can lead to preconceived notions and stereotypes about the kinds of people who work in cybersecurity or AI [61, 64, 65]. An open question for our work to investigate is whether Hewner's model is reflected in students' experiences with AI or cybersecurity, where they may have already decided their level of interest in the field based on prior cultural exposure.

2.2.3 Recruitment into AI and cybersecurity

In AI, prior research on recruitment into the specialization has focused on the need to broaden the participation of people from HUGs in computing. A survey of undergraduates by Barretto et al. [66] showed that students from HUGs were less likely to take an AI course: the explanation proffered by many participants was that they lacked an interest in the field, especially in the technical content. Another barrier to student recruitment into AI is that students do not always have a clear understanding of what the content of the field really is. For example, Ottenbreit-Leftwich et al. [67] interviewed 4th grade students about what they thought AI was, and common answers referred to "robotic vacuums" and "search engines". Kreinsen and Schulz [68] found similar themes in their interviews with students from grades 7 through 10, where many interviewees centered their view of AI as being "the brain of the robots". These studies suggest that, at least prior to any exposure at the college level, students are unlikely to have accurate or specific ideas of what work in AI looks like, which may make it harder for students to decide to pursue the specialization.

In cybersecurity, prior work suggests that a lack of sufficient cybersecurity course offerings may be contributing to the challenge of recruiting students to the specialization [69]. A study found that none of the 10 highest-ranked CS undergraduate programs require a cybersecurity course and that most CS programs ranked in the top 50 offered fewer than five cybersecurity electives [69]. This dearth of available courses makes it harder for students to develop an interest in the field or learn what kind of work cybersecurity entails. Increasing knowledge of cybersecurity may be a particularly important factor in attracting students to the field, as suggested by the National Initiative for Cybersecurity Education (NICE) Workforce Framework for Cybersecurity [70]. This framework was created to more precisely describe cybersecurity work in support of the education of students and the recruitment of employees. For example, the 52 "Work Roles" listed by the NICE Framework describe areas of cybersecurity that an individual may be responsible for, including occupations such as "Cyber Legal Advisor" and "Technical Support Specialist" [60]. Students may not be aware that such roles are examples of cybersecurity work, making it harder to attract those with cybersecurity skills to these positions.

2.3 METHODS

2.3.1 Participant Recruitment

To understand students' perceptions of AI and cybersecurity, we conducted a qualitative study using semi-structured interviews with undergraduate students majoring or minoring in CS. Following IRB approval, participants were recruited by asking leaders of CS-specific student clubs to share our recruitment materials with their constituents. In particular, we recruited from the institution's chapter of the Association for Computing Machinery and a club for women in CS. Of the respondents, we interviewed students who had completed more of the degree program because these students were more likely to be taking advanced electives and to have thought about which specialization they want to pursue. Students were incentivized to participate with a $20 Amazon gift card.

In total, we conducted 17 interviews; these were conducted over Zoom, lasted an hour, and were recorded. Of our participants, 16 were majoring and one was minoring in CS; 11 identified as women, five as men, and one did not share their gender identity; 12 identified as Asian or Asian American, two as Caucasian or White, and three did not share their racial/ethnic identity. We deliberately oversampled women due to our interest in broadening the participation of women in AI and cybersecurity. Each participant was assigned a pseudonym by the research team. Individual participant identities are not shared due to confidentiality concerns as well as because identifying differences between groups was not the focus of our work. Instead, we hope to have captured student perceptions that can later be used to identify such differences with the goal of broadening participation in AI and cybersecurity.

2.3.2 Interview Protocol

Consistent with recommended practice [71], interview questions were updated throughout data collection. Our initial protocol asked students about their CS background, their experiences with internships, and their perceptions about AI and cybersecurity, such as what they believe day-to-day work in these fields looks like. Based upon ongoing analysis, the interview protocol was refined three times to add questions where further detail was needed. We added questions about stereotypes students may have about AI and cybersecurity; whether they felt AI and cybersecurity should be required for all CS students; and beliefs about whether certain computing specializations were more difficult, rigorous, or prestigious than others, and their impact on society. The questions pertaining to AI and cybersecurity are provided in Appendix A.

2.3.3 Data Analysis

Interview audio recordings were transcribed by a third-party service, Rev.com, and were anonymized and corrected by the authors. Data analysis was conducted in Saturate, an application for text-based qualitative analysis [72], using inductive approaches with no codes determined ahead of time. After six interviews had been independently read by two team members, we created an initial list of codes (words or phrases capturing some aspect of the text) based upon patterns apparent in these interviews. Our codes used the principles of concept and descriptive coding, i.e., were based on high-level ideas and descriptions in the text [73]. The two team members used this initial list to each code three of the six interviews and then met to refine the application of the codes. This formed the basis of the codebook used to code all subsequent interviews. To increase the confirmability of the findings [74], the team met weekly while coding the rest of the interviews. During these meetings, we reviewed codes by category in order to ensure consistency in how codes were applied. We added and removed codes as necessary, as well as resolved any conflicts or confusion regarding the meaning of the codes. Finally, upon the completion of coding, the data for related codes were reviewed to identify themes and variations in participants' perceptions of the fields of AI and cybersecurity. The coding scheme can be found in Appendix B.

2.4 RESULTS

Our findings are organized into four categories relating to student perceptions of AI and cybersecurity that emerged from our analysis: their beliefs about (1) what skills are required in the field, (2) what the day-to-day work looks like, (3) what kind of people are in the field, and (4) what impact the field has on society. In all participant quotes, an ellipsis ("...") indicates omitted words, a dash ("—") pauses, square brackets added or amended words for clarity, and *sic* an error in the original quote.

2.4.1 Skills required

AI Many participants expressed the belief that AI requires mathematical skill and, perhaps for this reason, viewed the field as challenging and difficult. For example, Quincy believed that AI is heavily mathematical and stated, when discussing the importance of math in computing, that *"I mean, I think all of AI/ML is essentially just math"*. Other participants shared similar views, such as Gail's perception that *"people who are good at statistics and math and, you know, will be encouraged to go into the field [of AI]"*. Among those who endorsed this view, participants also seemed to agree that AI is difficult. They described AI as *"intimidating"* (Diane) and *"rigorous"* (Parth) and highlighted the difficulty of the AI course they had taken, which Mahir described as being *"considered one of the most challenging [courses], probably"*. Mahir further noted that it *"involved a lot of math and I think that's what contributed to its perception as being difficult"*. It seems likely, therefore, that the perception of AI as being difficult is tied to its perception as mathematical, and that some students are choosing not to take AI courses at all because they think it would be too hard. For instance, Kasey expressed a reluctance to take an AI course because of its perceived difficulty: *"I have [considered it], I'm just afraid that it will be too difficult"*.

Cybersecurity Much like AI, participants seemed to view cybersecurity as a difficult field, in this case due to the belief that it requires systems programming. When asked about what parts of computing were rigorous, Gail said, *"Just kind of with all the low-level, like assembly stuff or like security, right now I'm taking security, I feel like it's really rigorous to be honest"*. Other participants shared the perception that cybersecurity incorporates low-level programming and is difficult. When Nora was asked what stereotypes she had heard regarding work in cybersecurity, she expressed that *"system[s] programming are (sic) always really time consuming and tedious"*. The reputation of cybersecurity as low-level also appeared to deter Idris from exploring it further, as they note that this has made them less interested in it: *"I guess like, a lot of security is lower level than I like to work"*.

Beyond its use of systems programming, participants seemed to suggest cybersecurity has a reputation as unfriendly to newcomers. For instance, when asked about what kind of people do cybersecurity, Gail recounted an experience attending a cybersecurity student club meeting and feeling *"out of place"* because *"everyone else kind of know (sic) what they're doing"*, while she felt that *"I don't even understand what's happening"*. This experience left Gail feeling *"not good enough"*, highlighting how discouraging the experience was for her. Similarly, Diane, sharing her overall perspective of the field, explained that the cybersecurity course at the institution is *"notorious"*, and that she has *"felt that sense of* *intimidation and just never really wanted to explore it because it sounded like a heavy and difficult course".* This sense of difficulty and intimidation may discourage students from exploring cybersecurity.

2.4.2 Day-to-day work

AI When asked about what the day-to-day work of AI looks like, many participants described their understanding of AI work as training models. Parth described his experience during an internship, talking to coworkers who were engaged in AI work, as follows:

> *"They're developing some model or something or the other with machine learning. So they take a bunch of derivatives and then they use that to figure out, how should the model look like? And then they write that and then they train it or—yeah, they train it and then they try to vary the parameters of that model to best fit it. And then they call it a day."*

This general view, that AI work involves training models with data, was repeated by other participants, but very few of them had concrete explanations of what this process looked like. For instance, when asked what she thought the day-to-day work of AI looks like, Lillian suggested *"I guess it'd be like reading research papers or like just training models".*

Cybersecurity Participants' conception of cybersecurity was that the work often entails "hacking" into systems or defending against attacks. For example, when Henry was asked what he thinks the daily work of cybersecurity looks like, he explained that he did not have first-hand experience, but added his perception that *"I guess there's like white-hat hacking, that's a thing, right. Where hackers like try to find vulnerabilities for companies, not to actually hack into them, but so that they can fix them".* On the defensive side, Gail recounted having heard her cybersecurity professor talk about *"the adversarial mindset"* and *"think[ing] from the perspective of [an] attacker".* She went on to say that the specific skills depend on what kind of cybersecurity work one wants to do, but that *"in general, it's the ability to be able to think from the perspective of my attacker... com[ing] up with a way to defend against the attacker for [a] system".* These perspectives suggest that, among our participants, cybersecurity had a narrow, specific definition related to adversarial hacking, a definition they are likely using in deciding whether it is something they want to pursue.

2.4.3 Perception of people in the field

AI Participants expressed the belief that the people pursuing AI are seen as "cool" and highly intelligent. For example, while discussing which specializations within CS are considered prestigious, Nora noted that *"AI/ML is really trending. So if you do that, then people think you're really cool".* The idea of AI being "cool" is also seen in Quincy's response to the same question about prestigious specializations, where he listed the current "buzzwords": *"Today's buzzwords are AI, ML, VR, future, crypto, blockchain or whatnot. Yeah. Um, so obviously that's gonna attract a lot of buzz towards the AI/ML courses".* Nora and Quincy's comments suggest that, for some students, feeling that people who do AI are "cool" and prestigious may contribute to an interest in pursuing AI. People studying or working in AI were also viewed as particularly knowledgeable and capable, likely related to the belief of it being a highly challenging field. Kasey, when asked about any stereotypes of those who do AI, responded simply, *"That they're really smart".* Other participants expressed similar sentiments, such as Diane describing the field as *"intimidating".* In response to a question regarding what kind of person is encouraged to do AI, Idris identified that *"they've been coding for ages, and they know all the software things",* and Nora felt that *"strong-minded"* people are encouraged to go into AI. These quotes suggest the belief that, to pursue AI, one must be very intelligent and capable, a suggestion that may deter less confident students.

Cybersecurity Among our participants, the primary perception of the people who work in cybersecurity was that they are also very intelligent and are usually men. For instance, Farah noted that *"cybersecurity seems like one of those fields where you have to be super duper smart to succeed".* Students seemed to believe that you must be particularly intelligent to do cybersecurity because mastering the content of the field is inherently impressive. Gail suggests this when she reported hearing the stereotype that *"security people are like super smart because you can do security".* A second stereotype Gail mentioned encountering is that cybersecurity is *"very male-dominated".* Other participants shared this impression, such as Esha, who explained that her mother had told her that cybersecurity is *"in terms of gender, [an] imbalanced field".* Mahir also acknowledged having heard a similar stereotype when asked what kind of person is encouraged to go into cybersecurity, pointing out that cybersecurity has the same *"bias towards men"* that he described as existing in computing generally:

> *"Well, besides, you know, the bias towards men in general in computer science, I don't think any other biases exist [in cybersecurity] that I'm aware of... Yeah, I think there [is] this new connection between like, [a] hacker is a man usually,*

> *which is, I guess, unnecessary, cause it's really quite accessible or it should be accessible to anyone."*

In pointing out that cybersecurity *"should be accessible to anyone",* Mahir appears to be highlighting that this is not the current perception of the field.

Notably, cybersecurity was not as popular among our participants as AI and lacked the perception of being "cool".

2.4.4 Impact on society

AI Some of our participants cited the belief that they could use AI to solve a large variety of problems, indicating a large potential for impact on society. For Idris, this was a specific reason for their decision to take courses in AI, because the problems they wanted to work on *"can be solved with AI... I think AI has the ability to solve a lot of really interesting problems".* Although they do not specify which problems these are, it is clear that they see AI as a field that can be applied to many different contexts. Similarly, Parth explains that *"[AI and ML] is like a very applicable (sic) and like, every business is trying to see how they can leverage machine learning to make their lives easier and produce more profit".* In Parth's view, not only can AI be used in many different contexts, but this utility is the reason it is an in-demand skill among employers.

However, several participants also expressed concern about the societal impact AI could have. Henry described these as *"moral gray areas"* in his discussion of what stereotypes he had heard about AI:

> *"I don't know if it's really a stereotype, but there's a lot of negative effects that AI can have, whether it's reinforcing prejudice on like Google and stuff or, well, Tesla self-driving cars running into people, but yeah, there's a lot of moral gray areas when it comes to AI."*

In this quote, Henry pointed to bad societal outcomes that could happen as a result of working in AI, which are likely to factor into his decision-making when determining whether he wants to work in the field himself. Concern about AI's societal impact is also suggested by Farah's apparent frustration when sharing her perception of the ethical concerns with using AI and the people who work in it:

> *"I feel like people are, like, yeah, we want to eliminate bias from AI and all this stuff. But like, I think the other thing that they have to acknowledge is that humans are biased, and, if you're training an algorithm or model or whatever it is, it's also going to be biased."*

Other participants shared similar concerns about the use of AI, such as when Parth said that AI *"can and has, intentionally and inadvertently, been harmful"*. Although our participants did not cite these ethical concerns as a reason to avoid pursuing AI, it is likely that many of them are weighing the impact of AI on society in their own decision of whether to pursue the field.

Cybersecurity Multiple participants explained that cybersecurity is a useful skill because of the importance of secure data in the world today, highlighting cybersecurity's potential for societal impact. Lillian, when asked whether cybersecurity should be a required course in the undergraduate CS curriculum, responded that *"I do see its utility. In the future—I mean, already data privacy is so important"*, suggesting that cybersecurity's utility is tied to the value it can bring people or companies, by keeping their data private. This value is especially important as technology becomes more prevalent in daily life, as Gail suggested when asked about why she felt that cybersecurity is an important field. She noted that *"it's very important to make all the system[s] secure because a lot of things are moving online, like credit card[s], online transaction[s], e-commerce"*. Similarly, Henry believed that cybersecurity is a *"fairly well-regarded area because it's protecting the security of the internet"*. His suggestion that cybersecurity is valued *"because"* of *"the security of the internet"* further reinforces the idea that cybersecurity skills are valued because they can significantly impact society.

We note participants did not express ethical concerns about cybersecurity work as they did about AI, despite multiple references to "hacking".

2.5 DISCUSSION

Our findings extend prior work and indicate several ways in which student perceptions of AI and cybersecurity may hinder efforts to broaden the participation of students from HUGs.

Our results suggest that students have similar preconceptions about AI and cybersecurity as about computing in general. In particular, participants' beliefs reflected stereotypes about the innate brilliance and masculine culture needed to participate in AI and cybersecurity. These beliefs have been shown to deter students from HUGs from computing [13, 20, 27]. In addition, the perception that AI requires mathematical skill may have a particularly detrimental effect on the participation of women: prior research has documented that women have lower math self-efficacy than men and that this difference affects their interest in pursuing math [75]. Participants' concerns about AI's unintentional negative societal impact may align with the belief that work in computing does not satisfy the goal of having a positive impact on society, which is a goal more often endorsed by students from HUGs [11]. This suggests that students with goals to help society may be less likely to pursue AI due to their belief that working in AI could have a more negative societal impact. Taken together, our work extends our collective knowledge about field-specific perceptions and suggests a particular need to highlight the ways in which work in AI and cybersecurity extends beyond stereotypes of requiring innate brilliance, involving math, and having a negative impact on society.

Our findings align with prior work that students at the undergraduate level understandably do not have complete or accurate knowledge of specializations within computing, as some of their beliefs about AI and cybersecurity are not supported by experts. For example, participants strongly associated cybersecurity with systems programming, but the NICE Framework lists a number of cybersecurity "work roles" whose responsibilities include computing skills that are *not* systems programming, such as the Warnings Analyst, who "[c]ollects, processes, analyzes, and disseminates cyber warning assessments" [60]. Similarly, although participants suggested that the main work of AI is training models, AI hiring managers report that this is a relatively small subset of the work [76]. Expanding student perceptions of the skills required by and the work of AI and cybersecurity may be crucial to broadening participation in these fields, because systems programming and training models may seem disconnected from societal impact, a factor shown to be important to students from HUGs when selecting a field to pursue [11].

Participants' perceptions of AI and cybersecurity are inconsistent with student experiences reported by Hewner, wherein students did not seem to have preconceived notions about a field prior to taking a course in it [59]; rather, many of our participants had narrow views of what AI and cybersecurity were like even before taking a course in them. A possible explanation is that prior exposure to these fields in media may be shaping student perceptions, as documented in prior work [61, 64, 65]. Our results also suggest that expanding student views of cybersecurity and AI cannot only take place in courses covering those subjects, as not all students will enroll in those courses.

2.6 IMPLICATIONS FOR PRACTICE

The results of our work suggest that many of participants' beliefs about AI and cybersecurity are likely to differentially impact students from HUGs, as these views mirror existing stereotypes that have been shown to deter students from HUGs in computing. Thus, in order to broaden participation in these fields, there is a need to expand students' views of what

AI and cybersecurity entail. Our work suggests that interventions to expand these views cannot be restricted to AI or cybersecurity classrooms and indicates a need to explore venues such as student clubs, conferences, introductory CS courses, and other places where students who do not have a preexisting interest in AI and cybersecurity may be found. Strategies to enhance students' exposure to the breadth of career options in these two fields could include inviting speakers whose work does not fit the stereotype of AI or cybersecurity, showcasing projects that highlight how these fields can benefit society, and sharing the NICE Framework in courses. In addition, future work in research and teaching may aim to challenge the existing stereotypes about these fields and computing generally, such as an intervention in a CS1 course designed to encourage a growth mindset, the belief that intellectual ability is not fixed and can grow [77], which could dispel the belief that these fields require innate brilliance.

2.7 LIMITATIONS AND FUTURE WORK

We note that our work is limited due to the influence of the institution's specific course offerings, which are likely to shape students' perceptions. The cybersecurity course at this institution requires a systems programming course as a prerequisite, which may explain students' beliefs that cybersecurity work is closely tied to systems programming; this belief may not be shared by students at other institutions. The lack of sufficient cybersecurity course offerings and its impact on student perceptions is a previously documented concern [69], as most of these courses focus primarily on building secure systems rather than the full depth of cybersecurity such as risk analysis & management, policy & law, penetration testing, and secure software engineering [60].

An additional limitation of our work is that participant responses to our questions may have been influenced by their unwillingness to share a perspective that may not appear to be socially desirable, such as a stereotype based in racial/ethnic identities. In order to mitigate this, our interview questions allowed for the conflation of participants' own views with beliefs they had heard expressed by others; as a consequence, our findings may reflect participants' impression of popular beliefs, rather than their own perceptions of AI and cybersecurity.

Our work's relatively small sample size limits our ability to conduct comparisons between groups, such as comparing views held by men versus women. However, we note that the existence of these perceptions and stereotypes impacts broadening participation efforts regardless of who expresses these views.

Future work in this area may investigate whether the beliefs held by our participants are shared by students at different institutions, such as those attending institutions that offer a greater number and variety of cybersecurity courses.

2.8 CONCLUSION

In this study, we interviewed computing undergraduate students in order to investigate their perceptions of AI and cybersecurity, particularly regarding the work, people, and societal impact of these fields. We found that student views did not always reflect the reality of working in AI or cybersecurity as described by experts and often reinforced stereotypes about computing. Our work suggests that there is a need to expand students' views of what working in AI and cybersecurity could be like, such as by demonstrating that they can be used for societal good. This need is particularly important in any effort to broaden the participation in these fields of students from HUGs, as many of the beliefs students held have been shown to exacerbate the existing inequitable representation of gender and racial/ethnic groups in computing.

CHAPTER 3: PHD EXPECTATIONS

3.1 INTRODUCTION

In the field of computer science (CS), educational institutions are failing to produce enough computing graduates to fulfill predicted market demand. This is the case not only at the undergraduate level [4], but also at the doctoral level, where institutions are unable to hire as many faculty as they need [16]. As recently as 2017, a survey of 155 institutions found that 18% of tenure-track faculty searches in CS failed to hire any faculty at all [16], indicating that not enough PhD candidates are pursuing faculty positions, at a time when surges in computing course enrollments across the U.S. [17] require more faculty than ever. In part, this may be due to the fact that PhD programs in computing have higher rates of attrition than other STEM fields [78]; not enough students are completing their doctoral programs to meet the demand for faculty positions. In addition, women, persons with disabilities, and individuals from racial and ethnic minority groups are significantly underrepresented in CS at all levels [3]. This lack of representation among computing degree-holders exacerbates the challenge of producing enough computing doctorates to fill required faculty positions.

In this paper, we investigate the experiences of doctoral computer science students who self-identify as belonging to historically underrepresented groups (HUGs) in computing with the goal of understanding how to support their persistence. Our provided definition of students from HUGs includes, but is not limited to, individuals from racial minorities (in computing), women of all races, people with disabilities, and individuals across a variety of gender and sexual orientations.[2] Understanding these students' perspectives is crucial to retaining students from HUGs in doctoral CS programs and may offer insight into why the field fails to sufficiently support and retain these students. Although we focus on this particular group of students, our findings can be used to support all PhD students in CS, and thereby enable more students to complete their degrees.

Previous research shows that one key factor affecting the rate of degree completion among doctoral students is a clear understanding of what is expected of them in order to complete the program [22, 31, 32]. These expectations may concern requirements from the department or from a student's advisor, who has an important say in awarding a doctoral degree. We focus on two kinds of expectations: what tasks the student is expected to complete, and how they are expected to complete them. The former category encompasses higher-level

[2]Participants were asked whether they self-identify as a member of a group underrepresented in computing, and this definition was provided as a non-exhaustive guideline.

goals, such as the publication of a paper, while the latter focuses on the lower-level, daily tasks needed to complete the overarching goal, such as debugging the code being used in a publication. In this paper, we aim to investigate the different sources that students use to form their expectations, and how each of these sources provides different kinds of guidance.

To investigate these topics, our research team conducted a survey and semi-structured interviews with doctoral students from HUGs at a large institution with high research productivity. These methods of investigation were chosen given the exploratory nature of this work. In particular, conducting semi-structured interviews allowed us to gather student perspectives on a variety of issues that they considered to be relevant. In this paper, we present our analysis of the interviews. Although all our participants identified as members of HUGs, our results do not aim to compare their experiences to their peers or identify experiences unique to students from HUGs. Rather, our work elevates the experiences of students from HUGs because of the particular importance of broadening participation in computing from HUGs. In addition, we note that these experiences may not generalize to *all* CS doctoral students from HUGs, as they likely have a wide variety of different experiences. Our analysis found that students had three primary sources from which they deduce what they are expected to do and how to do it: research experience prior to beginning their program, their PhD advisor, and their peers. Each of these sources helps students understand different kinds of expectations, with advisors providing primarily high-level guidance on what tasks to accomplish, and peers helping each other with lower-level tasks. Many students began the program anticipating more hands-on support from their advisor, and instead found themselves relying more on their labmates and peers. We also find that students' initial expectations of what graduate school would entail did not always provide a complete picture, often giving them incomplete or inaccurate expectations. To address these issues, we suggest developing lab-based mentorship systems, and encouraging clear communication between advisors and students not only on high-level goals, but on the daily tasks needed to accomplish them.

3.2 PREVIOUS WORK

Prior research has identified a number of factors that lead to doctoral program completion, many of which involve clear communication of student expectations. In a meta-analysis of 163 empirical articles regarding completion, achievement, and well-being in PhD education, Sverdlik et al. [22] found that students' understanding of what their advisor (or supervisor) wanted them to do was an important contributor to student outcomes, with the authors noting that "open, supportive, and frequent communication with [their] supervisor was found to be essential for student success and satisfaction". The authors also noted that poor communication of departmental requirements may lead to a discrepancy between the student's and the department's expectations of what the student needs to do [22]. Such a discrepancy can lead to the student insufficiently integrating into their institution and discipline, resulting in a decreased likelihood of degree completion [31]. In addition, clear communication of expectations may impact students in more ways than simply completing the program. For example, a survey of graduate students by Fisher et al. [32] found that a clear understanding of their department's expectations contributed to lower levels of distress in women and minority (defined as Black, Latinx, and American Indian/Alaska Native) graduate students and indirectly affected students' publication rates.

Apart from communication of expectations, much of the existing body of work on this topic has found that advisors are one of the most important factors affecting doctoral degree completion [22, 79, 80, 81]. The role of the advisor may be so important in shaping doctoral student experiences because they are the primary source for students' understanding of what is expected of them. In addition to helping set expectations, good advising has many facets: according to Posselt [23], successful advisors provide holistic support, with "academic, psychosocial, and sociocultural dimensions, which faculty enact through specific behaviors". This kind of supportive advising is not experienced by all doctoral students [31, 81, 82], and the impact of poor quality advising may differentially impact students from groups underrepresented in computing due to the discrimination they may face [83].

Doctoral attrition rates are generally higher in CS than other STEM fields, with computing having the lowest 10-year degree completion rate and highest attrition rate among STEM fields for students beginning their program in the 90s [78]. In addition to the already low number of PhDs awarded in computing due, in part, to high attrition, few of these degrees are awarded to students from HUGs. The 2020 Taulbee survey found that only 21.7% of the PhDs awarded in computing by the surveyed institutions were earned by students who identified as women, and 13.4% of them by domestic students who did not identify as white [3]. Part of this disparity in doctoral degree attainment between majority and underrepresented groups may be due to the challenges students from HUGs continue to face in higher education. For instance, Cohoon et al. [84] found that computing graduate programs can feel unwelcoming to women and that their continued persistence in the face of adversity may depend upon the use of several coping mechanisms. Students from HUGs may also encounter a lack of supportive mentors [82], which may increase their likelihood of attrition. In a qualitative study of African American undergraduate and graduate computing students, Charleston [85] highlighted the importance of mentors such as advisors in navigating challenges, noting that many participants "described how they considered withdrawing from computing science programs if not for the intervention of a mentor".

Taken together, prior work suggests that computing doctoral students are less likely to complete their programs, and also that students from HUGs face different and additional challenges compared to their majority[3] peers, which may further decrease their likelihood of completion. In this paper, we seek to better understand these compounded challenges by investigating how doctoral computing students from HUGs understand what is expected of them and how to do it.

3.3 RESEARCH QUESTIONS

RQ1: How do students from HUGs form expectations of their CS PhD programs?

RQ2: What sources do students from HUGs rely on to form expectations of their CS PhD programs?

RQ3: How do students from HUGs in CS PhD programs learn how to meet these expectations in order to complete their degree?

3.4 METHODS

In order to answer the above research questions, our research team conducted a survey and follow-up interviews with CS doctoral students at a large institution with high research productivity. Participants for the survey were recruited by soliciting doctoral students in the Computer Science department who self-identified as being part of an underrepresented community in computing. We chose to recruit only students from HUGs in order to highlight their voices and experiences. As a result, we do not have equivalent data from majority students, and thus do not make any comparative claims.

Recruitment occurred through email lists and a graduate student Slack channel. Our survey received 29 respondents, of whom 19 identified as women, and 14 identified as racial and ethnic identities other than white. After completing the survey, participants were invited to participate in an optional follow-up interview. In this paper, we discuss the findings from the resulting 14 interviews.

The interviews were conducted and recorded over Zoom due to the COVID-19 pandemic. Interviews lasted no more than an hour. The protocol was semi-structured in nature and consisted of questions across three sections: the participant's relationship with their advisor,

[3] This refers to groups that are well-represented in computing, that is, white and Asian men.

their expectations during the PhD, and their experiences learning and acquiring new skills in graduate school. A copy of the interview questions is provided in Appendix C.

Of the 14 interview participants, eight were in either the first or the second year of their PhD program. Table 3.1 provides a list of the pseudonyms used for our participants, along with their seniority within their doctoral program. Due to the relatively small number of individuals involved, we do not provide demographic or otherwise identifying information about the interview participants.

Table 3.1: The interview participants by degree seniority, using researcher-assigned pseudonyms.

Early stage (1st or 2nd year)	**Middle to late stage (3rd year+)**
Anthony	Anna
Glen	Diane
Kurt	Hannah
Mia	Patricia
Nora	Robert
Phoebe	Tom
Tracy	
Yvonne	

After all of the interviews had been completed, the audio recordings were sent to a third-party service to be transcribed. These transcriptions were further reviewed by the authors to remove any identifying information. The authors then analyzed the anonymized transcriptions using a qualitative coding process. Due to the exploratory nature of the work, coding was conducted inductively, with no codes determined a priori [73]. Each team member independently coded one of the interviews in order to identify initial key ideas and themes. A preliminary codebook was developed based on these observations and used to code the remaining interviews. The authors met regularly during the coding process in order to continue iterating on the codebook, adding and removing codes as necessary, and resolving any conflicts or confusion regarding the application of the codes. At the end of the first pass of coding, the authors listed and discussed findings relating to themes and variations observed in the data. In this paper we discuss only those findings which are relevant to our research questions regarding expectations. A second round of coding was conducted on the subset of codes relevant to this paper in order to ensure all evidence regarding the findings of interest to this work had been identified. This subset of codes is provided in Appendix D.

3.5 RESULTS

We note that the results presented here are a subset of our findings, restricted to those that pertain to forming and fulfilling expectations. Within this scope, we find that participants did not typically refer to experiences specific to their social identities when discussing expectations. Thus, our results are not specific to students from HUGs, but can be used to better understand and improve the experiences of all PhD students.

Our analysis identified three key sources from which students appear to form their expectations: (1) students' prior research experiences, (2) their PhD advisor, and (3) their peers in the doctoral program. Notably, each of these sources provides different types of support and advice. We describe each type of source and what students learned from them below.

In the quotes used in this section, we have used square brackets (i.e., [added text]) to indicate words added for clarification, and three dots (i.e., ...) to indicate a removal of text.

3.5.1 Prior Research Experiences

Mentors and peers from before graduate school help shape expectations. Many participants reported engaging in research experiences prior to attending graduate school. During these research experiences, some participants encountered key mentor figures and peers that often formed their expectations of what graduate school would look like. For instance, when Anthony was asked about the source of his expectations for what he "should be doing" in graduate school, he discussed attending conferences as an undergraduate, and observing the graduate students "not only attending, but also actively involved in making some contribution, or organizing". Anthony went on to describe treating these individuals as "kind of a role model", indicating that he adopted these behaviors as a blueprint for what to do during the doctoral program and how to be "active" in the research community. Similarly, Mia reported that the professors at her undergraduate institution formed her view of what graduate school would look like, providing advice regarding "what you need to do in order to be successful in a PhD program". In Mia's case, mentors from her prior experience not only helped her understand what she would need to do in graduate school, but also shaped her perception of what success in the program might look like. Most of our participants discussed engaging in research before entering graduate school, and it is reasonable for those interactions to form their initial expectations regarding what they would need to do in a PhD.

Prior expectations can be incomplete. Despite these prior research experiences, many participants reported being surprised by aspects of research once in graduate school. For instance, a few participants discussed being surprised by the importance of writing and presenting the results of a study, and the relative lack of focus on programming. As Patricia explained, "it's not the system you build that is the big part of the research, but it's *after*, when you get the results back" (emphasis added). Phoebe realized that time management during the PhD was more challenging than she had anticipated. She expected coursework to play a minimal role in the degree program, but found instead that it was difficult to balance both research and coursework. For many students who felt that they knew *what* needed to be done coming in, learning *how* to complete those tasks still felt challenging. Yvonne struggled to find a compatible advisor and felt that although it was clear that finding an advisor was important, there was little guidance on how to go about it. She reflected that, in retrospect, it would have been better to have a particular advisor in mind before applying to the program. Thus, in Yvonne's case, even though she came in with expectations of graduate school based on prior experience, these expectations were incomplete, helping her understand what needed to be done, but not how to do it.

3.5.2 Advisors

Advisors tend to provide high-level guidance. Once in the PhD program, students often based their expectations of what to do on the recommendations of their advisors, particularly regarding what big-picture steps to take to complete their degree. Most participants indicated that their meetings with their advisors were conducted at an abstract level, discussing research directions, status updates, or degree progress, and that they received advice on how to proceed on those topics. For example, Tom described their meetings as including "career advice, internship advice, and just general research things", highlighting that he primarily received advice on the big-picture goals he needed to pursue. Discussions of what to do in the PhD often operated on the same level: for instance, when asked what his advisor expects him to achieve during the degree program, Kurt responded that his advisor's expectations are "totally the same requirements as the department's to obtain the PhD: pass the qual exam, do your thesis". Another participant, Robert, described his advisor as being "more like a sage rather than a teacher", explaining that the kind of guidance he received from his advisor did not help him understand how to do research on a daily basis. Milestones and advice of this kind represent a clear expectation of what the student needs to do, such as publish a paper or pass an exam, but does not necessarily inform students *how* these tasks are accomplished.

Advisors expect students to "figure it out on [their] own". Although their respective advisors guided our participants towards the milestones they were expected to achieve, most of our participants indicated that their advisor did not help them solve lower-level, everyday issues. Two of our 14 participants indicated that their advisors had helped them debug code; the others described their meetings as covering only big-picture topics. In fact, Glen reports that after the first few challenges, he realized that his advisor "really didn't have time to actually get into weeds to help me out with something". The relative independence advisors expect from their students was a surprise to some of our participants. For example, Patricia says that she came in thinking that she would be assigned to work on an existing project, but instead found that "it was my job to figure out a research idea and then work on that". Her advisor was willing to guide her, but was "pretty hands off" when it came to her understanding what she needed to do on a daily basis. Patricia, and other participants reporting similar experiences, had to find a different source to help them understand the daily tasks required to conduct research.

Advisors recommend asking peers for help. Of the participants that mentioned that their advisor did not have time to help with low-level issues, many noted that their advisors usually suggested talking to a labmate in order to resolve the issue. Diane describes her advisor pairing her with a senior PhD student in her lab as a first year in order to work on her first project, and describes the senior student as "sort of like a mentor". This mentor-like relationship included relying on the senior student for feedback and on project ideas and "general grad school things... like housing". Both through the recommendation of their advisor and their own initiative, many of our participants relied on peers' help in the PhD program, as discussed below.

3.5.3 Peers

Students often reach out to peers for help. Interactions with peers were another key source from which students formed expectations and understood how to fulfill them. Participants often reported turning to other students in the same lab for help. These instances of asking for help came both from their own initiative and from their advisors suggesting that they ask a labmate for help. For instance, Glen describes seeking out advice from his senior labmates about his timeline for achieving a PhD, and finding out that he "should probably get a paper by my second year". This piece of advice highlights peers' ability to help convey *what* students need to do, not just the how. Glen added that he felt more comfortable talking with other graduate students who had similar experiences instead of his advisor, who always seemed busy. In addition to receiving such high-level advice, students often sought out peers for more lower-level help. When planning out her thesis, Anna reached out to former labmates who had graduated for help in how to organize it and asked for copies of their theses to read. By doing so, she was able to develop her thesis structure without relying on her advisor. In another example, Nora notes that when she asked her advisor for help with a new concept, he suggested that she reach out to one of his other students who had experience with the tools she needed. Nora and Anna are both examples of cases where peers were able to help our participants figure out *how* to achieve a particular task.

Observing peers can help students form expectations. Although many students reported actively seeking out peers for help, it is also worth noting that many of them formed expectations for graduate school simply based on their observations of other students. These expectations often concerned the high-level goals students felt like they ought to be achieving. For example, Anna recollects that she felt pressured to publish more after "seeing how often they [peers] publish" and that this was a source of her "standards" for how often PhD students should publish. Other participants similarly mention using other students' research output and behavior as a comparison to their own. Hannah also describes that she better understood what kind of work graduate school entails by observing her peers, although these expectations did not necessarily concern a task required to complete the degree. She notes that after some time in graduate school, her understanding of what research work in academia looked like "really comes down to my peers and what I see them working on". The result was that "the vision that I had at the beginning of what academia would be like didn't really live up to what it actually was, for me, anyway". Observing her peers allowed Hannah to form a clearer picture of what academia looked like, and in her case, realize that it was something she did not want to pursue further. The phenomenon of students using peers' work as a tool for comparison and clarification highlights the importance of peer relationships in graduate school, as they are the first source students turn to when expectations from other sources are not clear enough.

3.6 DISCUSSION

Our research questions aimed to investigate how students' expectations of a PhD program are formed, what sources they tend to rely on to form these expectations, and how they learn to meet them. We were interested in expectations of two kinds: expectations surrounding *what* work they will do, and *how* they will do it. In the first category, we include discussions around what goals they need to achieve (e.g., publishing a paper, passing an exam) in order to complete the degree. The second category describes the actions students must take to achieve these goals, such as building software or writing the results of a study. Although students' own definition of success plays a role in determining expectations, we narrow our definition here to tasks participants perceived as necessary to complete the degree. Although all our participants identified as members of HUGs, our results do not aim to compare their experiences to their peers or identify experiences unique to students from HUGs. Rather, our work elevates the experiences of students from HUGs because of the particular importance of broadening participation in computing from HUGs. In addition, we note that these experiences may not generalize to *all* CS doctoral students from HUGs, as they likely have a wide variety of different experiences. We found that students have three sources that they use to form expectations of the doctoral program: prior research experiences, their advisor, and their peers.

One key finding from our work is that for our participants, although they used their prior research experiences to form expectations of the doctoral program, their experiences did not fully or accurately reflect what graduate school would look like. This finding is especially worth noting given that our participants attend a highly-ranked CS PhD program and had sufficient prior research experience to make them competitive enough to be admitted. Despite this, they expressed surprise about aspects of research in graduate school. In particular, our participants were surprised by the daily work that goes into research, such as the amount of writing and the time-management required. It is possible that the surprise our participants felt is because undergraduate research experiences do not aim to perfectly prepare students for graduate school, but rather give an introduction. However, prior research experience is highly valued in the PhD admissions process, which may be why students were surprised to find that previous research did not give them accurate or complete expectations of the doctoral program. This highlights the importance of advisors and students needing to calibrate their expectations together, rather than assuming prior experience will provide a complete picture.

Our results further suggest that students encountered more obstacles in understanding *how* to accomplish a particular task, rather than understanding what they needed to do. Some participants felt that they did not have enough support in how to do the things required of them, such as finding an advisor. These findings imply that clear communication between students, advisors, and the department requires not only stating what the goals are, but also providing step-by-step support to students in achieving that goal.

Another major contribution of our analysis is to emphasize the importance for students to have a support network. As discussed above, advisors tend to provide guidance at a high-level, and often do not have the time to "get into the weeds" with each of their students individually. Therefore, advisors may attempt to redirect their students to other resources, such as more senior students in their lab. This allows students to expand their support network and learn the independence needed to succeed in graduate school. In addition, even when advisors do not direct students to do so, we find that they often choose to turn to their peers for help. Taken together, these results point to the need for students to be able to support each other with smaller-scale issues in a systematic way, so that they do not have to spend time searching for those relationships themselves. In particular, advisors could create a mentoring system within their labs, wherein senior students support their more junior labmates. Such a system may be beneficial in developing the support structure a newer student may need without relying solely on their advisor. A mentoring system has the added benefit that students may be more comfortable talking to their peers than their advisor. It also provides each new student with a designated mentor to ask for help, rather than the new students wondering who to talk to. Given the qualitative nature of our study, we recommend that additional work is undertaken to further explore such a recommendation. Implementation of a mentoring system may require training students and faculty in how to be inclusive and effective mentors. For resources regarding inclusive mentorship, we suggest the National Academies of Science, Engineering and Medicine's Guide to Effective Mentoring [86], the MentorFirst project [87], and the Center for the Improvement of Mentored Experiences in Research [88].

3.7 LIMITATIONS & FUTURE WORK

Our work is limited in a number of ways. First, it is worth highlighting that this work was conducted with students currently attending a highly-ranked program and institution, meaning that they are more likely to have had more access to research opportunities before graduate school. In addition, since they are persisting in the program, we do not know what other factors may affect students' decisions to leave their programs, which may be a possible avenue for future work. A majority of our participants were early-stage PhD students (eight out of 14 were in their first or second year); as a result, experiences of students later in the degree progression are not as well-represented. While there is a heavy focus on early-stage students, we note that the early years are a key period during which students transition to and become socialized within PhD programs, and are thus of particular interest in understanding how their expectations are formed. We also note that for these early-stage students, a majority (if not all) of their experience in graduate school has occurred during the COVID-19 pandemic. This has likely impacted their interactions with their advisors, peers, and the department. One possible avenue for future work would be to conduct a longitudinal study, encompassing the changes in perspectives students may undergo over time, in order to better understand how their expectations may have changed throughout their graduate school career.

We also note that students' expectations are likely to be interwoven with their own perception of what it means to be successful in graduate school. For this reason, we have chosen here to focus primarily on expectations that relate to requirements set by either the department or the advisor, as these connect directly to degree completion. However, it is difficult to separate which expectations stem from the advisor and which of them are impacted by the student's own sense of what they need or ought to accomplish. In the same vein, it is conceivable that students' expectations of themselves are affected by their background and their social identities. Future work studying this question can engage specifically with the question of how students conceptualize success, whether this connects to aspects of their identity, how this conception interacts with their expectations of what they need to do, and how these notions may change over the course of the degree program.

A final limitation we wish to acknowledge is that the members of the research team identify as belonging to HUGs themselves and were also familiar with some of the participants. This may affect our analysis in ways we cannot accurately measure.

We focused the sample and analysis of this paper on student voices, in particular students from HUGs, in order to highlight their voices and identify strategies that can be used to support these populations and benefit all graduate students. Although our analysis does not identify experiences unique to students from HUGs, we feel that this approach is valuable in understanding the challenges these (and other) students may face. However, a more complete picture of the graduate school experience requires an understanding of the perspectives of the department and the faculty. While this was not the aim of this study, we believe future work in this area can investigate how expectations are being communicated from the department and faculty perspective, as well as the experiences of majority students.

3.8 CONCLUSION

Our study sought to investigate the experiences of students from HUGs in CS PhD programs and to better understand how they determine what is expected of them in order to complete the program. We were interested in what sources students rely on to form these expectations, as well as how they learn to meet them. We limited our work to those goals which were necessary in order to complete the degree, which includes tasks set by the department and by the student's advisor. Understanding how students from HUGs form and fulfill expectations is particularly important in the field of computing, where attrition rates are higher than those in other STEM fields, and the PhDs awarded do not proportionally represent the population of the US.

In our investigation, we found that our participants had three primary sources from which their expectations were formed: prior research experience, their PhD advisor, and their peers in the degree program. We find that students learn about different types of expectations from different sources, with advisors primarily helping them understand what their big-picture goals should be and peers helping them with smaller research tasks. However, we also find that students were surprised by certain aspects of how graduate school differs from their research experiences as an undergraduate, including their advisors being available largely for high-level support. To address this issue, we suggest that clear and successful communication between students, advisors, and the department should include not only what goals the students need to achieve, but also support on how to achieve them. Developing a lab-based mentorship system, where more senior students can provide lower-level help to their junior labmates, may also be one mechanism by which to provide such support.

CHAPTER 4: SELF-EFFICACY

4.1 INTRODUCTION

Computing self-efficacy is an important factor in shaping students' motivation, performance, and persistence in computer science (CS) courses [36, 37, 38, 39, 52, 89]. Self-efficacy refers to an individual's belief that they can act in ways that will lead to a desired outcome [34, 35] and has been shown to contribute to student retention in computing [36, 37, 38, 39]. Understanding factors that impact computing self-efficacy may help in improving the retention of students from historically underrepresented groups in CS. In the US, women, disabled people, Black, African American, Hispanic, Latina/o/x, and/or Native American people are underrepresented among CS students and professionals compared to their representation in the US population [3, 4, 5]. The lack of proportional representation of people from these groups exacerbates the dearth of skilled individuals to fill the growing number of computing positions [4] and, more importantly, represents the loss of important potential insights, innovation, and contributions. Our goal with this work is to deepen our understanding of computing self-efficacy in order to pursue greater diversity, equity, and inclusion in computing.

Our work aims to understand the factors that may predict computing self-efficacy in undergraduate students enrolled in CS courses. Previous work has shown that prior experience with computing contributes positively to students' self-efficacy [40, 41, 42, 43]. Students' identities have also been shown to relate to their computing self-efficacy. In particular, women tend to report lower self-efficacy in computing than men [38, 39, 44, 45, 46], which may be related to the fact that women tend to have less experience in and exposure to computing than men prior to college [19, 90, 91, 92]. A similar phenomenon may occur for students who identify as Black, Native, and/or Hispanic, collectively referred to in this paper as historically underrepresented racial/ethnic groups (HUREGs), who also tend to have less access to computing [2, 5, 93]. In this study, we aim to quantify the relationship between these identities and computing self-efficacy while controlling for prior experience.

We also aim to understand the unique experiences of those at the intersection of multiple of these marginalized identities, using the lens of intersectionality [24, 47, 94, 95, 96]. In addition, we know that institutional policies can affect students' computing self-efficacy [46]; in this paper, we investigate the policy of requiring a higher proportion of students to take a CS course. To this end, our research question is:

RQ: How is the computing self-efficacy of students enrolled in CS courses predicted by students' prior experience in computing, demographic identities, and institutional policies?

We answer this research question using data from 31,425 student responses across 262 institutions to a large, multi-national survey from 2018 to 2021. Using this multi-institutional sample, we performed linear regression in order to determine the role played by the different factors discussed above. The scale of our data permits analyses that may not otherwise be possible due to an insufficient sample size, such as including a variable representing non-binary gender identities. The results of our models confirm previous work that more experience with computing predicts higher computing self-efficacy. However, identifying as a member of a HUREG, Asian, non-binary, and/or a woman were all statistically significantly associated with lower computing self-efficacy. Our findings demonstrate that, even with recent data, gender- and race-based differences in student experiences continue to exist, and further investigation is required to understand why some of these patterns exist. We point to several avenues for future research in self-efficacy with the overarching goal of broadening participation in computing.

4.2 PREVIOUS RESEARCH

4.2.1 Computing Self-Efficacy

Self-efficacy is a measure of an individual's belief that they can act in ways that will lead to a desired outcome [34, 35]; in the context of CS, these outcomes may include, for example, writing a complete computer program from scratch or successfully passing a CS course. Scholars have studied different goals that students may be trying to achieve in computing, resulting in naming different types of computing self-efficacy. For instance, "computer" self-efficacy refers to an individual's perceived ability to use computers and software packages [97], while "programming" self-efficacy measures students' fluency with a particular programming language, such as C++ [98]. In a study by Lin [99] on different types of computing self-efficacy, "learning" self-efficacy referred to an individual's perceived ability to accomplish academic tasks in their CS courses. Learning self-efficacy is the type of computing self-efficacy most closely aligned with our work. In this paper, we are interested in student's perceptions of their ability to master computing concepts and pass their computing courses. For more information on how self-efficacy was measured in our study, see Section 4.4.2.

In his seminal work, Bandura [35] suggested that self-efficacy may correspond with students' persistence and final performance at a task. Indeed, subsequent research has found that self-efficacy positively predicts academic interest and persistence in a career in different science, technology, engineering, and mathematics (STEM) majors [89] and in computing specifically [52]. In computing, researchers have shown the importance of self-efficacy in students' persistence in a computing major, finding that high self-efficacy predicts increased motivation and desire to continue pursuing a CS degree [25, 36, 37, 38, 39]. As a result, studying computing self-efficacy may be valuable for retaining the participation of students from historically underrepresented groups, who tend to have higher rates of attrition in STEM fields than their peers [100]. Although most computing education research has focused on self-efficacy for students pursuing CS majors, we believe it is important to understand the self-efficacy beliefs of all students in computing courses, as the workplace demand for technology skills is growing substantially even in non-computing jobs [101].

4.2.2 Factors Related to Computing Self-Efficacy

Prior work has documented that women tend to have lower computing self-efficacy than men [38, 39, 44, 45, 46]. A study by Wilson et al. [102] highlighted that this gender difference is particularly notable in CS, being one of only three STEM fields where there was a statistically significant gap in self-efficacy between men and women. There are several possible explanations for this gender difference, such as the perception of computing as a masculine field [21, 103] or the pattern whereby women have less prior experience with computers and programming coming into college [19, 90, 91, 92]. Existing work investigating the relationship between gender and self-efficacy has made use of a gender binary, treating students as either men or women. With emerging societal understanding of the complexity of gender [104], we note the importance of additionally investigating the experiences of individuals who identify outside of the gender binary. Given that self-efficacy predicts motivation and persistence in a computing major [39], understanding the relationship between gender and computing self-efficacy may suggest opportunities for broadening the participation of women and non-binary individuals in computing.

Students from HUREGs in computing may also experience lower self-efficacy than their peers due to underrepresentation and marginalization in the field of CS [2, 93] as well as a lack of same-race mentors [105, 106]. For instance, Nguyen & Lewis [46] found that Black, Latinx, and Asian[4] students have lower computing self-efficacy than their white peers. Of particular interest to our work is the theory of intersectionality, which posits that individuals with multiple marginalized identities experience unique forms of discrimination and

[4]Asian students are not considered underrepresented in computing in the US, but may still experience discrimination and bias based upon their race.

bias [47]. In light of calls to investigate the complexities associated with multiple axes of oppression (e.g., [14]), our work investigates individuals who identify as either women or non-binary *and* as a member of a HUREG. Scholars have begun to study the experiences of students from HUREGs in computing through an intersectional lens [24, 94, 95]; our work contributes to this growing field by investigating the quantitative relationship between multiple marginalized identities and computing self-efficacy.

Importantly, evidence suggests that prior experience with computing may predict students' computing self-efficacy. Existing research has shown that students' computing self-efficacy is influenced by their familiarity with programming [40, 43] and with computers [41, 42]. Furthermore, more computing experience may create more opportunities for self-efficacy-enhancing experiences, such as interactions with faculty, shown by Blaney & Stout [39] to positively predict students' computing self-efficacy. It is likely that successful experiences with computing contribute to greater self-efficacy and to a student's choice to major in computing [20]. We add to this body of literature by examining the relationship between prior experience and self-efficacy while controlling for students' gender and racial/ethnic identities.

4.2.3 Institutional CS Course Requirement

There is evidence that institutional policies can also predict student outcomes such as self-efficacy. In particular, Nguyen & Lewis [46] demonstrated that competitive enrollment policies negatively predict students' self-efficacy. In our work, we investigate the relationship between students' computing self-efficacy and an institutional policy requiring a greater proportion of students to take a CS course. This is motivated by emerging research questions on the impact of mandatory computing courses on affective student outcomes such as sense of belonging and self-efficacy [107]. We hypothesize that requiring more students, not just CS majors, to take a computing course will contribute to the belief that everyone can do computing, which will in turn contribute to students' self-efficacy.

4.3 RESEARCH QUESTION AND HYPOTHESES

RQ: How is the computing self-efficacy of students enrolled in CS courses predicted by:

(i) the student's prior computing experience,

(ii) the student's gender identity,

(iii) the student's racial/ethnic identities,

(iv) the computing requirement at their institution, and

(v) the intersection of their gender *and* race/ethnicity?

Based on prior literature [39, 40, 42, 43], we hypothesize that students' prior experience in computing will have a positive relationship with their computing self-efficacy. Additionally, we hypothesize that women will tend to have lower self-efficacy than men [38, 39, 44, 45, 46]. Although we are not aware of research exploring the computing self-efficacy of non-binary students, we expect that they too will tend to have lower self-efficacy than men due to the masculine culture of computing [21]. We hypothesize that students who identify as members of a historically underrepresented racial/ethnic group (HUREG) in computing will have lower self-efficacy than their white or Asian peers because of their underrepresentation and lack of same-race mentors in computing [2, 93], a factor shown to be important in retaining students in a field [105, 106]. In the case of institutional computing requirements, we hypothesize that attending an institution that requires *more* of their students to take a computing course will positively predict computing self-efficacy. Finally, our work regarding intersectionality is exploratory. Existing work in intersectionality emphasizes the importance of investigating the unique experiences of students with multiple marginalized identities [24, 47, 94, 95].

4.4 DATA

4.4.1 Sources

Student Data Our student data is from the Computing Research Association's (CRA) Data Buddies survey for undergraduates, a large, multi-institutional survey of students enrolled in CS courses [108]. The CRA distributes the survey to computing departments across the US and Canada, who then distribute the survey to students enrolled in their CS courses. Most institutions have a response rate of less than 20%[5] [108]. Our dataset contained 31,425 student responses across 262 institutions from the years 2018-2021. Students may appear in the dataset multiple times, but our data does not permit tracking a student between years. We do not have data regarding students' year in their program when completing the survey.

Institutional Data Data on an institution's computing requirement policy was gathered from the Enrollment Survey administered by the CRA to computing departments at the beginning of the 2015-2016 academic year. Of the institutions attended by students in

[5]Although these response rates raise concerns regarding self-selection in our data, we are unaware of any other survey with comparable institutional participation that has a higher response rate.

our participant pool, 98 completed the department survey. In the dataset provided by the CRA, institutional data was appended to the individual student data. Institution type, size, location, and other identifying information were all removed by the CRA in creating the dataset.

4.4.2 Measures

Below, we describe how the variables relevant to this work were measured. Correlations between these variables are provided in Appendix F.

Students' Computing Self-Efficacy In order to create a measure of students' computing self-efficacy, we used student responses to the following survey items:

- I am confident that I can pass my computing courses.
- I am confident that I can learn the foundations and concepts of computing.
- I am confident that I can quickly learn a new programming language on my own.

Survey respondents rated each of these items on a 5-point Likert scale between "Strongly disagree" and "Strongly agree". We computed Cronbach's alpha for the three self-efficacy items and found that the items are internally consistent, with a Cronbach's alpha of 0.76 ($\alpha > 0.7$), indicating good reliability [109]. Prior to beginning our data analysis, we examined our data and found that on each of the three self-efficacy items, less than 14% of the responses were missing. We shifted the values such that "Strongly disagree" corresponded to a -2 and "Strongly agree" to a 2. A student's computing self-efficacy was then computed by taking an average of all the self-efficacy items they responded to. In our dataset, students' mean self-efficacy on a scale from -2 to 2 was 1.28, with a standard deviation of 0.72.

Student Demographics Responses to the student survey were also used to identify students' gender and racial/ethnic identities.

Respondents chose one of the following for their gender identity: Man; Woman; Genderqu eer/non-conforming; Non-binary; Agender[6]; Self-identify. Students who selected genderquee r/non-conforming, non-binary, agender, or self-identify were grouped together as non-binary. In our participant pool, 59% of students identified as men, 31% identified as women, and 2.5% identified as non-binary. The remaining 7.5% did not provide their gender.

[6]The Agender option was only available in the 2021 survey.

Students were able to indicate identifying as one or more races, which we grouped as: Asian, Black, Native[7], and white[8]. Further, respondents could indicate whether they identified as being of Hispanic ethnicity. In our dataset, 10% of students identified as Hispanic, 2% as Native, 6% as Black, 36% as Asian, and 50% as white. Each student may be represented in one or more of these groups.

Students' Prior Computing Experience To create a measure of students' computing experience before college, we used responses to the following question in the survey: *Which of the following experiences did you have prior to entering an undergraduate program? Select all that apply.* The research team categorized these experiences as evidence of *substantial* experience, *some* experience, or *no* experience. Experiences categorized as *substantial* prior experience included taking AP CS, whereas an example of *some* prior experience was attending computing workshops. Since students could select multiple options, their final categorization was based on the "highest" of the three categories that any of their selections were in. For a list of all possible responses and how each was categorized, see our pre-registration [111] or Appendix E. 77% of our respondents were categorized as having substantial prior experience, 4.1% as having some prior experience, and 19% as having none. Students who did not complete this question were not included in our analysis.

Major(s) In the student survey, students select their major(s) from a list of options. Options were coded into one of the following categories: a CS major, a non-CS STEM major, or a non-STEM major. 84% of our respondents were categorized as CS majors, 5.6% as non-CS STEM majors, 0.02% as non-STEM majors, with the remainder not having provided their major. The complete list of majors and their categorisation can be found in our pre-registration [111] or in Appendix E.

Institutional CS Course Requirement To determine an institution's CS requirement policy, we used the following question from the departmental survey: *Does your institution require a computing course for any non-computing major?* Institutions were classified into three groups based on which students were required to take CS: *all* students, *some* non-CS majors, or *no* non-CS majors. For the options available on the department survey and how each was categorized, see our pre-registration [111] or Appendix E. 4% of students attended

[7] We included those who identified as American Indian/Alaska Native, Indigenous or First Nations, and/or Native Hawaiian/Pacific Islander as Native.

[8] White students included those who identified as Caucasian/white and/or Arab/Middle Eastern. This is in accordance with the US Census, although many individuals of Arab/Middle Eastern descent may not identify as white [110].

institutions that require all students to take CS and 46% attended institutions where some non-CS majors were required to take CS.

4.5 METHODS

4.5.1 Regression Models

We used linear regression to examine whether the independent variables in our research question predict students' computing self-efficacy. We created six models and added new variables in each to be able to isolate the changes due to the inclusion of new variables. Variables were added based on the strength of the evidence regarding their relationship with computing self-efficacy or related constructs (see Section 4.3) in the following order:

Model 1: Prior computing experience

Model 2: Gender identity

Model 3: Racial/ethnic identities

Model 4: Institutional computing requirement

Model 5: Intersection of gender and racial/ethnic identities

Model 6: Major and response year (fixed effects)

Not all students in our dataset responded to every question. As a result, each subsequent model contains responses from fewer students. The equation for our complete model is shown below.

$$\begin{aligned} SE_s = {} & \beta_0 + \beta_1 subsExp_s + \beta_2 someExp_s \\ & + \beta_3 woman_s + \beta_4 nonBinary_s \\ & + \beta_5 black_s + \beta_6 native_s + \beta_7 hispanic_s + \beta_8 asian_s \\ & + \beta_9 allReq_i + \beta_{10} someReq_i \\ & + \beta_{11}(HUREG_s \times woman_s) + \beta_{12}(asian_s \times woman_s) \\ & + \beta_{13}(HUREG_s \times nonBinary_s) + \beta_{14}(asian_s \times nonBinary_s) \\ & + \gamma_y + \mu + \epsilon_s \end{aligned} \tag{4.1}$$

In this equation, SE_s refers to students' computing self-efficacy. $subsExp_s$ and $someExp_s$ indicate whether a student had substantial or some prior computing experience. Students with no prior computing experience were therefore the reference group. $woman_s$, $nonBinary_s$, $black_s$, $native_s$, $hispanic_s$, and $asian_s$ capture the student's gender, race(s), and ethnicity, with men and white students as reference groups. We note that students may be represented in multiple racial/ethnic groups. $allReq_i$ and $someReq_i$ represent whether the student's home institution required all non-CS majors to take a CS course, or only some; institutions with no CS requirement were the reference group.

Finally, we have four variables to represent the intersection between the student's gender and racial/ethnic identities. The abbreviation HUREG refers to historically underrepresented racial/ethnic groups, i.e., groups of students who identify as Black, Native, and/or Hispanic[9]. Therefore, any student who identifies as Black, Native, and/or Hispanic would have a value of 1 for the variable $HUREG_s$. By multiplying this binary value with either $woman_s$ or $nonBinary_s$, we construct binary variables that have a value of 1 only for those students who identify as *both* members of HUREGs and women/non-binary. These intersectional variables allow us to capture how the experiences of individuals with multiple marginalized identities differ from individuals who share only one of those identities. For example, a Black woman's experiences are not a simple sum of the experiences of Black people and women.

We accounted for which year the student's survey response is from (γ_y) and their major (as categorized in Section 4.4.2; μ) as fixed effects in our model. Fixed effects are variables included to represent different groups among our data (e.g., major and response year) in order to control for unobserved variation between the groups. ϵ_s is the error term per student. Standard errors were clustered by institution because responses from students attending the same institution may not be independent of one another.

4.5.2 Limitations

We note that linear regression indicates correlational relationships between variables and does not provide evidence for causal relationships. In addition, our findings are limited by the data, as we could not measure all the factors we believe may relate to computing self-efficacy. Notably, we think self-efficacy could be affected by students' (1) performance in their current CS course(s), (2) perception of their performance relative to their peers, and (3) years completed in their program and in CS. There is also the possibility of self-selection

[9]These groups were combined due to low numbers in their intersection with women and non-binary students.

Table 4.1: Regression models predicting computing self-efficacy.

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Substantial computing experience	0.36*** (0.04)	0.33*** (0.04)	0.33*** (0.04)	0.25*** (0.02)	0.25*** (0.02)	0.24*** (0.02)
Some computing experience	0.08* (0.04)	0.06+ (0.04)	0.06 (0.04)	0.01 (0.04)	0.01 (0.04)	0.01 (0.04)
Woman		−0.23*** (0.02)	−0.20*** (0.02)	−0.25*** (0.02)	−0.24*** (0.02)	−0.22*** (0.02)
Non-binary		−0.22*** (0.03)	−0.22*** (0.03)	−0.27*** (0.05)	−0.20*** (0.06)	−0.22*** (0.06)
Black			−0.17*** (0.03)	−0.18*** (0.04)	−0.19*** (0.05)	−0.19*** (0.05)
Native			−0.15** (0.04)	−0.18* (0.09)	−0.18* (0.09)	−0.20* (0.10)
Hispanic			−0.07** (0.02)	−0.09*** (0.02)	−0.10*** (0.03)	−0.10*** (0.02)
Asian			−0.15*** (0.02)	−0.16*** (0.02)	−0.15*** (0.02)	−0.15*** (0.02)
CS required for all				−0.00 (0.04)	−0.00 (0.04)	−0.00 (0.04)
CS required for some				−0.05 (0.03)	−0.05 (0.03)	−0.04 (0.03)
HUREG × woman					0.02 (0.04)	0.01 (0.04)
Asian × woman					−0.03 (0.04)	−0.04 (0.03)
HUREG × non-binary					0.04 (0.12)	−0.00 (0.12)
Asian × non-binary					−0.20* (0.10)	−0.22+ (0.11)
Major (fixed effect)						x
Data collection year (fixed effect)						x
Number of respondents	31,425	28,899	22,796	11,300	11,242	10,611
Adjusted R-squared	0.04	0.06	0.08	0.08	0.08	0.08

Note: Standard errors clustered by institution are given in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. An x indicates that the variable was included as a fixed effect. HUREG refers to historically underrepresented racial/ethnic groups and has a value of 1 for students who identify as Black, Native, and/or Hispanic. Each coefficient represents a change in self-efficacy along a 5-point scale.

in students who opt to take a voluntary survey about their experiences in computing. Our dataset also contains duplicate participants, who cannot be tracked from year to year, and uneven subgroup sizes. Social norms may impact students' responses; for example, women self-assess themselves as lower in computing ability than men with the same grades [25]. Despite these limitations, we believe our work points to patterns that are likely to affect students' persistence and performance, and merit further investigation.

4.6 RESULTS

Our regression model findings are shown in Table 4.1.

4.6.1 Prior Computing Experience

In line with our hypotheses, our results confirm previous work that prior computing experience positively predicts computing self-efficacy, even when controlling for students' gender, race, ethnicity, and major. Although having *substantial* prior experience remained a statistically significant predictor of self-efficacy in all models, having *some* prior experience was no longer a significant predictor once our model controlled for gender, race, and ethnicity. We note that only 4.1% of our respondents were categorized as having *some* prior experience. In our model, *some* prior experience was defined as participation in computing workshops, student groups, or projects, whereas *substantial* prior experience referred to completing a CS course or independently learning a programming language. Therefore, our findings suggest that limited exposure to computing concepts prior to attending college is not sufficient to significantly predict students' computing self-efficacy.

4.6.2 Gender Identity

Consistent with previous research, in all models, identifying as a woman had a statistically significant and negative relationship with students' computing self-efficacy. We know of no prior research investigating the computing self-efficacy of non-binary individuals. We found that the relationship between identifying as non-binary and computing self-efficacy was negative and statistically significant. The coefficients indicating identifying as non-binary had similar magnitudes to those indicating identifying as a woman.

4.6.3 Race and Ethnicity

Identifying as Black, Native, and/or Hispanic predicted significantly lower computing self-efficacy, supporting previous work. However, contrary to our hypotheses, identifying as Asian also predicted significantly lower computing self-efficacy in all models. This finding suggests that underrepresentation and a lack of same-race models is not a sufficient explanation for the relationship between race/ethnicity and computing self-efficacy, as Asian students are not typically considered underrepresented in computing in the US.

In addition, we note that the coefficients corresponding to Hispanic students were smaller in magnitude than those corresponding to Black, Native, and Asian students. Although we did not have hypotheses regarding the relative magnitude of the coefficients, we believe this to be a surprising result worth further investigation.

4.6.4 Institutional CS Course Requirement

Our results indicated that an institution's CS course requirement was not a significant predictor of students' self-efficacy. In addition, the coefficients for both variables, although not statistically significant, had the opposite sign from our hypotheses; we had hypothesized positive coefficients for students attending institutions that require more of its students to take a computing course.

4.6.5 Intersections of Gender and Race/Ethnicity

Only one variable representing intersectional groups, the one representing Asian non-binary students, was statistically significant in any of our models. The negative coefficient indicates that identifying as a non-binary Asian person predicts lower computing self-efficacy, *in addition* to the already demonstrated negative relationships of identifying as Asian and as non-binary separately.

4.7 DISCUSSION AND FUTURE WORK

Our work shows that prior computing experience positively predicts computing self-efficacy, while identifying as Black, Native, Hispanic, Asian, non-binary, and/or a woman all negatively predict computing self-efficacy. An important contribution of our work is the investigation of groups who are not typically represented in large enough numbers for inferential analysis, such as students who identify as non-binary or with multiple marginalized identities.

The positive relationships between prior computing experience and computing self-efficacy aligns with previous research [40, 41, 42, 43] and our hypotheses. Specifically, having *substantial* computing experience (e.g., taking AP CS or independently learning a programming language) predicts an increase in students' computing self-efficacy. Notably, however, having only *some* computing experience did not significantly predict self-efficacy after accounting for gender, race, and ethnicity. This is a concern as not all students have access to substantial computing opportunities due to existing inequities in the K-12 educational system [2, 19, 91, 92]. These findings underscore the importance of equity and access in computing, particularly if we hope to positively impact students' self-efficacy in computing and, consequently, their motivation and persistence [36, 38, 39].

Non-binary identities predicting lower computing self-efficacy confirms our hypothesis, but further work is required to understand what causes this relationship. This pattern may be related to the masculine culture of CS [21, 103] or to the discrimination and bias those with non-binary identities may face in STEM fields [112] and in daily life [113]. In addition, our models indicate that identifying as *both* non-binary and Asian negatively predicts computing self-efficacy, demonstrating that we cannot estimate the change in self-efficacy for non-binary Asian students by summing the coefficients for Asian students and for non-binary students. Future work may seek to investigate the unique experiences of these students in order to better understand and address this effect.

Our findings regarding race and ethnicity confirm some of our hypotheses and refute others: identifying as Black, Native, and/or Hispanic negatively predicts computing self-efficacy, as expected, but so does identifying as Asian, counter to what we had hypothesized. In the US, Asian students are not typically considered underrepresented in computing, meaning that a lack of same-race models may not be the explanation for their lower self-efficacy. However, we note that Asian students may identify with a large number of national and ethnic groups, some of which may be underrepresented [114]. Additionally, representation does not preclude Asian students from experiencing race-based discrimination and bias in computing spaces, which may contribute to their lower computing self-efficacy. Further investigation into the experiences of Asian students ought to incorporate a greater level of nuance in their national and ethnic origins than our data permitted, as well as explore their experiences of discrimination and bias.

Our findings regarding CS course requirements were not significant, but the negative coefficients suggest that requiring CS courses does *not* predict higher computing self-efficacy, contrary to our hypotheses. Additional exploratory work with students attending such institutions may explain why. One possible explanation is that at institutions where a CS course is not required, only students with relatively high computing self-efficacy would opt to take one.

While the results of our intersectional analyses were not all statistically significant, we encourage additional research into the experiences of students who identify as both members of HUREGs and as women/non-binary.

4.8 CONCLUSION

Using a large, recent, and multi-institutional dataset, this study investigated the extent to which prior computing experience, gender, race, ethnicity, and their institution's requirement of CS courses for non-CS majors predict the computing self-efficacy of undergraduate students. Self-efficacy is an important outcome for broadening participation in computing as it predicts improved academic performance, motivation, and persistence [36, 38, 39, 52]. We found that substantial prior experience with computing positively predicts computing self-efficacy. However, identifying as Black, Native, Hispanic, Asian, non-binary, and/or a woman had a significant, negative relationship with computing self-efficacy. Our results point to patterns in student self-efficacy that require further investigation to understand why these trends exist and how to address them. These findings are important for better understanding and improving the experiences of students identifying with these identities in computing spaces in order to promote their retention and well-being. Although our results were not all significant, we encourage further research into the unique experiences of individuals with multiple marginalized identities, as they may differ from their peers in important ways, and into the role institutional policies may play in computing self-efficacy and other student outcomes.

CHAPTER 5: INSTRUCTIONAL TRANSPARENCY

5.1 INTRODUCTION

Instructional transparency, or simply *transparency*, makes a course's learning goals, evaluation criteria, and path to success clear to students [49, 50]. Practices to support instructional transparency, known as transparent teaching practices, were developed by Winkelmes et al. [49] with the goal of improving equity by making expectations explicit and providing all students with clear guidelines for success in the course. Using such practices in CS courses may also contribute to improving equity by supporting the success of students from historically underrepresented groups: these students are likely to have less exposure to computing coming into college [5] and therefore may benefit most from making the path to success transparent.

Scholars have found that students' perception of transparency may predict two important outcomes: students' confidence of succeeding in the course [49, 51], which is related to *self-efficacy*, and students' *sense of belonging* [49, 51]. Self-efficacy, a measure of students' confidence that they can achieve their goals in computing, is itself a predictor of students' motivation, performance, and persistence [38, 39, 52]. Sense of belonging, the extent to which students feel that they are legitimate members of a field, has similar relationships with motivation and persistence [53, 54]. Instructional transparency could improve students' self-efficacy and sense of belonging in computing, and, correspondingly, contribute to students' persistence in computer science (CS).

By improving student persistence in computing, instructional transparency may play a role in efforts to broaden participation in computing. In the US, people who identify as women, Black/African American, Hispanic/Latina/o/x/e, Native American, Native Alaskan, Native Hawiian, Pacific Islanders, and/or having a disability are underrepresented in computing [1, 115, 116]. We refer to these groups as historically underrepresented groups (HUGs) in computing. Individuals from HUGs are underrepresented in computing at all levels [1, 115, 116], including in the K-12 educational system [5, 8]. Students from HUGs face barriers in accessing computing instruction and resources in K-12 [117, 118, 119, 120] and may, as a result, arrive in college less familiar with strategies for success in computing courses. Transparency in college CS courses aiming to make instructions and expectations clear to all students may, therefore, particularly benefit students from HUGs and encourage their persistence in computing. Because prior experience, knowledge, and background differ between students, it is important to understand whether students vary in their perception of instructional transparency. Therefore, we aim to understand whether there are group differences in how students perceive and benefit from instructional transparency. We are additionally interested in understanding whether perceiving instructional transparency can positively influence students' self-efficacy and sense of belonging and, therefore, contribute to the persistence of students from HUGs in computing.

Although existing evidence suggests a relationship between instructional transparency and both self-efficacy and sense of belonging, there remain gaps in our knowledge that this study aims to fill. First, there has not yet been a theoretical articulation of *how* transparency may contribute to students' self-efficacy and sense of belonging. Second, prior work has not investigated whether group differences exist in either students' perception of transparency or in any benefits that may result from greater instructional transparency. Third, the empirical relationships between transparency and self-efficacy or sense of belonging have not been examined at a large scale, across institutional contexts, while controlling for other variables known to impact self-efficacy and sense of belonging. Our work aims to address these gaps by answering the following research questions:

RQ0: What are the theoretical mechanisms through which instructional transparency may influence (a) self-efficacy and (b) sense of belonging?

RQ1: Does the extent to which students perceive instructional transparency in their CS courses differ by their gender, race, ethnicity, disability status, first-generation student status, and prior computing experience?

RQ2: To what extent is students' computing self-efficacy predicted by their perception of instructional transparency in their CS courses?

RQ3: To what extent is students' sense of belonging in computing predicted by their perception of instructional transparency in their CS courses?

To address RQ0, we provide a theoretical argument connecting a seminal set of transparent teaching practices, known as the purpose-task-criteria template (described in Section 5.2), to self-efficacy and sense of belonging in Section 5.4. By making this argument, we seek to illuminate *possible* pathways that may explain the influence of instructional transparency on self-efficacy and sense of belonging, but do not aim to articulate *all* ways in which this influence may occur or argue that all of these possible pathways will be equally relevant for all students. We include this argument because it provides theoretical support for the potential benefits of transparent teaching practices and may, therefore, further encourage adoption of these practices. To answer RQs 1-3, we conducted linear regressions using data acquired from a multi-institutional survey of undergraduate students enrolled in computing courses. Using 11,046 undergraduate student survey responses from 203 institutions, we created three regression models to answer research questions 1, 2, and 3, respectively.

We found that there are group differences in students' perception of instructional transparency in their CS courses. Students who identify as women, first-generation college students (i.e., their parents or guardians did not attain a college degree), and/or disabled reported perceiving less instructional transparency than their peers. We also found that perceiving more transparency positively predicts students' self-efficacy and sense of belonging in computing while controlling for important confounding variables, such as prior CS experience. We further demonstrated that this relationship is different for certain groups of students: first, for Black students and first-generation college students, transparency has a *larger* positive impact on their self-efficacy, and second, for Hispanic students, transparency has a *smaller* positive impact on their sense of belonging. We note that our results are limited by the lack of access to students' performance data and information regarding how transparent teaching practices were implemented, as addressed in Section 5.9. However, we are able to demonstrate patterns in students' perception of transparency, self-efficacy, and sense of belonging that highlight the value of transparent teaching practices as an intervention.

Our work constitutes one of the first empirical, multi-institutional investigations of the perceptions and benefits of transparency in CS classrooms that focuses on group differences. Our work also includes a theoretical articulation of the mechanisms through which transparent teaching practices may influence students' self-efficacy and sense of belonging in computing. Taken together, our empirical findings and theoretical argument provide important evidence for the benefits of instructional transparency in CS courses, particularly as it relates to improving equity in computing.

5.2 BACKGROUND: TRANSPARENT TEACHING PRACTICES

For our theoretical argument addressing RQ0, we use the *purpose-task-criteria* template, which is a common example of transparent teaching practices (e.g., [49, 51, 121, 122]). The template was created as part of the Transparency in Learning and Teaching in Higher Education (TILT Higher Ed) project by Winkelmes et al. [49][10], which aims to use transparent teaching practices to promote equity in higher education [123]. This template describes three components (the purpose, task, and criteria) that instructors can include in their

[10] We note that neither we nor our study are affiliated with the TILT Higher Ed project.

assignments as well as discuss with students to promote instructional transparency.

Transparency in *purpose* refers to sharing with students how the assignment supports the course learning goals. Specifically, Winkelmes et al. [49] recommend articulating which skills are being developed and what particular knowledge is learned by completing the assignment. Winkelmes [50] argues that sharing the purpose of an assignment before students begin work allows them to better understand how the assignment serves their learning goals and plan their approach.

Transparency of the *task* refers to providing students with clear instructions regarding "what to do" and "how to do it" [49, p. 33]. Winkelmes [50] notes that clear instructions help students to monitor their progress. In addition, Winkelmes [50] contends that students likely spend less time having to understand the task and more time learning, as they are told ahead of time what they are supposed to do.

Finally, providing transparent *criteria* entails giving students a detailed rubric and annotated examples of prior work for each assignment [49]. Winkelmes [50] argues that instructors should discuss the examples of prior work with students to determine how well these examples meet the criteria described in the rubric. According to Winkelmes [50], such a discussion aims to give all students the same starting knowledge to attempt the assignment and also demonstrates to them multiple ways of successfully fulfilling the criteria.

Researchers have demonstrated several benefits of implementing these transparent teaching practices, which we describe further in Section 5.3.1 below.

5.3 PREVIOUS RESEARCH

5.3.1 Transparent Teaching Research

The purpose-task-criteria template was developed by Winkelmes et al. [49] as part of a teaching intervention with the goal of improving equity in higher education. In this intervention, faculty at seven institutions used the template to alter assignments given to students in one section of a course [49]. These faculty additionally agreed to discuss the purpose, task, and criteria for each of the altered assignments in class prior to students working on the assignment [49]. Another section of the same course served as a control, where students received unaltered assignments, but Winkelmes et al. [49] noted that the faculty participants "struggled to keep the intervention cleanly out of their control courses" due to its perceived effectiveness. Their analyses reflect this concern: instead of analyzing differences between the control and treatment courses, the authors separated students into those who reported perceiving higher-than-average transparency in their course and those who perceived lower-than-average transparency in their course [49]. When comparing students who perceived high vs. low transparency, the results showed that students reporting high-transparency also reported a higher sense of belonging and higher academic confidence [49]. Winkelmes et al. [49] reported that students who were first-generation college students, not white, and/or from lower socioeconomic backgrounds (i.e., in the lowest quartile for income) additionally benefitted from instructional transparency. However, they did not conduct statistical comparisons *between* student groups in order to investigate whether transparency has a differential impact for students from HUGs. Our work aims to fill this gap by investigating the impact of students' identities on their perception of transparency, self-efficacy, and sense of belonging.

Other scholars have also found evidence that using the purpose-task-criteria template benefits students. For example, Peplow et al. [121] found that using the purpose-task-criteria template led to students asking fewer questions about assignments, reporting less test anxiety, and reporting improved personal organization skills in a variety of courses at their institution. Ou [51] also implemented the purpose-task-criteria template and found an improvement in students' ability to learn independently, their confidence in their success in the field (which is related to self-efficacy), and their sense of belonging. In another study, Howard et al. [122] found that assignments designed using the purpose-task-criteria template can help mitigate typical learning challenges associated with online learning. In particular, they found that learning outcomes for students in online courses that implemented the purpose-task-criteria template were comparable to those of students taking in-person courses with traditional teaching methods [122]. Howard et al. [122] highlight the value of this result for equity efforts, as students from HUGs tend to achieve lower grades in online courses [124, 125, 126].

Although such relationships have been identified empirically, previous work has not provided a theoretical explanation for the relationship between instructional transparency and self-efficacy or sense of belonging. We provide a possible explanation in Section 5.4.

5.3.2 Hidden Curriculum

Transparency may also address the need to make visible what scholars have called the *hidden curriculum.* Giroux and Penna [127] defined the hidden curriculum as the "unstated norms, values, and beliefs" that are conveyed to students but are not part of the official curricula. Understanding these norms and values provides "insider knowledge" that may affect students' ability to participate in the "formal" curriculum [128]. Therefore, it is important to make the hidden curriculum transparent to all students [128, 129, 130]; indeed, it has been shown that making expectations explicit and accessible to all students is beneficial to their learning [130, 131].

Students from HUGs may particularly benefit from revealing the hidden curriculum in computing. In K-12 education, students from HUGs experience barriers in accessing computing instruction and resources [117, 118, 119, 120], and so may arrive in college less familiar with the hidden curriculum in computing than their majority peers. Greater instructional transparency may be one way to reveal the hidden curriculum, as transparency, by design, aims to make the path to success in a course clear to all students.

5.3.3 Self-Efficacy

One of the outcomes of interest to this work is students' self-efficacy. In this section, we review prior research on self-efficacy and its implications for our study.

Self-efficacy is a measure of an individual's belief that they can act in ways that will achieve their goals [34]. According to the original conceptualization of self-efficacy by Bandura [34], an individual's perceived self-efficacy can impact outcomes and behaviors such as the effort they put into a task and their persistence in the face of challenges [34, 35, 132]. Computing self-efficacy has been shown to have a positive relationship with student retention and motivation in CS courses [37, 38, 39, 52]. The positive relationships between these outcomes and self-efficacy inspires our investigation into whether perceiving transparency can improve students' self-efficacy.

Prior research has also demonstrated that aspects of students' identities and experiences have a relationship with their computing self-efficacy. In particular, greater prior experience with CS predicts higher computing self-efficacy [33, 40, 43]. On the other hand, computing self-efficacy is negatively predicted by identifying as a woman [33, 39, 46], non-binary [33], a first-generation college student [39], and/or a member of an underrepresented racial/ethnic group [33, 46]. Previous work also suggests a relationship between self-efficacy and disability status[11]: scholars have found that students with physical [133] and learning disabilities [134] may have different perceptions of their self-efficacy than their peers without disabilities. We include all of the above aspects of students' identities in our models predicting computing self-efficacy, as described in Section 5.8.2.

[11] We note that people with disabilities encompass a wide variety of individual experiences and needs. As such, groups of disabled people should be expected to be heterogeneous. We provide a description of we categorized students as having a disability in Section 5.7.2.

5.3.4 Sense of Belonging

The second outcome of interest to this study is students' sense of belonging in computing. We review relevant literature and its implications for our work below.

Sense of belonging for undergraduate students is defined by Strayhorn [135] as their "perceived social support on campus... the experience of mattering or feeling cared about, accepted, respected, valued by, and important to the group" [135, p. 28-29]. Positive relationships between sense of belonging and positive outcomes have been demonstrated in the field of computing. For example, prior work has demonstrated a positive relationship between belonging and students' perceived CS ability [136], interest in pursuing computing [53], and performance in CS courses [54]. As a result, understanding whether transparency positively predicts students' sense of belonging in CS may be a valuable insight for promoting student success.

In higher education settings, students' sense of belonging has been shown to be related to aspects of their identities, such as gender, race, and disability status, and their prior experiences. Identifying as a woman or as a member of an underrepresented racial/ethnic group predicts a lower sense of belonging [137, 138], as is identifying as a first-generation college student [139, 140]. Similar patterns have been found for students with disabilities: Blaser and Ladner [141] found that students with disabilities reported "feeling like an outsider" more than their male peers without disabilities, indicating that disability status may also predict sense of belonging. Finally, in the field of computing, scholars have identified that prior experience in CS may contribute to students' sense of belonging in the field: Veilleux et al. [136] found that students' sense of belonging was closely related to their perception of their own ability, which is likely to be tied to their level of experience in computing. We include all of the above aspects of students' identities in our models predicting sense of belonging, described in Section 5.8.2.

5.4 RQ0: THEORETICAL FRAMEWORK

In this section, to address RQ0, we use existing theory from educational psychology to argue that transparency, operationalized as the purpose-task-criteria template, may positively influence (a) students' self-efficacy and (b) sense of belonging in computing. Existing work has not focused on providing a theoretical explanation for the benefits of instructional transparency. Thus, an important contribution of our work is an articulation of the mechanisms that connect instructional transparency to students' self-efficacy and sense of belonging. Our goal is to illustrate *possible* pathways that may explain the influence of instructional transparency on self-efficacy and sense of belonging, but we do not aim to articulate *all* ways in which this influence may occur or argue that all of these possible pathways will be equally relevant for all students. We include this argument because it provides theoretical support for the benefits of transparent teaching practices and may, therefore, further encourage adoption of these practices. We note that there may be other benefits to instructional transparency, such as decreasing students' cognitive load; however, our work focuses on self-efficacy and sense of belonging due to our focus on students' wellbeing.

5.4.1 Self-Efficacy

As introduced in Section 5.3.3, *self-efficacy* is a measure of an individual's belief that they can act in ways that will achieve their goals [34]. For students enrolled in CS courses, their goals might include mastering CS concepts or passing their computing courses. Our work aims to investigate whether there is a relationship between students' perception of transparency and their computing self-efficacy; below, we provide a theoretical argument for why this may be the case.

Bandura [34] detailed four possible sources of information individuals draw upon when determining their self-efficacy in a particular domain. One of these sources, "verbal persuasion" from others [34], does not feature in our argument. The first source of self-efficacy information relevant to our work comes from the individual's "performance accomplishments", i.e., succeeding at tasks in the domain of interest [34]. Bandura [34] argued that succeeding at a task increases an individual's expectations of continued success, while failing at tasks lowers their expectations of success in the future. Initial success is especially important as it can build sufficient self-efficacy for later failures to have a diminished negative impact [34]. The second source is "vicarious experience", which is the observation of peers succeeding at tasks in the domain [34]. By observing that others can persevere and be successful, an individual can believe that this success is possible for themself too [34]. The third source of self-efficacy information is described by Bandura [34] as "emotional arousal" and refers to the level of stress and anxiety felt by the individual during the tasks. Negative reactions such as excessive stress and anxiety may lead an individual to avoid the tasks and/or perform worse; on the other hand, Bandura [34] argues, lower stress and anxiety during a task correspond to an increase in self-efficacy.

We argue that a relationship between transparency and self-efficacy is supported by Bandura's theory of self-efficacy. Each element of the purpose-task-criteria template can act as one or more of the above sources of self-efficacy information and, therefore, may bolster self-efficacy. These relationships are pictured in Figure 5.1. First, knowing the explicit *pur-*

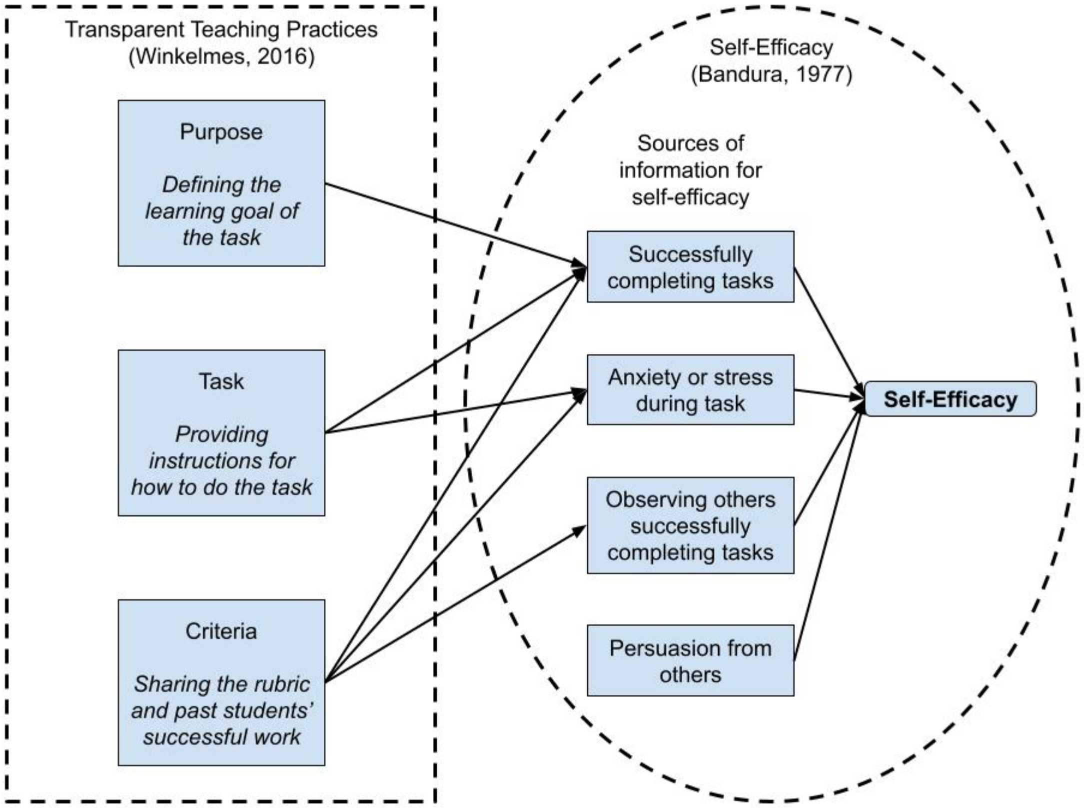

Figure 5.1: A graphical representation of our theoretical argument connecting the purpose-task-criteria template [49] to self-efficacy [34].

pose of each assignment may motivate students to put in the effort required to eventually complete the task. In particular, knowing how the task pertains to their overall goals in the course may increase the value students see in it, which may increase their motivation to complete the task. This relationship between perceived value and motivation is known as expectancy-value theory [142]. Increased motivation is likely to result in greater effort and thus a greater likelihood of achieving a "performance accomplishment" [34]. Second, providing detailed instructions on how to complete the *task* may affect self-efficacy by increasing the likelihood that students accomplish the assignment successfully and gain "performance accomplishments" [34], as well as by diminishing the stress and anxiety caused by not knowing what to do. Third, sharing the evaluation *criteria* may contribute to students' self-efficacy in three ways: sharing criteria may clarify to students what the standards are, making it easier for students to achieve them; the ability to use these criteria to self-assess may diminish students' uncertainty and therefore their anxiety; and showing examples of past students' work may act as "vicarious experience" of others' success [34]. In summary, based on the ways the purpose-task-criteria template contributes to the sources of self-efficacy information described by Bandura [34], we hypothesize that transparency may bolster students' self-efficacy.

5.4.2 Sense of Belonging

As discussed in Section 5.3.4, for undergraduate students, *sense of belonging* is defined by Strayhorn [135] as their "perceived social support on campus... the experience of mattering or feeling cared about, accepted, respected, valued by, and important to the group" [135, p. 28-29]. Belonging is not only important for students' well-being as a basic human need [135, p. 31], but also because of the role it plays in fostering positive academic outcomes [53, 54, 136]. Therefore, it is valuable to investigate what educators can do to bolster students' sense of belonging in computing, and we provide a theoretical argument below that instructional transparency may do so.

Vaccaro and colleagues have developed a theoretical framework that identifies four factors that contribute to sense of belonging [143, 144, 145, 146, 147, 148]. Two of the factors in their model, the campus environment and students' community involvement [148], are not directly relevant to our study; we describe the relevant factors below. The first factor relevant to our work is described by Vaccaro and Newman [148] as "academic success and/or mastery of the student role". This factor was described by students as including earning good grades as well as being treated as a "legitimate student" and "blending in", rather than being treated as different from their peers [148]. The second factor relevant to our work is the relationships students have with their peers and educators, which includes receiving "task-related" and affective support [148]. Vaccaro and Newman [148] noted that it is particularly valuable for students from HUGs to be able to build relationships that felt authentic and supportive.

We use the theoretical framework developed by Vaccaro and Newman [148] to argue that instructional transparency may influence students' sense of belonging. Specifically, we contend that instruction using the purpose-task-criteria template impacts two of the sources of sense of belonging discussed above: students' perceived mastery of the student role and their relationship with peers and educators. We illustrate these arguments in Figure 5.2. Firstly, by explicitly stating the *purpose* of each assignment in a course, instructors are providing students with clarity and an understanding of why the course is designed the way it is, potentially improving students' relationships with and trust in their instructors (i.e., educators). Secondly, detailed *task* instructions may make it easier for students to successfully complete the task, contributing to their "academic success and/or mastery of the student role". These instructions may also be considered "task-related" support, which Vaccaro and Newman [148] describe as a component of students' relationships with their educators.

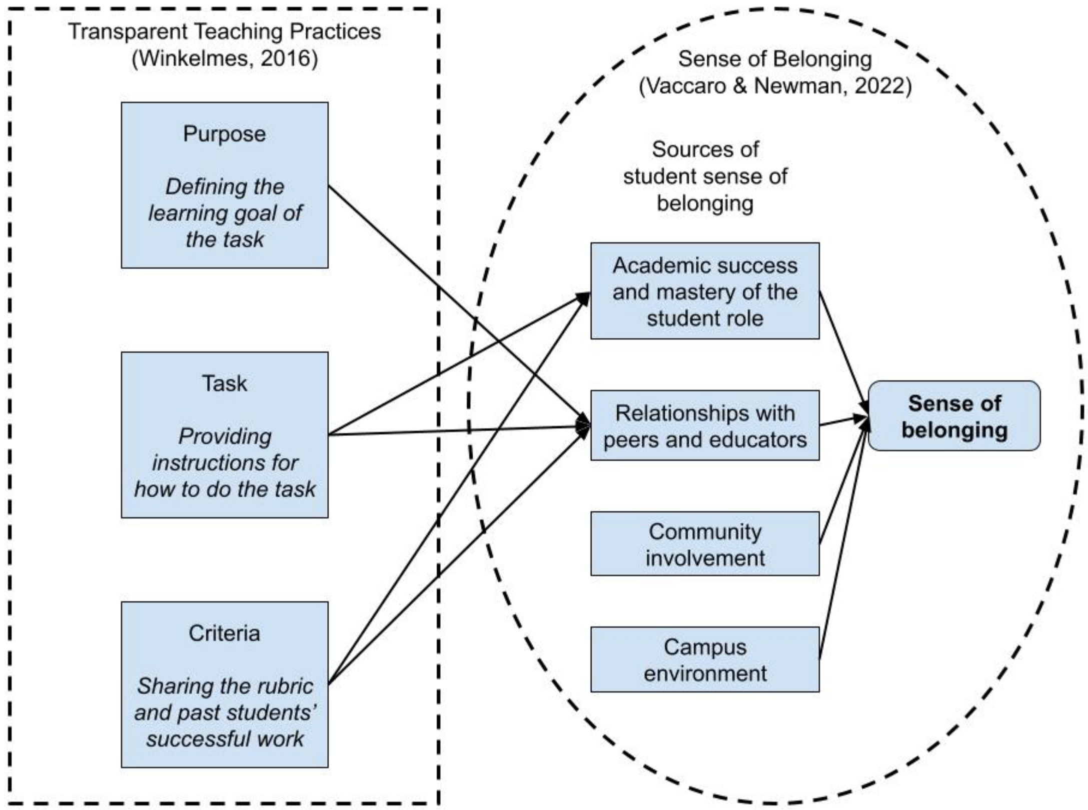

Figure 5.2: A graphical representation of our theoretical argument connecting the purpose-task-criteria template [49] to sense of belonging [148].

Thirdly, providing the assignment's *criteria* and examples of past work may increase the likelihood of students succeeding at the task, contributing to their "academic success" [148]. In addition, sharing the evaluation criteria may also enhance students' belief that their instructors are preparing them for success, contributing to a better relationship with their educators. Based upon these relationships between the purpose-task-criteria template and the sources of sense of belonging theorized by Vaccaro and Newman [148], we hypothesize that increased perception of instructional transparency will correspond with an increase in students' sense of belonging.

5.5 RESEARCH QUESTIONS AND HYPOTHESES

Building upon our theoretical argument above in response to RQ0, we address the following remaining research questions:

RQ1: Does the extent to which students **perceive instructional transparency in their**

CS courses differ by their gender, race, ethnicity, disability status, first-generation student status, and prior computing experience?

RQ2: To what extent is students' **computing self-efficacy** predicted by their perception of instructional transparency in their CS courses?

RQ3: To what extent is students' **sense of belonging in computing** predicted by their perception of instructional transparency in their CS courses?

Our first research question aims to explore whether any of the listed factors predict perceptions of instructional transparency. We hypothesize that prior computing experience may lead to students perceiving greater transparency in their CS courses, because familiarity with computing concepts may lead them to readily perceive learning goals and task requirements that their less experienced peers do not.

For our second and third research questions, we hypothesize a positive relationship between perception of instructional transparency and each outcome variable investigated, i.e., self-efficacy and sense of belonging in computing. We also hypothesize that students' identities will moderate the relationships between their perception of transparency and self-efficacy or sense of belonging; that is to say, that students from HUGs will have an additionally positive relationship between their perception of transparency and their self-efficacy or sense of belonging in computing. This hypothesis stems from the documented barriers students from HUGs encounter in participating in computing [5, 117, 118, 119, 120]: we hypothesize that students from HUGs will have had fewer opportunities to learn the hidden curriculum in computing, and will, as a result, benefit more from instructional transparency.

Our research questions, hypotheses, and methods were pre-registered [149].

5.6 DATA

The data used in this study was provided by the Computing Research Association from their Computing Education Research Pipeline's undergraduate Data Buddies survey [108]. The survey is provided to computing departments in the US and Canada, who share the survey with students in their computing courses. The response rate to the survey is below 20% at most participating institutions [108]. Data for this study is drawn from the surveys administered in the academic years ending in 2021 and 2022, encompassing 11,046 student respondents from 203 institutions. Because not all students responded to all relevant survey items, our models for self-efficacy and sense of belonging use different subsets of the data used in the larger model for instructional transparency. In particular, our model for self-efficacy uses data from 10,973 students, and our model for sense of belonging uses data from 10,982 students. While some students may have responded to the survey in both years, our data does not permit tracking individual students.

5.7 MEASURES

Below, we describe how the measures used in our models were created. Correlations between these measures are provided in Appendix H.

5.7.1 Perception of Instructional Transparency, Self-Efficacy, and Sense of Belonging

We used the procedure described below to create three measures, one for each of our three student outcomes of interest: perception of transparency, self-efficacy, and sense of belonging. Each measure was constructed using the relevant survey items, provided below in Sections 5.7.1, 5.7.1, and 5.7.1.

To verify that each set of items measured a related construct, we conducted Exploratory Factor Analyses. Based on a low factor loading, we excluded one of the items under perception of instructional transparency, "My instructors provide students with annotated examples of past students' work". All other items had factor loadings above 0.4 in their respective measure, which is a typical cutoff point for inclusion in a factor [150], and were therefore included in calculating that particular measure. (Factor loadings for all items are provided in Appendix G.) Cronbach's alpha for each set of the resulting items was above 0.7, which is considered acceptable reliability [151].

Students were excluded from a model if they did not respond to all of the items used in that particular model. Each student's value for each measure was then calculated by taking an average of a student's rating (i.e., a value from 1 to 5) to all the items pertaining to that factor. For ease of interpreting results, the values were then standardized to have a mean of 0 and a standard deviation of 1.

Perception of Instructional Transparency. To calculate students' perception of instructional transparency, we used the following question: *Think about the computing courses you are taking. Please assess the following statements and indicate how many of your computing courses they apply to.* For each item under this question, students could select one of the following options: (1) None of my courses; (2) Less than half of my courses; (3) About half of my courses; (4) More than half of my courses; (5) All of my courses. The items were:

- In these courses, I know the purpose of each assignment.
- My instructors identify a specific learning goal for each assignment.
- Each assignment includes a detailed set of instructions for completing it.
- I know how my work would be evaluated.

Prior to standardization, students' mean perception of transparency was 4.1 (out of 5) with a standard deviation of 0.83. Cronbach's alpha for these items was 0.71.

Computing Self-Efficacy. Students' computing self-efficacy was measured using their agreement with the three items given below. Students could indicate one of the following Likert scale options: (1) Strongly disagree; (2) Somewhat disagree; (3) Neither disagree nor agree; (4) Somewhat agree; (5) Strongly agree.

- I am confident that I can learn the foundations and concepts of computing.
- I am confident that I can quickly learn a new programming language on my own.
- I am confident that I can pass my computing courses.

Students' mean computing self-efficacy prior to standardization was 4.2 out of 5 with a standard deviation of 0.79. These items had a Cronbach's alpha of 0.77.

Sense of Belonging in Computing. To measure students' sense of belonging in computing, we used their agreement to to the following items on the same 1-5 Likert scale as for the self-efficacy items:

- I feel like I belong in computing.
- I feel like an outsider in computing. *(Reverse-coded)*
- I feel welcomed in computing.

Prior to standardization, students' mean sense of belonging was 3.6 out of 5 with a standard deviation of 0.94. Cronbach's alpha for this measure was 0.79.

5.7.2 Student Demographics

For details regarding how student responses were categorized for each category below, please refer to our pre-registration [149]. We provide details of how many students identified with each of these groups in Table 5.1.

Gender. For each student, the survey data included a binary variable indicating whether the respondent identified as either a man or a woman, and an additional variable indicating whether they identified as gender non-binary. Unfortunately, our participant pool included few non-binary students ($n = 68$); since our analysis would be underpowered, we did not include students who indicated only a non-binary gender identity in our models.

Race and Ethnicity. The data provided to us indicated each student identifying as one or more of the following racial/ethnic identities: Asian, Black, Native, Pacific Islander, Hispanic, Arab/Middle Eastern, and white. Due to low numbers, we combined students who identified as Native and those who identified as Pacific Islanders into one group, "Native". In accordance with the US Census [110], we combined students who identified as Arab/Middle Eastern with those who identified as white. Each participant may be represented in multiple of the above groups.

Disability Status. Our dataset included a variable where the respondent could indicate identifying as having one of 14 disabilities, more than one disability, or no disability. The 14 options included, for example, cognitive disorders (e.g., attention deficit disorder), learning disabilities, and sensory impairments (e.g., visual disability). Because we did not have specific hypotheses related to the kind of disability students indicated having, we grouped all students who indicated having one or more disabilities as disabled.

First-generation Status. The dataset also included a variable indicating whether students identified as the first generation in their family to attain a college degree.

5.7.3 CS Experience & Major

In response to the question *Which of the following experiences did you have prior to entering an undergraduate program? Select all that apply*, respondents were able to indicate having participated in computing experiences prior to college. Based on the activities they selected, our research team categorized students as having *substantial* prior computing experience, *some* prior computing experience, or *no* prior computing experience in high school (HS). Experiences that were categorized as substantial experience included taking an AP CS course and learning a programming language. Some experience included participating in computing-related student groups. For a complete list of possible responses and how they were categorized, see our pre-registration [149].

Table 5.1: Descriptive statistics of participants in our dataset.

	Number of students	**Percent**
Women	3,963	36%
Men	7,083	64%
First-generation	2,842	26%
Disabled	2,206	20%
Black	892	8%
Native	194	2%
Hispanic	1,293	12%
Asian	4,585	42%
White	5,663	51%
CS majors	10,029	91%
Intro CS completed	9,057	82%
Within 2 years of graduation	6,068	55%
Substantial HS CS experience	7,463	68%
Some HS CS experience	608	6%
All participants	11,046	100%

Note: These numbers represent participants included in our largest model, Model 1. The participants included in the other models are subsets of this group.

Students indicated their major(s) from a list of options. We categorized each of these options as either a CS major, a non-CS STEM major, or a non-STEM major. For a complete list of possible major options and their categorization, please see our pre-registration [149].

Descriptive statistics of students' CS experience categorization and their majors are provided in Table 5.1.

5.8 METHODS

We answered each of our research questions using linear regression models.

5.8.1 RQ1: Predictors of Perception of Instructional Transparency

RQ1 aimed to examine whether students' perception of instructional transparency in their CS courses is predicted by their identity or prior experience. To this end, the equation for Model 1 was as follows:

$$
\begin{aligned}
transparency_s = \; & \beta_0 + \beta_1 woman_s + \beta_2 firstGen_s + \beta_3 disabled_s \\
+ \; & \beta_4 black_s + \beta_5 native_s + \beta_6 hispanic_s + \beta_7 asian_s \\
+ \; & \beta_8 substantialExp_s + \beta_9 someExp_s \\
+ \; & \gamma_y + \mu + \iota_s + \epsilon_s
\end{aligned}
\tag{5.1}
$$

In this model, $transparency_s$ is the measure of a student's perception of instructional transparency in their CS courses. $woman_s$ indicates whether a student identifies as a woman, with students who identify as men being the reference group. $firstGen_s$ represents whether the student indicated being a first-generation college student (compared to continuing-generation students), and $disabled_s$ represents whether they indicated having a disability (compared to those who indicated not having a disability). $black_s$, $native_s$, $hispanic_s$, and $asian_s$ capture whether the student identifies with each of those racial and ethnic identities. Students who identify as white are the reference group. As students may identify with multiple racial and ethnic identities, each student may be represented in multiple of the categories above. $substantialExp_s$ and $someExp_s$ represent how much experience the student had with computing prior to college, as described in Section 5.7.3; students with no experience prior to college are the reference group. We used fixed effects to control for the differences between students completing the survey in each year (γ_y), students pursuing different categories of major (μ; as defined in Section 5.7.3), and students attending different institutions[12] (ι_s). ϵ_s represents the error term per student.

5.8.2 RQ2: Predictors of Self-Efficacy and RQ3: Predictors of Sense of Belonging

To answer RQ2 and RQ3, we created four linear regression models. Two of these models, which we called "a" models (Models 2a and 3a), included only the main effects we were investigating, while in the "b" models (Models 2b and 3b), we added variables representing possible moderating effects. The equation for the "b" models is given below; the equation for the "a" models differs in that it does not include any of the moderation effects (i.e., interaction terms that include transparency).

[12]Due to an oversight, we omitted including the institution the student attends as a fixed effect in our pre-registration [149]. We include it in our analysis regardless due to the importance of controlling for differences in the perception of instructional transparency among students at different institutions.

$$
\begin{aligned}
outcome_s = \quad & \beta_0 + \beta_1 transparency_s \\
+ \quad & \beta_2 woman_s + \beta_3 firstGen_s + \beta_4 disabled_s \\
+ \quad & \beta_5 black_s + \beta_6 native_s + \beta_7 hispanic_s + \beta_8 asian_s \\
+ \quad & \beta_9 substantialExp_s + \beta_{10} someExp_s \\
+ \quad & \beta_{11}(transparency_s \times woman_s) + \beta_{12}(transparency_s \times firstGen_s) \\
+ \quad & \beta_{13}(transparency_s \times disabled_s) + \beta_{14}(transparency_s \times black_s) \\
+ \quad & \beta_{15}(transparency_s \times native_s) + \beta_{16}(transparency_s \times hispanic_s) \\
+ \quad & \gamma_y + \mu + \iota_s + \epsilon_s
\end{aligned}
\tag{5.2}
$$

In each model in the form of the above equation, $outcome_s$ represents each of the outcome variables we are interested in predicting:

Model 2: students' self-efficacy in CS ($selfEfficacy_s$), and

Model 3: students' sense of belonging in computing ($belonging_s$).

The following variables, representing the same constructs as in Equation 1, were included in the model due to their previously documented relationship with computing self-efficacy and sense of belonging: $woman_s$, $firstGen_s$, $disabled_s$, $black_s$, $native_s$, $hispanic_s$, $asian_s$, $substantialExp_s$, and $someExp_s$.

We also anticipated the possibility of moderation effects between perceptions of transparency and students' identities. In particular, Winkelmes et al. [49] noted a *greater* effect of instructional transparency on self-efficacy and sense of belonging for students who identified as first-generation college students, of lower socioeconomic backgrounds, or not white, although the authors did not conduct between-group comparisons. Given the data we have available, we tested for whether identifying as a member of a HUG moderates the effect of perceiving transparency by including moderating effects between perception of transparency and students' identities as Black, Native, Hispanic, and/or a first generation college student. We further included moderating effects with women and with students with disabilities, as these groups are underrepresented in CS [115]. These moderation effects investigate whether perception of transparency has a differential impact on self-efficacy and sense of belonging for different groups.

5.9 LIMITATIONS

Our work constitutes a large-scale investigation of the relationships between students' perception of instructional transparency and their self-efficacy and sense of belonging, as well as of group differences in these relationships. However, our study and its implications have a number of limitations.

First, we do not know what, if any, transparent teaching practices were intentionally implemented in the courses students were enrolled in. The survey questions we use in our measure for transparency are specifically about students' *perception* of instructional transparency in their CS courses. Therefore, our claims are limited to concerning how students' perception of instructional transparency predicts their self-efficacy and sense of belonging, rather than the actual practices implemented in their courses.

Second, our methods indicate correlational relationships, but do not demonstrate whether our hypothesized relationships are causal. We mitigate this concern by providing a theoretical argument for the mechanisms by which transparent teaching may affect self-efficacy and sense of belonging in Section 5.4. Future work may seek to implement classroom interventions to further investigate the impact of instructional transparency.

Third, we could not control for all factors we believe may be relevant. In particular, students' self-efficacy and sense of belonging in computing may be affected by factors that we did not have data for, such as students' current performance in the CS course(s) they are enrolled in. There may also be self-selection in the students who agree to take the survey. Different students may also interpret and/or respond to the questions differently; for example, prior work suggests that women may report lower self-efficacy even when earning the same grades as men [25].

5.10 RESULTS

5.10.1 RQ1: Predictors of Perception of Instructional Transparency

In RQ1, we investigated whether students' gender, race, ethnicity, disability status, first-generation status, and prior CS experience predicted students' perception of instructional transparency in their CS courses. We hypothesized that prior computing experience may lead to students perceiving greater transparency in their CS courses. The regression table for Model 1 is shown in Table 5.2.

Our findings show that identifying as a woman, a first-generation college student, or disabled all predict a lower perception of instructional transparency. That is to say, compared

Table 5.2: Regression Model 1 predicting perception of instructional transparency.

Predictor of instructional transparency perception	**Model 1**
Woman	−0.08*** (0.02)
First-generation	−0.07** (0.02)
Disabled	−0.20*** (0.02)
Black	−0.06 (0.04)
Native	0.07 (0.07)
Hispanic	−0.05 (0.03)
Asian	0.01 (0.02)
Substantial experience	0.05* (0.02)
Some experience	0.05 (0.04)
Major (fixed effect)	x
Data collection year (fixed effect)	x
Institution (fixed effect)	x
Number of respondents	11,046
Adjusted R-squared	0.04

Note: Standard errors are given in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. An x indicates that the variable was included as a fixed effect. Each coefficient represents a change in perception of instructional transparency along a distribution with mean 0 and standard deviation 1.

to students who identify as men, continuing-generation, and not disabled, respectively, these students express lower agreement that their CS courses had instructional transparency. In addition, we confirm our hypothesis that a student's reporting of more CS experience prior to college predicts greater perception of transparency in their courses.

We note that the coefficient associated with students with disabilities had a magnitude almost three times as large as the other statistically significant relationships. Recall that this coefficient represents a difference in students' perception of transparency, as defined in Section 5.7.1, a distribution that was standardized to have a mean of 0 and standard deviation of 1. The coefficient for disabled students, −0.20, means that compared to students without disabilities, students with disabilities reported a lower perception of transparency by 0.20 standard deviations. Our definition of disabled students included those who identified with one or more of 14 disabilities, including but not limited to cognitive disorders, learning disabilities, and sensory impairments.

We also note that Model 1 has a relatively low adjusted R-squared, which is a measure of how much of the variation in our dependent variable, perception of transparency, can be explained by our independent variables. A low adjusted R-squared is not surprising, as we expect that most of the variation in students' perception of instructional transparency will be due to the many different CS courses they are enrolled in and the extent to which those courses implement transparent teaching practices. As our work does not aim to predict an individual student's perception of transparency, but rather to find statistically significant patterns in students' perceptions more broadly, we do not believe this R-squared value is a cause for concern.

Table 5.3: Regression Models 2a (main effects only) and 2b (with moderating effects) predicting computing self-efficacy.

Predictor of self-efficacy	**Model 2a**	**Model 2b**
Transparency perception	0.24*** (0.01)	0.22*** (0.01)
Woman	−0.21*** (0.02)	−0.21*** (0.02)
First-generation	−0.14*** (0.02)	−0.14*** (0.02)
Disabled	−0.12*** (0.02)	−0.12*** (0.02)
Black	−0.20*** (0.03)	−0.20*** (0.03)
Native	−0.15* (0.06)	−0.13* (0.06)
Hispanic	−0.06* (0.03)	−0.05 (0.03)
Asian	−0.23*** (0.02)	−0.23*** (0.02)
Substantial experience	0.29*** (0.02)	0.29*** (0.02)
Some experience	0.06 (0.04)	0.06 (0.04)
Transparency perception × Woman		−0.02 (0.02)
Transparency perception × First-generation		0.05* (0.02)
Transparency perception × Disabled		0.01 (0.02)
Transparency perception × Black		0.11*** (0.03)
Transparency perception × Native		0.10 (0.06)
Transparency perception × Hispanic		0.02 (0.03)
Major (fixed effect)	x	x
Data collection year (fixed effect)	x	x
Institution (fixed effect)	x	x
Number of respondents	10,973	10,973
Adjusted R-squared	0.16	0.16

Note: Standard errors are given in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. An x indicates that the variable was included as a fixed effect. Each coefficient represents a change in self-efficacy along a distribution with mean 0 and standard deviation 1.

5.10.2 RQ2: Predictors of Self-Efficacy

RQ2 sought to investigate the extent to which students' perception of transparency predicts their computing self-efficacy, while controlling for factors known to relate to self-efficacy. To answer RQ2, we created two models, Model 2a and Model 2b. Model 2a included only main effects, while Model 2b included moderating effects of underrepresented identities. We hypothesized that perceiving more transparency will predict higher computing self-efficacy, and that this effect will be stronger for students from HUGs, i.e., underrepresented identities will positively moderate the relationship between perception of transparency and self-efficacy. The results of the regression analysis are shown in Table 5.3.

Findings from Models 2a and 2b indicated that perception of instructional transparency has a significant, positive relationship with computing self-efficacy, as hypothesized. Specifically, Model 2b predicts that students whose reported perception of transparency in their CS courses was one standard deviation above the mean reported a higher computing self-efficacy by 0.22 standard deviations. We additionally found two significant moderating effects, demonstrating that for Black students and for first-generation college students, instructional transparency has a larger positive impact on their self-efficacy. These significant moderating effects partially confirm our hypotheses that perceiving greater transparency is additionally beneficial for students from HUGs, as we demonstrate the differential benefit only for Black and first-generation students, and not for all students from HUGs.

The variables representing students' gender, race, ethnicity, disability, and first-generation status confirm prior work in having a negative relationship with self-efficacy. Having substantial prior CS experience predicted higher computing self-efficacy, also confirming prior work.

5.10.3 RQ3: Predictors of Sense of Belonging

In RQ3, we examined the extent to which students' perception of instructional transparency predicted their sense of belonging in computing, controlling for known factors relating to sense of belonging. As with Model 2, we created Model 3a with only main effects, while Model 3b included moderating effects of underrepresented identities. We hypothesized a positive relationship between perception of transparency and sense of belonging, with positive moderation effects of underrepresented identities. The regression table for these models is shown in Table 5.4.

We found that students' perception of transparency is significantly positively predicted their sense of belonging in computing. Our model indicates that greater perception of instructional transparency by one standard deviation above the mean predicts a greater reported sense of belonging by 0.24 standard deviations above the mean. We found evidence for one significant moderating effect: identifying as Hispanic *negatively* moderated the relationship between perceiving transparency and sense of belonging. This negative moderation is contrary to our hypotheses that perceiving greater transparency would additionally benefit students from HUGs.

Other variables included in Model 3 confirmed results from prior work in their relationship to sense of belonging. We note, however, that even having *some* prior CS experience

Table 5.4: Regression Models 3a (main effects only) and 3b (with moderating effects) predicting sense of belonging in computing.

Predictor of sense of belonging	**Model 3a**	**Model 3b**
Transparency perception	0.24*** (0.01)	0.24*** (0.01)
Woman	−0.47*** (0.02)	−0.47*** (0.02)
First-generation	−0.06** (0.02)	−0.06** (0.02)
Disabled	−0.17*** (0.02)	−0.17*** (0.02)
Black	−0.15*** (0.03)	−0.15*** (0.03)
Native	−0.10 (0.07)	−0.10 (0.07)
Hispanic	−0.05 (0.03)	−0.06* (0.03)
Asian	−0.12*** (0.02)	−0.12*** (0.02)
Substantial experience	0.31*** (0.02)	0.31*** (0.02)
Some experience	0.15*** (0.04)	0.15*** (0.04)
Transparency perception × Woman		−0.01 (0.02)
Transparency perception × First-generation		0.01 (0.02)
Transparency perception × Disabled		0.03 (0.02)
Transparency perception × Black		0.02 (0.03)
Transparency perception × Native		−0.02 (0.06)
Transparency perception × Hispanic		−0.06* (0.03)
Major (fixed effect)	x	x
Data collection year (fixed effect)	x	x
Institution (fixed effect)	x	x
Number of respondents	10,982	10,982
Adjusted R-squared	0.19	0.19

Note: Standard errors, given in parentheses, are clustered by institution. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. An x indicates that the variable was included as a fixed effect. Each coefficient represents a change in belonging along a distribution with mean 0 and standard deviation 1.

predicted an increase in sense of belonging, unlike self-efficacy, where only *substantial* prior experience was a statistically significant predictor.

5.11 DISCUSSION & FUTURE WORK

Our work constitutes one of the first empirical, multi-institutional investigations of the perceptions and benefits of transparency in CS classrooms that focuses on group differences. We found that there are group differences in students' perception of instructional transparency in their CS courses. Students who identify as women, first-generation college students, and/or disabled reported perceiving less transparency than their peers. We also found that perceiving more transparency positively predicts with students' self-efficacy and sense of belonging in computing while controlling for important confounding variables, such as prior CS experience. We further demonstrated that this relationship is different for certain groups of students: first, for Black students and first-generation college students, transparency has a *larger* positive impact on their self-efficacy, and second, for Hispanic students, transparency has a *smaller* positive impact on their sense of belonging. In this section, we contextualize our findings in previous research, discuss its implications, and propose avenues for future research.

5.11.1 RQ1: Predictors of Perception of Instructional Transparency

Our results for RQ1 indicate that, compared to their peers, women, first-generation college students, and students with disabilities perceive less transparency in their CS courses. As far as we are aware, these are the first findings to illustrate group differences in students' perception of instructional transparency in their CS courses. We note that because we controlled for the institution students attend using fixed effects in Model 1, this pattern cannot be explained by people from different demographic groups attending institutions that employ more or less transparency in their CS courses. However, it could be the case that within an institution students from different demographic groups tend to select CS courses that employ more or less transparency.

The effect size corresponding to disabled students indicated a larger effect than the other statistically significant relationships. That is to say, students with disabilities report perceiving *less* transparency in their CS courses than their peers with no disabilities. This finding may signal that transparent teaching practices that "work" for non-disabled students are less effective for disabled students. For instance, one of the survey items asked whether students agree that they were provided a "detailed set of instructions" for an assignment (see Section 5.7.1); it is possible that what students without disabilities considered "detailed instructions" are less clear for those with particular disabilities. Prior work has documented the many challenges disabled students face regarding accommodations in higher education [152, 153]; disabled students being denied accommodations may make it more challenging for them to benefit from instructional transparency.

Our findings also suggest that instructional transparency may not be as apparent to students who identify as first-generation college students, i.e., students whose parents or guardians did not attain a college degree. A possible explanation may be that first-generation college students are less likely to be familiar with the "insider knowledge" that is part of the hidden curriculum at academic institutions [128] and may, therefore, need additional or different forms of transparency than their continuing-generation peers. We note that this finding cannot be explained by first-generation students tending to attend particular institutions that have high transparency in their CS courses, because our model controlled for students' institutions.

Women also reported perceiving less transparency in their CS courses. A possible explanation for this pattern may also relate to the hidden curriculum in computing: prior work has noted the implicit masculine culture, values, and norms in computing [53, 154, 155], which may lead to women perceiving the same instruction as less transparent than their peers who are men.

Consistent with our expectations, having substantial prior CS experience positively predicted students' perception of instructional transparency. This result suggests that prior familiarity with computing concepts helps students more readily perceive learning goals and instructions. Therefore, students with *less* prior CS experience may not perceive instruction as transparent even when their more experienced peers do. Students from HUGs are less likely to have had prior exposure to computing in their K-12 education [117, 118, 119, 120]; therefore, this patterns *compounds* the disadvantages students from HUGs face in accessing instruction in CS.

5.11.2 RQ2: Predictors of Self-Efficacy

The regression model for RQ2 demonstrated that students who perceive greater instructional transparency in their CS course also report higher computing self-efficacy, as hypothesized. Our findings in this regard are a notable contribution to the evidence for the benefits of instructional transparency, as we are able to demonstrate this relationship across institutional contexts and while controlling for confounding variables.

The results from these models also demonstrate that transparency has a differential positive impact on the computing self-efficacy of Black students and first-generation college students. The significant moderating effects indicate that although transparency alone has a positive relationship with self-efficacy, there is an even stronger relationship for Black students and first-generation college students in our participant pool. Winkelmes et al. [49] noted greater effects of transparency for students who were first-generation and for students who did not identify as white, but did not conduct comparisons between groups; therefore, we are able to fill this gap and demonstrate that instructional transparency may indeed have a greater impact on self-efficacy for Black students and for first-generation college students. However, we cannot confirm this relationship for students from other underrepresented racial/ethnic groups among our participants. An open question for researchers to consider is why instructional transparency may be particularly effective for Black students' and first-generation students' computing self-efficacy.

It is possible that students' overall attitude towards computing may positively influence both their self-efficacy and their perceived transparency, explaining the positive relationship between the two found in our model. However, we note that attitude towards computing is unlikely to affect the survey items used in our measure for perception of transparency, as these were specific to practices seen in students' courses, such as "each assignment includ[ing] a detailed set of instructions for completing it".

We note that the lack of other significant moderators may simply indicate that instructional transparency is equally useful for all other groups of students. Our results indicate that students who identify as women, Native, Hispanic, and/or having a disability do not *differentially* benefit from transparency, but for all students, perception of transparency still positively predicts self-efficacy.

5.11.3 RQ3: Predictors of Sense of Belonging

In response to RQ3, we found that students' perception of instructional transparency predicted their sense of belonging, confirming our hypothesis. These statistically significant results provide empirical evidence regarding the benefits of instructional transparency across institutions while accounting for confounding variables.

Our regression models further demonstrated one significant moderating effect, but not in the direction we hypothesized: identifying as Hispanic *negatively* moderated the relationship between perception of transparency and sense of belonging. This means that for Hispanic students, the positive relationship between perceiving transparency and sense of belonging has a smaller effect size than for other students[13]. This result is inconsistent with claims from prior work regarding the additional benefits of instructional transparency for students from HUGs [49]. We encourage future research to investigate other factors that may affect Hispanic students' sense of belonging in computing.

A notable result is the significant relationship we found between having *some* prior CS experience in high school and college students' sense of belonging in computing. (We note that *some* prior HS CS experience was defined as engaging in software/hardware projects, attending a computing workshop, or being part of a computing student group.) In our models for self-efficacy, we found that only *substantial* experience positively predicted self-efficacy, in line with prior work [33]. This suggests that while *some* prior experience may be insufficient to predict self-efficacy, it can be beneficial for students' sense of belonging in computing. Finding that even limited exposure to computing can predict sense of belonging

[13] The effect size of the main effect between perception of transparency and sense of belonging is 0.24, while the effect size for the moderating effect for Hispanic students is -0.06. Therefore, for Hispanic students, the overall effect remains positive, but is smaller than for other students.

is a promising result, as sense of belonging itself predicts positive outcomes such as perceived ability, interest, and performance [53, 54, 136].

5.12 CONCLUSION

This study aimed to investigate whether there are group differences in students' perception of transparency in their CS courses, and whether instructional transparency predicts students' self-efficacy and sense of belonging in computing. Self-efficacy and sense of belonging are important outcomes for equity in computing: both constructs predict persistence in computing and can, therefore, support efforts to retain students from HUGs in computing. Using undergraduate students' responses to a large, multi-institutional survey, we conducted linear regression to examine whether perception of transparency predicts self-efficacy and sense of belonging. Our work constitutes one of the first empirical, multi-institutional investigations of the perceptions and benefits of transparency in CS classrooms that focuses on group differences.

We found that there are group differences in students' perception of transparency in their CS courses. Students who identify as women, first-generation college students, and/or disabled reported perceiving less instructional transparency than their peers. We also found that perceiving more transparency positively predicts with students' self-efficacy and sense of belonging in computing while controlling for important confounding variables, such as prior CS experience. We further demonstrated that this relationship is different for certain groups of students: first, for Black students and first-generation college students, perceiving transparency has a *larger* positive impact on their self-efficacy, and second, for Hispanic students, perceiving transparency has a *smaller* positive impact on their sense of belonging. Finally, our work also includes a theoretical articulation of the mechanisms through which transparent teaching practices may influence students' self-efficacy and sense of belonging in computing. Taken together, our empirical findings and theoretical argument provide important evidence for the benefits of instructional transparency in CS courses, particularly as it relates to improving equity in computing.

CHAPTER 6: CONCLUSION

This dissertation consists of work investigating graduate and undergraduate students' perceptions and experiences in computing in order to promote efforts to broaden participation in computing. Below, I summarize this work's intellectual merit and broader impacts.

6.1 INTELLECTUAL MERIT

My research contributes to the body of literature investigating students' perceptions and experiences of CS in numerous ways.

In Chapter 2, I investigated students' perceptions about the specializations of AI and cybersecurity to extend previous research documenting stereotypes and beliefs about the field of computing. My findings demonstrate that students' stereotypes about computing subfields mirror stereotypes about computing more broadly and may dissuade students from those subfields [26]. This work extends our collective knowledge about field-specific perceptions and how they may relate to students' participation in computing. The findings suggest a particular need to highlight the ways in which AI and cybersecurity encompass more than the stereotypes of requiring innate brilliance, involving math, and having a negative impact on society [26]. It further suggests that more work is needed to understand how students choose to specialize within computing, particularly if we hope to attain equitable representation within impactful subfields such as AI and cybersecurity. While this study identifies perceptions students may have about these subfields, it also illuminates a gap in the existing research regarding how students form these perceptions and how those beliefs can be expanded.

Chapter 3 investigated how doctoral students from HUGs understand expectations such as what work they need to do and how to do it. Adding to prior findings that clear expectations contribute to student persistence [22, 31, 32], my work shows that students experience challenges in understanding what these expectations are and how to fulfill them [30]. Despite participating in research experiences as undergraduates, our participants expressed surprise about many aspects of the PhD program, prompting future research questions about how experiences in undergraduate programs can translate to preparation for graduate programs. These findings add to our understanding of the challenges students face in graduate programs, such as being surprised by the lack of guidance on how to complete tasks, and suggest possible strategies to address these challenges, such as formal mentorship programs [30].

The study in Chapter 4 contributes to the body of work on computing self-efficacy, an important facet of students' experiences that is positively correlated with their persistence and motivation, on a large scale and across institutional contexts. My analysis in this work replicated prior results, demonstrating that identifying as a woman, non-binary, Black, Native, Hispanic, and/or Asian correlates with lower self-efficacy when controlling for prior experience [33]. In addition, this study contributed to self-efficacy literature by including analyses of students who identified as non-binary and students who identified with multiple marginalized identities. As a result, we were able to demonstrate the novel results that identifying as non-binary also correlates with lower computing self-efficacy, as did identifying as both Asian and non-binary [33]. These findings highlight the value of an intersectional approach to understanding students' experiences; without this approach, experiences of students with multiple marginalized identities, such as Asian non-binary students, may go unexamined. In my future work, I hope to use qualitative research methods to investigate why some of these patterns may be occurring.

Finally, in Chapter 5, I found that students' perception of instructional transparency in their CS classroom predicts their self-efficacy and sense of belonging in computing, even when controlling for aspects of students' identities known to impact those outcomes [48]. Although prior work has shown evidence for these relationships in a limited context, this work contributes to the field by demonstrating the correlation across institutional contexts while controlling for confounding variables. In addition, this study demonstrated that students' perception of what teaching is transparent differs by their gender, first-generation status, disability status, and by their prior experience, all novel results in the literature [48]. By demonstrating these relationships, my work suggests possible questions to be investigated in future research, such as how transparent teaching can be made more accessible and impactful for all students.

6.2 BROADER IMPACTS

My PhD work has incorporated a variety of methods in conducting research across different levels of higher education with the overarching goal of improving the experiences of students from HUGs in computing. A greater understanding of these experiences can help support the creation of more inclusive and supportive environments within computing departments, which may help attract students and contribute to improved retention rates in CS programs. The need for such work is undeniable as the field of CS in the US continues to lose potential contributions from millions of students and struggles to meet the increasing market demand for qualified computing professionals. In addition, as technology becomes an increasingly common part of daily life as well as a lucrative career, there is a social justice imperative to ensure that computing education is inclusive for all. Faculty, staff, and administrators in CS departments across the country could use insights from my work to change policies and create better systems that serve all students.

At the undergraduate level, my work suggests a need to account for how experiences outside of the classroom may affect students and their decisions within computing. For instance, pre-existing conceptions of fields affect students' decisions to pursue subfields within computing which, in turn, exacerbate existing inequitable representation [26]. A key implication of this finding is the need to broaden students' perceptions of computing and its subfields, both in classrooms and outside of them, as students may form their opinions before taking a CS course. My work also demonstrates that students' experiences prior to college and their sociocultural identities influence important outcomes, such as their computing self-efficacy [33], which affect their persistence in the field of computing. As such, educators may consider implementing practices in their classrooms that are shown to positively correlate with self-efficacy and other important outcomes. Transparent teaching practices are one example, as my work illustrates a positive relationship between students' perceptions of transparency and their self-efficacy and sense of belonging in computing [48].

My work with doctoral students elevates the importance of creating clear expectations within graduate programs in order to better support student success [30]. Some practical implications of my work include highlighting the value of an official mentorship system within graduate student labs and of interventions outside of classrooms to shift undergraduate students' perceptions of AI and cybersecurity. My research also has the potential to provide more direct support to students by giving them the knowledge they may need to advocate for themselves.

Across the dissertation, the planned impact of my work is to inform policies and practices at CS departments across institutions. My work thus far has shown that policies and practices at different levels (e.g., affirmative action at a national level, lab-based mentorship at an institutional level, transparent teaching at a classroom-level) may impact students in a myriad of ways. Understanding these and other policies and practices may lead us to suggested best practices that bolster the recruitment and persistence of students in computing, particularly those from HUGs.

REFERENCES

[1] A. Lane, R. Mekonnen, C. Jang, P. Chen, and C. M. Lewis, "Motivating literature and evaluation of the teaching practices game: Preparing teaching assistants to promote inclusivity," in *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, 2021, pp. 816–822.

[2] J. Margolis, *Stuck in the Shallow End, updated edition: Education, Race, and Computing.* MIT Press, 2017.

[3] Computing Research Association, "The CRA Taulbee Survey," https://cra.org/resources/taulbee-survey/, 2021.

[4] National Center for Women and Information Technology, "By The Numbers," https://ncwit.org/resource/bythenumbers/, 2021.

[5] Code.org, "2022 State of Computer Science Education," https://advocacy.code.org/stateofcs, 2022.

[6] J. John and M. Carnoy, "Race and gender trends in computer science in the Silicon Valley from 1980-2015," 2017, unpublished. [Online]. Available: https://cepa.stanford.edu/sites/default/files/JohnCarnoy_Sept2017.pdf

[7] International Information System Security Certification Consortium, "Women in cybersecurity: Young, educated, and ready to take charge," https://www.isc2.org/-/media/ISC2/Research/ISC2-Women-in-Cybersecurity-Report.ashx, 2020, accessed 2022-07-31.

[8] Computing Research Association, Center for Evaluating the Research Pipeline, "Statistics & Data Hub, Postsecondary Computing Degrees Awarded (Version 1.3.5)," https://bpcnet.org/statistics/, 2022, accessed 2024-01-12.

[9] United States Census Bureau, "U.S. Census Bureau QuickFacts," https://www.census.gov/quickfacts/fact/table/US/LFE046222, 2022.

[10] H. B. Carlone and A. Johnson, "Understanding the science experiences of successful women of color: Science identity as an analytic lens," *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, vol. 44, no. 8, pp. 1187–1218, 2007.

[11] A. B. Diekman, E. R. Brown, A. M. Johnston, and E. K. Clark, "Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers," *Psychological Science*, vol. 21, no. 8, pp. 1051–1057, Aug. 2010.

[12] B. James DiSalvo, S. Yardi, M. Guzdial, T. McKlin, C. Meadows, K. Perry, and A. Bruckman, "African American men constructing computing identity," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2011, p. 2967–2970.

[13] S. Cheryan, S. A. Ziegler, A. K. Montoya, and L. Jiang, "Why are some STEM fields more gender balanced than others?" *Psychological Bulletin*, vol. 143, no. 1, pp. 1–35, 2017.

[14] S. L. Rodriguez and K. Lehman, "Developing the next generation of diverse computer scientists: the need for enhanced, intersectional computing identity theory," *Computer Science Education*, vol. 27, no. 3-4, pp. 229–247, 2017.

[15] R. Rashid, "Image crisis inspiring a new generation of computer scientists," *Communications of the ACM*, vol. 51, no. 7, pp. 33–34, 2008.

[16] C. E. Wills, "Outcomes of Advertised Computer Science Faculty Searches for 2017," https://cra.org/crn/2017/11/outcomes-advertised-computer-science-faculty-searches-2017/, November 2017.

[17] National Academies of Sciences, Engineering, and Medicine (NASEM), *Assessing and Responding to the Growth of Computer Science Undergraduate Enrollments.* Washington, DC: The National Academies Press, 2018.

[18] Bureau of Labor Statistics, "Computer and Information Technology Occupations," https://www.bls.gov/ooh/computer-and-information-technology/home.htm, 2022.

[19] E. Tsagala and M. Kordaki, "Critical factors influencing secondary school pupil's decisions to study computing in tertiary education: Gender differences," *Education and Information Technologies*, vol. 12, pp. 281–295, 2007.

[20] C. M. Lewis, K. Yasuhara, and R. E. Anderson, "Deciding to major in computer science: A grounded theory of students' self-assessment of ability," in *Proceedings of the 7th International Workshop on Computing Education Research*, 2011.

[21] S. Cheryan, A. Master, and A. N. Meltzoff, "Cultural stereotypes as gatekeepers: Increasing girls' interest in computer science and engineering by diversifying stereotypes," *Frontiers in Psychology*, vol. 6, 2015.

[22] A. Sverdlik, N. C. Hall, L. McAlpine, and K. Hubbard, "The PhD experience: A review of the factors influencing doctoral students' completion, achievement, and well-being," *International Journal of Doctoral Studies*, vol. 13, pp. 361–388, 2018.

[23] J. Posselt, "Normalizing struggle: Dimensions of faculty support for doctoral students and implications for persistence and well-being," *The Journal of Higher Education*, vol. 89, no. 6, pp. 988–1013, 2018.

[24] Y. A. Rankin and J. O. Thomas, "The intersectional experiences of Black women in computing," in *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, 2020, pp. 199–205.

[25] C. Hunt, S. Yoder, T. Comment, T. Price, B. Akram, L. Battestilli, T. Barnes, and S. Fisk, "Gender, self-assessment, and persistence in computing: How gender differences in self-assessed ability reduce women's persistence in computer science," in *Proceedings of the 2022 ACM Conference on International Computing Education Research - Volume 1*, 2022, p. 73–83.

[26] V. Ojha, C. Perdriau, B. Lagesse, and C. M. Lewis, "Computing specializations: Perceptions of AI and cybersecurity among CS students," in *Proceedings of the 2023 ACM SIGCSE Technical Symposium on Computer Science Education*, 2023.

[27] S.-J. Leslie, A. Cimpian, M. Meyer, and E. Freeland, "Expectations of brilliance underlie gender distributions across academic disciplines," *Science*, vol. 347, no. 6219, pp. 262–265, Jan. 2015.

[28] C. M. Lewis, R. E. Anderson, and K. Yasuhara, "'I don't code all day': Fitting in computer science when the stereotypes don't fit," in *Proceedings of the 2016 ACM Conference on International Computing Education Research*, ser. ICER '16. New York, NY, USA: Association for Computing Machinery, Aug. 2016, pp. 23–32.

[29] S. Cheryan, V. C. Plaut, C. Handron, and L. Hudson, "The stereotypical computer scientist: Gendered media representations as a barrier to inclusion for women," *Sex Roles*, vol. 69, no. 1, pp. 58–71, July 2013.

[30] V. Ojha, R. E. Platero, and B. B. Bullock, "Forming and fulfilling expectations: Perspectives of underrepresented computer science doctoral students," in *2022 CoNECD (Collaborative Network for Engineering & Computing Diversity)*, 2022.

[31] C. M. Golde, "Should I stay or should I go? Student descriptions of the doctoral attrition process," *The Review of Higher Education*, vol. 23, no. 2, pp. 199–227, 2000.

[32] A. J. Fisher, R. Mendoza-Denton, C. Patt, I. Young, A. Eppig, R. L. Garrell, D. C. Rees, T. W. Nelson, and M. A. Richards, "Structure and belonging: Pathways to success for underrepresented minority and women PhD students in stem fields," *PLoS One*, vol. 14, no. 1, 2019.

[33] V. Ojha, L. West, and C. M. Lewis, "Computing self-efficacy in undergraduate students: A multi-institutional and intersectional analysis," in *Proceedings of the 2024 ACM SIGCSE Technical Symposium on Computer Science Education*, 2024.

[34] A. Bandura, "Self-efficacy: toward a unifying theory of behavioral change." *Psychological review*, vol. 84, no. 2, p. 191, 1977.

[35] A. Bandura, "The explanatory and predictive scope of self-efficacy theory," *Journal of social and clinical psychology*, vol. 4, no. 3, pp. 359–373, 1986.

[36] I. T. Miura, "The relationship of computer self-efficacy expectations to computer interest and course enrollment in college," *Sex Roles*, vol. 16, no. 5, pp. 303–311, 1987.

[37] F. Weng, F. Cheong, and C. Cheong, "Modelling IS student retention in Taiwan: Extending Tinto and Bean's model with self-efficacy," *Innovation in Teaching and Learning in Information and Computer Sciences*, vol. 9, no. 2, pp. 1–12, 2010.

[38] S. Beyer, "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*, vol. 24, no. 2-3, pp. 153–192, 2014.

[39] J. M. Blaney and J. G. Stout, "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*, 2017, pp. 69–74.

[40] T. Busch, "Gender, group composition, cooperation, and self-efficacy in computer studies," *Journal of Educational Computing Research*, vol. 15, pp. 125–135, 1995.

[41] D. Potosky, "A field study of computer efficacy beliefs as an outcome of training: The role of computer playfulness, computer knowledge, and performance during training," *Computers in Human Behavior*, vol. 18, no. 3, pp. 241–255, 2002.

[42] B. Hasan, "The influence of specific computer experiences on computer self-efficacy beliefs," *Computers in Human Behavior*, vol. 19, pp. 443–450, 2003.

[43] H. Cigdem and O. G. Yildirim, "Predictors of C# programming language self efficacy among vocational college students," *International Journal on New Trends in Education and their Implications*, vol. 5, no. 3, pp. 145–153, 2014.

[44] S. Cassidy and P. Eachus, "Developing the computer user self-efficacy (CUSE) scale: Investigating the relationship between computer self-efficacy, gender and experience with computers," *Journal of Educational Computing Research*, vol. 26, no. 2, pp. 133–153, 2002.

[45] M.-J. Tsai, C.-Y. Wang, and P.-F. Hsu, "Developing the computer programming self-efficacy scale for computer literacy education," *Journal of Educational Computing Research*, vol. 56, no. 8, pp. 1345–1360, 2019.

[46] A. Nguyen and C. M. Lewis, "Competitive enrollment policies in computing departments negatively predict first-year students' sense of belonging, self-efficacy, and perception of department," in *SIGCSE Proceedings*, 2020, pp. 685–691.

[47] K. Crenshaw, "Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics," in *Feminist Legal Theories*. Routledge, 2013, pp. 23–51.

[48] V. Ojha, A. Watkins, C. Perdriau, K. Isenegger, and C. M. Lewis, "Instructional transparency: Just to be clear, it's a good thing," in *Under review*, 2024.

[49] M.-A. Winkelmes, M. Bernacki, J. Butler, M. Zochowski, J. Golanics, and K. H. Weavil, "A teaching intervention that increases underserved college students' success," *Peer Review*, vol. 18, no. 1-2, pp. 31–37, 2016.

[50] M.-A. Winkelmes, "Why it works: Understanding the concepts behind transparency in learning and teaching," *Transparent Design in Higher Education Teaching and Leadership*, pp. 17–35, 2019.

[51] J. Ou, "Board 75: Work in progress: A study of transparent assignments and their impact on students in an introductory circuit course," in *2018 ASEE Annual Conference & Exposition*, 2018.

[52] R. W. Lent, A. M. Lopez Jr, F. G. Lopez, and H.-B. Sheu, "Social cognitive career theory and the prediction of interests and choice goals in the computing disciplines," *Journal of Vocational Behavior*, vol. 73, no. 1, pp. 52–62, 2008.

[53] A. Master, S. Cheryan, and A. N. Meltzoff, "Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science." *Journal of Educational Psychology*, vol. 108, no. 3, p. 424, 2016.

[54] S. Krause-Levy, W. G. Griswold, L. Porter, and C. Alvarado, "The relationship between sense of belonging and student outcomes in CS1 and beyond," in *Proceedings of the 17th ACM Conference on International Computing Education Research*, 2021, pp. 29–41.

[55] LinkedIn Talent Solutions, "US Jobs on the Rise Report," https://business.linkedin.com/talent-solutions/resources/talent-acquisition/jobs-on-the-rise-us, 2022.

[56] Cyberseek, https://www.cyberseek.org/, accessed 2022-07-31.

[57] Department of Homeland Security, "DHS Launches Innovative Hiring Program to Recruit and Retain World-Class Cyber Talent," https://www.dhs.gov/news/2021/11/15/dhs-launches-innovative-hiring-program-recruit-and-retain-world-class-cyber-talent, Nov 2021.

[58] R. S. S. Kumar, M. Nyström, J. Lambert, A. Marshall, M. Goertzel, A. Comissoneru, M. Swann, and S. Xia, "Adversarial machine learning-industry perspectives," in *2020 IEEE Security and Privacy Workshops (SPW)*. IEEE, 2020, pp. 69–75.

[59] M. Hewner, "How CS undergraduates make course choices," in *Proceedings of the 10th Annual Conference on International Computing Education Research*. Association for Computing Machinery, July 2014, pp. 115–122.

[60] National Initiative for Cybersecurity Careers and Studies, "The Workforce Framework for Cybersecurity (NICE framework) Work Roles," https://niccs.cisa.gov/about-niccs/workforce-framework-cybersecurity-nice-framework-work-roles, Mar 2021.

[61] D. Gordon, "Forty years of movie hacking: Considering the potential implications of the popular media representation of computer hackers from 1968 to 2008," *International Journal of Internet Technology and Secured Transactions*, vol. 2, p. 59, 2010.

[62] Y. K. Hoh, "Using notable women in environmental engineering to dispel misperceptions of engineers," *International Journal of Environmental and Science Education*, vol. 4, no. 2, pp. 117–131, Apr. 2009.

[63] B. Dym, N. Pasupuleti, C. Rockwood, and C. Fiesler, "'You don't do your hobby as a job': Stereotypes of computational labor and their implications for CS education," in *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '21, Mar. 2021, pp. 823–829.

[64] L. Ouchchy, A. Coin, and V. Dubljević, "AI in the headlines: The portrayal of the ethical issues of artificial intelligence in the media," *AI & Society*, vol. 35, no. 4, pp. 927–936, Dec. 2020.

[65] E. Saffari, S. R. Hosseini, A. Taheri, and A. Meghdari, ""Does cinema form the future of robotics?": A survey on fictional robots in sci-fi movies," *SN Applied Sciences*, vol. 3, no. 6, p. 655, May 2021.

[66] D. Barretto, J. LaChance, E. Burton, and S. N. Liao, "Exploring why underrepresented students are less likely to study machine learning and artificial intelligence," in *Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education V. 1*, ser. ITiCSE '21, June 2021, pp. 457–463.

[67] A. Ottenbreit-Leftwich, K. Glazewski, M. Jeon, C. Hmelo-Silver, B. Mott, S. Lee, and J. Lester, "How do elementary students conceptualize artificial intelligence?" in *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, Mar. 2021, p. 1261.

[68] M. Kreinsen and S. Schulz, "Students' conceptions of artificial intelligence," in *The 16th Workshop in Primary and Secondary Computing Education*, Oct. 2021, no. 14, pp. 1–2.

[69] CloudPassage, "CloudPassage Study Finds U.S. Universities Failing in Cybersecurity Education," https://www.globenewswire.com/news-release/2016/04/07/1312702/0/en/CloudPassage-Study-Finds-U-S-Universities-Failing-in-Cybersecurity-Education.html, Apr. 2016.

[70] W. Newhouse, S. Keith, B. Scribner, and G. Witte, "National Initiative for Cybersecurity Education (NICE) Cybersecurity Workforce Framework," *NIST Special Publication*, vol. 800, no. 2017, p. 181, 2017.

[71] S. B. Merriam and E. J. Tisdell, *Qualitative Research: A Guide to Design and Implementation*. John Wiley & Sons, 2015.

[72] J. Sillito, "Saturate," http://www.saturateapp.com/.

[73] J. Saldaña, *The Coding Manual for Qualitative Researchers*, 4th ed. SAGE Publishing, 2021.

[74] A. K. Shenton, "Strategies for ensuring trustworthiness in qualitative research projects," *Education for Information*, vol. 22, no. 2, pp. 63–75, Jan. 2004.

[75] C. Good, A. Rattan, and C. S. Dweck, "Why do women opt out? Sense of belonging and women's representation in mathematics," *Journal of Personality and Social Psychology*, vol. 102, no. 4, p. 700, 2012.

[76] B. Lagesse and C. M. Lewis, "Key concepts of AI-enhanced cybersecurity," in preparation.

[77] D. S. Yeager, P. Hanselman, G. M. Walton, J. S. Murray, R. Crosnoe, C. Muller, E. Tipton, B. Schneider, C. S. Hulleman, C. P. Hinojosa et al., "A national experiment reveals where a growth mindset improves achievement," *Nature*, vol. 573, no. 7774, pp. 364–369, 2019.

[78] Computing Research Association, "Time to Degree in Computing," https://cra.org/crn/2014/04/time_to_degree_in_computing/, April 2014.

[79] B. J. Barnes and A. E. Austin, "The role of doctoral advisors: A look at advising from the advisor's perspective," *Innovative Higher Education*, vol. 33, no. 5, pp. 297–315, 2009.

[80] W. Lyons, D. Scroggins, and P. B. Rule, "The mentor in graduate education," *Studies in Higher Education*, vol. 15, no. 3, pp. 277–285, 1990.

[81] S. T. Charles, M. M. Karnaze, and F. M. Leslie, "Positive factors related to graduate student mental health," *Journal of American College Health*, pp. 1–9, 2021.

[82] G. Lichtenstein, H. L. Chen, K. A. Smith, and T. A. Maldonado, "Retention and persistence of women and minorities along the engineering pathway in the United States," *Cambridge Handbook of Engineering Education Research*, pp. 311–334, 2014.

[83] M. T. Nettles, "Success in doctoral programs: Experiences of minority and white students," *American Journal of Education*, vol. 98, no. 4, pp. 494–522, 1990.

[84] J. M. Cohoon, M. Nable, and P. Boucher, "Conflicted identities and sexism in computing graduate programs," in *2011 Frontiers in Education Conference (FIE)*, 2011.

[85] L. J. Charleston, "A qualitative investigation of African Americans' decision to pursue computing science degrees: Implications for cultivating career choice and aspiration," *Journal of Diversity in Higher Education*, vol. 5, no. 4, p. 222, 2012.

[86] National Academies of Science, Engineering, and Medicine, "The Science of Effective Mentoring in STEMM," https://www.nationalacademies.org/our-work/the-science-of-effective-mentoring-in-stemm, 2024, Accessed 2024-04-07.

[87] MentorFirst, "Mentor First Homepage," https://mentorfirst.org/, 2024, Accessed 2024-04-07.

[88] Center for the Improvement of Mentored Experiences in Research, "CIMER Homepage," https://cimerproject.org/, 2024, Accessed 2024-04-07.

[89] R. W. Lent, S. D. Brown, and K. C. Larkin, "Relation of self-efficacy expectations to academic achievement and persistence," *Journal of Counseling Psychology*, vol. 31, no. 3, p. 356, 1984.

[90] J. He and L. A. Freeman, "Are men more technology-oriented than women? The role of gender on the development of general computer self-efficacy of college students," *Journal of Information Systems Education*, vol. 21, no. 2, pp. 203–212, 2010.

[91] A. Fisher and J. Margolis, "Unlocking the clubhouse: The Carnegie Mellon experience," *ACM SIGCSE Bulletin*, vol. 34, no. 2, pp. 79–83, 2002.

[92] M. Papastergiou, "Are computer science and information technology still masculine fields? High school students' perceptions and career choices," *Computers & Education*, vol. 51, no. 2, pp. 594–608, 2008.

[93] S. Vakil, "Ethics, identity, and political vision: Toward a justice-centered approach to equity in computer science education," *Harvard Educational Review*, vol. 88, no. 1, pp. 26–52, 2018.

[94] L. J. Charleston, P. L. George, J. F. Jackson, J. Berhanu, and M. H. Amechi, "Navigating underrepresented STEM spaces: Experiences of Black women in US computing science higher education programs who actualize success." *Journal of Diversity in Higher Education*, vol. 7, no. 3, p. 166, 2014.

[95] A. Solomon, D. Moon, A. L. Roberts, and J. E. Gilbert, "Not just Black and not just a woman: Black women belonging in computing," in *2018 Research on Equity and Sustained Participation in Engineering, Computing, and Technology (RESPECT)*, 2018, pp. 1–5.

[96] M. Ong, C. Wright, L. Espinosa, and G. Orfield, "Inside the double bind: A synthesis of empirical research on undergraduate and graduate women of color in science, technology, engineering, and mathematics," *Harvard Educational Review*, vol. 81, no. 2, pp. 172–209, 2011.

[97] D. R. Compeau and C. A. Higgins, "Computer self-efficacy: Development of a measure and initial test," *MIS Quarterly*, vol. 19, p. 189, 1995.

[98] V. Ramalingam and S. Wiedenbeck, "Development and validation of scores on a computer programming self-efficacy scale and group analyses of novice programmer self-efficacy," *Journal of Educational Computing Research*, vol. 19, no. 4, pp. 367–381, 1998.

[99] G.-Y. Lin, "Self-efficacy beliefs and their sources in undergraduate computing disciplines: An examination of gender and persistence," *Journal of Educational Computing Research*, vol. 53, no. 4, pp. 540–561, 2016.

[100] K. M. Whitcomb and C. Singh, "Underrepresented minority students receive lower grades and have higher rates of attrition across stem disciplines: A sign of inequity?" *International Journal of Science Education*, vol. 43, no. 7, pp. 1054–1089, 2021.

[101] E. Brynjolfsson and A. McAfee, *Race Against the Machine: How the Digital Revolution is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy.* Brynjolfsson and McAfee, 2011.

[102] D. Wilson, R. Bates, E. P. Scott, S. M. Painter, and J. Shaffer, "Differences in self-efficacy among women and minorities in STEM," *Journal of Women and Minorities in Science and Engineering*, vol. 21, no. 1, 2015.

[103] R. A. Miller, A. Vaccaro, E. W. Kimball, and R. Forester, "'It's dude culture': Students with minoritized identities of sexuality and/or gender navigating STEM majors," *Journal of Diversity in Higher Education*, vol. 14, no. 3, p. 340, 2021.

[104] J. S. Hyde, R. S. Bigler, D. Joel, C. C. Tate, and S. M. van Anders, "The future of sex and gender in psychology: Five challenges to the gender binary," *American Psychologist*, vol. 74, no. 2, p. 171, 2019.

[105] C. A. Lindsay and C. M. Hart, "Exposure to same-race teachers and student disciplinary outcomes for Black students in North Carolina," *Educational Evaluation and Policy Analysis*, vol. 39, no. 3, pp. 485–510, 2017.

[106] S. Gershenson, C. M. Hart, J. Hyman, C. A. Lindsay, and N. W. Papageorge, "The long-run impacts of same-race teachers," *American Economic Journal: Economic Policy*, vol. 14, no. 4, pp. 300–342, 2022.

[107] Á. Fernandes and R. Duran, "When computing is mandatory: Sense of belonging and self-efficacy in elementary and secondary education," in *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 2*, 2023, pp. 1274–1274.

[108] Computing Research Association, "Data Buddies Survey: 2022 Annual Report," https://datavisualization.cra.org/DBS_annualReport_2022.html, July 2023.

[109] M. Tavakol and R. Dennick, "Making sense of Cronbach's alpha," *International Journal of Medical Education*, vol. 2, p. 53, 2011.

[110] Wang, Hansi Lo, "The U.S. census sees Middle Eastern and North African people as white. Many don't," https://www.npr.org/2022/02/17/1079181478/us-census-middle-eastern-white-north-african-mena, February 2022.

[111] V. Ojha, "Pre-registration: Computing self-efficacy in undergraduate CS students," https://osf.io/8r25g/?view_only=a9bdf1c6b91e46a59aeac26302902670, July 2023.

[112] C. D. Xavier Hall, C. V. Wood, M. Hurtado, D. A. Moskowitz, C. Dyar, and B. Mustanski, "Identifying leaks in the STEM recruitment pipeline among sexual and gender minority us secondary students," *PLoS One*, vol. 17, no. 6, 2022.

[113] N. Truszczynski, A. A. Singh, and N. Hansen, "The discrimination experiences and coping responses of non-binary and trans people," *Journal of Homosexuality*, vol. 69, no. 4, pp. 741–755, 2022.

[114] Z. Vue, C. Vang, N. Vue, V. Kamalumpundi, T. Barongan, B. Shao, S. Huang, L. Vang, M. Vue, N. Vang et al., "Asian Americans in STEM are not a monolith," *Cell*, vol. 186, no. 15, pp. 3138–3142, 2023.

[115] National Science Foundation, "Broadening Participation in Computing (BPC)," https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=13510, Accessed 2023-02-21.

[116] National Center for Education Statistics, "Degrees in computer and information sciences conferred by postsecondary institutions, by level of degree and sex of student: Academic years 1964-65 through 2020-21," https://nces.ed.gov/programs/digest/d22/tables/dt22_325.35.asp, 2022, Accessed 2024-03-27.

[117] J. Goode, "Increasing diversity in K-12 computer science: Strategies from the field," in *Proceedings of the 39th SIGCSE Technical Symposium on Computer Science Education*, 2008, pp. 362–366.

[118] K. Aguar, H. R. Arabnia, J. B. Gutierrez, W. D. Potter, and T. R. Taha, "Making CS inclusive: An overview of efforts to expand and diversify CS education," in *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, 2016, pp. 321–326.

[119] J. Wang and S. Hejazi Moghadam, "Diversity barriers in K-12 computer science education: Structural and social," in *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, 2017, pp. 615–620.

[120] S. Gretter, A. Yadav, P. Sands, and S. Hambrusch, "Equitable learning environments in K-12 computing: Teachers' views on barriers to diversity," *ACM Transactions on Computing Education (TOCE)*, vol. 19, no. 3, pp. 1–16, 2019.

[121] A. Peplow, J. Carter, J. Baumgartner, M. Hennessy, M. Greer, S. Schlembach, B. Mallory, and B. Refaei, "Transparent assignment design: A multidisciplinary survey assessing students' perceptions," *The Journal for Research and Practice in College Teaching*, vol. 6, no. 1, pp. 61–102, 2021.

[122] T. O. Howard, M.-A. Winkelmes, and M. Shegog, "Transparency teaching in the virtual classroom: Assessing the opportunities and challenges of integrating transparency teaching methods with online learning," *Journal of Political Science Education*, vol. 16, no. 2, pp. 198–211, 2020.

[123] M.-A. Winkelmes, "TILT Higher Ed," https://tilthighered.com/, 2023, accessed: 2024-02-05.

[124] R. Kaupp, "Online penalty: The impact of online instruction on the Latino-White achievement gap," *Journal of Applied Research in the Community College*, vol. 19, no. 2, pp. 3–11, 2012.

[125] D. Xu and S. S. Jaggars, "Performance gaps between online and face-to-face courses: Differences across types of students and academic subject areas," *The Journal of Higher Education*, vol. 85, no. 5, pp. 633–659, 2014.

[126] C. Wladis, K. M. Conway, and A. C. Hachey, "The online STEM classroom — who succeeds? An exploration of the impact of ethnicity, gender, and non-traditional student characteristics in the community college context," *Community College Review*, vol. 43, no. 2, pp. 142–164, 2015.

[127] H. A. Giroux and A. N. Penna, "Social education in the classroom: The dynamics of the hidden curriculum," *Theory & Research in Social Education*, vol. 7, no. 1, pp. 21–42, 1979.

[128] B. Smith, *Mentoring At-Risk Students through the Hidden Curriculum of Higher Education*. Lexington Books, 2013.

[129] J. V. Orón Semper and M. Blasco, "Revealing the hidden curriculum in higher education," *Studies in Philosophy and Education*, vol. 37, pp. 481–498, 2018.

[130] G. Koutsouris, A. Mountford-Zimdars, and K. Dingwall, "The 'ideal' higher education student: Understanding the hidden curriculum to enable institutional change," *Research in Post-Compulsory Education*, vol. 26, no. 2, pp. 131–147, 2021.

[131] V. Sellers and I. Villanueva Alarcón, "What strategies do diverse women in engineering use to cope with situational hidden curriculum?" in *American Society of Engineering Education*, 2021.

[132] A. Bandura and N. E. Adams, "Analysis of self-efficacy theory of behavioral change," *Cognitive Therapy and Research*, vol. 1, no. 4, pp. 287–310, 1977.

[133] S.-F. Tam, "Self-efficacy as a predictor of computer skills learning outcomes of individuals with physical disabilities," *The Journal of Psychology*, vol. 130, no. 1, pp. 51–58, 1996.

[134] R. Klassen, "A question of calibration: A review of the self-efficacy beliefs of students with learning disabilities," *Learning Disability Quarterly*, vol. 25, no. 2, pp. 88–102, 2002.

[135] T. L. Strayhorn, *College Students' Sense of Belonging: A Key to Educational Success for All Students*. Routledge, 2018.

[136] N. Veilleux, R. Bates, C. Allendoerfer, D. Jones, J. Crawford, and T. Floyd Smith, "The relationship between belonging and ability in computer science," in *Proceeding of the 44th ACM Technical Symposium on Computer Science Education*, 2013, pp. 65–70.

[137] K. Rainey, M. Dancy, R. Mickelson, E. Stearns, and S. Moller, "Race and gender differences in how sense of belonging influences decisions to major in STEM," *International Journal of STEM Education*, vol. 5, pp. 1–14, 2018.

[138] R. Campbell-Montalvo, G. Kersaint, C. A. Smith, E. Puccia, J. Skvoretz, H. Wao, J. P. Martin, G. MacDonald, and R. Lee, "How stereotypes and relationships influence women and underrepresented minority students' fit in engineering," *Journal of Research in Science Teaching*, vol. 59, no. 4, pp. 656–692, 2022.

[139] D. R. Means and K. B. Pyne, "Finding my way: Perceptions of institutional support and belonging in low-income, first-generation, first-year college students," *Journal of College Student Development*, vol. 58, no. 6, pp. 907–924, 2017.

[140] S. D. Museus and T.-H. Chang, "The impact of campus environments on sense of belonging for first-generation college students," *Journal of College Student Development*, vol. 62, no. 3, pp. 367–372, 2021.

[141] B. Blaser and R. E. Ladner, "Why is data on disability so hard to collect and understand?" in *2020 Research on Equity and Sustained Participation in Engineering, Computing, and Technology (RESPECT)*, vol. 1. IEEE, 2020, pp. 1–8.

[142] A. Wigfield and J. S. Eccles, "Expectancy–value theory of achievement motivation," *Contemporary Educational Psychology*, vol. 25, no. 1, pp. 68–81, 2000.

[143] A. Vaccaro, M. Daly-Cano, and B. M. Newman, "A sense of belonging among college students with disabilities: An emergent theoretical model," *Journal of College Student Development*, vol. 56, no. 7, pp. 670–686, 2015.

[144] A. Vaccaro and B. M. Newman, "Development of a sense of belonging for privileged and minoritized students: An emergent model," *Journal of College Student Development*, vol. 57, no. 8, pp. 925–942, 2016.

[145] A. Vaccaro and B. M. Newman, "A sense of belonging through the eyes of first-year LGBPQ students," *Journal of Student Affairs Research and Practice*, vol. 54, no. 2, pp. 137–149, 2017.

[146] A. Vaccaro, M. A. Marcotte, H. J. Swanson, and B. M. Newman, "Using familial funds of knowledge to transition and develop a sense of belonging: Reflections from first-generation college women of color," *Journal of the First-Year Experience & Students in Transition*, vol. 31, no. 2, pp. 29–44, 2019.

[147] A. Vaccaro, H. J. Swanson, M. A. Marcotte, and B. M. Newman, "Insights into the sense of belonging from women of color," *Journal Committed to Social Change on Race and Ethnicity (JCSCORE)*, vol. 5, no. 2, pp. 33–65, 2019.

[148] A. Vaccaro and B. Newman, “Theoretical foundations for sense of belonging in college,” in *The Impact of a Sense of Belonging in College*, E. Bentrim and G. W. Henning, Eds. New York: Routledge, 2022, ch. 1, pp. 3–20.

[149] V. Ojha, “Pre-registration: Transparent teaching practices in CS courses, self-efficacy, and sense of belonging,” https://osf.io/ebqdg/?view_only=0b1687e637034c1492c63e9d23331225, February 2024.

[150] M. C. Howard, “A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve?” *International Journal of Human-Computer Interaction*, vol. 32, no. 1, pp. 51–62, 2016.

[151] J. J. Vaske, J. Beaman, and C. C. Sponarski, “Rethinking internal consistency in Cronbach’s alpha,” *Leisure Sciences*, vol. 39, no. 2, pp. 163–173, 2017.

[152] L. Marshak, T. Van Wieren, D. R. Ferrell, L. Swiss, and C. Dugan, “Exploring barriers to college student use of disability services and accommodations,” *Journal of Postsecondary Education and Disability*, vol. 22, no. 3, pp. 151–165, 2010.

[153] C. Toutain, “Barriers to accommodations for students with disabilities in higher education: A literature review,” *Journal of Postsecondary Education and Disability*, vol. 32, no. 3, pp. 297–310, 2019.

[154] S. Cheryan, V. C. Plaut, P. G. Davies, and C. M. Steele, “Ambient belonging: How stereotypical cues impact gender participation in computer science,” *Journal of Personality and Social Psychology*, vol. 97, no. 6, p. 1045, 2009.

[155] N. Ensmenger, “‘Beards, sandals, and other signs of rugged individualism’: Masculine culture within the computing professions,” *Osiris*, vol. 30, no. 1, pp. 38–65, 2015.

APPENDIX A: PERCEPTIONS OF AI & CYBERSECURITY INTERVIEW QUESTIONS

Note: Below, we provide only those questions from the interview protocol that were relevant to our study.

A.1 MATH

- Which subfields of CS is math important in?

A.2 NARRATIVES WITHIN CS

- Which areas of CS have you heard called easiest and which are the hardest? Why?
- Which areas of CS have you heard called the most prestigious? Why?
- Which areas of CS have you heard called the most rigorous? Why?
- Which areas of CS are the most likely to help you have a positive impact on society? Why?

A.3 PERCEPTIONS OF AI

- Would you take or have you taken [AI courses at our institution]?
  - Why/why not?
- What does the day-to-day work of AI look like?
- What skills does it require?
- What stereotypes have you heard about the people who work in it?
  - What type of person do you think would be encouraged to pursue AI?
  - Where do you think you've heard this?
- Do you think AI should be a required course in your CS curriculum?
  - If yes, which class would you remove to put in AI?

A.4 PERCEPTIONS OF CYBERSECURITY

- Would you take or have you taken [cybersecurity courses at our institution]?
  - Why/why not?
- What does the day-to-day work of cybersecurity look like?
- What skills does it require?
- What stereotypes have you heard about the people who work in it?
  - What type of person do you think would be encouraged to pursue cybersecurity?
  - Where do you think you've heard this?
- Do you think cybersecurity should be a required course in your CS curriculum?
  - If yes, which class would you remove to put in cybersecurity?

APPENDIX B: PERCEPTIONS OF AI & CYBERSECURITY CODING SCHEME

Note: Below, we provide the categories and codes used in the qualitative analysis of interview transcripts. We limit ourselves to the categories and codes relevant to and used in this paper, rather than the broader study. For each code, we provide a brief description of the types of excerpts that would be categorized with that code.

B.1 CATEGORY: PERCEPTION OF AI

- *Chasing small gains:* Participants expressing the belief that progress in AI is made in very small increments.
- *Cool/Popular:* Participants expressing the belief that AI is cool, trendy, and/or popular.
- *Everyone should know a bit:* Participants expressing the belief that everyone would benefit from understanding the basics of AI.
- *Evil:* Participants expressing the belief that AI has potential for "evil".
- *Impact on society:* Participants discussing any aspects of how they perceive the impact of AI on society.
- *Intimidating:* Participants expressing the belief that AI is intimidating.
- *Lots of problems to solve:* Participants expressing the belief that there are many problems that can be solved using AI, or discussing which problems these might be.
- *Money:* Participants expressing the belief that professionals can make a lot of money working in AI.
- *Needs creativity:* Participants expressing the belief that work in AI requires creativity.
- *Not for everyone:* Participants expressing the belief that everyone cannot do AI.
- *Requires background knowledge:* Participants expressing the belief that studying AI requires already having a lot of background knowledge, or discussing the background knowledge they believe is required.

- *Smart:* Participants expressing the belief that studying AI requires being inherently smart.
- *Stats/math:* Participants expressing the belief that studying or working in AI entails using mathematical or statistical knowledge.
- *The average work day:* Participants discussing any aspects of what they believed was an average work day for an AI professional.
- *"Trash papers":* Participants expressing the belief that research in AI produces papers of low quality.
- *Useful:* Participants expressing the belief that AI can be useful, or discussing any aspects of how AI can be useful.
- *What is valuable:* Participants discussing any aspects of what aspects of AI are useful or valuable.
- *Who does AI:* Participants discussing any aspects of what kind of person (e.g., gender or personality traits) does work in AI.
- *Why it's now cool* Participants discussing why they think AI is cool/popular.

B.2 CATEGORY: PERCEPTION OF SECURITY

- *Cool/Popular:* Participants expressing the belief that cybersecurity is cool, trendy, and/or popular.
- *Everyone should know a bit:* Participants expressing the belief that everyone would benefit from understanding the basics of cybersecurity.
- *Impact on society:* Participants discussing any aspects of how they perceive the impact of cybersecurity on society.
- *Intimidating:* Participants expressing the belief that cybersecurity is intimidating.
- *Lots of jobs:* Participants expressing the belief that there are many jobs available in cybersecurity.
- *Low-level:* Participants expressing the belief that cybersecurity work is all low-level, systems programming.

- *Not fundamental:* Participants expressing the belief that cybersecurity knowledge is not fundamental to computing.
- *Requires background knowledge:* Participants expressing the belief that studying cybersecurity requires already having a lot of background knowledge, or discussing the background knowledge they believe is required.
- *Smart:* Participants expressing the belief that studying cybersecurity requires being inherently smart.
- *Stats/math:* Participants expressing the belief that studying or working in cybersecurity entails using mathematical or statistical knowledge.
- *The average work day:* Participants discussing any aspects of what they believed was an average work day for a cybersecurity professional.
- *Time-consuming:* Participants expressing the belief that work in cybersecurity is time-consuming and/or tedious.
- *Useful:* Participants expressing the belief that cybersecurity can be useful, or discussing any aspects of how cybersecurity can be useful.
- *Who does security:* Participants discussing any aspects of what kind of person (e.g., gender or personality traits) does work in cybersecurity.

B.3 CATEGORY: THE HARDEST SUBFIELD

- *AI:* Participants expressing the belief, or having heard others say, that AI is one of the hardest subfields in computing.
- *ML:* Participants expressing the belief, or having heard others say, that ML is one of the hardest subfields in computing.

B.4 CATEGORY: THE PRESTIGIOUS SUBFIELD

- *AI:* Participants expressing the belief, or having heard others say, that AI is one of the most prestigious subfields in computing.
- *ML:* Participants expressing the belief, or having heard others say, that ML is one of the most prestigious subfields in computing.

APPENDIX C: PHD EXPECTATIONS INTERVIEW QUESTIONS

C.1 INTRODUCTION

- What year in your PhD are you? What kind of research do you do?
- What made you pursue a graduate degree?

C.2 ADVISOR RELATIONSHIP

- How often do you meet with your advisor? What is the format of these meetings? (E.g. status updates, big-picture discussions about your degree progress, help with debugging)
- How has this changed during the pandemic?
- Do you talk about non-academic things? E.g. things happening in the world, things happening in your lives?
- In what ways does your advisor try to make your lab/research group an inclusive space?
- In what ways do you think aspects of your identity have affected your interactions with your advisor?
- Have you ever had to discuss mental health issues with your advisor? If so, how did that go?

C.3 EXPECTATIONS

- Are there differences between what you expect to achieve during your PhD, and what your advisor expects?
- What would you say is the source of your expectations? (E.g. parents' graduate degrees, meeting grad students in undergrad, movies/books)
- Have any of the things you are expected to do during the PhD surprised you? (E.g. didn't think I would be expected to get a publication every year)
- What do you think it means to be a "good researcher"? How does this compare to how your advisor defines it?

- How do you work towards becoming a "good researcher"?
- Are there differences in how you measure your own progress and how your advisor does the same?
- Have these "standards" changed over time, e.g. during COVID?

C.4 LEARNING/ACQUIRING SKILLS

- In what ways has your advisor helped you pick up new skills, such as teaching and research?
- Did they point you to specific resources?
- If not, how did you first learn about these resources?

C.5 CONCLUSION

- Is there anything else regarding your experience as an underrepresented CS PhD student that you would like to bring up?

APPENDIX D: PHD EXPECTATIONS CODING SCHEME

Note: Below, we provide the categories and codes used in the qualitative analysis of interview transcripts. We limit ourselves to the categories and codes relevant to and used in this paper, rather than the broader study. For each code, we provide a brief description of the types of excerpts that would be categorized with that code.

D.1 CATEGORY: POSITIVE ADVISOR ACTIONS

- *Academic:* Participants sharing an action their advisor took that was helpful academically.
- *Non-academic:* ..Participants sharing an action their advisor took that was helpful non-academically.

D.2 CATEGORY: NEGATIVE ADVISOR ACTIONS

- *Academic:* Participants sharing an action their advisor took that was unhelpful academically.
- *Non-academic:* Participants sharing an action their advisor took that was unhelpful non-academically.

D.3 CATEGORY: EXPECTATIONS

- *From yourself:* Participants sharing expectations they have for themselves.
- *From advisor:* Participants sharing expectations their advisors have for them, either explicitly stated, implied, , or based on observation.
- *From family:* Participants sharing expectations their family members have for them, either explicitly stated or implied.
- *From peers:* Participants sharing expectations their peers have for them, either explicitly stated, implied, or based on observation.
- *Surprises:* Participants sharing aspects of the PhD and what they were expected to do that surprised them.

- *Prior beliefs:* Participants sharing what they had expected of the PhD prior to beginning the program.

D.4 CATEGORY: PRE-GRAD SCHOOL EXPERIENCE

- *Research experiences:* Participants discussing research experiences they had prior to beginning the PhD.
- *Family role models:* Participants discussing role models they had among their family members as pertains to graduate school and/or research.
- *Key mentors:* Participants discussing important mentors they had prior to beginning the PhD.
- *Luck:* Participants discussing the role of luck in their experiences prior to beginning the PhD.
- *Work experience:* Participants discussing work/industry experiences they had prior to beginning the PhD.

D.5 CATEGORY: CHALLENGES

- *Academic:* Participants discussing academic challenges they have faced during the PhD.
- *Non-academic:* Participants discussing non-academic challenges they have faced during the PhD.

D.6 CATEGORY: MEASURING PROGRESS

- *Milestones:* Participants discussing measuring their own progress, or that their advisor measures their progress, by the milestone exams they have to take (qualifying, preliminary, and final).
- *Publishing papers:* Participants discussing measuring their own progress, or that their advisor measures their progress, by the publication of research papers.
- *Weekly tasks:* Participants discussing measuring their own progress, or that their advisor measures their progress, by the completion of weekly tasks assigned to them.

D.7 CATEGORY: IMPACT OF IDENTITY

A catch-all category used for any discussion of how students' identities impact their experiences in the PhD.

D.8 CATEGORY: ADVISOR RELATIONSHIP

- *Inclusive lab:* Participants indicating a belief that their lab is inclusive.
- *Meeting dynamics:* Participants discussing any aspects of the dynamics of their meetings with their advisor.
- *Shared identities:* Participants discussing how sharing an identity with their advisor (e.g., if student and advisor both identify as women) has affected their relationship.
- *Identity disconnect:* Participants discussing how aspects of their identity not shared with their advisor (e.g., a Black student and an Asian advisor) affect their relationship.
- *Changes over time:* Participants discussing how their relationship with their advisor has changed over time.
- *Comfort:* Participants discussing how comfortable (or not) they felt with their advisor.
- *"Like a sage":* Participants indicating that their advisor provides them with high-level guidance, rather than concrete support.
- *Pressure:* Participants indicating that they feel pressure from their advisor, both intentional and unintentional.

D.9 CATEGORY: STRATEGIES

- *Acquiring research skills:* Participants discussing specific ways they tried to acquire the skills to do research.
- *Asking for help:* Participants discussing reaching out to someone (advisors, peers, or others) for support.
- *Senior student mentor:* Participants discussing reaching out to a specific senior student who acted as a mentor for support.

APPENDIX E: CATEGORIZATION OF DATA BUDDIES SURVEY ITEMS

E.1 CATEGORIZATION OF PRIOR COMPUTING EXPERIENCE

Survey question: Which of the following experiences did you have prior to entering an undergraduate program? Select all that apply.

Table E.1: Survey options pertaining to students' prior computing experience and their categorization.

Survey options	**Category**
Took AP Computer Science A Took AP Computer Science AB Took AP Computer Science Principles Completed an online course related to computing (e.g., MOOC) Learned a computer programming language	Substantial experience
Engaged in software or hardware related projects Took part in student groups related to computing Attended a workshop or other training in computing (e.g., through your local library, community center, etc.)	Some experience
Took other AP courses Took dual enrollment courses None of the above	No experience

E.2 CATEGORIZATION OF INSTITUTIONAL CS REQUIREMENT

Survey question: Does your institution require a computing course for any non-computing major?

Table E.2: Survey options pertaining to departments' course requirement policies and their categorization.

Survey options	**Category**
All non-computing majors require a computing course	Required for all
Some (not all) non-computing majors require a computing course Computing course is not required, but is one of short list of options to meet a requirement Both of the above are true	Required for some
We have no such requirement	No requirement

E.3 STUDENT DISABILITY OPTIONS

Survey question: What type of disability do you have? Please check all that apply.
Survey options:

- No disability
- Attention deficity
- Autism Spectrum Disorder
- Intellectual Disability
- Deaf/Hard of Hearing
- Mental health disability
- Mobility or Orthopedic Disability
- Speech or Language Disability
- Learning/Specific Learning Disability
- Traumatic Brain Injury/Head Injury
- Visual Disability
- Nerve Damage
- Other
- More than one disability
- Unknown disability
- Chronic illness

E.4 CATEGORIZATION OF STUDENTS' MAJORS

Note: The Data Buddies survey contained different options for majors that students could select in the different years of survey data. The table below captures the options from the 2020 survey onwards.

Table E.3: Survey options pertaining to students' major(s) and their categorization.

Survey options	**Category**
Computer Science Computer Information Systems/Informatics Computing and Business Information Technology Computer Engineering Software Engineering Electrical Engineering and Computer Science (EECS) Electrical and Computer Engineering (ECE) Other computing & technology major	CS Major
Bioinformatics/Computational biology Data Science/Data Analytics Game Design A biological science major A mathematics major A physical sciences major Other engineering major	STEM Major
An arts and humanities major A business major A professional major A social science major An education major Other major	Non-STEM Major

APPENDIX F: SELF-EFFICACY CORRELATIONS

Below, we provide correlations between the variables used in our model. SE represents students' computing self-efficacy, a continuous variable. The remaining nominal variables represent whether a student:

- has substantial prior experience (sbE),
- has some prior experience (smE),
- has no prior experience (noE),
- identifies as a woman (wm),
- identifies as non-binary (nb),
- identifies as a man (mn),
- identifies as Black (bk),
- identifies as Native (nt),
- identifies as Hispanic (hp),
- identifies as Asian (as), and
- identifies as white (wt).

F.1 CORRELATIONS BETWEEN SELF-EFFICACY AND NOMINAL VARIABLES

In Table F.1, we provide point biserial correlations between self-efficacy, SE, and all the nominal variables used in our model.

Table F.1: Point biserial correlations between self-efficacy and nominal variables used in our models.

	SE
sbE	0.20
smE	−0.05
noE	−0.19
wm	−0.15
nb	−0.03
mn	0.16
bk	−0.06
nt	−0.04
hp	−0.02
as	−0.08
wt	0.11

F.2 OVERLAP BETWEEN NOMINAL VARIABLE CATEGORIES

In Table F.2, for each column x and row y, we provide the proportion of students in category x that are also in category y. For example, the cell in column wm and row sbE represents the proportion of *women* who indicated having substantial prior experience. On the other hand, the cell in column sbE and row wm represents the proportion of *students with substantial prior experience* who indicated identifying as women. Thus, the column for each category can be used to find the proportion of students in the intersection between that category and every other category. Note that racial and ethnic groups are not mutually exclusive, as students may identify with one or more races and ethnicities.

Table F.2: Overlap between nominal variable categories used in our models.

	sbE	*smE*	*noE*	*wm*	*nb*	*mn*	*bk*	*nt*	*hp*	*as*	*wt*
sbE	1	0	0	0.66	0.75	0.74	0.61	0.66	0.65	0.76	0.70
smE	0	1	0	0.05	0.05	0.05	0.08	0.05	0.07	0.04	0.06
noE	0	0	1	0.29	0.21	0.21	0.32	0.30	0.28	0.20	0.25
wm	0.32	0.34	0.42	1	0	0	0.36	0.25	0.30	0.41	0.30
nb	0.03	0.02	0.02	0	1	0	0.03	0.08	0.04	0.02	0.03
mn	0.65	0.63	0.55	0	0	1	0.61	0.67	0.67	0.57	0.66
bk	0.06	0.10	0.09	0.07	0.08	0.07	1	0.14	0.08	0.01	0.02
nt	0.02	0.02	0.03	0.02	0.07	0.02	0.04	1	0.09	0.01	0.02
hp	0.11	0.16	0.14	0.10	0.15	0.13	0.14	0.48	1	0.03	0.12
as	0.41	0.28	0.33	0.46	0.29	0.35	0.05	0.23	0.08	1	0.07
wt	0.54	0.61	0.58	0.49	0.64	0.58	0.17	0.56	0.58	0.10	1

Note: Each cell represents the proportion of students *in the column category* who *also* identified with the row category. Categories representing experience level and gender are mutually exclusive. For the names of each variable, please refer to the beginning of this appendix.

APPENDIX G: INSTRUCTIONAL TRANSPARENCY FACTORS

In Table G.1, we provide the survey items used in our measures of students' perception of instructional transparency, self-efficacy, and sense of belonging. For each item, we provide its mean and standard deviation on a 5-point scale as well as its factor loading in the relevant factor.

Table G.1: Survey items used to measure students' perception of instructional transparency, self-efficacy, and sense of belonging.

Item	**Mean**	**Standard deviation**	**Loading**
Perception of instructional transparency ($n = 11,046$)			
In these courses, I know the purpose of each assignment.	4.4	0.91	0.67
My instructors identify a specific learning goal for each assignment.	4.1	1.14	0.72
Each assignment includes a detailed set of instructions for completing it.	4.1	1.08	0.73
I know how my work would be evaluated.	4.1	1.11	0.68
My instructors provide students with annotated examples of past students' work.	2.1	1.38	0.24
Computing self-efficacy ($n = 10,973$)			
I am confident that I can learn the foundations and concepts of computing.	4.4	0.82	0.89
I am confident that I can quickly learn a new programming language on my own.	3.8	1.11	0.56
I am confident that I can pass my computing courses.	4.3	0.88	0.82
Sense of belonging in computing ($n = 10,982$)			
I feel like I belong in computing.	3.7	1.15	0.83
I feel like an outsider in computing. *(Reverse-coded)*	3.4	1.23	0.66
I feel welcomed in computing.	3.7	1.01	0.69

APPENDIX H: INSTRUCTIONAL TRANSPARENCY CORRELATIONS

Below, we provide correlations between the variables used in our models. TP represents students' perception of instructional transparency, SE represents students' computing self-efficacy, and SoB represents students' sense of belonging in computing, all of which are continuous variables. The remaining nominal variables represent whether a student:

- has substantial prior experience (sbE),
- has some prior experience (smE),
- has no prior experience (noE),
- identifies as a woman (wm),
- identifies as a man (mn),
- identifies as a first-generation college student (fg),
- identifies as a continuing-generation college student (cg),
- identifies as disabled (db),
- identifies as not disabled (nd),
- identifies as Black (bk),
- identifies as Native (nt),
- identifies as Hispanic (hp),
- identifies as Asian (as), and
- identifies as white (wt).

H.1 CORRELATIONS BETWEEN CONTINUOUS VARIABLES

In Table H.1, we provide the Pearson correlation between students' perception of instructional transparency (TP), their self-efficacy (SE), and their sense of belonging (SoB).

Table H.1: Pearson correlations between the continuous variables used in our models.

	TP	SE	SoB
TP	1		
SE	0.27	1	
SoB	0.24	0.55	1

H.2 CORRELATIONS BETWEEN CONTINUOUS AND NOMINAL VARIABLES

In Table H.2, we provide point biserial correlations between perception of transparency (TP), self-efficacy (SE), sense of belonging (SoB), and all the nominal variables used in our model.

Table H.2: Point biserial correlations between the continuous and nominal variables used in our models.

	TP	SE	SoB
sbE	0.03	0.24	0.25
smE	-0.00	-0.04	-0.02
noE	-0.03	-0.22	-0.25
wm	-0.04	-0.15	-0.27
mn	0.04	0.15	0.27
db	-0.09	-0.05	-0.10
nd	0.09	0.05	0.09
fg	-0.04	-0.10	-0.03
cg	0.04	0.10	0.03
bk	-0.02	-0.05	-0.02
nt	-0.01	-0.03	-0.01
hp	-0.04	-0.01	0.01
as	0.04	-0.02	0.01
wt	-0.00	0.08	0.02

H.3 OVERLAP BETWEEN NOMINAL VARIABLE CATEGORIES

In Table H.3, for each column x and row y, we provide the proportion of students in category x that are also in category y. For example, the cell in column wm and row sbE represents the proportion of *women* who indicated having substantial prior experience. On the other hand, the cell in column sbE and row wm represents the proportion of *students with substantial prior experience* who indicated identifying as women. Thus, the column for each category can be used to find the proportion of students in the intersection between that category and every other category. Note that racial and ethnic groups are not mutually exclusive, as students may identify with one or more races and ethnicities.

Table H.3: Overlap between nominal variable categories used in our models.

	sbE	*smE*	*noE*	*wm*	*mn*	*db*	*nd*	*fg*	*cg*	*bk*	*nt*	*hp*	*as*	*wt*
sbE	1	0	0	0.62	0.71	0.66	0.68	0.56	0.71	0.56	0.57	0.59	0.73	0.66
smE	0	1	0	0.05	0.06	0.07	0.05	0.07	0.05	0.09	0.07	0.09	0.04	0.06
noE	0	0	1	0.33	0.24	0.27	0.27	0.36	0.24	0.36	0.36	0.32	0.23	0.28
wm	0.33	0.33	0.44	1	0	0.42	0.34	0.34	0.37	0.37	0.29	0.31	0.42	0.32
mn	0.67	0.67	0.56	0	1	0.58	0.66	0.66	0.63	0.63	0.71	0.69	0.58	0.68
db	0.20	0.25	0.20	0.24	0.18	1	0	0.20	0.20	0.19	0.33	0.23	0.12	0.27
nd	0.80	0.75	0.80	0.76	0.82	0	1	0.80	0.80	0.81	0.67	0.77	0.88	0.73
fg	0.22	0.34	0.35	0.24	0.27	0.26	0.26	1	0	0.44	0.49	0.48	0.24	0.21
cg	0.79	0.66	0.65	0.76	0.73	0.74	0.74	0	1	0.56	0.51	0.52	0.76	0.79
bk	0.07	0.13	0.11	0.08	0.08	0.08	0.08	0.14	0.06	1	0.15	0.08	0.01	0.02
nt	0.01	0.02	0.02	0.01	0.02	0.03	0.01	0.03	0.01	0.03	1	0.07	0.01	0.01
hp	0.10	0.19	0.14	0.10	0.13	0.13	0.11	0.22	0.08	0.12	0.47	1	0.02	0.14
as	0.45	0.30	0.35	0.49	0.38	0.24	0.46	0.39	0.42	0.05	0.26	0.08	1	0.08
wt	0.50	0.57	0.53	0.45	0.55	0.68	0.47	0.41	0.55	0.12	0.43	0.61	0.10	1

Note: Each cell represents the proportion of students *in the column category* who *also* identified with the row category. Categories representing experience level, gender, disability status, and first-generation status are mutually exclusive. For the names of each variable, please refer to the beginning of this appendix.