Title

Compiling a Computing Identity: A Byte of Self-Efficacy, Belonging, and Other Predictive Factors

Author

Stormes, Kaitlyn Nicole

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Compiling a Computing Identity:

A Byte of Self-Efficacy, Belonging, and Other Predictive Factors

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Education

by

Kaitlyn Nicole Stormes

2024

ABSTRACT OF THE DISSERTATION

Compiling a Computing Identity:

A Byte of Self-Efficacy, Belonging, and Other Predictive Factors

by

Kaitlyn Nicole Stormes

Doctor of Philosophy in Education

University of California, Los Angeles, 2024

Professor Linda J. Sax, Chair

Despite efforts to increase representation among those enrolled, earning degrees, and working in the computing and technology industry, women across races/ethnicities and People of Color more broadly remain underrepresented in the field. Fortunately, extant literature has found that psychosocial factors like computing identity can help broaden participation for these groups by supporting their academic and career interests and persistence. However, a consistent and comprehensive measure of computing identity remains elusive.

This quantitative dissertation draws on a conceptual framework developed for Women of Color in computing and uses longitudinal survey data from 1,036 undergraduate students from one of 15 research universities across the U.S. to: (a) develop a reliable measure of computing identity and assess the role of self-efficacy and sense of belonging in that measure; (b) validate the measure across time and gender and racial/ethnic groups; and (c) identify college-related

factors influencing computing identity, considering variations across gender and racial/ethnic identities.

This dissertation provides a nuanced account of what computing identity is and how it forms among undergraduate students, noting the unique facilitators and impediments across gender and racial/ethnic groups. In particular, findings highlight that computing identity is a multidimensional measure that incorporates a student's self-identity as a computing person, their sense that computing is a part of their core personal identity, and their sense of belonging in the greater computing community. Hence, it is distinct from related concepts—interest, self-efficacy (competence/performance), and social recognition. While the construct of the computing identity measure is consistent across gender, racial/ethnic, and intersectional gender and racial/ethnic identities, it does not necessarily hold over time during college.

Further, while the baseline computing identity and academic major variables generally fostered greater computing identity scores, separate linear regression analyses revealed distinct predictors for different student groups, implying that a one-size-fits-all approach to fostering computing identity should be avoided. For instance, interest in computing was a salient predictor of men's computing identity but not women's, and self-efficacy was a positive predictor for underrepresented Students of Color but not for white students. This dissertation probes these and other findings, offering implications for enhancing practice and policy and advancing theory and research.

The dissertation of Kaitlyn N. Stormes is approved.

Jennifer M. Blaney

Mark Kevin Eagan

Kimberley Gomez

Kathleen J. Lehman

Linda J. Sax, Committee Chair

University of California, Los Angeles

2024

DEDICATION

This dissertation is wholeheartedly dedicated to the younger version of myself and all rising women in computing and first-generation scholars on their doctoral journeys worldwide.

A note from me to you: believe in yourself, even when it feels like no one believes in you. Consider all that you have already accomplished and all the possibilities that you can achieve. Recognize that while this dissertation is a big deal, you are much more of a person than your dissertation; it does not define you. You are whole with or without it, and you are already enough as you are.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ACKNOWLEDGMENTS

Without my greater community of support, which enabled me to successfully complete and defend this dissertation, I would not be where I am today. Your encouragement, support, and commitment to my success have been invaluable.

To my family, I am so grateful for every one of you. Mom, thank you for caring about each step of my PhD journey like you were on it, too. Dad, thank you for dropping everything at a moment's notice to support me when I needed it. To my sisters, Michelle and Shannon, I appreciate you showing up for me whether I asked for help or not. To my husband, Adam, thank you for your undivided love and support. I am so grateful to you for encouraging me to leave the desk occasionally, as going outside for fresh air and adventures allowed me to find a balance. I cannot wait to return the support you've given me as you embark on your own PhD journey.

To my mentors near and far, thank you for believing in me, encouraging me to go back to school to pursue this degree (Drs. Gold, Olson, and Manke), and helping me cross the finish line to complete it (Drs. Bolkan, Blazek, Lin, Gray, and George). A special shout out to my dissertation friends who became more like sisters—I know we will be lifelong friends.

To my dissertation committee at UCLA and beyond, your wisdom and expertise has been invaluable. To my PhD advisor and chair, Linda, I appreciate your generous support; you have challenged me to become a better researcher and helped shape my identity as a scholar. Kevin, your commitment to my success is unparalleled; thank you. Jennie, Kim, and Kate, I am grateful for your advice on how to make my work stronger, more critical, and better equipped to uplift the voices of women in this field. Finally, thank you to the UCLA Graduate Division, *Momentum* and Dr. Gershon in Ed&IS for helping me navigate funding for my education and to the UCLA CAPS team for helping me foster my well-being in my most difficult and rewarding journey.

VITA

EDUCATION

Master of Arts in Education, Higher Education and Organizational Change — 2020
University of California, Los Angeles (UCLA), Los Angeles, CA

Master of Arts in Psychology, Academic Research and Social Psychology — 2016
California Polytechnic State University, Humboldt (Cal Poly Humboldt), Arcata, CA
Thesis: *Advising Support for Students on Academic Probation*
Committee: Dr. Gregg Gold (Chair), Dr. Kathy Thornhill, Dr. Carrie Aigner, and Dr. Christopher Aberson

2015
Bachelor of Arts in Psychology
California Polytechnic State University, Humboldt (Cal Poly Humboldt), Arcata, CA
Honors: *summa cum laude*

RESEARCH EXPERIENCE

Research Analyst | University of California, Los Angeles (UCLA) — 2019–Present
Momentum: Accelerating Equity in Computing and Technology
Principal Investigator: Dr. Linda Sax; Co-PI: Dr. Kathleen Lehman

2020–2022
Senior Data Manager | California State University, Long Beach (CSULB)
Academic Affairs Faculty Research Teams
Principal Investigator: Dr. San Bolkan

Lead Research Technician | CSULB
Academic Affairs Faculty Research Teams
Principal Investigator: Dr. Kelly Young; Co-PI: Dr. Beth Manke

2018–2019
Senior Data Manager/Program Evaluator | CSULB
Building Infrastructure Leading to Diversity (BUILD) Program
Principal Investigator: Dr. Laura Kingsford; Co-PI: Dr. Guido Urizar

2016–2019
Data Analyst | CSULB
College of Liberal Arts and College of Education

2016–2019
Research Technician/Computer Programmer | CSULB
Institutional Research, Assessment, and Analytics

PUBLICATIONS

Refereed Journal Articles

Stormes, K. N., Streicker, N., Bowers, G., Urizar, G., & Ayala, P. (2022). Impact of undergraduate research-training programs: An illustrative example of finding a comparison group and evaluating academic and graduate school outcomes. *Scholarship and Practice of Undergraduate Research, 5*(3)*,* 25–36.

Young, K.A. & Stormes, K. N. (2020). The BUILD mentor community at CSULB: A mentor training program designed to enhance mentoring skills in experienced mentors. *Understanding Interventions, special issue, 11*(1), 1–20.

Bolkan, S., Pedersen, W. C., Stormes, K. N., & Manke, B. (2018). Predicting four-year graduation: Using social cognitive career theory to model the impact of prescriptive advising, unit load, and students' self-efficacy. *Journal of College Student Retention: Research, Theory & Practice, 22*(4) 1–21.

Research/Policy Briefs & Reports

Stormes, K. N., Rojas, L. (October, 2023). *Do graduation rates improve once universities become Hispanic-Serving Institutions?* Urban Institute. https://www.urban.org/research/publication/do-graduation-rates-improve-once-universities-become-hispanic-serving

Conference Proceedings

Lehman, K. J., Stormes, K. N., Smith, K. N., & Lapan, J. C. (2024). Sealing the Deal: Factors That Promote Computing Interns' Interest in Computing Careers. Proceedings of the ACM Special Interest Group on Computer Science Education (SIGCSE) Technical Symposium.

Sax, L. J., Nhien, C., & Stormes, K. N. (2024). A Quantitative Methodological Review of Research on Broadening Participation in Computing, 2005–2022. Proceedings of the ACM Special Interest Group on Computer Science Education (SIGCSE) Technical Symposium.

Book Chapters and Other Publications

Stormes, K. N., Lehman, K. J., Litzer, E., & Barker, L. (forthcoming) History of and Trends in Women's Participation in Academic Computing *in Women in Academia: Engineering and Science*. Edited by Jenna P. Carpenter. Springer Nature. Switzerland, AG.

Stormes, K. N., & Gold, G. J. (2016) Academic advising support for students on academic probation. *ideaFest: Interdisciplinary Journal of Creative Works and Research from Humboldt State University*, 1(5), 37–47.

Publications in Development

Sax, L. J., Stormes, K. N., & Pereyra, M. W. J. *The Role of Course Modality on Students' Computing Orientation and Course Satisfaction*.

Lehman, K. J., Stormes, K. N., Smith, K. N., & Lapan, J. C. *Exploring Participation in Computing Internships Among Undergraduates in Non-Computing Majors.*

Sax, L. J., Nhien, C., & Stormes, K. N. A *Quantitative Methodological Review of Research on Broadening Participation in Computing, 2005–2022*.

PREFACE

Take a moment to imagine a scientist. Does that image change when you picture a "computer" scientist? Decades of research show that when asked to "draw a _________ scientist," children and undergraduate students, regardless of gender and race/ethnicity, often depict scientists as white, intelligent, yet "nerdy" men with crazed white hair and glasses, often holding Erlenmeyer flasks (Miller et al., 2018). This general schema of what a scientist looks like tends to hold for all types of scientists, including those in computing (Public Broadcasting Service [PBS], 2021; Martin, 2004). Images such as these convey that the technology industry and broader STEM fields are not a place for women across races/ethnicities and People of Color more broadly (PBS, 2021). This is an important phenomenon because the ability to see oneself as a scientist is a key predictor of students' intent to major in STEM fields and various academic and career outcomes (e.g., Carlone & Johnson, 2007; Chang et al., 2011). Further, given the lack of gender and racial/ethnic diversity in some STEM disciplines, including computer science, fostering students' discipline-specific identity (i.e., computing identity) may be critical to increasing students' interest and persistence in the field (Lunn et al., 2021a; 2021c; Rodriguez et al., 2020; Rodriguez & Lehman, 2017).

Identity-based computing education research interests me due to personal experiences I faced early in my computing career. After becoming the first in my family to graduate from a four-year university, I worked as a data analyst and programmer. Excited for my new career, I was surprised to feel dissonance as I encountered misogynistic comments about my role as a woman in the workplace. I was talked over, ignored, and made to feel less intelligent than my male peers. I was frequently asked to take on administrative tasks and sit quietly at tables with university administrators. I was often not given credit for my work when it was shared with university administrative leaders. Finally, I learned that I earned a lower annual wage than my male peers with the same educational level and job title. Overall, my early career experiences made me feel frustrated that these acts of sexism in the workplace still occur today. Soon after, I began reflecting on other instances during my primary and secondary educations when I was made to feel (and later internalized) that I was not good enough as a student, let alone as a student of math, science, and computing.

Unfortunately, the experiences I encountered in that professional setting (and the resurgence of prior similar feelings from K-12 to higher education) made me question my ability to perform computing-related tasks and whether I could continue to see myself as a computing person. My privileged experience as a white woman also made me wonder about the experiences of Women of Color, particularly Black, Indigenous, and Latina women who have historically been underserved by higher education. From conversations with our only female computer science intern and through reading computing education literature, however, I learned that many women report similar experiences, and racially/ethnically underrepresented students in computing often also report experiences with racism and objectification. To support women in computing and serve as a mentor for how to combat these barriers, I began volunteering for a local Girls Who Code club. Despite my desire to fit in and support others through this trying climate, I ultimately convinced myself that I did not belong. In an effort for me (and others) to feel valued, respected, and heard, I left that role to research undergraduate women's experiences in STEM and, more specifically, computing.

In my next role as a senior evaluator and data manager for an undergraduate research-training program, I learned about the gender and racial/ethnic equity gaps in undergraduate STEM enrollment and degree attainment and the importance of diversifying pathways to graduate school and the workforce. I also learned how students' demographics and experiences before and during college affect their interest and persistence in STEM. With a better understanding of the data and the factors that facilitate (or impede) students' persistence in STEM, I was now able to name my experiences—a lack of sense of belonging, a decline in my computing self-efficacy, and the important role that others play in fostering my sense of identity as a computing person (i.e., my computing identity). Around the same time, I learned about the "draw a scientist" phenomenon, and I was both dismayed and intrigued to understand the phenomenon of how students come to think of themselves as scientists. With renewed excitement, but this time for studying STEM recruitment and retention efforts to foster students' psychosocial factors, I decided to pursue a PhD.

Given my gendered—yet-privileged—experience as a white woman in a programming role paired with my research experience, genuine interest in the topic, and commitment to STEM student success, I decided to dedicate my work to (a) understanding the experiences that women, especially those from marginalized racial/ethnic identities, face in male-dominant STEM disciplines like computing; and (b) exploring how to foster computing identity among students in these fields. Through my research platform, I want to help transform students' lives by promoting social justice and mobility.

This is my story, which led me to return to school and continues to fuel me to this day. In reflecting on my educational and career background that guided me to this point in my journey, I believe that anybody can solve problems in science, regardless of their gender identity, racial/ethnic identity, or any other social identity. As a researcher it is my job and my passion to study ways to foster greater student success in this regard. While this dissertation may be the start of my research agenda, I aspire to continue this work long after I become "Dr. Stormes."

Tapping into my background in psychology and my drive to better support women of all racial/ethnic identities and People of Color in general in computing, I bring you this dissertation. The forthcoming study focuses on fostering discipline-specific identity for all students in computing as a means to promote greater participation and persistence in computing and to serve as a reminder that “there is no problem in science that can be solved by a man that cannot be solved by a woman” (Rubin, 1996).

Positionality Statement

My intersectional identity as a white woman and personal experiences as a former computer programmer inherently inform the lens through which I approach and interpret this research. My identities and values inform the work that I do and how I situate findings for research and practice. Importantly, however, as I reconcile my own identity in this space, I am aware that I will never know what it is like to be an undergraduate Woman of Color in computing, so I continually practice interrogating the biases that I bring to this work. As described above, I stand alongside others who see the need to diversify computing and believe more must be done to create environments where all are welcome, especially those of marginalized gender and/or racial/ethnic identities (e.g., Daily & Eugene, 2013; Washington et al., 2019). In my research, I tend to take a critical feminist approach, aiming to unveil and dismantle the systems of oppression that women and Students of Color face in computing. I strive to avoid perpetuating the systems that default to white male dominance in computing education. In practice, this means that the questions I ask and the decisions I make regarding the inclusion of the study sample and how the analyses are carried out are and will continue to be informed by my personal experience and from scholarship that aligns with my approach to conducting said research. In doing so, methodologically, I opt for critical quantitative analytical methods whenever I have adequate sample sizes in the data. Admittedly, given the disproportion representation of non-binary students and students from historically marginalized racial/ethnic groups in undergraduate computing, this is not always feasible, even with data from multiple institutions. As I aim to honor students' gender and racial/ethnic identities, I strive to disaggregate social identity variables wherever possible.

Additionally, I examine important facets of students' pre-college and undergraduate experiences that foster their computing identity development separately by gender and race/ethnicity. Nevertheless, the lived experiences of gender nonbinary and racially/ethnically minoritized students deserve to be heard, and research that centers on their experiences is valued and needed. In this dissertation, I do what I can to honor all students' experiences. Still, additional research may be required to understand how all groups of students develop their identities in computing. Following this study's findings, I provide ideas for how other researchers and I can further this scholarship. Ultimately, I believe we can achieve equity in computing, and I remain hopeful that this dissertation is one step toward that goal.

CHAPTER ONE: INTRODUCTION

While the field of computing is traditionally thought of as male-dominated, previous research has in fact shown that women outnumbered their male counterparts in the field of computing prior to World War II (Shetterly, 2016; Thompson, 2019; Washington et al., 2019). In the post-war era of the 1950s, however, men's computational and programming skills were more highly regarded, which ultimately drove women out of these jobs[1]. Today, women in the aggregate comprise only 27% of technologists in this industry, and Women of Color make up only 14% (Anita Borg Institute, 2022; Ashcraft et al., 2016). Given that computer and information technology fields account for almost one in ten jobs in the U.S. workforce (Ariella, 2022; Flynn, 2022), these gender and racial/ethnic equity gaps in technology (U.S. Bureau of Labor Statistics [BLS], 2021) have serious implications. For example, the median annual wage for individuals in computer and information technology occupations is $97,430—a value more than double the median annual pay for those in all other occupations (BLS, 2022b). Additionally, employment in the technology sector is projected to grow by 15% in the next eight years, resulting in 682,800 new jobs (BLS, 2022b). The forthcoming growth in the industry's employment rates and the field's lucrative wages provide an important opportunity to increase the social mobility of those minoritized by their gender and/or racial/ethnic identities (Lockard & Wolf, 2012).

[1] During World War II, female "coders"—also known as "female 'human computers'"—outnumbered their male counterparts and often performed calculations by hand"'" (Ensmenger, 2012; Henn, 2014). Even as the first computers were created, Women of Color often did the early computer programming for private companies, NASA, and the U.S. military (Ensmenger, 2012). However, as the war ended and men returned to the workforce, computing began to be seen as a profession that relied upon critical thinking skills. As such, white women and Women of Color, while once needed for their expertise, were no longer considered equipped for the work, and the field of computing soon evolved to exclude them (Ensmenger, 2012).

The equity gaps in the U.S. tech workforce present both societal and individual-level implications. For instance, equitable representation can help lead to diverse perspectives and innovations, which in turn is an important way to help foster a boost in the economy and be competitive with the rest of the world (Ong, 2011; Peckham et al., 2007; Rodriguez et al., 2020; Washington et al., 2019). However, the lack of diversity in the tech industry has important implications and real-world ramifications when our inventions do not mirror the needs of our increasingly diverse population (Noble, 2018; Peng et al., 2019). Further, when women across races/ethnicities and People of Color more broadly are not considered in these technological advances, the innovations, in turn, limit the safety and well-being of all who rely on them. For instance, seatbelt technology was originally designed for the height and weight of males, and automatic hand dryers, soap dispensers, and facial recognition software did not originally acknowledge those with darker skin tones (Noble, 2018). There remains a critical need for the individuals who design technology to represent the population they serve; otherwise, technology will continue catering to certain groups and being inaccessible to all who rely on it (Peng et al., 2019).

Some scholars interested in addressing the equity gap in computing and technology have justifiably focused their research on workplace issues, including bias in hiring practices, promotion processes, and hostile work climates (e.g., Ashcraft et al., 2016; Whitney, 2021). However, these workplace inequities are, in fact, indicative of problems that begin much earlier on the pathway to the profession (Lane et al., 2021; Lunn et al., 2021c). As such, looking at undergraduate degree attainment rates is useful for understanding these pathways.

While earning a degree in computing[2] is a viable way to pursue a career in technology (BLS, 2022a), there are gender and racial/ethnic disparities in undergraduate enrollment and degree attainment rates as well (Zweben & Bizot, 2021). For instance, across all racial/ethnic groups, there is a greater representation of men than women among those who earn degrees in computing (see Figure 1.1 below; Lane et al., 2021; Lunn et al., 2021c; Ong, 2011; Zweben & Bizot, 2021). In 2020, 21% of all computing degrees were awarded to women, and only 11% were awarded to Women of Color[3] (WOC; NCSES, 2021). Notably, the percentage of degrees conferred in computing to *non-Asian* WOC is only 5% (Lane et al., 2021; Lunn et al., 2021c; NCSES, 2020; Zweben & Bizot, 2021).

As stark as these current equity gaps are in computing, they are not novel. In fact, over the last two decades, the share of men's degrees across all fields that were earned in computing has nearly doubled, but the rates for women have remained fairly stagnant (National Center for Education Statistics [NCES], 2018, 2020; Lunn et al., 2021a; Barr, 2018). Thus, from the early 1990s to today, the equity gaps in computing continue to grow.

[2] The term "computing" in the broad sense is used to represent subdisciplines within the field that incorporate programming components. The operational definition of this term varies across institutions and published studies. However, for this dissertation, computing represents a variety of students' undergraduate academic majors in technology (e.g., bioinformatics, computer science, computer information systems/informatics, business information management, management information systems, computer engineering, software engineering, data science, game design, information science/studies, information technology).

[3] The term "Women of Color" includes women doubly marginalized by race/racism and gender/sexism. Specifically, this term includes women who identify in the following racial/ethnic social identities: Asian American, American Indian or Alaska Native, Black or African American, Hispanic or Latina/o/x/e, Middle Eastern, Native Hawaiian or Other Pacific Islander, and those who marked two or more racial/ethnic categories (National Academies of Sciences, Engineering, and Medicine, 2017).

Figure 1.1

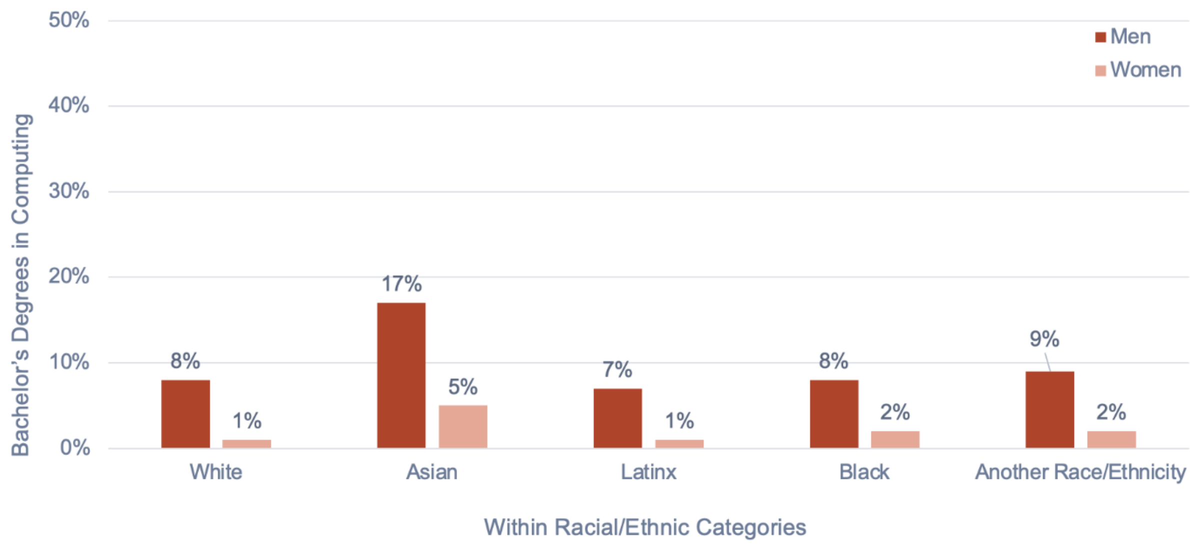
Traditional Analysis: Percentage of All Bachelor's Degrees in Computing Awarded to Students by Gender and Race/Ethnicity, 2019–20

Note. Calculations represent the percent of bachelor's degrees earned in computing by a particular gender and racial/ethnic group divided by all bachelor's degrees earned in computing in the given year. "All Other Racial/Ethnic Identities" represents Pacific Islander, American Indian/Alaska Native, and Two or More Races, excluding Non-resident students. Data derived from NCES IPEDS 2020, Digest of Education Statistics

Another way to consider the severity of these equity gaps in computing is to examine the number of bachelor's degrees earned in computer science (CS) for each group relative to all bachelor's degrees earned by that group (Barr, 2018). Figure 1.2 below depicts this approach for the most recently available degree completion data (NCSES IPEDS, 2020). Specifically, across all racial/ethnic groups, the share of degrees awarded in computing fields (relative to all fields) is far larger for men than for women. Among women, Asian women earn 5% of their bachelor's degrees in computing, the highest share reported for any group of women.

Figure 1.2

Cohort Analysis: Percentage of Bachelor's Degrees in Computing by Gender and Racial/Ethnic Group Relative to All Bachelor's Degrees by That Group



Note. Calculations represent the percent of degrees earned in computing to a particular gender and racial/ethnic group divided by all bachelor's degrees earned in all fields to that gender and racial/ethnic group in the given year. Data derived from the NCES IPEDS Completions Survey. "All Other Racial/Ethnic Identities the following groups: non-Hispanic American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Two or More Races, Other or unknown race or ethnicity, excluding temporary visa holders.

Taken together, the increase in the demand for computing professionals to meet the needs of the growing workforce and the low representation of certain groups in computing has created a paradox with important considerations that are yet to be addressed (Daily & Eugene, 2013). Indeed, despite long-term job prospects in tech, problems will persist if the computing graduates, job pool, industry, and inventions continue to cater to dominant social groups and their perspectives. These equity issues in computing underscore a critical need to increase the enrollment and degree attainment of undergraduate students who are marginalized via sexism (as women) and/or racism (as People of Color; Rodriguez et al., 2020). Thus, the need to broaden participation in computing (BPC) has long been a focus of researchers and policymakers due to the individual and societal consequences of persistent gender and racial/ethnic equity gaps (BPCnet, n.d.).

Research and Funding Efforts to Address the Equity Gaps in Computing

Historical trends demonstrate a shift in representation among those interested in, studying, earning degrees in, or working in computing; these trends sustain the interest of researchers and policymakers alike. However, despite BPC efforts and the boom in enrollment seen nationally, gender and racial/ethnic disparities remain in computing (CRA, 2017; Zweben & Bizot, 2021). As a result, several organizations have funded efforts to inspire, engage, and encourage more women and Students of Color in computing. These important organizations allocate resources and funding to support efforts to broaden participation and close equity gaps but vary in the communities they target (e.g., school-age children[4], undergraduate college students[5], graduate and doctoral students[6], the computing workforce[7], broadening participation for all women[8], Black women[9], Hispanic/Latine women and men[10], Native American Women[11], and Asian women[12]. One such effort was the Building Recruiting and Inclusion for Diversity

[4] Google, Code.org, Girls Who Code.
[5] CRA, Melinda French Gates' Pivotal Ventures, The National Center for Women & Information Technology [NCWIT], AnitaB.Org, and Harvey Mudd's BRAID initiative.
[6] CRA's Widening Participation Grad Cohort for Women.
[7] AnitaB.org.
[8] Systers, Status of Women in Computing Committee, Grace Hopper Celebration of Women in Computing, PWERE Program, NCWIT, Girls Who Code, and she++, PyLadies, Girl Develop It, and the not-for-profit Women in Technology organization later known as the Anita Borg Institute.
[9] Historically Black Colleges and Universities, Black Women in Computing, ABI, NSF-funded African American Mentoring in Computing Sciences, National Society for Black in Computing, BlackComputeHER conference, Black Girls Code.
[10] Hispanic-Serving Institutions, Latinos in Science and Engineering, the Society of Hispanic Professional Engineers, Hispanic Technology and Telecommunications Partnership, TechLatino and Latino Tech Summit, the NSF's Computing Alliance of Hispanic-serving Institutions initiative, and the Grace Hopper Celebration's Latinas in Computing group.
[11] Tribal Colleges and Universities and the Tribal Colleges and Universities Program, NSF's Tribal Colleges and Universities Program, and the South Dakota School of Mines and Technology's Tiospaye program.
[12] Asian American Native American and Pacific Islander Serving Institutions, Society of Asian Scientists and Engineers, and the China Computer Federation, STEM Women Asian; Chinese Women in Computing.

(BRAID) initiative. Drs. Telle Whitney (former CEO and president of Anita Borg Institute for Women and Technology) and Maria Klawe (former president of Harvey Hudd College)—two women computer scientists who led and continue to lead groundbreaking efforts to diversify technology—developed the BRAID initiative as an effort to broaden the participation of undergraduate women and Students of Color majoring in computer science. The BRAID initiative included 15 research universities across the U.S. that opted to participate and diversify their computing departments by providing more inclusive-centered pedagogy and student-centered support. Data from these institutions was used for this study.

Suffice it to say that diversifying technology matters to many, as evidenced by the abundance of and funding behind organizations supporting these efforts, yet the equity gaps remain. See Lunn et al. (2021c) for a recent publication on the historical overview of the known efforts to diversify computing for all women (both nationally and abroad) and a thorough timeline of when these groups were established. While more must be done to diversify technology in industry and academia (Washington et al., 2019), we can learn from existing research to expand our understanding of computing students and curate better support systems for these students.

What We Know from Existing Research

Various societal, structural, and individual-level explanations have been offered for the continued gender gap in STEM and, more specifically, in computing. Through myriad reasons—such as interactions with others, misrepresentations in the media, inequitable access and exposure to computers and technology, and uncomfortable racist and/or sexist experiences from primary through postsecondary school—marginalized students are often made to feel that computing spaces are unwelcoming and unsafe environments to pursue a degree and career in

(Barker et al., 2002; Blum & Freize, 2005; Daily et al., 2022; Diekman et al., 2010, 2011; Gallup-Amazon, 2021; Margolis & Fisher, 2002; NCES, 2021; Jhally & Kilbourne, 2010; PBS, 2021; Rodriguez et al., 2020; Tung, 2016; Scott et al., 2022; Sherman & Zurbriggen, 2014). Decades of scholarship document that students' perceptions of computing and the unwelcoming experiences mentioned above contribute to equity gaps among those who are interested and those who persist in the field (CRA, 2017; Daily et al., 2022). In addition to documenting the role played by environments, research in this area also highlights the important role of students' affective or psychosocial characteristics (e.g., Barker et al., 2002; Beyer, 2014; Robnett, 2013; Rodriguez & Lehman, 2017; Rodriguez et al., 2020; Sax et al., 2018). Namely, computing self-efficacy[13], sense of belonging[14] in computing, and computing identity[15] have been identified as important in predicting various educational and career-related computing outcomes.

The aforementioned psychosocial constructs vary in terminology and use. In particular, **computing self-efficacy**—or students' confidence regarding their ability to skillfully perform tasks needed to succeed in the field of computing—has been deemed valuable in that it is linked to undergraduate student persistence (Lehman et al., 2022), students' graduate school aspirations (Wofford et al., 2022), and their computing career aspirations (George et al., 2022). Additionally, students with a greater **sense of belonging in computing**—or the extent to which students feel they fit among those in the greater computing community—are more likely to enroll and persist in computing (e.g., Sax et al., 2018; Strayhorn, 2012). Finally, **computing identity** (the primary

[13] Computing self-efficacy is the extent to which a student believes in their ability to skillfully perform the necessary computing-related tasks required of someone in the field of computing (Bandura, 1977, 1986a, 1986b; Beyer, 2014; Compeau & Higgins, 1995; Lin, 2016).
[14] Sense of belonging in computing is the extent to which a student feels that they belong or fit in the computing community (Sax et al., 2018).
[15] Computing identity is the extent to which a student sees themselves as a computing person, based on how they make meaning of their core identity and navigate the world around them (Abes et al., 2007; Carlone & Johnson, 2007; Lunn et al., 201b; Gee, 1999; Rodriguez et al., 2020, Rodriguez & Lehman, 2017).

focus of this study) is an essential psychosocial factor and key determinant of students' sustained continuance in computing through college and into graduate school and the workforce (Aschbacher et al., 2010; Lunn et al., 2021a, 2021c; Perez et al., 2014; Rodriguez et al., 2020; Rodriguez & Lehman, 2017; Taheri et al., 2019; Wofford et al., 2022).

Unfortunately, our understanding of these important psychosocial variables is muddied because while the definitions of identity, self-efficacy, and sense of belonging differ, scholars often use these phrases interchangeably in CS education (McGill et al., 2019), and items measuring these different constructs vary little across studies. Indeed, some computing education researchers have used the same survey items to create an *identity* factor that others have used to create a *self-efficacy* or *sense of belonging* factor (Beyer, 2014; Blaney et al., 2022; Blaney & Barrett, 2022; Cohoon & Aspray, 2006; George et al., 2022; Lehman, 2017; Lehman et al., 2022). Additionally, these latent variables are also combined *within* studies. For instance, Lewis and her colleagues (2019) did not differentiate between belonging and identity as they used both "I feel like I belong in computing" and "I see myself as a computing person" as survey items in their measurement of students' sense of belonging in computing.

Even still, we know from existing literature that there are some recognized gender and racial/ethnic differences in these psychosocial measures that compel further examination (Beyer, 2014; Kvasny et al., 2011; Rodriguez & Lehman, 2017; Rodriguez et al., 2020; Sax et al., 2018; Seymour et al., 2004). Examinations of racial/ethnic and gender differences in identity, self-efficacy, and sense of belonging across institutional settings tend to reveal that the salience of these measures varies over time and by student demographics (Cohoon & Aspray, 2006). For example, men and white or Asian students, who are among the majority of those enrolled in computing, tend to score higher on these concepts; findings have mostly held consistent across time, but more current research that investigates this issue is certainly needed (Beyer, 2014; Sax et al., 2018; Seymour et al., 2004; Smith et al., 1994).

Of relevance to this study, however, is research that supports a deep-dive investigation into the concept of computing identity. In response to Rodriguez and Lehman's (2017) call for researchers to nuance our understanding of computing identity and factors that support such a measure, this dissertation centers on supporting students' perceptions of themselves as a computing person by fostering their sense of "computing identity." While there are various operational definitions and terminologies used in the field to describe the concept of computing identity—professional identity, discipline-specific identity, and even other psychosocial phases such as self-efficacy and belonging—for this study, **computing identity** is operationalized as the extent to which students see themselves as a computing person. The conceptualization of this measure is informed by several scholars who suggest that computing identity comprises five elements, including items related to self-efficacy and belonging. As such, the core concepts of computing identity include (a) interest, (b) competence, (c) performance, (d) recognition, and (e) sense of belonging (Abes et al., 2007; Carlone & Johnson, 2007; Lunn et al., 2021a; Gee, 1999; Rodriguez & Lehman, 2017).

While more research is needed to further our understanding of how computing identity is developed among undergraduate students (Rodriguez & Lehman, 2017), there is some research that highlights potential factors that play a pivotal role in determining the strength of students' computing identity that is worth mentioning. For example, in their conceptual framework of computing identity (which is used as a foundation for this study), Rodriguez et al. (2020) describe the important role of students' early computing identity experiences (e.g., interactions with peers, family, teachers, and science and computing-related coursework) as factors that support or deter the development of one's computing identity. Additionally, related psychosocial constructs such as self-efficacy, early computing identity, and sense of belonging are said to promote greater computing identity in college (Chemers et al., 2011; Leary et al., 2013; Lunn et al., 2021a; Margolis et al., 2000; Robnett et al., 2015; Seymour et al., 2004). Students' computing-related experiences within and outside of their introductory computing courses also play a role in facilitating their computing identity. Of note are students' interactions with peers, faculty, and the community writ large via conference attendance and participation in undergraduate research experiences or internships and co-op opportunities (Aschbacher et al., 2010; Charney et al., 2007; Eagan et al., 2013; Hunter et al., 2007; Kapoor & Gardner-McCune, 2019; Peters, 2014; Peters & Pears, 2013).

Gaps in the Literature

Although prior literature has aided in our collective understanding of computing identity, research to date has yet to identify a consistent and psychometrically sound measure (Rodriguez & Lehman, 2017) that is validated across time and student demographics, such as gender and racial/ethnic identity (Lent et al., 2011; Rodriguez et al., 2020). Moreover, computing self-efficacy and identity are two key variables often used interchangeably, and there is mixed research on whether related psychosocial constructs such as self-efficacy (Robnett et al., 2015) predict computing identity or vice versa (Graham et al., 2013; Maxwell & Cole, 2007). Therefore, there needs to be more consensus on how to better differentiate these psychosocial constructs and whether they should be operationalized as subcomponents of computing identity (e.g., Lunn et al., 2021a; Mahadeo et al., 2020; Taheri et al., 2019) or separate factors that may instead predict computing identity (Graham et al., 2013; Maxwell & Cole, 2007; Robnett et al., 2015). Thus, given the importance of computing identity in predicting students' persistence in computing, Rodriguez and Lehman (2017) report that there remains a critical need to enhance our conceptual understanding and measurement of this construct.

Study Objectives

This dissertation focused on students' undergraduate experiences in computing as they relate to shaping their computing identity. In particular, I aimed to (a) explore if a psychometrically sound measure of computing identity could be developed, (b) determine the extent to which self-efficacy and sense of belonging were necessary components of one's computing identity, (c) examine if the new computing identity factor held over time and for different gender and racial/ethnic groups, and (d) identify key college-related variables that facilitate the development of computing identity among undergraduate students who took an introductory computing course. This study was informed by existing literature, my theoretical frameworks (described in the next chapter), and the following three key research questions.

Research Questions

1. To what extent can a psychometrically sound measure of computing identity be developed for students who took an undergraduate introductory computing course?
    a. Are self-efficacy and sense of belonging necessary components of computing identity?
2. To what extent does computing identity hold over time during college, and how does this vary by gender, race/ethnicity, and intersections of gender and race/ethnicity?
3. What variables predict computing identity in the year following students' participation in an introductory computing course?
    a. To what extent do these predictors differ for women and men?
    b. To what extent do these predictors differ across racial/ethnic categories?

Study Significance

This study advances BPC research by using existing literature and theory (Rodrigez et al., 2020; Lunn et al., 2021a) to inform the creation of a computing identity measure that can also be uniquely applied to student groups holding systemically minoritized social identities in computing (i.e., white and Asian women, Women of Color, and Men of Color). Additionally, this study furthers existing research by ensuring that the measure of computing identity holds over time, by gender, and within gender groups by race/ethnicity. Finally, the study contributes to our limited understanding of the factors that foster students' computing identity, how those factors differ by gender, and how the predictive power of each racial/ethnic group varied among women and men, respectively—an approach that other researchers have called for in response to the changing demographics in the U.S. and in education (Crenshaw, 1991; Collins, 1986; Howard, 2019; Johnson, 2007; Wilson & Kittleson, 2013). The early college experience, particularly in introductory courses, is a critical opportunity to promote student retention in computing and diversification of the field, and by better understanding the pre-college and early college experiences that strengthen students' identities in computing, we are now better equipped to support student persistence (Hill et al., 2010; Margolis & Fisher, 2003; Seymour & Hewitt, 1997; Walker, 2017; Washington et al., 2019). Taken together, the findings from this study can help inform how we design, implement, and adapt interventions to strengthen the development of these constructs, with particular attention to students' gender and race/ethnicity, which in turn has implications for diversifying the representation of students in computing.

CHAPTER TWO: LITERATURE REVIEW

The primary focus of this study is developing computing identity among undergraduate students in computing. In particular, in this dissertation, I (a) explore if a measure of computing identity could be developed, (b) determine the extent to which self-efficacy and identity are necessary components of one's computing identity, (c) examine if the new computing identity factor held over time and for different genders and racial/ethnic groups, and (d) identify key college-related variables that facilitate the development of computing identity among undergraduate students who took an introductory computing course. While there is research on equity gaps in computing as well as information regarding the importance of computing identity in predicting various outcomes for computing students, to my knowledge at the time of this study, a measure of computing identity with strong psychometric properties has not yet been developed and tested across students' intersecting gender and racial/ethnic identities, nor is there enough existing research that describes the factors that promote identity development among computing students from differing gender and race/ethnic identities.

Importantly, this chapter highlights the breadth and depth of what is known about these concepts broadly in STEM and, more specifically, in computing. As such, this chapter is divided into two parts. Part I includes the literature review[16], in which I broadly provide an in-depth overview of identity-related research in STEM before I review the literature on computing identity. This section is followed by a brief summary of research on self-efficacy and belonging as they relate to computing identity. Part II provides an overview of the theoretical framework

[16] The literature referenced in this chapter spans the last few decades, focusing especially on publications since 2010 while acknowledging foundational studies published before then. It should be reiterated that as computing has largely been dominated by white and Asian men, there is limited research that focuses on the experiences of Women of Color in computing. Most of the intersectional research that does exist is qualitative, which has been included in this literature review.

guiding this study and concludes with a precursor to the general methodology I use to address my research questions, which is discussed more thoroughly in Chapter Three.

Part I: Literature Review

This section includes a brief review of the science identity literature related to science, technology, engineering, and mathematics (STEM) writ large, before delving further into STEM subdiscipline-specific identity in the field of computing and technology.

History and Definitions of Identity

Several identity theorists whose work originated in the early 1960s are still highly cited in higher education research today (Mayhew et al., 2016). The first notion of identity formation is attributed to Erik Erikson, a neo-Freudian developmental psychologist whose work centered on the psychobiological life stages an individual goes through as they age (Erikson, 1959, 1963, 1968). Erikson's psychosocial stages built upon Freud's psychosexual stages to incorporate three adult stages. Of note is Erikson's fifth developmental stage, which occurs as an individual enters and progresses through their undergraduate experience. At this point, students reconcile the dissonance between their views of the world and themselves from childhood and their experiences and lessons learned during college to form their identity (Berk, 2015; Erikson et al., 1959; Festinger, 1962).

Since Erikson's conceptualization of identity, a variety of researchers have expanded upon this work. Marcia (1966, 1980) extended psychological identity in research by centering that identity development requires both exploration and commitment. In doing so, they introduced a series of identity typologies and offered that individuals go through stages of identity development throughout their lifespan. During roughly the same time, Tajfel (1979) took a social-psychological approach to create social identity theory, which posits that identity development relies heavily on (un)conscious cues from one's social environment, thereby focusing less on the individual. Moving beyond general identity theory, Chickering and Reisser (1993) drew upon the original theorists to create an identity model unique to college students. In Chickering's seven vectors of identity development theory (1969), students cycle through several phases before conceptualizing their identity. For a full discussion of Chickering's student identity development model, see Chickering (1969; Chickering & Reisser, 1993).

Moving forward to the early 2000s, a grounded model of multiple dimensions of identity (MMDI), which has roots in intersectionality and critical race theory, was created and has since evolved. Jones and McEwen (2000) first devised the MMDI, which incorporates the environmental context alongside an individual's personal identity development. Their original model formation centered on the notion that students have a core identity and multiple social identities (e.g., gender, race/ethnicity) that they navigate simultaneously. In 2004, Abes and Jones revised the framework and created a reconceptualized MMDI, including a meaning-making filter. The idea behind this aspect is that students filter through influences from their environment to make sense of the world around them and themselves. As a result, the salience of their social identities varies based on the context of their surroundings. Once again, this model was updated to include a self-authorship component, which highlights that students have the autonomy to sift through environmental cues to form how they choose to see themselves and the world but holds that contextual influences remain interconnected with their meaning-making and perceptions of their multiple social identities (Abes et al., 2007).

Around the same time, in their study of undergraduate Women of Color in engineering, Tate and Linn (2005) established that students have multiple identities (i.e., academic, social, and intellectual) and that these identities are influenced by their environments before and during college. Building on the notion that identity is established based on various individual and environmental factors, Gee (1999, 2000) developed the theory of identity, which illuminates the important role others play in identity development. Specifically, Gee determined that identity represents “the kind of person one is seeking to be and enact in the here and now,” but that identity cannot be created by one’s lonesome (1999, p. 13). Instead, Gee declared that the creation of one’s identity relies upon the demonstration of one’s proficiency with others and that, in return, one’s own contributions are recognized.

Common to all the vast research on general identity and student identity theory is that students’ identities are conceptualized through their lived experiences (before and during college). Likewise, while students have some autonomy over their social identity formation, their identities are influenced by their social and environmental surroundings. What remains less understood, however, are the factors that influence a sense of identity among students in STEM and, more specifically, those in computing.

Building upon existing identity literature incorporating personal attributes, students’ social identities, and contextual influences from one’s environment, Carlone and Johnson created a grounded model of science identity (2007). Carlone and Johnson describe *science identity* as how people make meaning of their science experiences alongside societal structures that influence their ability to construct possible meanings of themselves. Although the model focuses on the development of science identity—which is discussed more in the next section—and the focus of this study is on the development of *computing identity*, this model is informative for two reasons. First, for its exemplary application of identity development among students in the sciences (which includes computing), and second, for its centering on those marginalized by their intersecting social gender and racial/ethnic identities.

In their pivotal study, Carlone and Johnson (2007) conducted ethnographic interviews with 15 undergraduate Women of Color in STEM and followed up with them six years later as they entered their careers to understand the lived experiences that formed their scientific identities. Through qualitative data analysis, the authors identified that those in their study had three science identity typologies: research scientist, altruistic scientist, and disrupted scientist. In short, *research scientists* achieved equilibrium as they were confident in their ability to do science (i.e. self-recognition) as they were recognized by others for their contributions to the field. *Altruistic scientists* did not seek recognition; instead, they were called to science and saw their work as a mechanism to help others. Finally, *disrupted scientists* sought after but did not receive recognition for their scientific contributions, and they reported feeling their science journeys were disrupted and unsatisfying. Moreover, the authors of this study found recognition from others to be critical in the development of science identity and these typologies.

Across these three science identity trajectories, the authors found three interrelated dimensions that were foundational to one's science identity development: (1) competence, or conceptualized knowledge and understanding of scientific content; (2) performance, or social performances of science practices; and (3) recognition, or seeing oneself and being seen by others as a "science person" (Carlone & Johnson, 2007). The authors argue that students need all three components to foster their science identity, but that recognition is the cornerstone of their development.

While science identity is much broader than computing identity, which is central to this study, the model offers utility toward our understanding of computing education literature. First, the theoretical underpinning of this model speaks to the innate and structural systems Students of Color face in many STEM fields, of which computing education researchers have also found

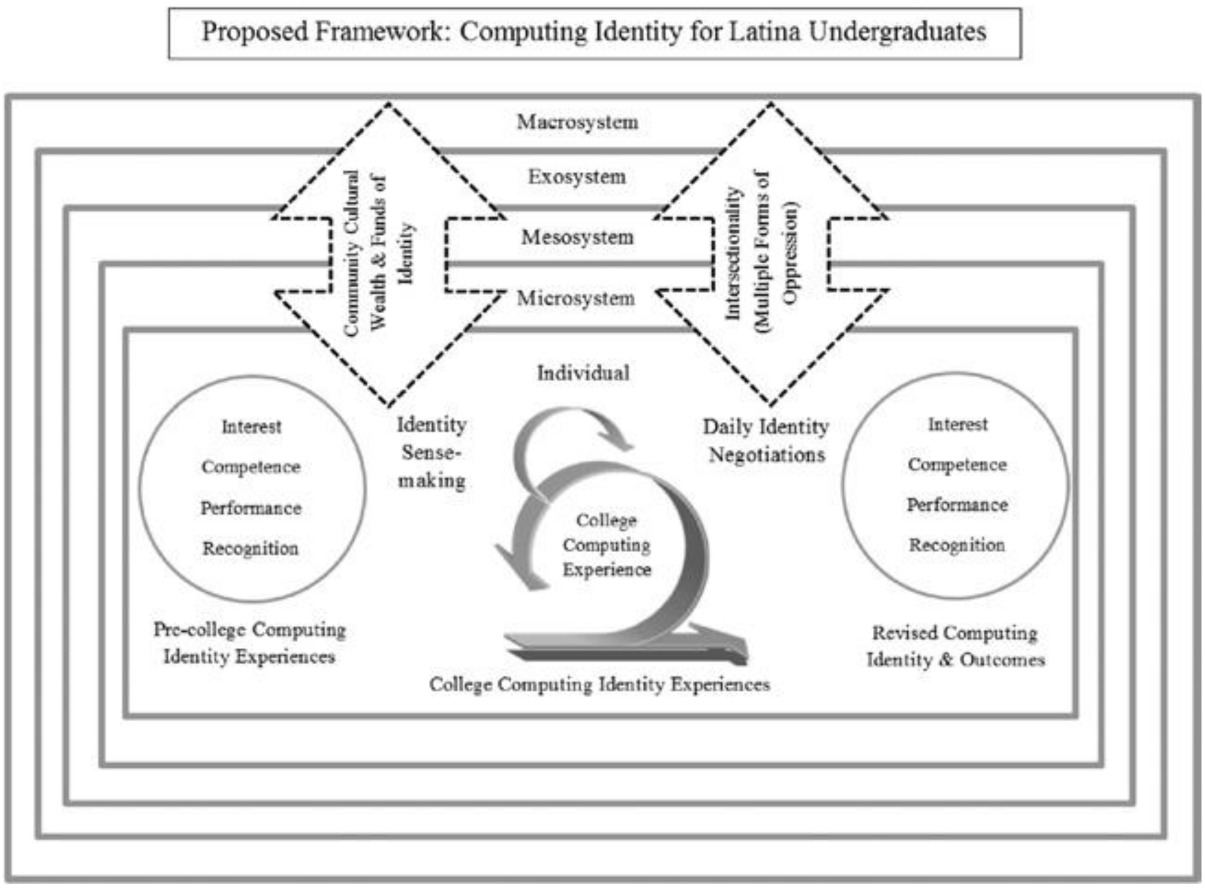
(Beyer, 2014; Cheryan et al., 2009). Also, the model centers on the intersectional experiences of those marginalized by their gender and race/ethnicity—a key aspect this dissertation aims to examine. As such, the model provides a framework for naming and quantitatively accounting for the systems of oppression that students in computing face while honoring the multiple, intersecting social identities present as they construct their computing identities. Finally, the model attests to the idea that identity is malleable, which supports the literature that is expanded upon in the sections below regarding the nature of one's identity changing as one's skills, experiences, and interactions evolve during college. As central as Carlone and Johnson's (2007) work is examining identity development in STEM writ large and in computing more narrowly, computing educational researchers have argued for identity frameworks central to the computing discipline (Rodriguez & Lehman, 2017). Perhaps in response to their call, two research studies emerged proclaiming to do just that, offering computing-specific lenses to examine identity development in computing.

First, Mahadeo et al. (2020) used Carlone and Johnson's (2007) research to develop their computing identity framework. In their research, they use survey data from multiple institutions to develop and test a measure of computing identity, which is a similar methodology to the one I use in this dissertation. While these authors still identify recognition, performance, and competence as core to students' computing identity, they combine performance and competence into one category and add a third core component they name interest. They even assess whether their computing identity model with subconstructs holds by gender (i.e., female/non-female) and race/ethnicity (i.e., Black/non-Black and Hispanic/non-Hispanic). Their model serves as an important starting point on factors central to conceptualizing students' computing identity, but it is also limited in several key ways. It is unclear if they combined men with nonbinary students, and they did not test the measure equivalence across all categories of racial/ethnic identities (i.e., white and Asian categories were left out, among other minority subgroups), investigate students' intersectional gender and racial/ethnic identity, or look at the equivalence of their measure over time.

Second, and around the same time, Rodriguez et al. (2020) created a conceptual framework for understanding students' computing identities (see Figure 2.1). In their conceptual study grounded in qualitative data, Rodriguez and her colleagues explain how supporting students' identity development in computing may need to be tailored to the needs of students with varying racial/ethnic and gender demographics, as their lived experiences—based on their social identities—vary person to person (2020). To that end, the authors create a framework for Latinas in computing that shares three core tenets with Carlone and Johnson's grounded model of science identity (i.e., performance, competence, and recognition). Like Mahadeo et al. (2020), Rodriguez et al. (2020) also add *interest* as another tenet of computing identity (see Figure 2.1).

Figure 2.1

Computing Identity Framework (Rodriguez et al., 2020)



At the core of their model, Rodriguez and her peers theorize that undergraduate students' individual identities are informed by students' navigating and making meaning of the world around them through daily negotiations. The authors describe the identity-making process as marked by three pivotable yet recursive markers. For instance, students' pre-college computing experiences inform whether they are interested in and decide to major in computing. As students enter their computing major and navigate introductory courses, their college computing experiences shape whether they strengthen their computing identity and, ultimately, whether they stay a computing major. Finally, those computing experiences during college are pivotal in shaping students' revised computing identities and other computing-related outcomes (2020).

However, the authors also move beyond naming the core components of computing identity to acknowledge pre- and during-college factors and the five interweaving systemic levels of oppression that may hinder or support the development of Latinas' computing identities. Informed partly by sociological theories, this model includes an individual, microsystem, mesosystem, exosystem, and macrosystem levels (Rodriguez et al., 2020). Here, Rodriguez and her coauthors explain that identity development is curated at the individual level, as demarcated as the model's center, but that students' identities are informed through bidirectional relationships, interactions, and socialization in the world around them—thus, the acknowledgment of the other four systems. Moving from the innermost to the outermost influence, the microsystem includes individuals with whom students often interact in computing (e.g., faculty, peers, and advisers); the mesosystem signifies influences from students' environments, such as their family, community, and off-campus work and extracurricular relationships; the exosystemic represents influences from the mass media, politics, and industry; and the most extant influence is attributed to cultural attitudes and values, which the authors depict occurring at the macrosystem level.

Additionally, this model also takes an asset-based approach to highlight that students bring unique skills and assets to college (based on their learned experiences growing up and throughout their science-educational journeys) and names the multiple forms of oppression Women of Color face in computing and the impact of these aforementioned aspects that lead students to revise their incoming computing identities over time (Rodriguez et al., 2020). In sum, applying this computing-specific model for the present study exceeds what the models from the literature on general identity or science identity offer. Due to its contributions to the literature, Rodriguez et al.'s (2020) framework was selected as the lead theoretical model to both inform the development and prediction of computing identity for undergraduate students in computing for this quantitative study. Additionally, this framework aids in a deeper understanding and meaning-making of this study's findings, with a particular lens toward understanding computing identity for Women of Color and other students marginalized by their intersecting gender and racial/ethnic identities in computing. The applicability of this model and how it helps inform important implications to disciplines within STEM, such as computing, is unparalleled for this dissertation.

The theoretical framing section in Part II of this chapter provides additional information regarding how the Rodriguez et al. (2020) framework is applied to this study, and Chapter Three provides information about how this framework is used to inform the variables for regression analyses. In the following section, I briefly summarize the importance of science identity within the context of STEM before diving deeper into the computing identity literature.

Research on Identity Within the Context of STEM

The identity-related STEM studies that have been published to date span various educational levels and student demographics. However, particular attention is given to STEM literature on undergraduate college students in general and underrepresented student populations more specifically. Below, I set the stage for the importance of understanding students' science identities in the context of STEM and then provide context on identity development in the context of computing specifically.

Science Identity

As mentioned above, since its conception, *science identity* has been a key concept in STEM education research that extends to various disciplines, education levels, and populations of varying racial/ethnic and gender representation (Carlone & Johnson, 2007; Hazari et al., 2013). Most foundational research exploring science identity has been qualitative and is often centered on K-12 education (Brickhouse & Potter, 2001; Brown, 2004). Within approximately the last decade, some researchers have explored science identity through quantitative research methods, though rarely among college-age students (Hazari et al., 2013). In the existing quantitative research on undergraduate students in STEM, science identity has been identified as an important factor in STEM persistence (Chang et al., 2011), improved course grades (Chen et al., 2020), sense of belonging (Chen et al., 2021), graduate school aspirations in STEM (Estrada et al., 2011), and a likelihood of students entering a science occupation (Stets et al., 2017).

How Has It Been Measured/What Does it Mean?

Perhaps due to limitations in quantitative researchers' ability to quantify science identity, research on this topic measures and operationalizes this concept in several ways. In her work, Avraamidou (2020) refers to science identity as a process of *becoming* a science person rather than solely a means to an end (i.e., *being* a science person). In other words, some may believe that students either have or do not have a science identity, whereas others see science identity as a belief students must continually strengthen as they encounter dissonance and navigate questioning themselves and their place in science (Holland & Lave, 2001). Just as there is no concrete definition of science identity, there are also different terms used to describe this overarching concept (e.g., professional identity, scientific identity, researcher identity) and discipline-specific measures (e.g., biology identity, physics identity, engineering identity) (Rollins et al., 2021). Because there is no consensus on the operationalization and terminology, I use the literature referring to the concept of "science identity" wherever possible. For this dissertation, **science identity** relies on the definition provided by Carlone and Johnson (2007), which centers on students' perception of the extent to which they identify as a science person.

The strength of students' science identities is increased with greater interest, performance, competence, and recognition in the field of science (Carlone & Johnson, 2007; Rodriguez et al., 2020).

While science identity has been used in the literature in various ways, one widely used measure of science identity comes from Estrada et al. (2011), whose work expanded upon Chemer's unpublished manuscript (2006). In their longitudinal research on college students nationwide, the Higher Education Research Institute at UCLA used item-response theory to create and validate a measure of science identity, which centers on the extent to which students conceive themselves as scientists (HERI, 2017). Building from the work of Estrada et al.'s five-item measure of scientific identity (2011), the HERI measure of science identity includes four items, each with a 5-point Likert scale ranging from one (*Strongly Disagree*) to five (*Strongly Agree*). An example item includes, "*To what extent are the following statements true of you: I think of myself as a scientist"* (HERI, 2017).

What Is Known to Predict It?

Across the science identity education literature, researchers generally agree that science identity is fostered through (a) positive pre-college experiences in science, (b) undergraduate science-related experiences, and (c) personal interest in and connection to the field of science. Students' preparation during high school, emotional support from peers, and exposure to science also play important roles in shaping students' development of science identity before college (Azmitia et al., 2009; Chen et al., 2020; Chang et al., 2014; Robnett et al., 2015). Additionally, student participation in undergraduate research experiences, science coursework, and professional development opportunities to practice science and present research at conferences during college supports their science identity development (Brickhouse et al., 2000; Chemers et al., 2011; Hunter et al., 2007; Kapoor & Gardner-McCune, 2019). Finally, students' science self-efficacy, early development of science identity, and a continued sense of belonging are important psychosocial factors that foster students' sense of identity as a science person (Chemers et al., 2011; Leary et al., 2013; Lunn et al., 2021a; Margolis et al., 2000; Robnett et al., 2015; Seymour et al., 2004).

Just as science identity is developed in a variety of ways, it can also be diminished through unpleasant (i.e., racist, misogynistic) experiences or "chilly climates,"[17] especially for undergraduate women and Students of Color in STEM fields (Hughes, 2014; Hall & Sandler, 1982; Seymour & Hewitt, 1997; Steele, 1997). Experiences such as these are unfortunately common for women and Students of Color in some more male-dominated STEM disciplines, such as computing (Aschbacher et al., 2010; Carlone & Johnson, 2007; Chang et al., 2014). Fortunately, however, students have some autonomy in "authoring" their identity by placing meaning on certain experiences and input from others in their community (Aschbacher et al., 2010; Le et al., 2019).

Are There Gender or Racial/Ethnic Differences?

Given the inherent systems of power that students marginalized by their gender and/or racial/ethnic identities encounter, it is reasonable to expect equity gaps in the extent to which individuals develop a science identity; however, research in this regard appears to be somewhat

[17] Hostile environments experienced by women in STEM are constantly referred to as "chilly climates;" the usage of the word "chilly" masks the actual misogyny women face, as well as the racialized misogyny Women of Color experience, and further perpetuates these systems of domination by not actually acknowledging these systems at play. In his critical analysis of problematic language, Harper (2012) encourages researchers to stop perpetuating racist norms and call racist experiences and acts of racism what they are. This approach should also be extended to hostile environments of (racialized) misogyny because continuing to refer to oppressive experiences in STEM for women as "chilly" affords comfort to those who benefit from the exclusion of women, particularly Women of Color, rather than critiquing the racism and sexism which occur daily. By not naming the power dynamics that adversely impact women, Women of Color continue to experience oppression uniquely due to the dual occurrence of racism and sexism (Ong et al., 2011; Rodriguez et al., 2020). Researchers must name the systems of domination and utilize frameworks that allow us to do so.

mixed. In studies exploring science identity in the formative developmental stage (e.g., high school level), some findings highlight gender differences but no racial/ethnic differences (Vincent-Ruz & Schunn, 2018). In studies examining science identity among undergraduate students, however, most findings suggest gender and race/ethnicity differences (Hazari et al., 2013) and differences among students by their racial/ethnic identity (Carlone & Johnson, 2007).

Regarding gender, women tend to have lower perceptions of their scientific identities than their male peers. Shapiro and Sax (2011) call upon researchers to expand upon cross-sectional research to investigate gender differences in students' science identities over time with longitudinal analyses. Importantly, Brickhouse et al. (2000) highlighted the need to understand differences within gender and not rely too heavily on differences between gender identities. They explain that research that compares students' gender across a variety of key variables can be valuable, but it can also exaggerate differences without recognizing the diversity within these gender groups. To address this, researchers might consider examining identity among women across different racial/ethnic groups.

Regarding race/ethnicity, science identity was found to be "important for minority students because it confers a sense of belonging in science that might otherwise be negatively targeted due to their marginalized group status" (Chen et al., 2020, p. 10). In another study, Hazari et al. (2013) found that white men had higher physics identity than white women or Men of Color. They also found that Students of Color reported lower identification in biology, chemistry, or physics. Overall, this research demonstrated that Hispanic women had the lowest science identities, which is consistent with Rodriguez et al.'s (2020) identity research for Latinas in computing.

Importantly, research exploring the intersectional experiences of Women and Men of Color is limited, yet scholars have called for additional studies incorporating intersectional lenses and analyses (Avraamidou, 2020). In fact, they claim that science identity cannot be studied either conceptually or methodologically without considering the multiple aspects of one's identity using an intersectional lens. Avraamidou's research calls upon researchers to name the systems of power and oppression at play to challenge the "broader agenda of science education for promoting equity and social justice" (2020, p. 327). As mentioned in Chapter One, there were and continue to be a variety of inequitable systems and power issues at play to keep certain students away from certain STEM fields (e.g., engineering and computing). I am hopeful that researchers' interest in and efforts toward interrogating these systems and acknowledging the impact of Whiteness as a container (Stewart & Nicolazzo, 2018) can aid in supporting more critical research on this topic to foster more equitable and supportive environments for all students. For example, Rodriguez et al.'s 2020 research and conceptual framework on Latinas' computing identities allow researchers to interrogate these systems in our understanding of the lived experiences of students marginalized by their gender and racial/ethnic identity in the subfield of computing within STEM.

The systemic, institutional, and curricular forces at play, along with the stigma surrounding masculinity associated with STEM (Simon et al., 2017) and students' incoming characteristics (e.g., demographics, academic preparation) contribute to the extent to which students come to see themselves (and to be seen by others) as scientists. However, more research is needed to systematically investigate these changes at various points before, during, and after students' experiences in their introductory courses.

How Does Science Identity Change During College?

The notion that science identity is fluid and can change over time aligns with the views of many identity researchers (e.g., Aschbacher et al., 2010; Avraamidou, 2020; Carlone & Johnson, 2007; Marcia 1966, 1980; Rodriguez et al., 2020). Peters (2018) states that as students enter college, they hone their interests, and their identities become narrower. Further, as Aschbacher et al. (2010) describe this process, “As students develop knowledge, competence, and meaning from these social interactions, they begin to construct their identities, or who they are and wish to be, in relation to these communities” (pp. 565–566). Therefore, it is natural that students’ science identities evolve, especially during their formative primary and secondary educational journeys. Shapiro and Sax (2011) also call upon researchers to investigate psychosocial factors over time.

While there is much existing literature that incorporates measures of identity in the prediction of students’ undergraduate outcomes (e.g., college/major persistence, time-to-degree, GPA, major change), graduate school outcomes (e.g., admission to masters and doctoral programs), and career-related outcomes (e.g., persistence into career aligned with major of study), there is still inconsistency in how the measures are operationalized and what factors predict them. Specifically, there remains a need to investigate the extent to which an identity measure can be created that holds over time across student demographics in STEM subfields (Sax & Newhouse, 2019), such as in the field of computing (Rodriguez & Lehman, 2017), and explore what aspects of a student’s educational journey or personal demographics contribute to their development of such a measure.

Identity Within the Context of Computing

Computing education researchers have explored the concept of computing identity in a variety of ways. Indeed, in 2022, Kapoor and Gardner-McCune conducted a review of 55 papers that explored computing identity. In their analysis of these studies, they found that there was an uptick in publications over time since 1996 and that most of the papers in their corpus sought to either describe the important role of computing in a qualitative manner or explore the relationship of computing on another outcome. The following year, Große-Bölting and their colleagues (2023) built upon this work to understand its theoretical underpinnings by conducting their own systematic literature review of 41 identity-focused research articles published in postsecondary computing education outlets since 2005. Through their analysis, Große-Bölting et al. (2023) found that publications have increased over time and that while most of the articles were not empirical research studies, of those that were, only a handful focused on the development of identity-based instruments and measurements.

Therefore, the field of computing is still lacking a reliable and valid discipline-specific identity measure (Rodriguez & Lehman, 2017; Kapoor & Gardner-McCune, 2022). Existing literature highlights the importance of *computing identity* in predicting a variety of student-success-related outcomes, but in its current state, there is not a comprehensive measure that has been operationalized in a consistent manner and tested across time and among different demographics of undergraduate students. As the demography of college students is changing and becoming more diverse (Howard, 2019), we must find ways to better support the unique needs of all our students, especially those with identities from marginalized gender and racial/ethnic groups (Rodriguez et al., 2020). The research community also lacks a comprehensive understanding of what computing identity is and what factors contribute to students' development of this measure. Conflicting terminology and operational definitions may be at fault, or there may not be enough research on identity development in the higher education computing sphere to have the answers we need. Given the importance of computing identity serving as a positive predictor of students' persistence in computing, graduate school aspirations, and interests in computing careers, creating a measure of identity specific to computing is an important goal. Additionally, finding ways to foster the development of computing identity among college students in the United States may be one important step toward supporting students' academic and professional success in computing—ultimately serving as a potential gateway toward diversifying the technology field.

What is Computing Identity?

The idea of a discipline-specific computing identity is more nuanced to the subfield than the general term science identity (Carlone & Johnson, 2007), which represents identity development in the broad field of *science*. Computing falls into the technology realm (i.e., the "T" in STEM) but is itself a science (Denning, 2005). Thus, there are substantial differences that must be considered in scholarship within STEM subfields, such as the context of the field and the general student demographic representation (Sax & Newhouse, 2019). To that end, developing a stronger science identity among students in certain STEM subdisciplines cannot be a one-size-fits-all approach, as disciplines that make up STEM are not a monolith. Indeed, there is a need to develop a field-specific computing identity measure to fill that gap (Rodriguez & Lehman, 2017).

For this dissertation, the concept of **computing identity** is a multidimensional measure that is defined as the extent to which a student sees themselves as a computing person informed by how they make meaning of their core identity and navigate the world around them (Abes et al., 2007; Carlone & Johnson, 2007; Lunn et al., 2021a, Gee, 1999; Rodriguez & Lehman, 2017). This concept is highly regarded as an important aspect of student success; however, there remains little consensus on the operational definition of computing identity to date. Computing education researchers tend to use different theoretical frameworks that center on unique aspects of identity, often leading to different conceptualizations of the term in published literature. For instance, recognition is a focal point in Gee's 2000 identity lens, whereas interest is a key factor in Mahadeo et al.'s (2020) and Rodriguez et al.'s (2020) conceptual frameworks. Thus, the vague yet commonly used term "computing identity" is a key facilitator in predicting outcomes of retention, interest, and motivation and is linked with a variety of identity-related phrases. Still, no agreement remains on this term or how to measure it in the computing field. More recently, computing identity has been defined as the extent to which students identify as a computing person (Rodriguez & Lehman, 2017), which Rodriguez et al. (2020) further illustrated is more of an identity spectrum than an "all or nothing" identity configuration.

As mentioned, several theoretical frameworks—each with unique tenets—guide scholars' exploration of computing identity. Perhaps because there is not a single theoretical framework highlighting key components of computing identity development, there is no consensus on the conceptual understanding of computing identity (Rodriguez & Lehman, 2017). Thus, there remains no standard for measuring computing identity nor in the terminology used to refer to this concept. For example, Kapoor & Gardner-McCune (2019) and Rollins et al. (2021) refer to this phenomenon as "professional identity." Because research examining the concept of computing identity is limited (Rodriguez & Lehman, 2017), this literature review incorporates research on *computing identity* and *professional identity* among undergraduate students in computing, with greater emphasis on "computing identity."

What Does It Predict?

While there remains no consensus on the operational definition of computing identity, STEM subdiscipline identity is, however, highly regarded as a predictor of a variety of positive undergraduate computing outcomes such as actual and aspirational persistence, graduate school, and career outcomes (Aschbacher et al., 2010; Perez et al., 2014; Taheri et al., 2019; Wofford et al., 2022). Researchers have also found field-specific identity to positively predict students' graduate school aspirations, career commitment, and career outcomes (Chemers et al., 2011; Hazari et al., 2010; Wofford et al., 2022). Further, in their systematic review of identity-based higher education computing literature, Große-bölting and their colleagues (2023) report the most common motivation for researchers using identity in their work was primarily to promote retention and greater gender or racial/ethnic diversity, followed by investigating its relationship with recruitment, motivation, and increased performance outcomes.

How Has It Been Measured?

As there is no consistent operational definition of the affective/psychosocial concept of computing identity, there is also no consistent way to measure computing identity that has been validated across student demographic groups, contexts, and time. Instead, there are many key phrases used to describe computing identity. For example, some studies refer to this concept as self-concept, self-perception, self-efficacy, sense of belonging, or professional identity (e.g., Bailey, 2003; Baumeister, 1997; Große-bölting et al., 2023; Kapoor & Gardner-McCune, 2019; Parker, 2018; Rollins et al., 2021). Again, while some researchers (e.g., Lunn et al. 2021a; Mahadeo et al., 2020; Taheri et al., 2019) underscore the importance of multiple subconstructs (or core components) of students' computing identities (i.e., interest, competence/performance, recognition, sense of belonging), other scholars think of these latent variables in different ways (e.g., as precursors/predictors of computing identity or as entirely separate constructs altogether). Across the literature, scholars use different terminology when referencing the same topic both across and within their empirical works. Not only do the terms to describe computing identity differ, but so do the variables used to measure this construct. For example, in one study exploring undeclared students' undergraduate experiences in introductory computing courses, the author created a latent factor of computing identity that comprised several variables such as, "I feel like I 'belong' in computing," "Using computers to solve problems is interesting," and "I am confident that I can complete an undergraduate degree in computing" (Lehman, 2017). Using the same master dataset, other authors used items from that factor in their studies to represent students' fit in computing (Lehman et al., 2022), self-efficacy (Blaney & Barrett, 2022), or sense of belonging (George et al., 2022; Sax et al., 2018). My argument is not that prior literature was too constrained, but rather that it was perhaps not constrained enough. More specifically, while other scholars have used similar items to measure computing identity (or related concepts), I am seeking to investigate if these larger latent factors that were included in prior studies' measures of computing identity are part of some broader latent construct (e.g., self-recognition and belonging, identity, self-authorship) or if are they different ideas entirely.

While acknowledging that previously created factors may or may not hold across new samples being studied with the same dataset, more research is needed to serve as a guide to identify what computing identity is and if self-efficacy and belonging are necessary components of this measure (Lunn et al., 2021a; Rodriguez et al., 2020; Rodriguez & Lehman, 2017). We can work toward clearer definitions and measurements by developing a greater understanding of computing identity.

How Does It Differ from Other Psychosocial Factors?

By definition, computing identity (Abes et al., 2007; Lunn et al., 2021a; Rodriguez et al., 2020; Rodriguez & Lehamn, 2017) is neither self-efficacy (Bandura, 1977, 1986a, 1986b; Beyer, 2014; Compeau & Higgins, 1995; Lin, 2016) nor sense of belonging (Sax et al., 2018), and the measurements and terminology used to describe these psychosocial factors should not be used in such a manner that implies that they are one and the same. While Rodriguez et al. (2017; 2020) recognize that computing identity is multifaceted, she and her coauthors join other scholars (Kapoor & Gardner-McCune, 2022; Große-Bölting et al., 2023) in urging for more research that supports the field's understanding of how these psychosocial variables work together (or separately) to inform students' computing identities.

Computing Identity vs. Computing Self-Efficacy. Like identity, research has emphasized that increased self-efficacy is related to (a) one's greater future involvement in a task, (b) the effort one is willing to expend to learn and perform a task, and (c) one's commitment to carrying through a task (Bandura, 1977, 1986a, 1986b; Lin, 2016). The operationalizations and terminology used to describe self-efficacy vary. Dating back to the late 1970s, Albert Bandura (1977, 1986a, 1986b) coined the term "self-efficacy," referring to "people's beliefs in their ability to perform specific behaviors or courses of action" (pp. 184–185). Other terms in computing to describe students' self-efficacy are "learning self-efficacy" (Lin, 2016), "computer self-efficacy" (Teo & Ling Koh, 2010), "technology self-efficacy" (Bertozzi, 2007), "programming self-efficacy" (Lin, 2016), and "computing self-efficacy" (Blaney & Stout, 2017). Marakas et al. (1998) label these types of computer self-efficacies as either *task-specific* or *general*. For this dissertation, **computing self-efficacy** is operationalized—in the general sense—as undergraduate students' belief in their ability to skillfully perform computational tasks required of them in the field of computing (Bandura, 1977, 1986a, 1986b; Beyer, 2014; Compeau & Higgins, 1995; Lin, 2016). Thus, a critical distinction between computing self-efficacy and computing identity is that the latter is a broader term representing students' perception of their sense of self as a computing person, and the former is a term that measures their perceived performance ability. While self-efficacy is a known predictor of identity (Robnett et al., 2015), based on the evolution of computing identity scholarship that situates students' recognition of themselves as a computing person alongside their competence and performance in computing (Rodriguez et al., 2020), this dissertation explores the extent to which computing self-efficacy is a core *component* of computing identity rather than a *predictor* of computing identity. However, as the literature in this regard is mixed, and because self-efficacy is not a component of computing identity (as demonstrated in Chapter Four), computing self-efficacy is used as a predictor to analyze the extent to which it contributes to one's computing identity (Graham et al., 2013; Maxwell & Cole, 2007; Robnett et al., 2015).

Computing Identity vs. Sense of Belonging in Computing. Like self-efficacy, the latent construct measuring students' sense of belonging in computing is a critical predictor of various student success measures and may also be a core component of computing identity. *Sense of belonging*—often described as a measure of environmental "fit" (Strayhorn, 2012)—is a key predictor of STEM retention (Espinosa, 2011) and students' interest and persistence in computing (Veilleux et al., 2012). Just as computing self-efficacy might be a core component of identity, so too might students' sense of belonging in computing. Indeed, Große-bölting et al. even said that identity and belonging are so intertwined that "identities are operationalized as a sense of belonging or membership to and within the discipline" (2023; p. 5). While Rodriguez and her colleagues acknowledge the importance of students' belonging, especially for Women of Color, in their research on Latinas in computing, sense of belonging is not a stated facet of their conceptual model (grounded in qualitative data; 2020). However, several recent quantitative scholars discovered through factor analysis and structural equation modeling that a sense of belonging was a core component and, in fact, a subconstruct of computing identity (e.g., Lunn et al., 2021a; Mahadeo et al., 2020; Taheri et al., 2018). Therefore, for this dissertation, **sense of belonging in computing** is operationalized as the extent to which a student feels that they are part of or "fit" in the computing community (Blaney & Stout, 2017; Cheryan et al., 2009; Taheri et al., 2019; Sax et al., 2018), and for the reasons mentioned above, sense of belonging is treated as a potential component of computing identity and not as an independent predictor.

How Does It Change?

While Carlone and Johnson (2007) suggest *science identity* should grow over time, few researchers have studied *computing identity* over time—despite its documented fluidity. In one study by Kapoor and Gardner-McCune (2019), the authors found that the 14 undergraduate computer science majors they interviewed formed their computing professional identities during the first three years of their degree programs. The authors mention that the identities were commonly formed between students' second and third years of their computing programs. This timeframe is particularly important as it is typically *after* students participate in their introductory computing course experiences. However, research has also indicated that students' introductory course experiences also play an important role in students computing identity development (Rodriguez et al., 2020; Peters, 2014; Peters & Pears, 2013). I am not aware of research that has looked at students' computing identities over time (e.g., before, during, and after their introductory computing courses).

What is Known to Predict It?

Most research to date focuses on qualitative methodologies to understand the development of students' computing identities, resulting in few studies that have explored computing identity as an outcome measure (e.g., Lunn et al., 2021a). Even fewer studies have explored the extent to which students' experiences during college predict their computing identities or have used longitudinal data to account for students' incoming computing identity scores, which Astin and antonio (2012) suggest is an important consideration to control. Given the limitations in the existing literature and the mixed findings pertaining to students' gender and racial/ethnic identity in predicting computing identity, there remains a need to further our understanding of the student demographic, academic background, and college environmental factors that facilitate or impede the development of computing identity. Of the existing literature on this topic, there are several key elements that may contribute to the development of computing identity for college students. Some important elements of computing identity include factors related to students' backgrounds, scores on other related psychosocial measures, and experiences within and outside of college.

Background Factors. Well before college, students' early childhood experiences play a role in their computing identity development (Rodriguez et al., 2020). Some of these experiences include family roles, cultural expectations, and experiences in elementary and high school that help to form how students identify when they get to college. Rodriguez et al. (2020) describe these influences as "pre-college computing identity experiences," listing encounters with peers, family, teachers, and early experiences in science and computing courses. Lunn et al. (2021a) describe students' feelings based on their interactions with others as their internalized recognition, which is a core component of their computing identity development. An example of these influences is students' negative experiences in math and science or experiencing racism in a classroom, which Chang et al. (2011) documented as impediments to women and racial/ethnic minority students' science identity and persistence. In fact, Hazari et al. (2010) found that early high school experiences predicted identity for students in physics and whether they persisted in the field (i.e., conceptual understanding, experience connecting computing concepts to the real world and students' communities, answering questions, teaching classmates, and encouragement from a high school teacher). While disciplines in STEM differ (Sax & Newhouse, 2019), we may extrapolate that early computing experiences likely play an important role in computing identity development as well. In fact, pre-college experiences such as how students make meaning of science and themselves based on their pre-existing knowledge formed in their childhood by family members in science (e.g., family support and social capital) are assets and contributing factors to their development of computing identity students can bring to computing spaces (Rodriguez et al., 2020; Wofford & Gutzwa, 2022). Further, researchers describe this phenomenon as "students' funds of identity directly influenc[ing] their computing identity and vice versa" (Rodriguez et al., 2020, p. 34). While these pre-college data may be harder to measure in quantitative studies, Lunn and her colleagues (2021a) found high school math grades, students' gender and race/ethnicity, the extent to which students worked outside the home, and whether they came from a supportive family to be predictors of students' computing identity.

Psychosocial-Related Factors. As described in the above literature review, there are a variety of psychosocial factors related to students' scientific identities that may foster growth in this regard. Similar to facets that develop science identity, students' interest in computing, self-computing-related self-efficacy, early development of computing identity, and a continued sense of belonging are likely important psychosocial factors that have the potential to foster students' sense of identity as a computing person (Carlone & Johnson, 2007; Chemers et al., 2011; Leary et al., 2013 Margolis et al., 2000; Mahadeo et al., 2020; Robnett et al., 2015; Rodriguez et al., 2020; Seymour et al., 2004). However, as some scholars have considered computing identity with subcomponents as their dependent variable (e.g., Lunn et al., 2021a), more research is needed to uncover the latent factors that contribute to students' computing identity development during college because there is little to no research examining the salience of these related psychosocial factors (e.g., students' interest in computing, self-efficacy, sense of belonging) in enhancing students' computing identity.

College Computing Identity Experiences within the Introductory Computing Course. Identity-related factors also shape the salience of undergraduate students' computing identities during college. For example, Rodriguez et al. (2020) highlight transformative encounters students have during their introductory computing courses—one aspect that other researchers (e.g., Peters, 2014; Peters & Pears, 2013) have also found to be important in early computing identity development. Additionally, these authors found that students' experiences in computing spaces and interactions with faculty and peers within computing may facilitate or impede their computing identity. Further, in their qualitative research on undergraduate Latinas in computing, Rodriguez et al. (2020) note that students who reported they experienced sexism and/or racism informed the researchers that these experiences were deterrents to their ability to identify as a computing person. Moreover, when faced with negative computing-related encounters, students negotiate their positionality and identity within the field (Rodriguez et al., 2020). I am not aware of literature that has investigated the salience of experiential intro course factors on quantitative studies predicting computing identity. While the sample of their study was not necessarily students in introductory courses, in their regression analysis predicting computing identity, Lunn et al. (2021a) found that students tended to have stronger computing identities when they were in a club, mentored others, worked with their advisor on classwork, and had friends in computing. Additionally, in their study, the more often students were mentored and reported that other students in class helped with their classwork, the lower their computing identities tended to be.

College Computing Identity Experiences outside of the Introductory Computing Course. Students' engagement in the broader computing context through other computing experiences and professional development opportunities (e.g., participating in internships, research experiences, and attending conferences) are also known to support students' identity development in computing (Lunn et al., 2021a; Kapoor & Gardner-McCune, 2019). In fact, some of the largest contributions to professional identity development in computing were among coursework from other computing courses, informal activities like hackathons, and professional development opportunities like internships and conferences that support the engagement in "communities of practice" (Lave & Wenger, 1991; Kapoor & Gardner-McCune, 2019). Importantly, the authors argue that students' engagement in the computing community has implications for their computing identity development beyond that of students' classroom environments. These findings are supported by other researchers who also articulate the importance of undergraduate research participation, presenting material other than coursework, and networking at conferences as important predictors of increased student discipline-specific identity (Aschbacher et al., 2010; Charney et al., 2007; Hunter et al., 2007; Lunn et al., 2021a).

Are There Gender and/or Racial/Ethnic Differences?

While Große-Bölting et al. (2023) reported that scholars named lack of diversity in computing as an important reason to study computing identity, they also reported that few studies (only three of the 13 focusing on diversity) mentioned students' intersectional gender and racial/ethnic or other identities—despite Rodriguez and Lehman's (2017) related critique several years prior. In line with the limited nature of research on undergraduate computing identity in this regard, few scholars have explored the extent to which there are gender and racial/ethnic differences in students' computing identity, and there are mixed findings from the studies that do exist on this topic. For instance, Parker did not find gender differences in their study of computer science professional identity. Additionally, Taheri and his colleagues (2018) did not unveil gender differences in their measure of computing identity; however, they did confirm four subconstructs of computing identity (i.e., interest, sense of belonging, recognition, and competence/performance). In a later study by some of the same authors, they again found that students' sense of belonging was a core component of their computing identity, but this time, they discovered gender and racial/ethnic differences, where white students and men had higher computing identity scores than their peers (Lunn et al., 2021a). However, while Lunn's research was conducted on multiple institutions, the study sample was limited to students from public universities in the state of Florida. As such, their results cannot be generalized to all students in computing, and additional research is needed to see if these findings hold for different institution types and a larger national sample. Other scholars have taken great strides to understand the lived experiences of how Latinas develop their computing identities (Rodriguez et al., 2020), but additional research is still needed to adequately respond to Rodriguez and Lehman's (2017) call to action.

While limited research exists on this topic, it is expected that men would score higher on computing identity—as they do, broadly speaking, in other affective psychosocial constructs such as science identity, sense of belonging, and self-efficacy (e.g., Hill et al., 2010; Corbett & Hill, 2015; Lunn et al., 2021a; Seymour et al., 2004). However, scholars must expand research in this area to determine whether the same findings hold true when examining students' computing identity and investigating what changes occur over time once controlling for students' incoming computing identity scores. There are currently no studies that have controlled for students' baseline computing identity scores in predicting longer-term computing identity, so we cannot be certain if students' scores strengthen more over time once pre-test measures are controlled and if there would be gender differences if they do change. To do so, several factors must be considered. Given that not all women and not all racialized people experience education and identity development in the same ways, we cannot apply a blanket approach to the measurement and study of computing identity. There is no monolith for the experiences of all women or the experiences of any one racial/ethnic group (Rodriguez et al., 2020). Researchers must do better to include women from racial/ethnic identities in their studies and aim to disaggregate gender and race/ethnicity and study their intersections in their efforts to explore factors that contribute to the development or deterrence of computing identity (Ong et al., 2011).

The Present Study

Taken together, there is a need to better understand what computing identity is (Rodriguez & Lehman, 2017) and how we can foster the development of this important concept for all students, especially those doubly oppressed in computing because of their gender and race/ethnicity (Rodriguez et al., 2020). In developing such a measure, researchers must consider students' demographic representation and determine if a psychometrically sound (i.e., reliable and valid) computing identity construct can be created. The extent to which self-efficacy and sense of belonging are necessary components of identity must also be explored. After creating the measure of computing identity, researchers must examine if the factor holds across time and student's racial/ethnic and gender identity categories. Specifically, if we want to diversify the student body of those in computing undergraduate programs as a step toward increasing the representation of all women and Men of Color in the computing and technology workforce, we must consider these aspects of their multiple social identities in the development of an identity measure in computing. Thus, there remains a continued need to examine computing identity development across and within student gender and racial/ethnic groups (Rodriguez & Lehman, 2017).

Objectives

Research exploring the nature of students' identity development in the context of undergraduate computing is limited, especially regarding how these constructs relate to historically marginalized groups, and more research is needed to fill this gap (Rollins et al., 2021). Importantly, this dissertation responds to the need identified for a discipline-specific computing identity measure (Rodriguez & Lehman, 2017) by creating a factor of computing identity and testing its (in)variance over time and across students' gender and racial/ethnic identities. In doing so, this study expands upon existing literature to provide the computing higher education field with a discipline-specific measure of computing identity and explores a variety of factors that predict this measure.

This study is guided by the following three research questions and their subcomponents.

Research Questions

1. To what extent can a psychometrically sound measure of computing identity be developed for students who took an undergraduate introductory computing course?
    a. Are self-efficacy and sense of belonging necessary components of computing identity?
2. To what extent does computing identity hold over time during college, and how does this vary by gender, race/ethnicity, and intersections of gender and race/ethnicity?
3. What variables predict computing identity in the year following students' participation in an introductory computing course?
    a. To what extent do these predictors differ for women and men?

b. To what extent do these predictors differ across racial/ethnic categories?

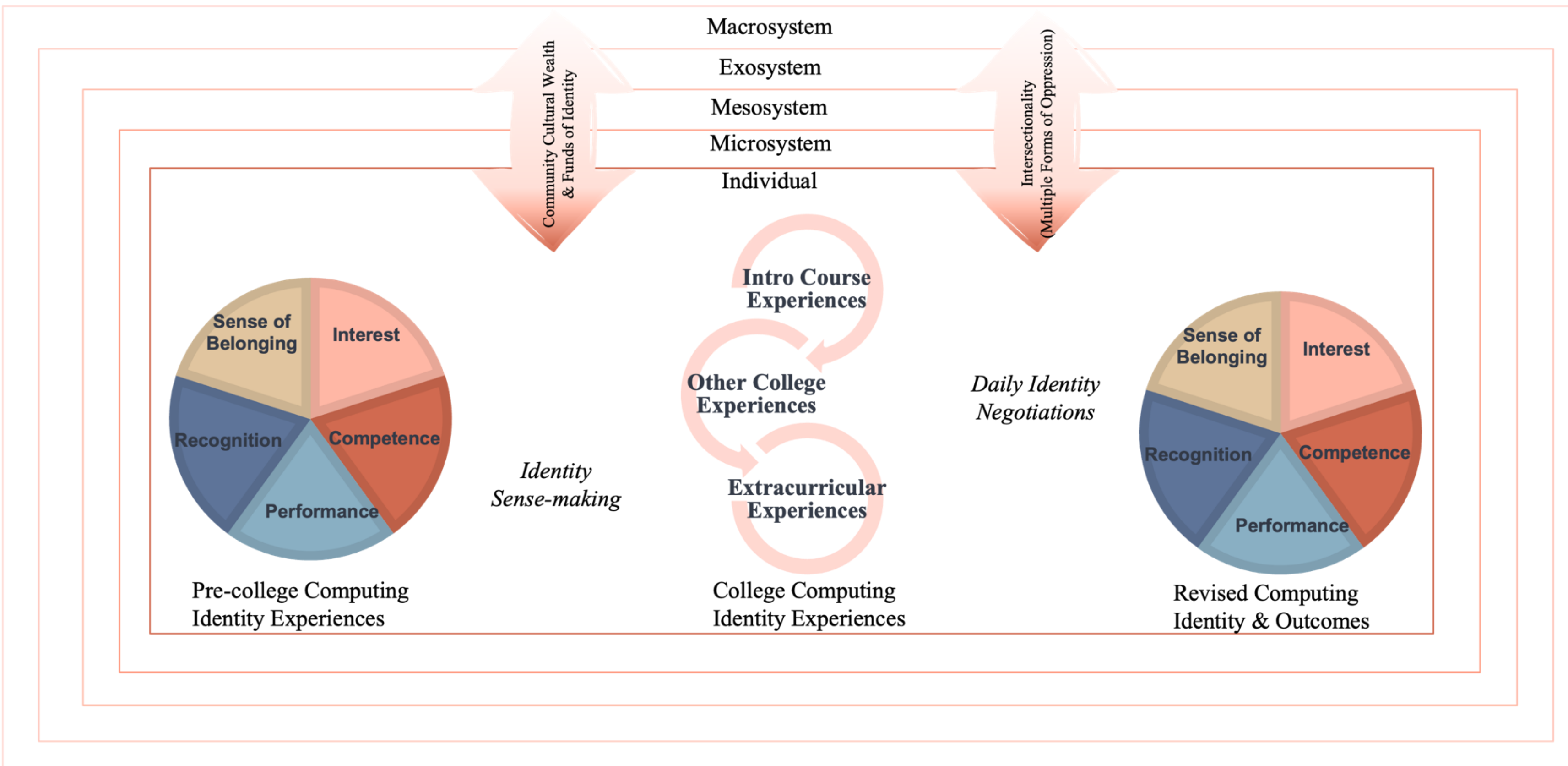
Part II: Theoretical/Conceptual Framing

Despite much research on equity gaps in computing, there are no prior higher education studies that validate a measure of computing identity, test the reliability and equivalency of this measure for students of different gender, racial/ethnic, and intersectional gender and racial/ethnic identities, or that examine the predictive nature of college environmental variables in the development of this construct. In this study, I seek to gain a deeper understanding of how computing identity is conceptualized, confirm whether self-efficacy and sense of belonging were core components of this measure, examine if the factor held over time and by gender and racial/ethnic groups, and finally study the key factors that support computing identity development among undergraduate students in computing. In doing so, Rodriguez et al.'s (2020) conceptual model of computing identity for Latinas is used to situate the context of supporting students' identity development to promote greater diversity and persistence in computing. Guided by literature that acknowledges the influence of individuals and their environments, this theoretical perspective lends utility toward developing and testing measures of students' psychosocial computing identity, identifying variables that predict this measure, and understanding the implication of these findings for the computing discipline.

Earlier in this chapter, I argue that models focusing on discipline-specific identities, such as computing identity, should naturally be more field-specific than the general concept of science identity. As such, this dissertation relies upon a revised version of Rodriguez et al.'s (2020) conceptual model of computing identity. The revision includes students' sense of belonging in computing as a core component of their identity development (Taheri et al., 2018; Lunn et al., 2021a), uses items related to computing self-efficacy as a proxy for students' perceived competence/performance (Godwin et al., 2016), and reiterates the importance of their internal recognition of themselves as a computing person (Carlone & Johnson, 2007). For this dissertation, the model depicted in Figure 2.2 below is used to inform items that are used in the development of the computing identity factor and serves as a guide for the independent (or predictor) variables used in the model predicting computing identity due to the important role student's personal characteristics, background contexts, and educational experiences play. The use of Rodriguez et al.'s 2020 model for variable selection is described in greater detail in Chapter Three. This model's development originated as a way to call attention to the multilayered experiences Women of Color face in computing, situated within the larger context of an individual's core identities (i.e., gender, race, and ethnicity), which is pertinent to this study. As this model is rooted in the experiences of Women of Color in computing, its usefulness extends beyond variable selection; it provides an important foundation for me to interpret this study's findings and contribute to computing literature, theory, and practice. See Figure 2.2 below for an overview of the revised computing identity framework. The difference between the prior models and the revised model below is that I added sense of belonging as a core component of computing identity and centered the college-level factors to focus on students' computing-related experiences.

Figure 2.2

Revised Computing Identity Framework



CHAPTER THREE: METHODOLOGY

Given the importance of computing identity for undergraduate students' success in computing and the notable gaps in the literature described in Chapter Two, the scope of this study is threefold. First, I develop a measure of computing identity. In doing so, I determine whether self-efficacy and sense of belonging are necessary components of one's computing identity. Second, I examine the extent to which the psychometric properties of the new computing identity measure are equivalent over time and by student demographics. Third, I study students' pre- and during-college factors and the predictive power of these variables on developing their computing identity and explore how these factors differ by students' gender and racial/ethnic identity. Together, findings from this study inform how to foster computing identity among students from varying demographics in introductory computing courses to promote greater recruitment and retention, ultimately diversifying technology. Informed by prior literature on this topic and Rodriguez et al.'s (2020) conceptual framework of computing identity described in Chapter Two, I use an existing longitudinal dataset collected from *Momentum*'s BRAID research team at UCLA to conduct a combination of quantitative analyses (i.e., descriptive statistics, factor analysis, measure invariance tests, ordinary least squares regression) to address the three key research questions below.

Research Questions

1. To what extent can a psychometrically sound measure of computing identity be developed for students who took an undergraduate introductory computing course?
    a. Are self-efficacy and sense of belonging necessary components of computing identity?

2. To what extent does computing identity hold over time during college and by gender, race/ethnicity, and intersections of race/ethnicity?
3. What factors predict computing identity in the year following students' participation in an introductory computing course?
    a. To what extent do these predictors differ for women and men?
    b. To what extent do these predictors differ across racial/ethnic categories?

The methodological plan for this study is driven by existing literature that documents gender and racial/ethnic equity gaps in computing while maintaining the importance of computing identity in predicting a variety of college, career, and graduate school outcomes for undergraduate students who took an introductory computing course. To systematically develop and test a computing identity measure and explore the factors that predict this measure when treated as an outcome, this study interrogates the conceptual understanding and universal applicability of a computing identity measure for students of varying gender and racial/ethnic identities. To honor students' intersectional social identities and not perpetuate systems of oppression for marginalized communities in computing, I work to apply criticality in my quantitative methodological approaches wherever possible (e.g., disaggregated racial/ethnic and gender categories, intersectional analyses; Carter & Hurtado, 2007; Mayhew & Simonoff, 2015; Rodriguez et al., 2020; Rodriguez & Lehman, 2017). More information on these methodological techniques, decisions, and rationales is provided in the remainder of this chapter.

Research Design

This study uses existing student-level survey data collected from the BRAID research team at UCLA. To understand students' experiences in computing courses on short- and long-term outcomes, the BRAID research team at UCLA conducted a longitudinal student-level study surveying two cohorts of students from 2014 to 2021. These data were collected starting in the fall of 2015 and 2016 from students enrolled in an undergraduate introductory computing course across 15 U.S. doctoral universities with either high or very high research activity[18], of which two were private and 13 were publicly controlled (The Carnegie Classification of Institutions of Higher Education, n.d.). The UCLA research team administered the first (pretest) survey at the beginning of the introductory course, a second (posttest) survey at the end of the introductory course, and a third (first follow-up) survey was administered in the year following students' completion of the introductory computing course. The remaining follow-up surveys were administered each fall term until 2021 (see Figure 3.1 for the surveys used in this study). In support of analyses on important marginalized subpopulations in computing, the data used for this study aggregates across cohorts to increase the sample sizes needed to reduce the effects of student survey attrition and increase statistical power (Aberson, 2019; Barrow et al., 2016; Stout et al., 2018).

Specifically, for this dissertation, the pretest survey provides data on students' incoming intentions and demographics, serving as an important control for the direct pretest computing identity measure captured upon students' entry to the introductory course (Astin & antonio, 2012). In addition, the posttest survey provides information about students' experiences during the introductory course (e.g., perceptions of faculty and department). Finally, the first follow-up survey provides information about students' experiences in college (e.g., college involvement). Combining these data sources into a matched panel database—where the sample is restricted to

[18] The 15 BRAID institutions include: Arizona State University, Missouri University of Science and Technology, New Jersey Institute of Technology, University of California, Irvine, University of Illinois at Chicago, University of Maryland, Baltimore County, College Park, University of Nebraska-Lincoln, University of North Texas, University of Rochester, University of South Carolina, University of Texas at El Paso, University of Vermont, University of Wisconsin-Milwaukee, and Villanova University.

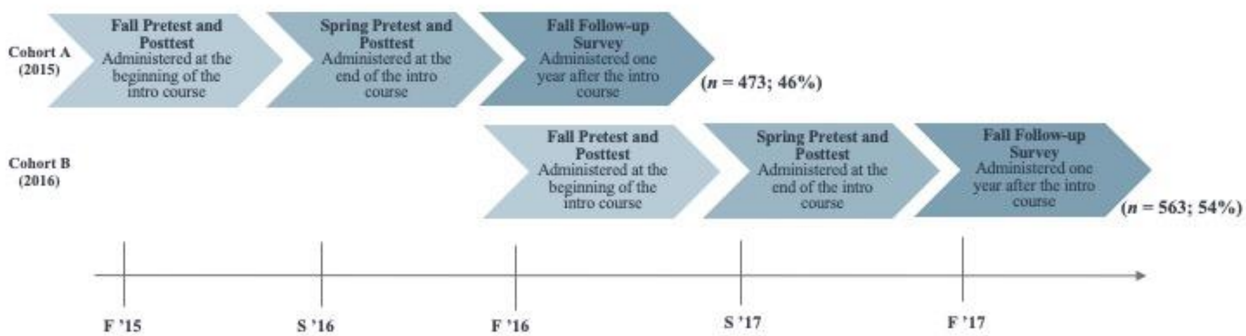
only include students who completed the pretest survey, posttest survey, and first follow-up survey—supports longitudinal research by allowing for a detailed investigation of the factors that facilitate or impede students' computing identity throughout their early college experiences.

Survey participants were given incentives to increase survey response rates and show appreciation for the time spent taking surveys. The survey incentives remained the same for both cohorts but varied by survey administration. For the pretest survey, the first 400 respondents received a $15 Amazon gift certificate, and all students were entered into a raffle for a chance to win one of two $125 Amazon gift certificates. A similar incentive was offered for students who participated in the posttest survey at the end of the intro class. However, all survey respondents who participated in the first follow-up survey received a $10 Amazon gift certificate, and no additional incentives were raffled. Response rates varied by student cohort and survey but ranged from 19% to 42%.

While representatives from computing departments at each of the BRAID institutions self-selected to participate in the BRAID initiative, the undergraduate student enrollment and degree attainment trends at these institutions are largely representative of the norms seen across research institutions nationwide (Lehman et al., 2022; Zweben & Bizot, 2015, 2016, 2017, 2018, 2019, 2020, 2021). Further, the reported computing experiences for the overall sample are akin to the experiences reported by students in computing at other BRAID institutions and, more broadly, at other research institutions (Lehman et al., 2022).

Figure 3.1

Overview of Survey Data Sample and Research Design (n = 1,036)



Study Sample

Across both cohorts, 8,067 students completed the pretest survey, 5,333 completed the posttest survey, and 4,412 completed the first follow-up survey. However, the sample for this study was restricted to students who had completed all three surveys, students who had not graduated with a bachelor's degree at the time of the first survey, and students who indicated they were computing majors, computing minors, or were undeclared but expressed interest in majoring in computing on any survey time point (*n* = 1,036)[19]. By the final survey, 765 students

[19] Momentum has already provided me with access to these data for use in sample exploration and preliminary sample descriptives to determine the extent to which the proposed analytical approaches can be achieved with the data available. However, the official use of this dataset for the proposed analyses is pending the successful completion of my dissertation proposal along with UCLA IRB approval. Therefore, the data provided in this section are from the curated merged dataset for this study.

(74%) were computing majors, of whom 690 had persisted in computing since the initial survey. The remaining 75 students had become computing majors since the initial survey time point.

Of the 1,036 students in the sample, there is nearly equal representation across cohorts (i.e., 46% from cohort A and 54% from cohort B). More than three-quarters of these students (83%) were in their first or second year of college at the time of the first survey (n = 863). Table 3.1 summarizes survey respondents' intersecting gender and racial/ethnic identities. In sum, 63% of students identified as men, and most students identified as white (42%) or Asian (28%). Additionally, more than half of the women in this sample are Women of Color (66%; Anita Borg, 2021). Importantly, those in this study closely represent those broadly enrolled in computing, as described in Chapter One.

Table 3.1

Sample Demographics by Gender and Racial/Ethnic Identity (n = 1,036)

	Men		Women		Nonbinary or Other Self-Identified		Gender Not Reported		Total	
Race/Ethnicity	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
White	303	46.3	123	34.1	10	52.6	1	100.0	437	42.2
Asian	178	27.2	113	31.3	2	10.5	0	0.0	293	28.3
Black	34	5.2	30	8.3	3	15.8	0	0.0	67	6.5
Latine	70	10.7	40	11.1	2	10.5	0	0.0	112	10.8
Indigenous, Multiracial, or Other	32	4.9	23	6.4	1	5.26	0	0.0	56	5.4
Race/Ethnicity Not Reported	38	5.8	32	8.9	1	5.26	0	0.0	71	6.9
Total	655	63.2	361	34.8	19	1.83	1	0.1	1,036	100.0

Note. Nonbinary included those identifying as gender non-conforming or genderqueer. Indigenous, Multiracial, or Other, including "Other" in race/ethnicity represents Arab, Middle Eastern, or Persian or indicated they identified as two or more racial/ethnic identities. Percentages are provided within columns, apart from the total gender where percentages are provided within the row.

Table 3.2

Additional Student Demographics and Background Characteristics (n = 1,036)

Variable	*n*	%
First-Generation Status		
First-Generation	294	20.3
Continuing Generation	694	47.9
Unknown or missing	48	3.3
Citizenship Status		
U.S. Citizen	880	60.7
Non-U.S. Citizen with permanent residency	61	4.2
Non-U.S. Citizen with temporary visa	51	3.5
Other Citizenship status	6	0.4
Unknown or missing	38	2.6
Socioeconomic status		
Poor	48	3.3
Below Average	163	11.2
Average	444	30.6
Above Average	325	22.4
Wealthy	25	1.7
Unknown or missing	31	2.1
Parent Career		
At least one parent with computing career	231	15.9
No parent with computing career	805	55.6
Transfer Status		
Student transferred from another institution	184	12.7
Student did not transfer	852	58.8
Major at the beginning of the introductory computing course		
Computing	732	50.5
Engineering	149	10.3
Biological Sciences	22	1.5
Business	34	2.3
Physical Sciences	9	0.6
Math/Statistics	23	1.6
Other	13	0.9
Health Professions	5	0.3
Social Sciences	25	1.7
Unknown or missing	24	1.7

Additional Student Demographics and Background Characteristics (Continued)

Variable	*n*	%
Undecided students' aspired majors		
Computing	79	7.6
Non-computing, other major or unknown	2	0.2
Non-undecided student	955	92.2
High School GPA		
A or A+	403	27.8
A−	285	19.7
B+	175	12.1
B	101	7.0
B−	21	1.4
C+	13	0.9
C	7	0.5
C−	3	0.2
D	2	0.1
Unknown or missing	26	1.8

Additionally, most students in this sample had parents who had earned a bachelor's degree, were born citizens of the United States, reported they were raised as part of an average socioeconomic class, shared that they did not have a parent with a career in computing, yet were either majoring in computing, minoring in computing, or were undecided but aspiring to become a computing major[20], and were high-achieving students in high school (see Table 3.2).

Regarding students' academic characteristics, many reported their average high school grade was a B or higher, and 51% of students were computing majors at the time of the first survey (see Table 3.2). Additionally, 84% of students reportedly had only one academic major,

[20] Computing majors/minors and aspirants included students in the following aggregated academic fields: Bioinformatics, Computer Science, Computer Information Systems/Informatics, Computing and Business (including Business Information Management and Management Information Systems), Computer Engineering (including Computer Engineering and Software Engineering), Data Science, Game Design, Information Science/Studies, Information Technology, and Other Computing.

8% reported having more than one major, and the remaining 8% indicated they were undecided about their major or did not report their major. Furthermore, all BRAID institutions are represented in the dataset, and 87% of students came from a publicly controlled university, with the majority of students having attended Arizona State University (17%); the University of Maryland, Baltimore County (15%); or the University of California, Irvine (11%; see Table 3.3).

Table 3.3

Frequency of Student Enrollment, by BRAID Institution

BRAID Institution	*n*	%
Arizona State University	174	16.8
College Park	75	7.2
Missouri University of Science and Technology	71	6.9
New Jersey Institute of Technology	76	7.3
University of California, Irvine	110	10.6
University of Illinois at Chicago	36	3.5
University of Maryland, Baltimore County	155	15.0
University of Nebraska-Lincoln	78	7.5
University of North Texas	39	3.8
University of Rochester	69	6.7
University of South Carolina	44	4.2
University of Texas at El Paso	38	3.7
University of Vermont	29	2.8
University of Wisconsin-Milwaukee	14	1.4
Villanova University	28	2.7
Total	1,036	100.0

Development of the Computing Identity Measure(s)

The primary dependent variable for this study was a new, latent, affective psychosocial measure of undergraduate students' computing identity. Guided by literature and the conceptual frameworks for each construct, I selected items that were present on all three surveys that conceptually aligned with this study's definition of computing identity[21]. While prior scholars have used the same surveys to develop factors of self-efficacy, sense of belonging, and other identity-related constructs (as mentioned earlier), I sought to create the most comprehensive computing identity factor possible. In alignment with the work of prior scholars (e.g., Carlone & Johnson, 2007; Rodriguez et al., 2020), my measure of computing identity includes items related to interest, competence, performance, and recognition. Additionally, as I am inspired by the work of Taheri et al. (2018) and Lunn et al. (2021a), my measure of computing identity also includes items related to sense of belonging to determine whether this construct is the cornerstone of our conceptualization of computing identity. Thus, while I am guided by existing literature and theory, the creation of this measure is still somewhat exploratory, so I use both exploratory and confirmatory factor analysis to achieve this goal—ensuring adequate factor loadings, eigenvalues, and internal consistency (Fabrigar et al., 1999; Streiner, 1994; Tabachnick & Fidell, 2007).

Importantly, while prior studies have created single measures of computing identity, there is limited research that has explored the extent to which the computing identity measure varies across time and student demographics. Therefore, before finalizing the dependent computing identity variable for this study, I examined the measure invariance at each survey time point and across categories of students' gender, race/ethnicity, and intersections of students' gender and racial/ethnic identities. This analytical process of assessing measure equivalence is explained in greater detail in this chapter's Analytical Plan and Critical Approach section.

[21] Computing identity represents the extent to which a student sees themselves as a computing person, based on how they make meaning of their core identity and navigate the world around them (Abes et al., 2007; Carlone & Johnson, 2007; Lunn et al., 2021a; Rodriguez et al. 2020; Rodriguez & Lehman, 2017).

Additionally, the first research question supported my quest to determine whether a psychometrically sound measure of computing identity could be created. I added items related to self-efficacy (i.e., performance and competence items) and sense of belonging to see if they were necessary components in how we measure one's computing identity. Then, I assessed if the new computing identity measure held over time and across students' gender, race/ethnicity, and gender and racial/ethnic identities combined. The computing identity measure created as a result of the first research question was then used as the dependent variable for the final research question. In the following section, I describe the operationalization of computing identity, but the detailed process for the factor development is provided later in this chapter and in the results of the first research question in Chapter Four.

Operationalizing Computing Identity

I anticipated that although computing identity would not be measured the same as science identity, it would have similar foundational components that held across students' racial/ethnic and gender identities but with a heavier emphasis on computing-related experiences and interactions during college. For example, just as Carlone and Johnson (2007) emphasize the fluidity of identity and the importance of students' performance, competence, and recognition paired with their racial, ethnic, and gender identities, I expected these to also be important in the conceptualization of computing identity. However, as the field of computing identity differs from that of other science disciplines (Sax & Newhouse, 2019), this study uses a computing-specific lens developed by Rodriguez et al. (2020) to create and test the computing identity measure resulting in a construct with items specific to the field of computing. Additionally, as there may be differences in the makeup of the computing identity factor across race/ethnicity and gender, there may also be differences in the variables that influence the predictability of computing identity by gender and race/ethnicity.

For this study, **computing identity** is operationalized as the extent to which students see themselves as a computing person. The computing identity measure relies upon a body of literature that suggests computing identity comprises five elements: interest, competence, performance, recognition, and sense of belonging (Abes et al., 2007; Carlone & Johnson, 2007; Gee, 1999; Lunn et al., 2021a; Mahadeo et al., 2020; Rodriguez et al., 2020; Rodriguez & Lehman, 2017; Taheri et al., 2018). Given the availability of items on the survey data being used for this study, I anticipated the concept of computing self-efficacy to be interwoven with students' perceived competence and performance characteristics.

This study's measurement of computing identity builds upon existing research by including students' interest in computing (Mahadeo et al., 2020; Rodriguez et al., 2020), social identification as a computing person (Tajfel, 1979), and sense of belonging as a computing person (Lunn et al., 2021a; Taheri et al., 2018). As my review of the literature revealed that it may be difficult to parse self-efficacy from identity, items related to computing self-efficacy are included in the computing identity measure as proxies for students' recognition of themselves and their ability to "do" computing. Specifically, the computing identity factor is informed, in part, by the work of prior scholars that measured computing identity with three self-report agreement items measured on a 5-point Likert scale (1 = *Strongly Disagree* to 5 = *Strongly Agree*): "I see myself as a computing person," and "Computing is a big part of who I am," as well as a reverse-coded item, "I do not have much in common with other students in my computing courses" (e.g., Lehman et al., 2022; Wofford, 2021).

While prior scholars developed their computing identity measures using the above survey items, this study moves their research forward by examining the extent to which self-efficacy and sense of belonging are core components of computing identity and by studying the equivalence of this new measure across survey time points and student demographics. See Table 3.4 for a list of proposed computing identity items, their coding schemes, and subconstructs that were explored in developing this study's computing identity measure. Only items that loaded onto their single factor (or subconstruct) at 0.5 or above, did not load onto another factor at 0.30 or above, were represented by at least three items, and that produced an overall Cronbach's alpha reliability of 0.70 were considered (Bolkan, 2017). Given that I include 18 items related to areas prior scholars found as subconstructs (e.g., Lunn et al., 2021a), I explore whether the best-fitting model for measuring computing identity is as a single factor with related but separate factors, a broad single-factor (unidimensional) measure, a bifactor, or a hierarchical (second order) factor (see Appendices C–E).

Table 3.4

The 18 Proposed Computing Identity Items

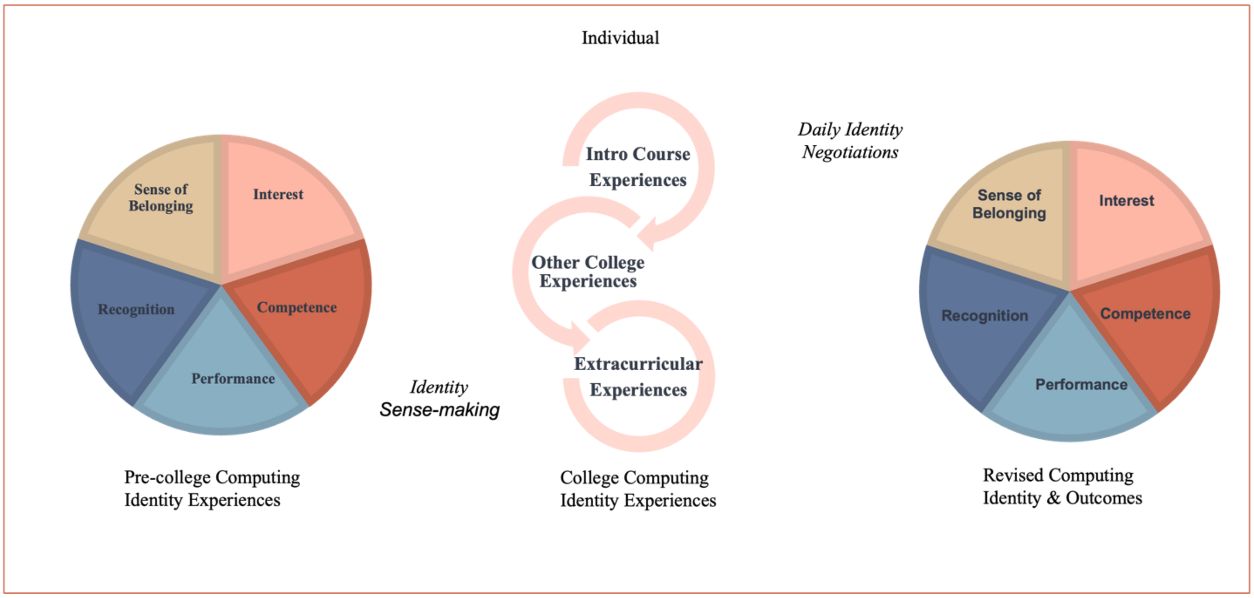
Item	Response Options
Interest How much do you agree or disagree with the following statements?	*1=Strongly Disagree; 2=Somewhat Disagree; 3=Neither Agree nor Disagree; 4=Somewhat Agree; 5=Strongly Agree*
—I am interested in learning more about what I can do with computing	
—Using computers to solve problems is interesting	
—I care about doing well in computing.	
Competence I am confident that I can…	*1=Strongly Disagree; 2=Somewhat Disagree; 3=Neither Agree nor Disagree; 4=Somewhat Agree; 5=Strongly Agree*
—Find employment in an area of computing interest.	
—Become a leader in the field of computing.	
—Win a computing-related contest (e.g., programming/robotics contest or hackathon).	
—Get admitted to a graduate computing program.	
—Complete an undergraduate degree in computing.	
—Quickly learn a new programming language on my own.	
—Clearly communicate technical problems and solutions to a range of audiences.	
Performance How would you rate yourself in the following areas compared to the average person your age?	*1=Lowest 10%; 2=Below Average; 3=Average; 4=Above Average; 5=Highest 10%*
—Computer skills	
Recognition How much do you agree or disagree with the following statements?	*1=Strongly Disagree; 2=Somewhat Disagree; 3=Neither Agree nor Disagree; 4=Somewhat Agree; 5=Strongly Agree*
—Computing is a big part of who I am.	
—My family emphasizes the value of earning a computing degree.	
Belonging How much do you agree or disagree with the following statements?	*1=Strongly Disagree; 2=Somewhat Disagree; 3=Neither Agree nor Disagree; 4=Somewhat Agree; 5=Strongly Agree*
—I feel like I "belong" in computing.	
—I feel like an outsider in the computing community.	
—I feel welcomed in the computing community.	
—I do not have much in common with the other students in my computing classes.	

Computing Identity Predictive Model

The conceptualization and factor creation of computing identity, the variables selected for the regression, and the order in which they entered the model relied upon the individual level of the guiding computing identity framework described in Chapter Two (Rodriguez et al., 2020). In Rodriguez et al.'s 2020 book chapter, the authors propose a conceptual computing identity framework for Latinas. The model is noteworthy for several reasons. First, it focuses on different levels of influence on Latina's computing identity development. Second, it centers on individual student assets to the field and names the systems of power and oppression they must navigate during their undergraduate experience. Finally, building upon Carlone and Johnson's 2007 framework, this model acknowledges the importance of students' interest, competence, performance, and recognition; speaks to the temporal influence of pre-college and during-college experiences; names intersectionality, which attends to Rodriguez and Lehman's (2017) call to action for more research on this topic; and has roots in the field of technology, which computing as a discipline is situated within (Rodriguez et al., 2020). While the model was created with Latinas in mind and its theoretical underpinnings stem from qualitative research, it still serves as an important guide for this study in documenting potential factors at play and determining their influence on developing students' computing identity, particularly for Underrepresented Students of Color in computing (USOCC). See Figure 3.2 below for an overview of the key variable categories and Appendix A for a complete list of variables that were included in the models, including their operationalizations and coding schemes. All variables were measured using a 5-point Likert scale ranging from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*) unless otherwise noted.

Figure 3.2

Theoretical Framework Informing the Computing Identity Measure and the Regression Model



Note. The difference between this figure and those provided in the prior chapters is that this visual is more nuanced in that it focuses particularly on the factors influencing students' computing identity at the *individual* level.

Pretest Computing Identity

A measure of students' pretest computing identity (α = 0.878) was used as a baseline measure to control for their incoming computing identity (Astin & antonio, 2012). This three-item factor was identical to the dependent variable but was measured at the beginning of students' introductory computing course. The time interval between this pretest survey measure and the first follow-up survey outcome measure (taken in the fall term following students' completion of their introductory course) was approximately six months to one year, depending on whether students completed their initial surveys in the fall or spring term of the previous year.

Personal Characteristics

While personal characteristics were not documented in Rodriguez et al.'s 2020 model for Latinas in computing, this study encompasses students from different racial/ethnic and gender backgrounds. As such, documenting their demographics is important. Variables in this section include students' gender identity, racial/ethnic identity, and an indicator of whether they would be the first in their family to earn a bachelor's degree. These demographic variables came from the pretest survey, unless missing, in which case an individual students' information was acquired from their responses to other survey timepoints.

Background Contexts

Prior literature suggests that students' preparation during high school is an important factor in promoting their identity development during college (Azmitia et al., 2009; Chang et al., 2014; Robnett et al., 2015). Consequently, an indicator of whether students took an AP computer science course in high school (Sax et al., 2022) and a variable measuring their high school GPA were considered. Additional predictors include a measure of students' class standing upon entry to the introductory computing course (e.g., as a first- or second-year student), an indicator of whether the student transferred from another college or university (Blaney, 2020; Blaney et al., 2022), and a measure of their perceived socioeconomic status as a way to capture students' family background (Rodriguez et al., 2020). These data came from the pretest survey.

Finally, while one of the assumptions of linear regression is that all variables in the model must be at the same "level" (i.e., a student-level model only), as this is a multi-institution study, I also investigated if institutional-level variables needed to be included in the model (PremalMatalia, 2020). To determine whether this was necessary, I ran a one-way analysis of variance (ANOVA) to explore the extent to which computing identity varies by institution. Then, I examined the intra-class correlations (ICCs) for the independent variables on the computing identity outcome to get a more detailed understanding. While one ICC exceeded the 10% threshold (the pre-test computing identity independent variable), the ICC for the dependent variable (a later time point for this variable) for this study was only 8%. Thus, I determined that a fixed-effects model was not necessary. In other words, despite the sample containing a different proportion of students across the 15 BRAID institutions (see Table 3.1), students' computing identities did not vary by institution, so it was not necessary to account for institutional fixed effects. Therefore, I did not include dummy-coded BRAID institutional variables in the regression models.

Early Computing Identity Experiences

Rodriguez and her colleagues (2017, 2020) underscore the importance of students' pre-college computing-related experiences on their computing identity development. This block of items comes from the pretest survey and includes an indicator of whether students' parents/guardians had a career in computing in order to signify whether this was an important predictor. A measure of students' familial support served as a proxy for students' pre-college identity-affirming experiences and to describe their funds of identity (Rodriguez et al., 2020; Wofford & Gutzwa, 2022). Specifically, I created a factor measuring students' family support (α = 0.687) consisting of three items. For these items, students were asked their level of agreement with statements such as, "My family encourages me to pursue a computing degree."

Additionally, because 69% of the sample reported at least some prior programming experience and programming experience is linked with computing self-efficacy, which is linked to computing identity (Chemers et al., 2011; Seymour et al., 2004), I included an aggregate measure of the amount of prior programming experience students had upon entry to the introductory computing course. This variable was measured using a sum of seven binary items where students responded "yes" or "no" to learning programming through a course in high school, at a computer camp, in an online course, by taking a college-level programming course at their current college, by taking a college-level programming course at another four-year college, by taking a course at a community college, or by learning programming on their own. As such, the aggregate measure of students' prior programming experience ranged from zero to seven.

Additional Psychosocial Computing Identity Factors

While broader science identity literature suggests that some psychosocial factors (e.g., self-efficacy, interest in computing) predict computing identity and that sense of belonging may be a component of identity, any items that did not load on my computing identity measure were added in this block to investigate the extent to which they are precursors to computing identity, *predicting* it instead of being components of it (e.g., Chemers et al., 2011; Leary et al., 2013; Lunn et al., 2021b; Margolis et al., 2000; Robnett et al., 2015; Seymour et al., 2004). Based on the results (detailed in Chapter Four), these measures included students' interest in computing and two self-efficacy factors. Each of these measures came from the posttest survey measured at the end of students' introductory course (i.e., the second survey timepoint) and included three survey items. A sample item from the interest measure ($\alpha = 0.846$) includes, "I am interested in learning more about what I can do with computing." The first computing self-efficacy measure related to students' professional competency ($\alpha = 0.817$) with one item including, "I am confident I can quickly learn a new programming language on my own." The other computing self-efficacy measure related to students' professional milestones ($\alpha = 0.847$) and included the following item: "I am confident I can find employment in an area of computing interest."

College Computing Identity Experiences Within the Introductory Course

Students' experiences in introductory courses are pivotal to their continued success and persistence in computing. Therefore, all items in this block came from the posttest survey that allowed me to capture these experiences. Given that culturally relevant teaching pedagogy is encouraged by faculty to create more identity-inclusive and equitable classroom environments (Daily et al., 2022; Lunn et al., 2021a), I created a four-item measure representing students' perception of the extent to which their instructor provided relevant social and personal examples ($\alpha = 0.795$). These items inquired about whether examples involved women across race/ethnicity, People of Color regardless of their gender identity, or relevant social issues, or whether they addressed misconceptions about the computer science field.

Additionally, in response to the continued need for faculty to promote relevant teaching pedagogy in classrooms (Ladson-Billings, 2021), I created a six-item measure of students' perception of instructional inclusivity ($\alpha = 0.906$). Items that make up this factor included students' level of agreement with items that asked if their instructors and teaching assistants (TAs) were inclusive and supportive of women or Students of Color, if they were interested in helping when the student came to them with questions, and if they were responsive to questions in class and via email (Blaney & Stout, 2017). In addition, given the well-documented hostility that some marginalized students face in computing (e.g., Ashcraft et al., 2016; Whitney, 2021), the final measure in this section included a factor of feeling dismissed ($\alpha = 0.832$) that included three items measured on a 5-point Likert scale ranging from 1 (*Never*) to 5 (*All of the Time*). In this measure, students indicated the extent to which their "ideas or opinions [were] minimized or ignored."

College Computing Identity Experiences Outside of the Introductory Course

Just as students' experiences in the introductory courses are important for their computing identity, so are their experiences outside of their introductory courses. To account for the importance of students' social interactions and experiences outside of the classroom, I added five variables. The first two variables represented latent factors: a four-item measure of peer support ($\alpha = 0.879$) and a three-item measure of mentor support ($\alpha = 0.916$), both measured using a 5-point Likert scale to which students reported their level of agreement from 1 (*Not at All*) to 5 (*Very Much*). The peer support prompt included, “To what extent is each of the following kinds of support available to you from other computing students if you need it?” A sample item from this measure is, “Someone to confide in or talk to about your problems.” For the mentor support measure, students were asked to rate their level of agreement to the following question: “To what extent do you have a mentor who-,” and a sample item is, “shares personal experiences as an alternative perspective to your problems.”

Next, a variable that indicated whether students were computing majors or minors at the last survey time point was also included. Kapoor and Gardner-McCune (2019) describe the important role that students’ participation in professional development opportunities and hackathons plays in fostering their identity development. Thus, I included two binary variables to indicate whether students participated in an internship or co-op, or if they attended a technical conference in computing (Aschbacher et al., 2010; Charney et al., 2007; Hunter et al., 2007).

Analytic Plan and a Critical Approach

My primary analytical approaches were threefold. First, I relied upon factor analysis to create a measure of computing identity in response to the first research question. Then I attended to the second research question by assessing the measure invariance of the new computing identity factor over time during college and across student demographics. Finally, in response to my third research question, I used multiple regression to identify pre- and during-college factors that informed the development of computing identity for all students and explored the unique predictors for women and USOCC in computing compared to their peers from other social identity groups.

In approaching these analyses, and in alignment with my research positionality, I applied aspects of quantitative criticalism (Carter & Hurtado, 2007) where possible to avoid perpetuating systems of power and oppression in the academy (Garvey, 2020; Wells & Stage, 2015). In the field of higher education, quantitative studies deemed *critical* are becoming increasingly popular; however, there are a variety of perspectives on the extent to which studies are indeed quantitatively "critical" (Garvey, 2020; Stage & Wells, 2014; Wells & Stage, 2015). In this study, I took quantitatively critical approaches in several ways. Conceptually and theoretically, I took an equity approach that promotes social justice through my research inquiries, which supports a nuanced understanding of how we conceptualize and make meaning of students' computing identities. I also applied criticality throughout my methodology (e.g., assessing measure invariance, honoring students' gender and racial/ethnic identities, and using a comparative group approach toward understanding how the factors predicting computing identity vary by gender and race/ethnicity). The following sections describe these approaches in greater detail.

Research Question 1: Developing a Computing Identity Measure

This section describes the processes for developing the computing identity measure, assessing model fit, and identifying the appropriate model structure. I used factor analysis to develop the measure of computing identity. In particular, factor analysis offers statistical strength by reducing the number of items that "hang" together on a similar latent construct (Tabachnick & Fidell, 2007). In addition, it offers practical strength as it allows researchers to make subjective decisions informed by literature that support greater interpretability and implications for practice (Streiner, 1994). As discussed in Chapter Two, the terminology referring to computing identity, self-efficacy, and sense of belonging have been used interchangeably in the existing literature. Therefore, because the operational definitions for this study (as informed through literature and theories) do differ, but I identified a need for more clarification to parse these constructs from one another, I utilized both exploratory and confirmatory factor analysis (EFA and CFA) to identify items that loaded onto—and the structure of—a computing identity measure. Further, only survey items that corresponded to computing identity and were informed through literature and theory were examined, and the analytical decisions to create the factor were informed by scholarship from the fields of statistics (Wells, 2021), higher education (Eagan, 2020), and computing education (e.g., Lunn et al., 2021a). See Table 3.4 for a curated list of the 18 proposed survey items that were explored in developing this computing identity measure.

Broadly speaking, the primary target sample for the development of this factor was the population of students from both cohorts who both completed the pretest survey and met the remaining study restrictions (n = 4,269). However, I used SPSS to randomly split this sample into two nearly equivalent groups, in alignment with literature advising the use of different samples for EFA than is used CFA and to achieve sample sizes of at least a 10 to 1 ratio of students per item and that the overall sample exceeds 300 cases (DeVellis & Thorpe, 2021; Fabrigar et al., 1999; Kline, 2023; Tabachnick & Fidell, 2007). Approximately 50% of the sample was included in an initial EFA to determine which items would be appropriate for the scale, and the remaining students were included in the CFA to confirm the factor structure (DeVellis & Thorpe, 2021). Then, the model structure was confirmed using the combined sample of students who completed the pretest survey. See Table 3.5 for a summary of the study analyses and corresponding sample sizes.

Table 3.5

Summary of Study Analyses and Sample Sizes

Analysis	Sample	Actual Sample Analyzed (excluding missing cases)
Assessment of Items		
—Initial EFA —Targeted EFA	First ½ of students who completed the pretest survey (n = 2,142)	(n = 2,068)
Confirmation of Factor Structure		
—CFA	Second ½ of students who completed the pretest survey (n = 2,134)	(n = 2,091)
Assessment of Model Structure		
—CFA —Unidimensional —Bifactor —Second Order	All students who completed the pretest survey (n = 4,269)	(n = 4,158)
Assessment of Demographic Measure Equivalence		
—Measure invariance testing	All students who completed the pretest survey (n = 4,269)	*Gender* (n = 4,017) *Race/Ethnicity* (n = 3,502) *Gender & Race/Ethnicity* (n = 3,460)
Assessment of Longitudinal Measure Equivalence		
—Measure invariance testing	Students who participated in all three survey time points (n = 1,036)	(n = 1,036)
Assessment of Factors that Predict Computing Identity		
—Hierarchical blocked multiple regression analyses	Students who participated in all three survey time points (n = 1,036)	*All Students* (n = 660) *Men* (n = 417) *Women* (n = 243) white students (n = 307) Asian students (n = 244) USOCC (n = 173)
—Stepwise multiple regression analyses		*All students* (n = 768) *Men* (n = 492) *Women* (n = 276) white students (n = 375) Asian students (n = 268) USOCC (n = 193)

Note. Underrepresented Students of Color in computing (USOCC) includes students from the following racial/ethnic identities: Black or African American; Hispanic or Latina/o/x; American Indian; Native Hawaiian or Pacific Islander; Arab, Middle Eastern, or Persian; and students from two or more racial/ethnic groups (Wofford et al., 2022). The racial/ethnic makeup of the sample groups for the blocked model and the stepwise model, respectively, were: Black, 20% and 21%; Latine, 39% and 37%; Native American, 1% and 1%; Native Hawaiian or Pacific Islander, 1% and 1%; Arab, 8% and 8%; and two or more race/ethnicities, 31% and 32%.

Exploratory Factor Analysis. The first step of my data analysis was determining the appropriate number of factors and corresponding items for the new measure of computing identity. To do this, I conducted an EFA with robust maximum likelihood estimation using Mplus version 8.7 (Muthén & Muthén, 2017). The purpose of the initial EFA was to assess which of the 18 proposed items related to computing identity loaded onto one or more latent constructs. Only those without missing data were included, yielding a sample of 2,068 students who completed the pretest survey. I used goemin rotation in Mplus to allow for correlations between factors, as prior literature suggests these items may be related (e.g., Beyer, 2014; Blaney et al., 2022; Blaney & Barrett, 2022; Cohoon & Aspray, 2006; George et al., 2022; Lehman, 2017; Lehman et al., 2022) and tested the models with between one and five latent variables. I then ran a second, targeted EFA with the same pretest sample retaining only the 12 items that loaded on a latent variable at 0.50 or higher without a cross-loading of 0.30 or higher (Bolkan, 2017) and had at least two items represented.

Confirmatory Factor Analysis. Next, I used CFA to confirm the structure of the items that loaded onto the computing identity factor in a second sample of students (i.e., the remaining half of students who completed the pretest survey), excluding those with missing data, yielding a sample of 2,091 students.

Research Question 1a: Assessing Subcomponents of Computing Identity

After establishing the computing identity factor through the steps above, it was important to determine whether the computing identity measure had one or more subcomponents (Tabachnick & Fidell, 2007), considering both the extant literature that describes how these measures are multifaceted (Marakas et al., 1998; Jones & McEwen, 2000) and the various ways computing identity has been measured (Taheri et al., 2018; 2019; Lunn et al., 2021a). Therefore, I ran a second CFA to determine whether the measure was unidimensional or multidimensional (i.e., multifactorial). For this additional CFA, I used the full sample of 4,158 students who took the pretest survey to assess the model fit, comparing a multidimensional CFA with unidimensional, bifactor, and second-order factor structures (DeVellis & Thorpe, 2021; see Appendices C–E). More specifically, if computing identity were a simple single-factor/unidimensional model, all items would have strong loading on a single computing identity measure. Whereas, if computing identity were best represented as a bifactor model, sub-items would load unidimensionally within several subcomponents or latent variables (e.g., interest, recognition, performance, competence, sense of belonging) and on a global computing identity factor. However, if computing identity were best represented as a second-order/hierarchical factor, all sub-items would first load on unique first-order factors (e.g., interest, recognition, performance, competence, sense of belonging) and then on a second-order factor (i.e., computing identity). Here the sub-items would not have a direct relationship with the overarching measure of computing identity, but instead would be related through their subfactors (DeVellis & Thorpe, 2021). See Appendices C–E for illustrations of these possible model structures.

After confirming the factor structure in Mplus, I computed Cronbach's alpha (α) and McDonald's omega (ω) to ensure that the reliability of the items that comprised the factor(s) was greater than 0.70 and did not exceed 0.90.

Assessing Model Fit. Based on the work of prior scholars (e.g., Bolkan, 2017; DeVellis & Thorpe, 2021), I relied on a combination of the Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), Comparative Fit Index (CFI), and Root Mean Square Error of Approximation (RMSEA) to assess the model fit for both the exploratory and confirmatory factor analyses. Specifically, I looked for the lowest AICs and BICs in the comparison of model options with a CFI ≥ 0.95 and an RMSEA ≤ 0.08 (Fabrigar et al., 1999; Hu & Bentler, 1999; Wells, 2021).

Research Question 2: Measuring (In)variance in the Computing Identity Measure

This section describes the process of using factorial measure invariance to assess whether the new computing identity measure functions equivalently over time and across students from varying demographic (gender and racial/ethnic) social groups (Litson & Feldon, 2021). Several scholars encourage researchers to run additional inferential analyses (beyond assessing reliability with Cronbach's alpha and McDonald's omega) to ensure equivalency of the new measure over time (longitudinal factor model) and across the various social identities represented in their samples (multigroup confirmatory factor model; Geiser, 2022, 2023; Litson & Feldon, 2021; Marcoulides & Yuan, 2017; Wells, 2021). Therefore, I examined the invariance of the computing identity measure over time and across student demographics by using longitudinal and multigroup confirmatory factor analysis in Mplus (Wells, 2021).

While several scholars have identified a variety of approaches to assess measure invariance, I was guided by Putnick and Bornstein's (2016) structural equation modeling (SEM) framework for CFA for this study. In this framework, the authors identify a four-step process to assess for configural, metric, scalar, and residual invariance whereby the tests become more stringent as you progress along the spectrum. *Configural invariance,* the least stringent of the four steps, allows researchers to assess if the pattern of parameters (factor loadings and intercepts) for all the items that load onto the measure of computing identity are equivalent across subgroups (Putnick & Bornstein, 2016). For example, configural invariance allows for the investigation of whether an item such as "I see myself as a computing person" consistently loaded onto the measure of computing identity for each subgroup of interest. *Metric invariance* (i.e., weak measurement equivalence) assesses if the strength of the relationship between the factor and each item is the same for each subgroup of interest by constraining factor loadings (i.e., making them equivalent) but allowing intercepts to range freely (Wells, 2021). In practice, this means comparing the magnitude of loadings for each item related to computing identity for each subgroup of interest. *Scalar invariance* (i.e., strong measurement equivalence) constrains both the factor loadings and intercepts for each item of the computing identity measure to assess if making them equivalent is a worse fit than metric invariance for all subgroups of interest. In other words, with this approach, we assess whether the mean differences in the computing identity factor account for all the subsequent mean differences across the items by constraining the factor loadings and intercept measurement parameters (Wells, 2021). *Residual invariance* (i.e., strict invariance) is the practice of assessing the equivalence of item residuals of metric and scalar invariant items. However, Wells (2021) describes it as overly restrictive and Putnick and Bornstein (2016) mention that many researchers omit this step. Therefore, because I was unable to assess configural invariance and residual invariance, which is often seen as too strict, I focused instead on the two midpoints described above—metric and scalar invariance.

In addition to these measure invariance (equivalence) assessments, I used paired samples *t*-tests in order to assess how computing identity mean scores changed between the pretest and the first follow-up survey for each group: (a) all students, (b) women, (c) men, (d) USOCC, (e) Asian students, and (f) white students. I also ran independent samples *t*-tests (for gender groups) and an ANOVA (for race/ethnicity groups) to assess between-group differences of students' computing identity scores within a given survey timepoint.

Assessing Model Fit. Assessing whether a model adequately fits the data involves examining the chi-square fit statistic and its corresponding *p*-value; a non-significant chi-square fit statistic is preferable, indicating a good model fit such that there is no significant difference between the model predictions and the data (Litson & Feldon, 2021). While chi-square is a typical indicator of fit, its use as such has been critiqued because of its extreme sensitivity to large sample sizes (Meade, 2005)—highlighting differences where they may be otherwise less meaningful (Chen, 2007). Therefore, I first investigated if the chi-square fit statistic was significant using a more stringent significance threshold (i.e., $p < .001$; Litson & Feldon, 2021). Then I investigated other accepted fit indices, namely, CFI and RMSEA, where a well-fitted model is identified when CFI ≥ 0.95 and RMSEA ≤ 0.08 (Fabrigar et al., 1999; Hu & Bentler, 1999; Wells, 2021). However, there are some considerations worth mentioning. First, CFA is often only possible with large sample sizes (Kline, 2023), as the chi-square goodness-of-fit indices produced from the maximum likelihood (ML) estimation allow adequate power to reject the null hypothesis even when the model does not fit the population (Wells, 2021). However, as further explained by Wells (2021), "Just because a model does not fit perfectly in the population does not mean it is not useful or that it does not provide a good enough approximation" (p. 261). Thus, as parameter estimates tend to be statistically significant with large datasets used in CFAs, I also examined factor (co)variances, factor loadings, intercepts, and residual (co)variances to interpret the magnitude of the parameters (Wells, 2021).

Overall, by clarifying these psychosocial factors, findings from this research question allow me to address a gap in the literature as to whether a strong single-factor psychosocial discipline-specific measure of computing identity can be created that holds across student demographic (gender and racial/ethnic) groups.

Research Question 3: Predicting Computing Identity

I used a combination of quantitative methods to address the third research question. First, before running the regression analyses, I ran frequency distributions to assess normality and variance and tested for multicollinearity among the independent variables—ensuring that no two variables exceeded a correlation of $r \geq .70$. While two variables—undergraduate research participation and technical conference participation—had little variance (most students had not indicated involvement in either computing-related activity), I chose to keep both variables in the model. This decision was informed by prior literature suggesting they were important in fostering students' persistence in computing because computing identity is a predictor of persistence. Thus, I wanted to see if these variables predicted computing identity.

Next, I ran a series of 12 multiple regression analyses, first using a blocked forward multiple regression approach then using a stepwise multiple regression approach. As opposed to analyses that force all variables to enter the model at the same time, I started with running six *hierarchical blocked multiple regressions* among each group of students: (a) all students, (b) women, (c) men, (d) USOCC, (e) Asian students, and (f) white students. The blocked approach allowed me to document how each of the seven conceptually related blocks of variables accounts for variation in the outcome, with the most predictive variables entering first, within their theoretical temporal blocks informed by the theory that guides the corresponding model (Astin & Dey, 1996; Field, 2013).

However, because findings varied across students' social groups and some variables were initially significant but lost significance as other sets of variables entered, the approach of entering variables *together* within blocks made it impossible to track how the entry of certain variables may have affected the predictive power of other variables. Therefore, with interest in further investigating the important changes within those particular blocks and diagnosing which variables were driving this said change, I decided to rerun the regressions using a *blocked stepwise multiple regression* (Astin & antonio, 2012), again among each of the six student groups: (a) all students, (b) women, (c) men, (d) USOCC, (e) Asian students, and (f) white students. In these revised models, however, only the variables that entered at any step for at least one social group from the initial regressions were included. This approach allowed me to better track the step-by-step changes in regression coefficients as each new variable was added to the model, providing greater insight into the relative predictive power of each independent variable on the computing identity dependent variable as additional variables were being controlled.

For both approaches to multiple regression, the pretest computing identity factor was force-entered at the first step to highlight the simple correlation between the pretest and outcome computing identity variables and the significance between the two variables (Sax & Harper, 2007). As mentioned earlier in this chapter, I also included independent variables that were not components of the computing identity outcome (i.e., the interest in computing factor and the two self-efficacy factors) since they may serve as predictors of computing identity.

Analytically, both blocked stepwise individual variable regression modeling approaches allowed me to study the relationship between key independent/predictor variables (e.g., gender, race/ethnicity, introductory course experiences) on the dependent variable of interest (i.e., computing identity) while controlling for student demographic, background, psychosocial, and other college-related variables as informed by the respective theoretical framing (Astin & Dey, 1996).

Research Questions 3a and 3b: Predicting Computing Identity Separately by Social Group

Informed through critical quantitative research methodological approaches, education researchers can work to center the experiences of certain groups of students and look for differences among these students in their analyses (e.g., limit the sample to only women and see what differences arise by race). Importantly, Carter and Hurtado (2007) posit that comparative group approaches, as opposed to group-specific approaches whereby all students are entered into the model together, are appropriate for use in studies where prior literature documents expected differences in the outcomes of interest. In this case, there is a variety of literature that highlights gender and racial/ethnic equity gaps in terms of students' computing identity. Contrary to a single model "group-specific approach," running analyses *separately* by gender and race/ethnicity in the "comparative group approach" allows me to examine how the predictive power of the variables—which would otherwise be masked or only illustrated through many conditional effect/interaction terms added to the model—vary among each group (Astin & antonio, 2012; Carter & Hurtado, 2007; Sax, 2008).

Assessing Model Fit. To determine the regression model that best fits the data in predicting computing identity for students in this study, I assessed the model's goodness of fit using R^2 as a measure of variance explained by the model (Field, 2013). In other words, R^2 is a measure of the overall strength of the regression model, which represents the proportion of variance accounted for by the variables in the regression equation (Astin & antonio, 2012). Ranging from 0 to 1, the R^2 statistic is the multiple correlation (i.e., the correlation between actual and predicted score; Astin & antonio, 2012). While the size of the R^2 depends on many factors including sample size, the number of variables, and the variability and usefulness of independent variables in predicting the outcome (Astin & antonio, 2012), the higher the value, the better the model. While a model that explains 100% of the variance is unrealistic, I aimed for models that explain around 60% of the variance that exists in the development of students' computing identity, which is considered "high" in higher education research (i.e., $R^2 \geq .60$; Eagan, 2020). While R^2 does not imply causality, it is useful in describing the model's strength in predicting the outcome of interest (Field, 2013).

By documenting which variables predict these psychosocial outcomes and for which demographic groups these predictors remain true, findings from these research questions contribute to the literature on what is known about computing identity and how to better foster this among all students, especially those minoritized by their gender and racial/ethnic identities in computing, to support persistence in the field.

Taken together, the critical quantitative approaches described above provide me with a mechanism for accounting for differences across gender within racial/ethnic categories and provide for a deeper understanding of differences by race/ethnicity within gender categories. Additionally, criticality in this research study provides me with a platform to acknowledge and inform the computing community of potential individualized approaches that may be needed to tailor and ultimately transform the support provided to the development of computing identity among all students in order to increase equity in computing.

Study Strengths and Limitations

It should be noted that quantitative research studies are inherently unable to provide the same level of in-depth understanding of students' experiences as qualitative interviews and focus groups. Still, quantitative research inquiries such as this offer several notable strengths. First, quantitative research that relies on large sample sizes allows researchers to explore statistical and practical significance, which is particularly useful for university administrations that determine policy based on “strength in numbers” approaches. Second, while quantitative studies are limited to existing data, quantitative methods allow for the inclusion and testing of an effect and the magnitude of that effect from multiple factors on an outcome of interest. Finally, these robust approaches allow quantitative studies to hold constant factors that institutions and departments cannot change (e.g., student demographic, institutional characteristics) to examine the unique effect of variables on a given outcome.

Findings from this study must be interpreted with caution, however, as institutions self-selected to participate and all students were enrolled at research-intensive universities, meaning that their experiences may be unlike those of computing students at other institutions. While the demographics of those who participated in the BRAID surveys are largely representative of students enrolled in computing at similar research institutions, as mentioned earlier, it is important to acknowledge potential non-response bias, in that those who responded to the survey and their survey responses may be inherently different from those that did not respond to the survey (Blaney et al., 2020). However, given how diverse all higher education systems are becoming, examining students (especially those marginalized in computing) at research-intensive institutions is important, too. There are also many differences within computing disciplines (Cohoon & Aspray, 2006) that this study does not address. Future research may consider disaggregating both STEM subdisciplines (Sax & Newhouse, 2019) and the computing fields (Cohoon & Aspray, 2006) in their examinations. Unfortunately, these data do not allow for the measurement of students’ experiences concerning racism, sexism, hostility, and microaggressions, which Rodriguez et al. (2020) and Harper (2012) argue are critical steps in measuring and naming these systems of power and oppression and in understanding the entirety of systemic influences on students’ computing identities.

Additional threats to the validity of these findings include criticisms of the proposed analyses. Factor analysis is critiqued for its subjective nature and infinite number of available rotations (Tabachnick & Fidell, 2007). Analyses with a comparative-group approach are often critiqued because while they explore nuances between groups, they still assume, to some extent, homogeneity of those within the groups being compared (e.g., in this case, presenting all women or all Asian students as monoliths). However, I counter that argument by examining the experiences among women by race/ethnicity and across race/ethnicity by gender, thus examining the experiences of students who hold multiple social identities.

Thus, although the strengths of this study outweigh any threats to validity, future researchers still need to replicate and extend these findings using other datasets and samples, as is the case with most research.

CHAPTER FOUR: RESULTS

Informed by the importance of broadening the participation of all women and students from marginalized racial/ethnic social identities in computing described in Chapter One, the review of extant literature presented in the first part of Chapter Two, and the research methodology described in Chapter Three, the analytical scope of this study is threefold. First, I aim to develop a psychometrically sound measure of computing identity. In doing so, I identified the measure structure and determined the extent to which self-efficacy and sense of belonging are necessary components of computing identity. Second, I explore how well the computing identity measure holds across each of the three survey time points and for students from various gender and racial/ethnic identities. Finally, I examine the key variables that facilitate or impede the development of computing identity among undergraduate students who took an introductory computing course. I also investigate if these factors vary by gender and race/ethnicity in accordance with my theoretical and conceptual frameworks described in the second part of Chapter Two.

This chapter begins with an overview of my research questions. I then briefly describe the analytical procedure before providing the findings corresponding to each research question listed below.

Research Questions

1. To what extent can a psychometrically sound measure of computing identity be developed for students who took an undergraduate introductory computing course?
    a. Are self-efficacy and sense of belonging necessary components of computing identity?

2. To what extent does computing identity hold over time during college and by gender, race/ethnicity, and intersections of race/ethnicity?
3. What pre-college and college-related variables predict computing identity in the year following students' participation in an introductory computing course?
    a. To what extent do these predictors differ for women and men?
    b. To what extent do these predictors differ across racial/ethnic categories?

Research Question 1: Developing a Computing Identity Measure

In response to my first research question, I used both exploratory and confirmatory factor analysis to assess whether a psychometrically sound measure of computing identity could be created. The following sections provide a high-level overview of the steps I took to run these analyses (see Chapter Three for more details), emphasizing the results I discovered along the way.

Exploratory Factor Analysis

I ran two exploratory factor analyses (EFAs). Using a random sample of roughly half of the pretest survey data among students who met all other study criteria (n = 2,068) and the 18 items that aligned with the literature on the five subcomponents of computing identity, I ran the first EFA to assess the number of items and factor(s) onto a computing identity measure. Based on the results, I retained a five-factor solution as findings from this initial EFA indicated this was the best-fitting model (see Table 4.1). However, because several items cross-loaded on different factors or had low loadings across all factors and only two reverse-coded items were represented for one factor, I only retained 12 variables representing four latent factors (see Table 4.2). I then included these 12 items in a second, targeted four-factor EFA to test how the data fit the proposed model. Results from this EFA indicated that the data fit the model relatively well, χ^2

(24) = 45.608, $p < .01$, *SRMR* = 0.008, *CFI* = 0.997, *RMSEA* = 0.021 (CI: 0.011, 0.030) with all variables loading significantly on their respective factors (see Table 4.4).

Table 4.1

Summary of Model Fit, the Initial EFA with Five Factors

Model	Chi-square	AIC	BIC	*df*	*p*	*CFI*	*SRMR*	*RMSEA*	*CI*
One-factor	4,372.01	88,702.55	89,006.81	135	< .01	0.63	0.10	0.12	(0.12, 0.13)
Two-factors	2,551.35	86,282.53	86,682.56	118	< .01	0.79	0.06	0.10	(0.10, 0.10)
Three-factors	1,320.95	84,867.33	85,357.52	102	< .01	0.89	0.04	0.08	(0.07, 0.08)
Four-factors	736.75	84,206.00	84,780.71	87	< .01	0.94	0.03	0.06	(0.06, 0.06)
Five-Factors	347.53	83,771.72	84,425.30	73	< .01	0.98	0.02	0.04	(0.04, 0.05)
Models compared									
1 vs. 2	1,410.73			17	< .01				
2 vs. 3	1,240.41			16	< .01				
3 vs. 4	578.05			15	< .01				
4 vs. 5	373.11			14	< .01				

Note. chi-square = Satorra-Bentler scaled chi-square. AIC = Akaike Information Criteria. BIC = Bayesian Information Criteria. CFI = Comparative Fit Index. SRMR = Standardized Root Mean Square Residual. RMSEA = Root Mean Square Error of Approximation. CI = 90 percent confidence interval. Adequate model fit to the data is determined by low AIC and BIC and CFI ≥ 0.95 and RMSEA ≤ 0.08 (Hu & Bentler, 1999).

Table 4.2

EFA, a Five-Factor Solution with GEOMIN Rotated Loadings

Item	Factor 1	2	3	4	5
1. I am interested in learning more about what I can do with computing	0.06	**0.70**	0.05	−0.02	−0.05
2. Using computers to solve problems is interesting	−0.01	**0.75**	−0.01	0.12	0.06
3. I care about doing well in computing.	0.01	**0.64**	0.17	−0.08	−0.02
4. I am confident I can—find employment in an area of computing interest.	0.05	0.05	**0.71**	0.05	−0.01
5. I am confident I can—become a leader in the field of computing.	−0.01	−0.03	0.12	**0.76**	−0.06
6. I am confident I can—win a computing-related contest (e.g., programming contest, robotics contest, hackathon).	0.01	−0.08	0.02	**0.80**	−0.02
7. I am confident I can—get admitted to a graduate computing program.	0.03	−0.05	**0.62**	0.17	0.02
8. I am confident I can—complete an undergraduate degree in computing.	−0.02	0.13	**0.72**	0.01	0.04
9. I am confident I can—quickly learn a new programming language on my own.	−0.01	0.05	0.15	**0.54**	0.05
10. I am confident I can—clearly communicate technical problems and solutions to a range of audiences.	0.02	0.05	0.23	0.48	0.01
11. Self-rating: computer skills	0.31	0.00	0.05	0.12	0.17
12. I see myself as a "computing person."	**0.88**	−0.02	0.09	−0.05	−0.02
13. Computing is a big part of who I am.	**0.71**	0.05	−0.13	0.19	0.00
14. My family emphasizes the value of earning a computing degree	0.06	0.11	0.01	0.20	−0.00
15. I feel like I "belong" in computing.	**0.81**	0.02	0.05	0.01	0.05
16. I feel like an outsider in the computing community. (R)	0.10	−0.05	0.03	−0.02	**0.78**
17. I feel welcomed in the computing community.	0.18	0.24	−0.08	0.32	0.23
18. I do not have much in common with the other students in my computing classes. (R)	−0.10	0.04	0.02	−0.00	**0.68**

Note. (R) represents items that should be reverse coded. Eigenvalues for the four factors are: (1) 6.248, (2) 2.002, (3) 1.496, (4) 1.151, (5) 0.938.

Table 4.3

EFA, a Targeted Four-Factor Solution GEOMIN Rotated Loadings

Item	Factor 1	Factor 2	Factor 3	Factor 4
Interest in Computing				
1. I am interested in learning more about what I can do with computing.	**0.76**	0.02	−0.03	0.01
2. Using computers to solve problems is interesting.	**0.71**	0.06	0.00	0.06
3. I care about doing well in computing.	**0.67**	−0.02	0.12	−0.07
Computing Identity				
1. I see myself as a "computing person."	−0.03	**0.87**	0.07	−0.04
2. Computing is a big part of who I am.	0.05	**0.71**	−0.14	0.18
3. I feel like I "belong" in computing.	−0.00	**0.85**	0.04	−0.01
Computing Self-Efficacy: Professional Milestones				
1. I am confident I can—find employment in an area of computing interest.	0.04	0.03	**0.73**	0.03
2. I am confident I can—get admitted to a graduate computing program.	−0.04	0.03	**0.64**	0.16
3. I am confident I can—complete an undergraduate degree in computing.	0.14	−0.01	**0.71**	0.00
Computing Self-Efficacy: Professional Competencies				
1. I am confident I can—win a computing-related contest (e.g., programming contest, robotics contest, hackathon).	−0.03	0.01	−0.01	**0.83**
2. I am confident I can—become a leader in the field of computing.	0.01	−0.02	0.11	**0.76**
3. I am confident I can—quickly learn a new programming language on my own.	0.06	0.05	0.16	**0.47**

Note. Eigenvalues for the four factors are: (1) 4.972, (2) 1.761, (3) 1.270, (4) 0.578.

Confirmatory Factor Analysis

I also ran two confirmatory factor analyses (CFAs). The first CFA was used to confirm the four-factor structure; findings indicated that the data fit the proposed four-factor model well, χ^2 (48) = 448.04, $p < .01$, *SRMR* = 0.03, *CFI* = 0.97, *RMSEA* = 0.05 (CI: 0.04, 0.05). See Figure 4.1 for the correlations between factors and factor loadings. In running the second CFA, I evaluated the fit of several alternative models, including a single (unidimensional) factor, a bifactor model, and a hierarchical (i.e., second order) factor structure. Results indicated the multidimensional CFA with four separate factors—each with three subsequent items—was the best fit for these data (see Tables 4.3 and 4.4).

Research Question 1a: Assessing Subcomponents of Computing Identity

As existing literature is mixed on whether computing self-efficacy (Blaney & Barrett, 2022) and sense of belonging (George et al., 2022; Sax et al., 2018) are core components of computing identity (e.g., students' "fit in computing"; Lehman et al., 2022), this study seeks to explore whether they are distinct, stand-alone measures or components of a broader computing identity measure. Based on the findings from the second CFA reported above, the present study's findings suggest that the measure of computing identity does not contain items related to *self-efficacy* (a proxy for competence and performance) but that a *sense of belonging in computing* is embedded within the measure of *computing identity* (and its items did not load onto their own unique factor; see Table 4.3). Despite adding four belonging-related items to the exploratory factor analysis tests, a unique measure for belonging in computing did not emerge; this finding is different from the work of prior scholars (Lunn et al., 2021a; Mahadeo et al., 2020; Taheri et al., 2019). Instead, for this study, a single belonging item, "*I feel like I belong in computing*," loaded onto the measure of computing identity—suggesting that a sense of belonging is a core component of how one sees oneself as a computing person. Therefore, operationally, these findings indicate that **computing identity** is a multidimensional measure that incorporates a student's self-identity as a computing person, a sense that computing is a part of their core personal identity, and a sense of belonging in the greater computing community.

Importantly, instead of the proposed performance-related items loading onto a sub-measure of self-efficacy, or self-efficacy being a subcomponent of computing identity, instead the results from the CFA suggest two separate self-efficacy factors. The first computing self-efficacy measure relates to students' confidence in their ability to achieve professional milestones in computing, and the other refers to students' confidence in their ability to achieve professional competencies. Both measures included three items, and for each item students were asked to rate their level of agreement on a 5-point Likert scale ranging from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*); however, the items for these factors differed in scope. The items related to computing self-efficacy professional milestones that were included pertained to students' confidence in their ability to achieve computing milestones such as (a) "complet[ing] an undergraduate degree in computing," (b) "get[ting] admitted to a graduate computing program," and (c) "find[ing] employment in an area of computing interest." These items differed from those in the other computing self-efficacy: professional competencies measure because the latter items pertained to students' confidence in their ability to perform professional computing skills. The items in this measure included students' confidence in their ability to (a) "become a leader in the field of computing," (b) "win a computing-related contest," and (c) "quickly learn a new programming language on [their] own."

Table 4.4

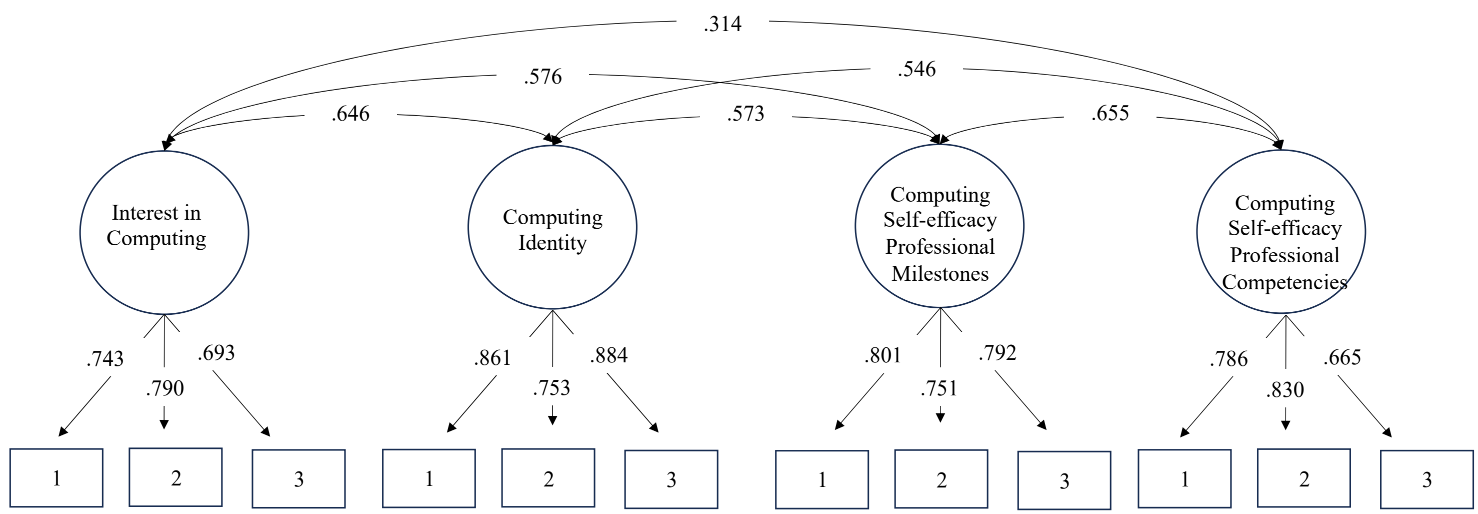
Summary of Model Fit, CFA Analysis of Factor Structure

Model	Chi-square	AIC	BIC	*df*	*p*	*CFI*	*SRMR*	*RMSEA*	*CI*
CFA, four factors	448.04	107,525.48	107,791.46	48	< .01	0.97	0.03	0.05	(0.04, 0.05)
Unidimensional	5,635.48	114,545.51	114,773.49	54	< .01	0.64	0.12	0.16	(0.15, 0.16)
Bifactor	741.39	107,896.41	108,200.38	42	< .01	0.96	0.05	0.06	(0.06, 0.07)
Second Order	859.69	108,045.40	108,298.71	50	< .01	0.95	0.06	0.06	(0.06, 0.07)

Note. Chi-square = Satorra-Bentler scaled chi-square. AIC = Akaike Information Criteria. BIC = Bayesian Information Criteria. CFI = Comparative Fit Index. SRMR = Standardized Root Mean Square Residual. RMSEA = Root Mean Square Error of Approximation. CI = 90 percent confidence interval. Adequate model fit to the data is determined by low AIC and BIC and CFI ≥ 0.95 and RMSEA ≤ 0.08 (Hu & Bentler, 1999).

Figure 4.1

Four Separate Factors; CFA



Note. Loadings are standardized. Variables correspond to items listed in Table 4.3.

Research Question 2: Measuring (In)variance in the Computing Identity Measure

After confirming the factorial structure and determining an adequate model fit for the computing identity measure, I then assessed the extent to which the new computing identity factor held over time and by students' social identities because in scale development, ensuring these measures are valid across students' diverse social identities is important (Wells, 2021). Therefore, despite the strong fit statistics of the CFA and the initial investigations of Cronbach's alpha and McDonald's omega indicating reliability and the strength of the factors over time and across student demographics (see Table 4.6), additional inferential analyses were necessary to assess the goodness of fit for the factor over time (longitudinal factor model) and across these social identities (multigroup confirmatory factor model; Geiser, 2022, 2023; Litson & Feldon, 2021; Marcoulides & Yuan, 2017). Unfortunately, I could not assess configural invariance because, as stated above, the new computing identity measure only consisted of three items, which is the bare minimum for a latent factor, leading it to be a perfect fit with zero degrees of freedom. Had the computing identity factor had four items, configural invariance could have been assessed, but because it does not, we can assume configural invariance is met for computing identity across the subgroups of interest. Therefore, to address this research question, I focused on *metric invariance* and *scalar invariance*. Due to the constrained nature of my three-item computing identity measure, *configural invariance* was not relevant (Geiser, 2022; 2023), and *residual invariance* was too stringent (Putnick & Bornstein, 2016; Wells, 2021).

For this dissertation, I had a mixture of large and small sample sizes. After excluding some groups with either very small sample sizes or ambiguous group identity[22], the sample size

[22] Several of the smaller groups were not included in this analysis, including students who identified as (a) non-binary, genderqueer, gender non-conforming (*n* = 37), or did not report their gender (*n* =208); or (b) Indigenous, Multiracial, as an [o]ther racial/ethnic identity (*n* = 156), or did not report their race/ethnicity (*n* = 605).

for invariance testing ranged from as small as 71 to as large as 2,964. See Table 4.5 below for the sample sizes and groups compared in the invariance tests.

Table 4.5

Reliability of Computing Identity over Time during College and across Students' Social Identities

	n	Cronbach's α	McDonald's ω
Computing Identity Over Time, During College (n = 1,036)			
—Pretest survey	1,036	0.878	0.879
—Posttest survey	1,036	0.899	0.900
—First Follow-up survey	1,036	0.893	0.894
Pretest Computing Identity, by Gender (n = 4,017)			
—Men	2,964	0.850	0.851
—Women	1,053	0.873	0.874
Pretest Computing Identity, by Race/Ethnicity (n = 3,502)			
—White	1,772	0.858	0.860
—Asian	1,045	0.874	0.877
—Black	251	0.868	0.870
—Latine	434	0.822	0.822
Pretest Computing Identity, by Gender & Race/Ethnicity (n = 3,460)			
—White women	400	0.865	0.865
—White men	1,346	0.850	0.852
—Asian women	323	0.874	0.877
—Asian men	713	0.867	0.871
—Black women	71	0.881	0.883
—Black men	178	0.839	0.841
—Latina/e women	108	0.871	0.873
—Latino/e men	321	0.783	0.784

Note. Both Cronbach's alpha (α) and McDonald's omega (ω) are reported. While the field appears to be slowly moving toward Omega for its robustness, as alpha is more restrictive due to its assumptions (Goodboy & Martin, 2020; McNeish, 2017), I report both to aid in comparison with existing and future scholarship that may only investigate alpha.

Measure Equivalence Across Time, During College

To assess the equivalence of the three-item computing identity measure over time, I used the full longitudinal sample with three time points (n = 1,036) to assess metric and scalar invariance using a longitudinal confirmatory factor model in Mplus (Geiser, 2023). Both metric and scalar invariance models had significant chi-square goodness-of-fit values, indicating a significant difference between the observed data and each specified model according to this fit index (see Table 4.7 for model fit and model fit comparisons; Litson & Feldon, 2021). Based on other fit indices, while the CFI was 0.96 for the metric model and 0.95 for the scalar model, the RMSEA for both models was 0.09 and 0.10, respectively, and thus, despite trending in the right direction (i.e., approaching non-significance based on the p-value of the chi-square, the CFI ≥ 0.95, and RMSEA being close to 0.08 as desired), neither the metric nor scalar model met the necessary fit model fit criteria (Hu & Bentler, 1999). This suggests students' responses to items in the computing identity measure may look different over time, which aligns with prior literature speaking to the fluid nature of discipline-specific identity (Aschbacher et al., 2010; Avraamidou, 2020; Carlone & Johnson, 2007; Marcia 1966, 1980; Rodriguez et al., 2020). Importantly, this finding is somewhat expected as it is typical that as students gain more experience, their psychosocial characteristics also evolve (Kapoor & Gardner-McCune, 2019). Therefore, suggestions for additional research in this regard is discussed more in Chapter Five.

However, because the sample sizes for each time point were quite large (possibly producing significant chi-square values where differences may otherwise be trivial) and due to the fact that the CFI and RMSEA were near the cutoff points (Hu & Bentler, 1999), I opted to conduct post hoc confirmatory factor analysis at the latter two survey time points to compare the standardized parameters (factor loadings and intercepts). These additional analyses allowed me to determine whether specific items led to differences in the factor structure over time (see Table 4.6). Findings show similar factor loadings and intercepts, despite the differences being identified in the measure invariance longitudinal test, revealing that researchers need to consider both practical significance and statistical significance when examining equivalency in the strength of factor loadings and factor means over time.

Table 4.6

Parameters of the Three-Item Computing Identity Measure over Time and By Subgroup

	1. I see myself as a "computing person."		2. Computing is a big part of who I am.		3. I feel like I "belong" in computing.	
Analytical Subgroups	Standardized Factor Loadings	Standardized Intercepts	Standardized Factor Loadings	Standardized Intercepts	Standardized Factor Loadings	Standardized Intercepts
Over Time, During College						
—Pretest Survey	0.95	−0.11	0.95	−0.12	0.95	−0.13
—Posttest Survey	0.95	−0.11	0.89	−0.13	0.87	−0.14
—First Follow-up Survey	1.00	−0.04	0.93	−0.05	0.80	−0.07
Gender Identity						
—Women	1.00	3.87	0.98	3.39	1.08	3.74
—Men	1.00	4.22	1.04	3.85	1.12	4.12
Racial/Ethnic Identity						
—White	1.00	4.21	1.12	3.77	1.12	4.09
—Asian	1.00	3.95	0.95	3.62	1.15	3.83
—Black	1.00	4.08	0.97	3.76	1.26	3.97
—Latine	1.00	4.25	1.03	3.83	1.01	4.20
Intersectional Gender & Racial/Ethnic Identity						
—White women	1.00	3.97	1.01	3.39	1.05	3.85
—White men	1.00	4.28	1.13	3.88	1.14	4.16
—Asian women	1.00	3.77	1.01	3.37	1.18	3.63
—Asian men	1.00	4.03	0.91	3.73	1.14	3.92
—Black women	1.00	3.70	0.95	3.42	1.08	3.52
—Black men	1.00	4.23	1.03	3.90	1.15	4.15
—Latine women	1.00	4.35	1.12	3.94	1.01	4.30
—Latine men	1.00	3.95	0.90	3.51	1.04	3.91

Measure Equivalence by Gender

Computing identity was also assessed among students' gender identities, comparing women to men. To assess the equivalence of the three-item computing identity measure by gender, I ran a multigroup confirmatory factor analysis in Mplus using students' gender identities (n = 4,017; see Table 4.7; Geiser, 2022). Non-significant findings for these tests indicate no significant difference between observed data and each specified model according to this fit index. These findings were confirmed with the alternative fit criteria (i.e., CFI ≥ 0.95 and RMSEA ≤ 0.08 for all models; Hu & Bentler, 1999). Therefore, findings from these analyses confirm the computing identity measure is equivalent for men and women. Despite the equivalence of this measure by gender, see Table 4.6 for subtle differences in standardized factor loading and intercepts.

Measure Equivalence by Race/Ethnicity

To assess if the three-item measure of computing identity varies by race/ethnicity, I ran a multigroup confirmatory factor analysis in Mplus among four of the main student racial/ethnic identities (n = 3,502; Geiser, 2022). The racial/ethnic groups that were compared included white, Asian, Black, and Latine students. The scalar model had significant chi-square goodness-of-fit values (see Table 4.7). Non-significant findings for the metric model indicate no significant difference between observed data and each specified model according to this fit index, such that the factor structure of the computing identity measure is equivalent between white, Asian, Black, and Latine students. These findings were confirmed using the alternative fit criteria, where all models produced a CFI ≥ 0.95 and RMSEA ≤ 0.08 (Hu & Bentler, 1999). Again, like the equivalence in computing identity by gender groups, despite the equivalence of this measure by race/ethnicity, Table 4.6 illustrates minor differences in standardized factor loading and intercepts.

Measure Equivalence by Intersections of Gender and Race/Ethnicity

Finally, because students carry multiple interwoven social identities, invariance testing was also assessed across students' intersectional gender and racial/ethnic identities (e.g., for Latina/e women vs. Black men). Like the above approaches for gender and race/ethnicity, I again ran a multigroup confirmatory factor analysis in Mplus using larger student gender and racial/ethnic groups (n = 3,460; Geiser, 2022). The groups that were compared included white women, white men, Asian women, Asian men, Black women, Black men, Latina/e women, and Latino/e men. Findings indicated the scalar model had a significant chi-square value but was in alignment with the alternative fit criteria CFI $\geq$ 0.95 and RMSEA $\leq$ 0.08, as was the same with metric model (see Table 4.7). According to this fit index, the metric model's non-significant findings indicate that the three-item computing identity measure is equivalent across white women and men, Asian women and men, Black women and men, and Latine women and men. Akin to the measure equivalence found in computing identity by gender and race/ethnicity, Table 4.6 provides small differences in standardized factor loading and intercepts by students' intersectional gender and racial/ethnic identities.

Table 4.7

Computing Identity Measure Equivalence over Time and by Students' Social Identities

Model	x^2 (*df*)	*p*	*CFI*	*TLI*	*RMSEA* (90% CI)	*SRMR*	Difftest x^2 (*df*)
Over Time, During College (n = 1,036)							
Metric	284.68 (28)	*	0.96	0.95	0.09 (0.08, 0.10)	0.04	13.42 (4)
Scalar	385.21 (32)	*	0.95	0.94	0.10 (0.09, 0.11)	0.04	100.53 (4)
Gender Identity (n = 4,017)							
Metric	1.74 (2)		1.00	1.00	0.00 (0.00, 0.04)	0.01	1.74 (2)
Scalar	15.03(4)		1.00	0.99	0.04 (0.02, 0.06)	0.01	13.81 (2)
Racial/Ethnic Identity (n = 3,502)							
Metric	20.867 (6)		0.99	0.99	0.05 (0.03, 0.08)	0.05	20.87 (6)
Scalar	39.559 (12)	*	0.99	0.99	0.05 (0.03, 0.07)	0.06	18.65 (6)
Intersections of Gender & Racial/Ethnic Identities (n = 3,460)							
Metric	25.89 (14)		1.00	0.99	0.04 (0.02, 0.07)	0.06	25.89 (14)
Scalar	65.723 (28)	*	0.99	0.99	0.06 (0.04, 0.07)	0.07	40.03 (14)

Note. x^2 = chi-square; *df* = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; CI = confidence interval; SRMR = Standardized Root Mean Square Residual. The chi-square x^2 difftest (difference test) was conducted using multiple-group CFA in Mplus. Adequate model fit to the data is determined by evaluating a chi-square fit statistic *p*-value, where * = $p < .001$ (Litson & Feldon, 2021). Other accepted fit indices include CFI and RMSEA, where a well-fitted model is identified when CFI ≥ 0.95 and RMSEA ≤ 0.08 (Hu & Bentler, 1999).

To provide additional context for whether the factors varied over time and by students' gender and racial/ethnic group, I ran additional analyses that revealed both within and between-group differences. Despite Carlone and Johnson (2007) suggesting students' computing identity scores should increase over time with greater exposure and skillset obtained, the paired samples *t*-test analyses comparing the average computing identity scores within groups over time revealed that computing identity scores decreased over time for all groups, with only some differences being statistically significant. For instance, students' pretest computing identity scores (M = 9.80, SD = 2.48) were far higher than their scores on the first follow-up survey computing identity scores (M = 9.12, SD = 2.77; $t(447) = 1.38$, $p < .001$, d = 0.07). The same was true for women, men, and USOCC. However, there were no significant mean score differences between the pretest computing identity and the first follow-up survey computing identity for Asian students (Pretest M = 9.36, SD = 2.22; First Follow-up Survey M = 9.12, SD = 2.60) nor for white students at either time point (Pretest M = 9.80, SD = 2.48; First Follow-up Survey M = 9.65, SD = 2.77). These analyses helped to uncover that students' scores decreased from the pretest to the first follow-up survey for all groups, though effect sizes were small across all social groups (Cohen's d ranged from 0.07 to 0.26; Cohen, 1988; see Table 4.8). The decline we see for all groups may be attributable to not all students in the sample being computing majors in the first follow-up survey. However, these decreases are akin to the findings by Sax and her colleagues in relation to their exploration of sense of belonging in computing in 2018.

The independent sample's *t*-test and the one-way ANOVA revealed some interesting findings that aligned with some literature and differed from other literature. As indicated by the superscripts in Table 4.8 below, there were several differences between groups within a given survey time point. Specifically, the findings for between-group differences by gender aligned with Sax and her colleague's research (2018) on sense of belonging in computing. Indeed, findings for this study revealed that men's computing identity scores were higher than women's computing identity scores on both the pretest survey (Men: $M = 10.15$, $SD = 2.13$; Women: $M = 9.05$, $SD = 2.40$) and the first follow-up survey (Men: $M = 10.07$, $SD = 2.39$, Women: $M = 9.05$, $SD = 2.40$), respectively (Pretest: $t(996) = 7.514$, $p < .001$, $d = 0.50$; Follow-up survey: $t(1{,}010) = 10.14$, $p < .001$, $d = 0.67$).

However, ANOVAs revealed there were also statistically significant differences seen across racial/ethnic groups (Pretest: $F(2, 998) = 6.82$, $p < .001$, partial $\eta^2 = 0.014$; Follow-up: $F(2, 999) = 5.18$, $p = .006$, partial $\eta^2 = 0.010$). Post hoc analyses revealed that these between-group differences did not align with what prior scholars found (Lunn et al., 2021a). While there were no differences seen between Asian and white students' initial computing identity scores, USOCC had higher pretest computing identity test scores ($M = 10.04$, $SD = 2.09$) than their white ($M = 9.88$, $SD = 2.43$, 95% CI [−0.19, 0.52]) and Asian peers ($M = 9.38$, $SD = 2.22$, 95% CI [0.28,1.04], partial $\eta^2 = 0.014$). Based on benchmarks by Cohen (1988) for the range of partial eta squared (i.e., the effect size for ANOVAs), this finding meets the criteria for a large effect, meaning there is a substantial difference between groups and that the practical implications of this difference should strongly be considered (Lakens, 2013). On the first follow-up survey, there were no differences between USOCC and their white and Asian peers. However, white students scored significantly higher on the follow-up survey than Asian students at this later timepoint, as indicated by the statistically significant difference between white students ($M = 9.73$, $SD = 2.73$) and their Asian counterparts ($M = 9.10$, $SD = 9.60$, 95% CI [0.24, 1.02], partial $\eta^2 = 0.010$). Again, based on the criteria presented in Lakens (2013), which is based on Cohen's 1988 effect size indicators, this effect is between medium (partial $\eta^2 = 0.06$) and large (partial $\eta^2 = 0.14$), meaning this difference is not trivial. Therefore, these differences in computing identity by gender and race/ethnicity illuminate not only statistical significance but also practical significance.

Taken together, the notion that computing identity decreases over time is somewhat expected as college students come to learn more about the world around them and realize how much they have yet to learn about the field they are studying (Aschbacher et al., 2010; Avraamidou, 2020; Carlone & Johnson, 2007; Marcia 1966, 1980; Rodriguez et al., 2020). However, the fact that there are differences between groups at different points in time suggests that we need to provide more equitable support to foster students' computing identities over time. Understanding how students' computing identities are developed and how to maintain or enhance the degree of computing identity students come to college with is the focus of analysis for the next research question. Recommendations to address this and other findings from this study are discussed in Chapter Five.

Table 4.8

Within and Between-Group Mean Differences between Students' Pretest and First Follow-up Computing Identity Scores

	Pretest Computing Identity		First Follow-up Computing Identity					
Social Group	*M*	*SD*	*M*	*SD*	*df*	*t*	*p*	Cohen's *d*
All Students	9.80	2.48	9.65	2.77	447	1.38	< .001	0.07
Gender								
Women	9.03^{a}	2.40	8.40^{a}	2.80	354	4.91	< .001	0.26
Men	10.13^{b}	2.12	10.06^{b}	2.40	630	0.90	< .001	0.04
Race/Ethnicity								
USOCC	10.01^{a}	2.09	9.41^{ab}	2.67	250	4.10	< .001	0.26
Asian	9.36^{b}	2.22	9.12^{a}	2.60	307	1.87	.062	0.11
White	9.80^{b}	2.48	9.65^{b}	2.77	447	1.38	.168	0.07

Note. Superscripts with different letters indicate between-group differences (down columns, within social groups), whereas statistics reported indicate within-group differences (across rows, within social group subcategories).

Research Question 3: Predicting Computing Identity

Having established a computing identity measure and demonstrated equivalence across students' gender, race/ethnicity, and intersectional gender and racial/ethnic identities above, the remaining analysis used a combined sample of 1,036 students who completed the pretest, posttest, and first follow-up survey to see what background, psychosocial, and college variables predict computing identity. With a particular interest in how computing identity is developed among students from different social groups, I ran a series of ordinary least squares (OLS) regression analyses using the measure of computing identity from the third survey time point ($\alpha = 0.893$) as the dependent variable. As a reminder, the computing identity factor was measured on a 5-point Likert scale ranging from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*) and included the following three items: "I see myself as a 'computing person,'" "Computing is a big part of who I am," and "I feel like I 'belong' in computing." Further, this outcome was used to assess the extent to which various independent variables predict stronger computing identity among (a) all students in computing, (b) women, (c) men, (d) white students, (e) Asian students, and (f) a combined group[23] of USOCC (i.e., Black or African American; Hispanic or Latina/o/x, American Indian; Native Hawaiian or Pacific Islander; Arab, Middle Eastern, or Persian; and students from two or more races/ethnicities; Wofford et al., 2022).

For the hierarchical blocked regression, variables included in the initial main effects models accounted for between 43% and 55% of the variance in computing identity—depending

[23] Importantly, I want to recognize that I understand students from different backgrounds do not experience college in the same ways and that even students who share social identities may still not experience college similarly. Yet, despite my best intentions to disaggregate racial/ethnic groups for reasons mentioned previously, due to small sample sizes and the need for enough statistical power to detect meaningful differences, I ultimately excluded the small group of students who identified as nonbinary and I combined Black, Latine, and all other students from one or more racial/ethnic groups together.

on the subpopulation examined (see Table 4.9). The direct pretest computing identity variable explained most of the variance in computing identity across the regressions (i.e., R^2 ranged from 0.29 in the regression among Asian students to 0.47 in the regression among white students). Still, several other variables also played a small but statistically significant role.

Table 4.9

Initial Hierarchical Linear Regressions Predicting Computing Identity One Year after Students' Intro Computing Courses

Dependent Variable —Computing Identity at First Follow-up	*REGRESSION 1* **ALL** (n = 660) Final Model (Block 7)		*REGRESSION 2* **MEN** (n = 417) Final Model (Block 7)		*REGRESSION 3* **WOMEN** (n = 243) Final Model (Block 7)		*REGRESSION 4* **WHITE** (n = 307) Final Model (Block 7)		*REGRESSION 5* **ASIAN** (n = 244) Final Model (Block 7)		*REGRESSION 6* **USOCC** (n = 173) Final Model (Block 7)	
Independent Variables	*b*	Sig	*b*	Sig	*b*	Sig	*b*	Sig	*b*	Sig	*b*	Sig
Block 1: Pretest Item												
—Pretest Computing Identity	0.41	***	0.41	***	0.41	***	0.51	***	0.36	***	0.40	***
Block 2: Personal Characteristics												
—Gender: Women	−0.15	***	---	---	---	---	−0.04		−0.16	**	−0.30	***
Race (White)												
—Race: Asian	−0.10	**	−0.09	*	−0.11		---	---	---	---	---	---
—Race: Black	−0.07	**	−0.03		−0.12	*	---	---	---	---	---	---
—Race: Latine	−0.03		0.00		−0.07		---	---	---	---	---	---
—Race: Other	−0.04		−0.03		−0.06		---	---	---	---	---	---
—First-Generation Status	0.01		−0.02		0.07		0.02		0.01		−0.02	
Block 3: Background Contexts												
—Class Standing	−0.05		−0.08		0.01		−0.09	*	0.02		−0.05	
—Socioeconomic Status	−0.02		−0.07		0.04		0.03		−0.06		−0.04	
—Transfer Status	−0.04		−0.04		−0.07		−0.04		−0.02		−0.15	
—HS CS Course	0.03		0.02		0.07		0.06		−0.02		0.04	
—HS GPA	0.05		0.06		0.04		0.07		0.01		−0.02	
Block 4: Early Computing Identity Experiences												
—Parent CS Career	0.02		0.03		−0.01		0.00		0.09		0.03	
—Family Computing Support+	−0.04		−0.03		−0.02		−0.04		−0.11		0.04	
—Amount of Prior Programming Experience	0.07	*	0.07		0.10		0.03		0.18	**	0.03	
Block 5: Additional Psychosocial Computing Identity Factors												
—Interest in Computing+	0.13	***	0.20	***	0.07		0.14	**	0.12	*	−0.04	

—Computing Self-Efficacy: Professional Competencies$^+$	0.06		0.05		0.07		0.08		0.08		0.08	
—Computing Self-Efficacy: Professional Milestones$^+$	−0.01		−0.03		−0.01		−0.12	*	−0.03		0.15	
Block 6: College Computing Identity Experiences *Within* the Intro Course												
—Relevant Social/Personal Examples$^+$	0.00		0.03		−0.07		0.03		0.03		−0.12	
—Perception of Inclusive TAs/Faculty$^+$	0.00		−0.03		0.05		0.00		0.09		0.00	
—Feeling Dismissed$^+$	0.05		0.10	*	0.00		0.00		0.05		−0.03	
Block 7: College Computing Identity Experiences *Outside* of the Intro Course												
—Peer Support$^+$	0.05		0.10	*	−0.02		0.08		0.10		−0.02	
—Mentor Support$^+$	0.00		−0.04		0.07		0.00		−0.06		0.03	
—Computing Major/minor	0.25	***	0.25	***	0.27	***	0.26	***	0.26	***	0.16	*
—Undergraduate Research Participation	0.04		0.01		0.07		0.01		0.06		0.04	
—Technical Conference Participation	0.00		−0.02		0.03		−0.01		−0.02		−0.01	
Model R^2		0.50		0.47		0.49		0.59		0.47		0.50

Note. $^+$ denotes variable is a factor; Sig = *p*-value; *** = $p < .001$; ** = $p < .01$; * = $p < .05$; *b* = unstandardized regression coefficient.

For the blocked stepwise regression approach, the pretest computing identity measure still served as the primary independent variable, entering it first, and all remaining variables entered next one by one only if they were significant at some point in the prior model for at least one subgroup (even if they were not significant at the final step when all variables were entered together). The following nine variables were excluded because they were not significant predictors for any group at any point:

1. **Five background variables:** first-generation status, socioeconomic status, parent computing career status, a factor of family computing support.
2. **One psychosocial variable:** a measure of computing self-efficacy: professional competencies.
3. **Four college variables:** a factor measuring the inclusivity of the instructional team (TA/Faculty), a mentor support factor, and two separate variables measuring whether students participated in undergraduate research or attended a technical conference.

For the revised stepwise OLS regression models, when all variables were entered at the final step, between 43% and 55% of the variance in computing identity was explained—depending on the subgroup examined (see Appendix G). Further, findings from these models predicting computing identity one year after students' introductory computing courses indicate several predictors and that these key variables differ to some degree across the different student social groups (see Table 4.10).

In the remainder of this section, I review findings from the regression analysis among all students and then pivot to discuss the more nuanced findings and value-added of comparing predictors among students by gender and race/ethnicity across the separate regression models.

Table 4.10

Linear Regression Predicting Computing Identity among All Students One Year after Students' Intro Computing Courses (n = 768)

Dependent Variable —Computing Identity at First Follow-up			Pretest (Step 1)		Background (Step 11)		Psychosocial (Step 13)		College (Step 17)	
Independent Variables	*r*	Sig	*β*	Sig	*β*	Sig	*β*	Sig	*β*	Sig
Pretest Item										
—Pretest Computing Identity	.61	***	0.61	***	0.54	***	0.46	***	0.41	***
Personal Characteristics										
—Gender: Women	−.31	***	*−0.16*	***	−0.16	***	−0.17	***	−0.16	***
Race (White)										
—Race: Asian	−.08	**	*−0.03*		−0.05		−0.04		−0.08	**
—Race: Black	−.09	***	*−0.07*	*	−0.07	*	−0.07	*	−0.06	*
—Race: Latine	.04		*−0.03*		−0.03		−0.03		−0.04	
—Race: Other	−.01		*−0.02*		−0.02		−0.02		−0.03	
Background Contexts										
—Class Standing	−.02		*0.00*		−0.01		−0.01		−0.04	
—Transfer Status	.01		*−0.04*		−0.04		−0.03		−0.04	
—HS CS Course	.18	***	*0.07*	*	0.03		0.04		0.02	
—HS GPA	−.03		*0.03*		0.04		0.03		0.03	
Pre-College Computing Identity Experiences										
—Amount of Prior Programming Experience	.28	***	*0.09*	***	0.08	*	0.07	*	0.07	*
Additional Psychosocial Computing Identity Factors										
—Interest in Computing+	.38	***	*0.17*	***	*0.18*	***	0.18	***	0.14	**
—Computing Self-Efficacy: Professional Milestones+	.30	***	*0.03*		*0.05*		0.01		0.00	
College Experiences *Within* the Intro Course										
—Relevant Social/Personal Examples+	.03	0.18	*0.00*		*−0.01*		*0.00*		0.00	
—Feeling Dismissed+	−.08	***	*−0.03*		*−0.02*		*0.03*		0.04	
College Experiences *Outside* of the Intro Course										
—Peer Support+	.16	***	*0.07*	*	*0.07*	*	*0.05*		0.05	
—Computing Major/minor	.40	***	*0.25*	***	*0.25*	***	*0.22*	***	0.23	***
Model R^2				0.368		0.410		0.437		0.480

Note. r = simple correlation; + denotes variable is a factor; *Sig* = *p*-value; *** = $p < .001$; **= $p < .01$; * = $p < .05$; β = standardized regression coefficient.

Predictors of Computing Identity for All Students

The 17 variables (including gender and race) used in the regression among all students (*n* = 768) explain 48% of the variance in computing identity one year after students' introductory course experience ($R^2 = 0.483$). As expected, the direct pretest computing identity measure alone accounted for most of the variance in the dependent variable when it was entered first (37%; $\beta = 0.61$). Because pretest computing identity was measured at the beginning of the introductory course, it captured students' incoming self-identity as a computing person, a sense that computing is a part of their core personal identity, and a sense of belonging in the greater computing community. However, as expected, the effect of the pretest variable was attenuated each time other independent variables were added to the model because the variables share predictive power (as indicated by a reduced beta coefficient at each step). Therefore, several additional variables beyond the pretest variable contribute to the remaining 11% of the variance in explaining the computing identity outcome.

Importantly, student background variables (i.e., personal characteristics, background contexts, and prior computing experiences) are among the majority of the predictive variables in explaining computing identity. For instance, being a woman (vs. man; $\beta = -0.08$), Asian (vs. white; $\beta = -0.16$), or Black (vs. white; $\beta = -0.06$) are negative predictors of computing identity, such that identifying in any of these social groups is associated with lower computing identity. Beyond demographics, students who took a computing course in high school are more likely to have a stronger computing identity than those who did not take such a course. However, the positive effect of high school computing courses is reduced and loses significance when the variable representing prior programming experience enters the model because high school computing courses and prior programming share predictive power. Further, students' prior computing experience is a positive background variable in predicting computing identity (β = 0.07).

By the time all variables entered the model, only one psychosocial variable predicted computing identity among all students. Specifically, students' interest in computing is a positive predictor of the computing identity outcome and is therefore associated with greater computing identity (β = 0.14). While this factor was measured at the end of students' introductory computing courses and their incoming interest in computing was not accounted for, it may be safe to assume some of the differences seen for these groups were due to their incoming interest in computing that they generated before college. Interestingly, the measure of students' computing self-efficacy: professional milestones did not end up being a facilitator or an impediment toward developing one's computing identity. This was likely because while the measure of computing self-efficacy: professional milestones on its own is a predictor of computing identity (r = .30), it is also correlated with the direct pretest computing identity measure (r = .45); when the pretest computing identity variable is accounted for first, the positive predictive power of computing self-efficacy: professional milestones disappears.

Once accounting for students' backgrounds, psychosocial factors, and college experiences, computing major was the only college environmental variable to remain significant, contributing positively toward computing identity (β = 0.23). This suggests that students majoring in computing have greater computing identity than their peers in other majors. However, there are two other college environmental variables that have significant correlations (p < .05) with the outcome but were not predictors at the final step of the model: a measure of feeling dismissed (r = −.08), and peer support (r = .16). This signifies that controlling for no other variables, students were more likely to report higher computing identity scores the less they felt dismissed and the more they experienced greater peer support. The former suggests that students who felt dismissed had low computing identity from the start of the course, but that feeling dismissed was not uniquely associated with lower computing identity scores one year after the introductory course. Similarly, the measure of peer support might have been a significant positive predictor had it not been for the fact that the measure of students' interest in computing entered first, as when this occurred, the main effect disappeared because these two variables share predictive power in explaining the variance of the computing identity outcome. In summary, and as expected, the more students perceive their peers support them, the more likely they are to experience higher computing identity scores, though this relationship may be mediated by students' interest in computing.

Non-Significant Predictors of Computing Identity, for All Students. The findings above revealed seven key variables (out of 17 total variables in the model) that predicted computing identity among all students; however, we must also take note of the variables that did not contribute to computing identity after all variables were controlled. Six of the 10 non-significant predictors were background contexts, one was a psychosocial variable, and three variables were related to college experiences (see Table 4.10).

There are two primary reasons why variables were not significant at the final step of the model: either they were never significant even at the simple correlation level, or they lost their significance over time as other variables entered the model due to multicollinearity. For instance, six of the 10 non-significant predictors were never significant. These variables included students' identifying as Latine or Other (both compared to white), class standing, transfer status, high school GPA, and perception of their introductory course instructor, providing relevant social/personal examples in the introductory computing course. The main variables that, once accounted for, drove other variables to lose significance included the pretest measure of computing identity, the amount of prior programming experience, and the measure of students' interest in computing. Further, the variables that lost significance due to their relationships with other variables included whether students completed a computing course in high school, a measure of computing self-efficacy: professional competencies, the more students felt dismissed, and the more they perceived having peer support.

Research Question 3a: Comparing Predictors of Computing Identity by Gender

The second part of the third research question focuses on how predictors of computing identity differ by gender. To examine whether the predictors are the same or different by gender, I ran two stepwise OLS multiple regressions predicting the computing identity outcome, first among women and then among men. The variables and coding schemes remained the same across models, allowing for clearer comparisons in the magnitude and direction of the unstandardized regression coefficients (*b*). After accounting for the direct pretest computing identity measure, students' backgrounds, other psychosocial factors, and college experiences, the final model explained 46% of the variance in computing identity for women and 43% of the variance among men.

Below, I report the salient predictors seen among both gender-related regressions and then provide separate summaries of the unique predictors for women and men. See Table 4.11 for simple correlations and standardized regression coefficients (β) used to identify the most important predictors within gender models and the significance level of each variable in predicting computing identity.

Salient Predictors of Computing Identity for Both Genders. Three (two positive and one negative) key variables emerged with the same significance and direction across the separate regression models for women (*n* = 276) and men (*n* = 492). The positive predictors seen among both regressions included the direct pretest computing identity measure and the variable representing whether students were majoring or minoring in computing. These findings suggest that having a greater incoming computing identity (β for women = 0.43; β for men = 0.39) and being a computing major or minor at the point of the first follow-up survey (β for women = 0.21; β for men = 0.27) are associated with higher levels of computing identity one year after students' introductory computing courses. Importantly, findings also indicate that for both men and women, identifying as Asian (compared to white) is associated with lower scores on computing identity one year after students' introductory computing courses (β for women = −0.11; β for men = −0.08).

Table 4.11

Linear Regression Findings Predicting Computing Identity One Year after Students' Intro Computing Courses, By Gender

Dependent Variable —First Follow-up Computing Identity	Among Women (n = 276)				Among Men (n = 492)			
	r	Sig	*b*	Sig	*r*	Sig	*b*	Sig
Pretest Item								
—Pretest Computing Identity	.57	***	0.43	***	.58	***	0.39	***
Personal Characteristics								
Race (White)								
—Race: Asian	−.04		−0.11	*	−.09	*	−0.08	*
—Race: Black	−.13	*	−0.11	*	−.04		−0.04	
—Race: Latine	.05		−0.07		.03		−0.02	
—Race: Other	−.02		−0.07		.02		−0.02	
Background Contexts								
—Class Standing	.02		0.02		−.03		−0.08	
—Transfer Status	.05		−0.02		−.02		−0.06	
—HS CS Course	.22	***	0.04		.11	*	0.01	
—HS GPA	.02		0.06		.01		0.01	
Pre-College Computing Identity Experiences								
—Amount of Prior Programming Experience	.30	***	0.10		.25	***	0.07	
Additional Psychosocial Computing Identity Factors								
—Interest in Computing$^+$	.34	***	0.09		.43	***	0.21	***
—Computing Self-Efficacy: Professional Milestones$^+$	.28	***	0.01		.32	***	−0.01	
College Experiences *Within* the Intro Course								
—Relevant Social/Personal Examples$^+$	−.04		−0.03		.04		0.02	
—Feeling Dismissed$^+$	−.16	***	−0.02		−.01		0.10	*
College Experiences *Outside* of the Intro Course								
—Peer Support$^+$	.14	***	0.01		.17	***	0.07	
—Computing Major/minor	.44	***	0.27	***	.34	***	0.21	***
Model R^2				0.457				0.428

Note. r = simple correlation; $^+$ denotes variable is a factor; Sig = p-value; *** = $p < .001$; ** = $p < .01$; * = $p < .05$; b = unstandardized regression coefficient

Salient Predictors of Computing Identity Only for Women. One variable was a unique predictor among women: identifying as Black (compared to white) was a significant negative predictor of women's computing identity ($\beta = -0.11$). This suggests that, like the finding regarding Asian women referenced above, women who identified as Black tended to score lower on the computing identity measure than women who identified as white. These differences among women (i.e., that being Black or Asian negatively predicts computing identity relative to their white women peers) reaffirm the need to understand the intersectional perspectives seen among gender groups by race/ethnicity.

Salient Predictors of Computing Identity Only for Men. While there was only one unique predictor of computing identity for women, there were two significant predictors of computing identity for men. In particular, the unique predictors for men were both positive: interest in computing ($\beta = 0.21$) and feeling dismissed ($\beta = 0.10$). These findings can be interpreted as the more men are interested in computing and feel dismissed, the higher their computing identity scores tend to be. The latter concept sounds counterintuitive but aligns with existing literature, as I discuss more in Chapter Five.

Non-Significant Predictors of Computing Identity across Gender Groups. There were 11 of the 16 total variables in the regression models that did not predict computing identity in *either* the regression among women or among men (see Table 4.11). Five of these variables were background variables, including identifying as Black, Latine, or Other racial/ethnic identities (both compared to white); whether students were in their first or second year (class standing); whether students transferred to their institution (transfer status); students' high school GPA; and the amount of prior programming experience students brought with them into their introductory course. There was roughly an even number of variables that were never significant across the regressions (even at the simple correlation level) to those variables that were significant initially but lost significance over time. The variables that were never significant across either gender-based regression included: students identifying as Latine or Other race/ethnicity (compared to white), class standing, transfer students, high school GPA, and a measure of relevant social/personal examples.

Despite being significant at the correlation level, many variables were correlated with the pretest computing identity measure, which, when entered first, consumed a lot of the variance in explaining the dependent variable. The variables that lost statistical significance due to multicollinearity included: the measure of computing self-efficacy: professional milestones, students' perception of the intro course instructor providing relevant social/personal examples during class, and having a network of peers that support them.

Research Question 3b: Comparing Predictors of Computing Identity by Race/Ethnicity

The final component of this research question seeks to investigate how predictors of computing identity vary by students' racial/ethnic identities (see Table 4.12). The analyses for this section mirror those used in the above regressions by gender, but this time, gender was added as an independent variable, and the four variables representing race/ethnicity were excluded. Using an OLS stepwise multiple regression approach with the same variables and order of entry, I ran three regressions to predict computing identity one year after students' introductory courses for the following groups: (a) USOCC, (b) Asian students, and (c) white students. Upon the final step, when all 13 variables were in the model, 45% of the variance in computing identity was explained among USOCC, 44% among Asian students, and 55% among white students.

In the following sections, I give an overview of the salient predictors seen *across* all racial/ethnic groups that have the same predictive power and direction. I then share the variables whose salience across racial/ethnic groups is different (e.g., a variable was significant for one group but not others) and share instances across the models where a variable is significant for more than one group but differs in terms of direction (i.e., one positive and one negative). See Table 4.12 for an overview of the magnitude and direction of key predictors within and across each of these analyses.

Table 4.12

Linear Regression Findings Predicting Computing Identity One Year after Students' Intro Computing Courses, By Race/Ethnicity

Dependent Variable —First Follow-up Computing Identity	USOCC (*n* = 193)				Asian (*n* = 268)				White (*n* = 375)			
Independent Variables	*r*	Sig	*b*	Sig	*r*	Sig	*b*	Sig	*r*	Sig	*b*	Sig
Pretest Item												
—Pretest Computing Identity	.55	***	0.40	***	.55	***	0.38	***	.67	***	0.52	***
Personal Characteristics												
—Gender: Women	−.36	***	−0.26	***	−.30	***	−0.16	***	−.29	***	−0.08	*
Background Contexts												
—Class Standing	.01		−0.05		.05		0.00		−.05		−0.07	
—Transfer Status	−.07		−0.13		.01		−0.02		.04		−0.04	
—HS CS Course	.11		−0.01		.16	**	−0.02		.24	***	0.05	
—HS GPA	−.03		−0.01		−.02		0.01		−.06		0.03	
Pre-College Computing Identity Experiences												
—Amount of Prior Programming Experience	.13	*	0.01		.27	***	0.16	***	.36	***	0.04	
Additional Psychosocial Computing Identity Factors												
—Interest in Computing+	.28	***	−0.01		.29	***	0.10		.42	***	0.16	***
—Computing Self-Efficacy: Professional Milestones+	.34	***	0.18	*	.29	***	0.01		.31	***	−0.12	*
College Experiences *Within* the Intro Course												
—Relevant Social/Personal Examples+	−.06	***	−0.10		.11	*	0.04		.01		0.03	
—Feeling Dismissed+	−.24		−0.04		.01		0.01		−.09	***	0.02	
College Experiences *Outside* of the Intro Course												
—Peer Support+	.13		0.00		.19	*	0.09		.22	***	0.07	*
—Computing Major/minor	.36	***	0.16	*	.35	***	0.24	***	.46	***	0.24	***
Model R^2				0.450				0.440				0.553

Note. USOCC = Underrepresented Students of Color in computing; *r* = simple correlation; + denotes variable is a factor; Sig = *p*-value; *** = $p < .001$; ** = $p < .01$; * = $p < .05$; *b* = unstandardized regression coefficient.

Variables That Are Significant for All Racial/Ethnic Groups. Three significant variables had the same direction of effect across the separate racial/ethnic regression models for USOCC (*n* = 193), Asian (*n* = 268), and white students (*n* = 375). Two of these variables were positive predictors of computing identity, and one was a negative predictor. The common positive predictors included the direct pretest computing identity measure and the variable representing whether students were majoring/minoring in computing. These findings mirror that of the above analyses (for all students and by gender) and reiterate the importance of students' incoming computing identity (β for USOCC = 0.40; β for Asian = 0.38; β for white = 0.52) and being a computing major/minor (β for USOCC = 0.16; β for Asian = 0.24; β for white = 0.24) in predicting students' computing identities one year after their introductory course experience. The gender variable served as a negative predictor, such that across all racial/ethnic groups, students who identified as women (vs. identifying as men) tended to have lower computing identities (β for USOCC = −0.26; β for Asian = −0.16; β for white = −0.08).

Variables with Inconsistent Effects across Racial/Ethnic Groups. Three variables did not yield similar results across the racial/ethnic regression across models—two variables were background characteristics (prior programming experience and interest in computing) and two were associated with students' college experience (peer support and self-efficacy: professional milestones). The first background variable was prior programming experience. While this variable was significant at the simple correlation level for all models, it only remained significant at the final step for Asian students. We can interpret this to mean that while, at the simple correlation level, prior programming experience is significantly associated with computing identity for all racial/ethnic groups, when we control for other variables—namely, the direct pretest for USOCC and high school computing courses for white students—the effect of prior programming experience disappears for USOCC and white students. However, having greater prior programming experience retained its predictive power on Asian students' computing identity development ($\beta = 0.16$). This finding means that at the simple correlation level, prior programming experience strongly correlates with computing identity for everyone. However, as other variables were added, the magnitude of that strength diminished due to multicollinearity.

The next background variable that was significantly correlated with computing identity but did not remain significant for all groups after controlling for other variables was the psychosocial variable measuring interest in computing. For USOCC, once I controlled for the pretest, the predictive power of students' interest in computing disappeared. For Asian students, it was not until the last step, when students' majors entered, that the predictive power of students' interest in computing disappeared. For white students, having greater interest in computing tended to foster greater computing identity even when controlling for all variables ($\beta = 0.16$).

The first college environmental variable that served as a uniquely positive predictor was a measure of peer support, which was seen only for white students. Here, white students who reported greater feelings of peer support also tended to have greater computing identities one year after their introductory computing course. In other words, peer support enhances computing identity, but only for white students. Interestingly, although this variable was not significant at the simple correlation level for USOCC, it was significant for Asian students. For Asian students, it was not until interest in computing entered that the effect of peer support became non-significant. This might suggest interest is more predictive of computing identity than peer support for these underrepresented students.

The findings from the regression models highlight only one salient variable that was significant but in different (i.e., opposite) directions across students' social groups: the measure of self-efficacy: professional milestones. At the simple correlation level, this measure was positive for all groups. For USOCC, this measure remained a significant positive predictor of computing identity ($\beta = 0.18$), even after controlling for other pretest, background, and psychosocial and college-related variables—implying that with increased computing identity scores, they tended to have higher computing identity scores. For Asian students, computing self-efficacy: professional milestones would have been a salient positive predictor had it not been for the relationship between this variable and the direct pretest computing identity measure sharing variance. However, white students with greater confidence in their ability to achieve computing outcomes tended to experience the opposite effect; indeed, for white students, this measure of computing self-efficacy: professional milestones was associated with lower scores on computing identity. For white students, the psychosocial measure of computing self-efficacy: professional milestones became negative when the direct pretest of computing identity entered and remained a negative predictor of computing identity at the final step ($\beta = -0.12$). This suggests that, for white students, once the other variables are controlled, the predictive power of self-efficacy: professional milestones reverses direction, though an explanation for this is not clear from the data (and may be a function of multicollinearity among these variables). However, these findings may be an anomaly because the beta coefficients are small and this variable was significant and positive at the simple correlation level. Thus, more research is needed to parse the unique role of computing self-efficacy, specifically concerning computing self-efficacy surrounding professional milestones, for students from different racial/ethnic groups such as these.

Non-Significant Variables across Racial/Ethnic Groups. After controlling for all 13 variables, six variables were not significant in predicting computing identity among any of the racial/ethnic groups. Four of these variables were background characteristics, with the remaining two being variables representing college experiences within the introductory computing course.

Similar to the regressions above among all students, and those among groups, I followed the betas to see what contributed to variables being insignificant in predicting computing identity on the first follow-up survey. The variables that were not significant for any of the racial/ethnic groups even at the simple correlation level included class standing, transfer status, and high school GPA. The key variables that became non-significant due to multicollinearity issues were items related to introductory course experiences; namely, the more instructors provided relevant social/personal examples and the more students felt they were dismissed. This is not to say these experiences do not matter, because ample literature implies the computing climate does matter. Rather, these findings suggest that there are a variety of other variables that are stronger predictors of computing identity.

Summary of Key Results

In the psychometric measure development, it became clear that computing identity is not the same as computing self-efficacy and that computing identity and belonging could not be separated. Further, while the new psychosocial measure of computing identity is representative of students from various genders, races/ethnicities, and intersectional gender and racial/ethnic backgrounds, we cannot assume a one-size-fits-all approach to fostering computing identity among our increasingly diverse group of computing students.

In particular, the findings from the six primary regressions reveal several salient predictors of computing identity for computing students in the year following their introductory computing course. Across the regressions, positive predictors tend to include greater pretest computing identity, greater amounts of prior programming experience (for Asian students only), greater interest in computing (for men and white students only), being a computing major, feeling dismissed (for men only), greater peer support (for white students only), and greater computing self-efficacy: professional milestones (for USOCC only). Negative predictors include identifying as a woman (compared to a man), identifying as Asian or Black (both compared to white), and computing self-efficacy: professional milestones (for white students only).

In the next (and final) chapter, I discuss these findings in conversation with existing literature and theory and highlight implications for future research and practice.

CHAPTER FIVE: DISCUSSION AND IMPLICATIONS

This study responds to the need for more research that thoroughly examines how *computing identity* is operationalized, measured, and developed among undergraduate computing students from varying demographic social groups in their introductory computing courses. In particular, the study relies upon longitudinal student survey data from 1,036 undergraduate students who completed three surveys during approximately two academic years (i.e., a pretest survey at the beginning of the introductory course, a posttest survey at the end of the introductory course, and a follow-up survey in the fall of the next academic year). The study sample was further restricted to students who (a) had not yet graduated at the time of the follow-up survey, (b) were either computing majors or undeclared majors who aspired to become computing majors, and (c) were enrolled in an introductory computing course at one of 15 research institutions across the United States. The following questions framed this inquiry.

Research Questions

1. To what extent can a psychometrically sound measure of *computing identity* be developed for students who took an undergraduate introductory computing course?
    a. Are self-efficacy and sense of belonging necessary components of computing identity?
2. To what extent does computing identity hold over time during college and by gender, race/ethnicity, and intersections of gender and race/ethnicity?
3. What pre-college and college-related variables predict *computing identity* in the year following students' participation in an introductory computing course?
    a. To what extent do these predictors differ for women and men?
    b. To what extent do these predictors differ across racial/ethnic categories?

Overall, findings from this research offer the computing education community more clarity on how to operationalize and measure the psychosocial *computing identity* concept in a way that applies to students from varying social groups and provides an opportunity to propel our collective understanding of the unique ways in which computing identity is developed among students from differing gender and racial/ethnic identities. The remainder of this chapter summarizes the high-level findings corresponding to each of the three research questions presented in Chapter Four and discusses the implications of these findings for future practice, policy, theory, and research.

Research Question 1 Finding: Computing Identity Is Distinct from Computing Self-Efficacy but Is Not Distinct from Sense of Belonging

The first research question was designed to provide clarity as to what the psychosocial concept of computing identity is (and is not) in partial response to Rodriguez and Lehman's (2017) call for more research that systematically parses computing identity from its related concepts—namely, from computing self-efficacy and sense of belonging in computing. Findings from the factor analyses lead me to conclude that while the concepts are related, computing self-efficacy is a distinct measure from computing identity, whereas sense of belonging is indeed integral to measuring computing identity. In fact, the best-fitting structure for these data was four separate measures, each with three items: *interest in computing*, *computing identity*, *computing self-efficacy: professional milestones*, and *computing self-efficacy: professional competencies*. Of note for this dissertation is the measure of computing identity, which includes three items related to students' self-identity as a computing person, their sense that computing is a part of their core identity, and their sense of belonging in the greater computing community.

Specifically, these three items are: "*I see myself as a 'computing person,'*" "*Computing is a big part of who I am,*" and "*I feel like I 'belong' in computing.*"

These findings highlight several important contributions to the literature. First, this is the first known study to assess whether computing identity is best measured as a second-order factor with subconstructs or as a distinct factor in and of itself. Understanding that self-efficacy is different from computing identity and that it can be broken down into two separate measures offers computing education researchers more clarity on what computing identity is and provides self-efficacy researchers with more insight into the nuance of this concept in computing. Second, the finding that computing identity cannot be separated from students' sense of belonging in computing is relatively new. Interestingly, this finding conflicts with the work of some scholars, such as Werner & Chen (2024), who recently argued that belonging and identity are separate measures and more closely supports the work of scholars who have argued that sense of belonging is a core aspect of computing identity (Lunn et al., 2021a; Mahadeo et al., 2020; Taheri et al., 2019; Rodriguez et al., 2023).

Importantly, while a measure of computing identity with subconstructs could be used and would align more with existing literature, the fact that four separate factors provided the best fit for these data does not mean the work of others should be overlooked. Instead, we must remember that other studies mostly used different populations of students and were driven by different theoretical frameworks, leading to different approaches to data analysis and subsequent findings. The implications of these findings and recommendations for future research is discussed later in this chapter.

Research Question 2 Finding: The Measure of Computing Identity Held Consistently for Students from Different Gender, Racial/Ethnic, and Intersectional Identities but Did Not Necessarily Hold Consistently over Time

Given the need to broaden the participation of students in computing paired with the mixed research on what is known about how computing identity varies for students by gender and race/ethnicity, I investigated the equivalence of the new computing identity measure for students from different social groups and across each survey time point (Litson & Feldon, 2021). Findings from these analyses indicated computing identity did not vary by gender, race/ethnicity, or intersections of students' gender and racial/ethnic identities but did vary over time.

Overall, these findings are important because they expand on the work of Mahadeo and their colleagues (2020), who investigated the equivalence of their computing identity measure and also found that it did not differ by gender, race, and ethnicity. While Mahadeo et al. (2020) assessed invariance for their computing identity measure with subconstructs, my study measured computing identity as a distinct factor without subconstructs. Additionally, Mahadeo and their coauthors treated women, Black, and Hispanic students as binaries in their analyses, and my study disaggregated gender and race/ethnicity to assess measure equivalence over more categories in addition to looking at it for intersections for students' gender and race/ethnicity. Because one year elapsed for students who took their pretest in the fall and the follow-up survey the following fall and less than one year elapsed for students who took their pretest survey in the spring before their final survey the following fall, more research may be necessary to explore whether computing identity should be measured in the same way over time, especially as scholars use data with longer time periods between survey time points. For instance, a researcher seeking to study computing identity among early career or more seasoned computing professionals might add survey items related to belonging in the tech industry or perception of being a computing person at work versus at technical conferences.

In practice, findings from this study suggest that researchers can use the single latent measure of computing identity in their computing education research studies, including research among samples of students from these student backgrounds. However, scholars should still take heed of the applicability of the three-item computing identity measure among different groups that this study did not assess (e.g., international students, gender nonbinary students, and racial/ethnic student subgroups).

Research Question 3 Finding: Across Gender and Racial/Ethnic Identities, there Exist Both Commonalities and Differences in Computing Identity Development

Findings from the regression analyses highlights several key variables that predict computing identity and how the predictive power varies not only among (a) all students, (b) women, (c) men, (d) a combined group of underrepresented Students of Color, (e) Asian students, and (f) white students, but also across gender and racial/ethnicity. In accordance with Miller's Generalization, Example, Exception (GEE) guide to writing effectively about findings from quantitative analyses (2013), I begin by reviewing the key variables that share similar direction and magnitude in predicting computing identity across two or more of the regression models. Then, I share variables with exceptions to those patterns, augmenting the unique relationships of these variables for students from varying gender and racial/ethnic identities.

Discussion of Key Variables with Shared Salience and Direction

Variables that played a similarly positive role in two or more models included the pretest measure of computing identity, amount of prior programming experience, interest in computing, and being a computing major/minor at the point of the first follow-up survey. Also consistent across multiple models was that when controlling for all other variables, women, Asian, and Black students have lower computing identity scores relative to men and white students. In this section, these common findings are discussed with the extant literature.

The Importance of Incoming Computing Identity and Prior Programming Experience. Findings from the regression analyses indicate the importance of students' pre-college experience—namely, students' incoming computing identity and prior programming experience—as critical components to fostering students' computing identity during college. In fact, the pretest computing identity measure was unsurprisingly the strongest predictor of students' computing identity in the fall of the academic year following students' introductory computing course for all student groups. Importantly, because many of the students surveyed were in their first or second year of college, students' scores on this measure were likely a result of their experiences prior to college, such as their family background, their experiences in primary and secondary school, and their exposure to the computing field to date (Kapoor & Gardner-McCune, 2019; Rodriguez et al., 2020; Taheri et al., 2019).

The salience of prior programming experience in the regression from the full model with all students aligns with existing literature on the importance of preparation during high school (Azmitia et al., 2009; Chang et al., 2014; Robnett et al., 2015). However, the gender and racial/ethnic exploration of this variable shows that previous programming experience was only predictive of computing identity in the regression among all students and for Asian students, not for other subgroups in particular. This finding indicates that the value of prior programming experience is not universal (especially when comparing Asian students with their white and USOCC peers), and thus merits further exploration—particularly to ascertain why early programming experiences appear to be particularly salient to Asian students' sense of computing identity.

In general, these findings align with existing literature highlighting the important role of preparation during high school in fostering students' psychosocial factors such as computing identity (Azmitia, Cooper, & Brown, 2009; Chang, Sharkness, Hurtado, & Newman, 2014; Robnett et al., 2015), which emphasizes the importance of early education (i.e., elementary, middle, and high school) in generating that interest for young students. As such, we must continue investigating ways to bolster students' incoming computing identities and focus our investments on resources that support these endeavors, as will be discussed later in this chapter.

The Value of Initial and Sustained Interest in Computing. College-related variables also contributed to students' development of computing identity, including students' interest in computing at the end of their introductory course and majoring or minoring in computing in the fall after completing their introductory computing course. These findings are an important contribution to the literature because prior studies have only included interest as a component of computing identity measures, not as a separate independent variable, and have not used longitudinal data to explore the role of major in fostering students' STEM disciplinary identities (e.g., Lunn et al., 2021a; Mahadeo et al., 2020; Taheri et al., 2019).

In alignment with literature suggesting that interest in computing is an essential facilitator of computing identity, the fact that the measure of interest in computing was a strong positive predictor in four out of the six regression models was unsurprising (e.g., Carlone & Johnson, 2007; Lunn et al., 2021a). In fact, the factor for interest was one of the strongest predictors for forming computing identity among men and white students. One possible explanation for these findings is that men and white students may find it easier to "fit in" with the CS stereotype (e.g., Cheryan et al., 2015), so seeing themselves as a computing person may come easier to them than it does for other students. Further, the alignment with stereotypical image of a computer scientist might make it easier for students to reinforce their self-perception as a computing person and sense of belonging in the computing domain (Leary et al., 2013). For example, students from these groups (i.e., men and white students) might observe more individuals in computing who share their gender and/or racial/ethnic identity, making it easier for them to internalize that they belong in the field.

Additionally, being a computing major or minor in the fall of the academic year following a student's completion of their introductory course was a significant predictor of their computing identity, meaning that computing majors and minors tended to have stronger computing identities than the sample of students who changed their major from computing to another field or who had not yet declared their major. While this variable was significant for all groups, being a computing major was the least important predictor among the key variables predicting computing identity for USOCC, suggesting that while important, there are other variables that are more salient in predicting computing identity for this population that need further investigation.

The Role of Students' Gender and Racial/Ethnic Identities. The separate regression analyses among gender and racial/ethnic groups help us to understand the variables that predict computing identity *across* the groups while offering insights into unique ways to foster computing identity for *particular* groups. For instance, my findings highlight that both Asian women and Black women had lower computing identity scores than white women, raising questions about why this may be. While literature documenting students' experiences with microaggressions, sexism, racism, and misogyny broadly in STEM (and computing more specifically) are quite dated, recent research from Rodriguez et al. (2020) provides evidence that Latinas in their study attributed their lower computing identity to ongoing racism and sexism in the academy. While the survey data used in this study did not include questions related to these important concepts, my findings highlight differences in women's computing identity by race/ethnicity; therefore, I encourage more researchers to study the environmental mechanisms that may contribute to these differences in computing identity for women. I return to this and provide recommendations for future survey design and analytic studies in this chapter's "Limitations and Future Directions" section.

Overall, the separate regression analyses across gender and racial/ethnic groups illuminates key variables that help to nuance our understanding of how computing identity is developed among students from different social identities during college that the analysis among all students combined could not reveal. Additionally, and perhaps more importantly, these analyses allowed for the comparison of predictive variables across gender and race/ethnicity models and the advancement of our understanding of differences in the magnitude of predictive variables within these groups (Brickhouse et al., 2000). However, we must continue to disaggregate gender and race/ethnicity data wherever possible (i.e., while maintaining sample sizes large enough to protect the anonymity of study participants) to better our understanding of these nuances and interrogate what it means that certain underrepresented populations in computing continue to be unable to see themselves or to adopt computing identities in the same ways as their majority male and majority white peers.

Discussion of Key Variables with Different Salience and/or Direction

I also found some variables that had different relationships to computing identity that were otherwise exceptions to the aforementioned patterns, further implying there is not a one-size-fits-all approach to fostering computing identity among our students. One such example is the finding that men's computing identity scores tended to be greater the more they felt dismissed. This finding is somewhat unanticipated, yet some recent scholarship has also found students' perceptions of feeling dismissed being a positive predictor of computing outcomes (Wofford et al., 2022; Lunn et al., 2021b). In these studies, the authors propose a few possible theories to explain this phenomenon—one such explanation suggesting that the dissonance students feel in these discouraging encounters could, rather than dissuade them, serve as fuel in their efforts to double down to prove others wrong. However, the important contribution of my work is that this variable was only a predictor for men. One explanation for the salience of this variable being seen only for men may be found in how masculinity is taught through sports and other activities, where men may be socialized toward being motivated by negative experiences so that they "fit the mold" of what it means to be successful in computing (Cheryan et al., 2015).

I would be remiss if I did not mention that there are other variables with interesting and sometimes counterintuitive effects on computing identity for some groups, such as the differential role that self-efficacy: professional milestones plays for USOCC versus for white students and the fact that the peer support variable is only significant in the model for white students. However, the unique role these variables play for some groups and not others was not easily interpretable, as these anomalous occurrences may be due to uniqueness among these groups for this dataset or multicollinearity. Therefore, further investigation beyond the scope of the present study is needed.

Implications and Recommendations

Given the significance of these findings and my interest in equitably supporting all students in computing, this dissertation provides new insights into computing identity and offers valuable recommendations for its development among the diverse range of students in the field. I center these recommendations around enhancing practice and policy and advancing theory and research.

Implications for Practice and Policy

In particular, my findings show four key student *experiences* that promote computing identity. Upon beginning their introductory course, students' preexisting computing identities and prior programming experience are the strongest predictors of their later computing identities. During their introductory course, however, the most salient variables of students' later computing identities are their computing interest and whether or not they major in computing. Therefore, in order to promote computing identity we must examine the educational spaces and opportunities that students have both before *and* during college, which research suggests promotes these factors. Put simply, providing greater computing exposure to more students can help encourage students to pursue computing and ultimately promote their computing identity. In doing so, given the persistent lack of diversity in the field, attracting a diverse range of students is particularly important. Below are some suggestions that speak to this.

Cultivate Interest before College. Several studies examining how to foster greater interest in and exposure to computing before college show that beyond issues of inequitable access to computing (2023 State of CS Report, n.d.), we must start early to promote positive and gender-neutral stereotypes that combat long-held beliefs about what computing is and the demographic of those in the field. For instance, masculine stereotypes about computing have been shown to develop as early as third grade, and these stereotypes tend to predict lower interest in computing among girls (Masters et al., 2021). Suggestions for combating these stereotypes include creating gender-neutral marketing content (Jhally & Kilbourne, 2010; PBS, 2021);

highlighting women technologists from all racial/ethnic backgrounds and People of Color from various gender identities in computing who can serve as role models (Gallup-Amazon, 2021; Kapor Center, 2023); and strengthening the encouragement young students receive from their counselors (NCWIT, n.d.a), their teachers (Kapor Center, 2023), and their families (e.g., Sax et al., 2017). For example, instructors can make learning computing fun, show students that what they are already doing in their classrooms is considered coding/programming, and explore the potential impact these skills could have on advancing technology for our society (Kapor Center, 2023). With more exposure to computing and a greater belief in one's ability to be successful, students will likely gain a stronger sense of not only belonging (Sax et al., 2018) but also computing identity, as these concepts are intertwined.

Even if students do not get this exposure in their K-12 education, they still have opportunities to cultivate an interest in computing during college. Given that women and other historically underrepresented minority groups have less early exposure to computing and their interests in computing tend to develop later in life (Margolis et al., 2017; CRA, 2017), there is a particularly important opportunity to foster interest among these groups in introductory computing courses (Lehman et al., 2022; Sax et al., 2017). Specifically, researchers have found that inclusive teaching practices can effectively garner students' interest in computing and promote the formation of strong computing identities (Kapor Center, 2017). In particular, Ladson-Billings (2021) highlights that there have been nearly three decades of research advocating for culturally responsive, sustaining, and congruent pedagogy that encourages educators to take asset-based approaches to support students from marginalized backgrounds. These approaches are critical to fostering and sustaining students' interest in computing during college because they recognize the strengths students bring with them and meet students where they are through inclusive curricula that students can relate to and instruction environments designed to foster computing identity. For instance, computing instructors can remind students that there are multiple ways to solve problem sets and allow students the freedom to find the answer in their own unique ways (NCWIT, n.d.). Further examples of these efforts include using socially relevant computing examples or assignments relating to things students care about or aspects of their identity (Kapor Center, 2017). Indeed, these practices have been proven useful for fostering interest even for students exploring majors (Lehman et al., 2022) and those seeking to acquire computing skills through an interdisciplinary computing major (Newhouse et al., 2024).

Importantly, while this study does not detect the pedagogical style and teaching philosophy of computing instructors or the unique role of structured programs tailored to support students from marginalized groups in computing, some studies suggest these factors may be helpful to nurturing computing identity, particularly for marginalized students in STEM (e.g., Chang et al., 2011, 2014). Therefore, awareness of these resources and support dedicated to members of these groups (e.g., Society of Asian Scientists and Engineers; BlackComputeHER) is necessary and may need to be scaled up. While I acknowledge that there is not enough literature that nuances the experiences of all individuals marginalized in computing—for instance, Asian women who are "overrepresented" in the field compared to their representation in the U.S. population but still experience sexism and racism as Women of Color (Lane et al., 2021; Ong et al., 2011)—there are still valuable resources for helping to understand the distinct experiences of other marginalized Women of Color, such as Black women, in computing. For instance, the Modern Figures podcast (McMullen & Waisome, 2019-present) seeks to inspire the next generation of Black girls in computing by uplifting the voices of successful Black women with PhDs in the field of computing. The podcast shares these women's various perspectives, successes, and past and present hurdles. Additionally, there are other important resources related to Latinas in computing (e.g., Latinas in Computing community, LATINITY conference) that may also be valuable. Of note, however, is that in this study, the Latine variable was not a significant predictor of students' computing identity during college.

Review and Revise Curriculum and Policy. In addition to interest in computing being a key factor of students' computing identity development, the variable representing students' computing major emerged as another key research finding. To address the important role of being a computing major or minor one year after students' introductory computing courses on students' computing identity, computing departments should critically examine the equity of their admissions and enrollment practices for students in computing at various transition points (CRA, 2017; Lehman et al., 2020). For example, institutions should (re)consider if their policies and practices around student admission (direct versus transfer) and major declaration inadvertently skew who is represented in their computing departments. For example, recent research shows that community college transfer students tend to be more diverse than their non-transfer peers but experience stigma as they navigate university systems (Blaney et al., 2024). Additionally, Newhouse and her colleagues (2024) demonstrate that online resources for those seeking interdisciplinary majors are vague at best and often unavailable on department websites. Therefore, computing departments may attract a more diverse student body by providing more transparency in their policies and processes and offering different pathways into and through the field.

Beyond these different points of entry into computing, it is also important for schools to consider making curricular changes and even offering different academic degree plans that are more tailored to student's goals and reflect the diversity of students' experiences. For instance, some institutions offer a bachelor of arts (BA) and a bachelor of science (BS) degree option in computing. Here, these institutions work toward redefining what it means to be a computing person by offering undergraduate courses such as "Computing and the Humanities" or interdisciplinary degree programs that bring together computing and humanities studies (e.g., Evrard & Guzdial, 2023). While these efforts are newer, computing programs have seen an increase in the percentage of women and historically marginalized students who opt for BA degrees in computing (Lapan, 2023). Therefore, to leverage these new degree offerings to increase the representation of women across race/ethnicities and People of Color more broadly in computing, it will be crucial for the technology industry to recognize the rigor of these emerging degrees. However, more research would be needed to determine whether students' computing identities in these programs differ, as found in one study (Parker, 2019).

Despite the importance of prior programming experience in predicting computing identity, not all students have the resources to acquire these skills and access these experiences before college, especially if their pre-college institutions do not offer foundational computing course options (2023 State of CS Report). Institutions aiming to address equity and cultivate interest and participation in computing across diverse student groups may consider restructuring their curriculum to accommodate students' varying levels of computing experience when they begin college. Such a curricular policy was used to help broaden participation in computing at Harvey Mudd College, which successfully increased the number of women in computing in the early 2010s. By creating different sections of their introductory computing course (i.e., "Green," "Gold," and "Black"), students were able to learn with others who had a similar level of prior programming experience. All courses prepared students equally for the next course in the sequence (Alvarado et al., 2012).

Taken together, combating stereotypes to foster greater interest in computing, adapting the computing curriculum and pedagogy in introductory computing courses to be more inclusive and formative, and even generating alternative pathways to computing degrees can help us to not only recognize the diverse needs of our students who may not have had computing-related opportunities prior to college but also be ready to support them during their college career (Kapor Center, 2023). Further, research shows these best practices can also help to foster a greater sense of belonging among students (Sax et al., 2018), which my study finds to be integral to students' computing identities. Therefore, elevating and investing more in these efforts can ultimately help us move the needle toward recruiting and retaining a more diverse range of students to create a more socially just computing environment.

Limitations and Future Research

Despite the significance of these research findings, there are also important limitations to consider. Although the limitations of this study are detailed in Chapter Three, it is important to briefly reiterate them here in order to explore how future research can address issues surrounding generalizability and sample size, extend what we know about differences by groups, and strengthen survey design and measurement.

Ultimately, the results of this study may not be generalizable to all students in computing writ large. For instance, I cannot be sure that the reason certain variables are not significant in the regression analyses predicting computing is due to the unique makeup of the students from research institutions that are driven (and funded) to diversify computing or because these variables simply are not predictors of computing identity. As a result, one key avenue of future research would be to corroborate these findings with additional research that replicates and extends the work of this study using samples of students from other higher education contexts to reflect a wider range of student demographics and institution types.

Although this dissertation contributes to the breadth of what we know about computing identity and provides evidence that computing identity forms differently among various student groups, it does not provide a deeper understanding of why these differences exist or the mechanisms by which computing identity is formed among each student group. While I applied an intersectional gender and racial/ethnic lens in my analyses, I ultimately opted to exclude or combine groups with small counts. Even though I made these decisions for this study, I recognize that not all students experience college in the same ways (even among these student groups); therefore, I recommend more research to more fully understand students' unique experiences with computing identity.

To this end, I have several recommendations to expand analyses to provide more nuanced understandings of students' intersectional computing identity experiences. One way to expand this work is by integrating a qualitative stream of research that contextualizes undergraduate students' lived experiences and how they make meaning of their computing identities. For instance, it would be valuable to conduct semi-structured interviews with a phenomenological approach to understand how students make meaning of the phenomenon of computing identity. Additionally, qualitative research that uses portraiture or case studies can help illuminate these differences further and avoid a homogenous approach to promoting students' computing identity (Lawrence-Lightfoot & Davis, 2002). Future research could explore the nuanced salience of the key variables unique to particular groups, as has been done for Latinas in computing (Rodriguez et al., 2020, 2023). Similarly, quantitative research might accomplish this by studying how computing identity is developed among more nuanced student groups (e.g., post hoc analyses among Asian and USOCC subgroups) or through simple descriptive disaggregation (e.g., frequencies, crosstabulations, mean comparisons). Studies of this nature would also benefit from comparative analyses focusing on understanding students' other intersectional identities, such as class or citizenship (Sax et al., 2024).

While using a longitudinal design and data on students from multiple institutions adds strength to contextualize these findings, the surveys these data were drawn from were not designed to study computing identity. While this study is longitudinal (and includes three survey time points), it only assesses students' computing identities in the year directly following their introductory course experiences. Therefore, additional research that examines the concept of computing identity over an extended period of time (e.g., either later during college or into students' early or later graduate school or career journey) is needed to document the development of students' computing identities and to articulate the ebb and flow of these identities during and after college (Mahadeo et al., 2020). Relatedly, while I treat computing identity as an outcome for this study, I conceptualize it as a process. Therefore, other scholars might consider computing identity as a malleable trait rather than as a discrete outcome. In doing so, research should explore if and how what constitutes our measurement of computing identity varies over time and if this study's results are more applicable to computing identity among *undergraduates* rather than among individuals at later educational and career stages.

Despite the value of being able to include data from before, during, and just following students' introductory computing course, findings from this study indicate that we need more research that uses student demographics, academic background contexts, and prior college experiences to help us further understand the factors that predict students' early college computing identities. Additionally, I found it difficult to identify variables on the survey with enough data related to components of computing identity that prior literature suggested were critical (i.e., recognition from others, performance, and sense of belonging). Therefore, studies that investigate students' computing identities might seek to develop survey questions around (a) recognition (how one is recognized and whose recognition matters); (b) performance (subjective measures, beyond grades and perceived confidence, of students' ability to do computing); (c) sense of belonging (both at the individual and environmental levels); (d) the quality and nature of students' interactions with peers, faculty/TAs, and computing department personnel; and (e) nuanced classroom experiences related to bias, hostility, racism, and sexism that may affect students' interests. Fortunately, many of these topics and others are being explored by the Center for Inclusive Computing longitudinal surveys developed by *Momentum* at UCLA (Momentum, 2021; 2023).

Future research should also provide transparency in operationalizing and assessing computing identity measures. Researchers using the same dataset and subsample can rely on the factor analyses and measure invariance tests done in this study and therefore use the three-item measure of computing identity I propose. However, as scholars continue to explore the phenomenon of computing identity in their studies of students in computing writ large (e.g., Mahadeo et al., 2020; Lunn et al., 2021a) or research focused on subpopulations such as Latinas in computing (Rodriguez et al., 2020), they should be clear in how they are defining their measure of computing identity and in how they have ensured adequate fit for their measure for their unique sample. Clear operationalization of one's computing identity measure includes naming whether related concepts (like sense of belonging or self-efficacy) are considered subcomponents or are intertwined with their theory and conceptualization. In terms of measurement, I encourage using confirmatory factor analysis to validate the number of items, factor(s), and model structure to elucidate the extent to which one's measure differs from prior scholarship. I also recommend that researchers assess how their computing identity measure functions among students from different demographic populations in their samples to combat the monolithic, one-size-fits-all approach to understanding and fostering computing identity.

Finally, findings from this study illuminate several important opportunities for future scholars to further research computing identity as a means to promote greater persistence among marginalized students in computing. There are a variety of different analytical approaches that I recommend to help extend this research. Given that computing identity is multifaceted (Rodriguez et al., 2017; 2020), we may need additional scholarship to identify meaningful subtypes or subprofiles of students' computing identities. Fortunately, statistical techniques such as latent class analysis and latent profile analysis, though less prominent in education research, take a more holistic, person-centered (rather than a variable-centered) approach that would allow researchers to understand how variables interact within and across student populations (Lanza et al., 2013). Similar to Carlone & Johnson's (2007) three computing identity typologies (i.e., research scientist, altruistic scientist, and disrupted scientist), these latent profile analyses with distal outcomes could help validate these "types" of computing identities among students in computing and unveil distinct profiles of students with these distal outcomes—thereby further nuancing our understanding of computing identity and its unique functions for each student (Carlone & Johnson, 2007). Alternatively, we can use this approach to explore the relationship between computing identity and self-efficacy to help us identify the profiles of those students who might have high self-efficacy but not identify with computing. We could also incorporate more scholarship that utilizes path analysis, such as SEM, to build upon the regression findings to better understand the indirect effects of self-efficacy, identity, and interest in supporting the development of students' computing identities. Further, similar to the work of Taheri (2019), multigroup SEMs could help clarify not only the influence of these variables upon students' educational or career-related outcomes but also how these distinct pathways vary for students by gender, race/ethnicity, class standing, or over time. Unlike Taheri's examination of how the latent constructs of interest, performance/competence, recognition, and sense of belonging (as subcomponents of disciplinary identity) contribute to academic persistence, however, future studies may consider whether and how a distinct, latent computing identity measure predicts other outcomes related to success in computing.

Conclusion

In response to the need for more research that thoroughly examines the operationalization and measurement of computing identity for marginalized student groups (Rodriguez & Lehman, 2017), the purpose of this study is to (a) create a robust measure of computing identity; (b) explore the extent to which the new computing identity measure holds over time, during college, and across student's gender, racial/ethnic, and intersectional social groups; and (c) identify the key demographic, pre-college, and during-college variables that predict the new computing identity measure and explore how the predictive power of these variables differs by gender and race/ethnicity. Using longitudinal survey data from 1,036 undergraduate students from 15 research universities across the United States, this study finds that the three-item measure of computing identity is distinct from measures of self-efficacy and interest but not from the concept of sense of belonging. Additionally, findings indicate that the measure of computing identity held across students' gender identities, racial/ethnic identities, and intersections of students' gender and racial/ethnic identities but did not necessarily hold over time during college.

Finally, this study highlights the importance of students' incoming computing identities and prior programming experiences, the value of an initial and sustained interest in computing, and the role of students' gender and racial/ethnic identities in the development of their computing identities in the academic year following their introductory course experience. Illuminating these unique factors that contribute to computing identity for students from different social groups can help educators serve their students more equitably.

In essence, we know from prior research that computing identity equips educators to better support student persistence, as computing identity is such a strong predictor of undergraduate and career outcomes in computing (George et al., 2022; Lehman et al., 2022; Wofford et al., 2022). In investigating computing identity for this study, I find that while there are some common contributing factors, there are also differences across these different gender and race/ethnicity groups that need to be taken into consideration if we are to broaden participation in computing. Further, I find that this comprehensive understanding of computing identity and the pre-college and college-related experiences that bolster students' identities in computing by gender and race/ethnicity is crucial to promoting student retention and much-needed diversification in the computing field. Therefore, this study offers not only important implications for enhancing students' interest in and experiences of programming before college but also inclusive curricula and policy recommendations that can help strengthen the development of students' computing identities when they get to college. Moving forward, we must also consider tailoring our approaches toward marginalized gender and racial/ethnic groups in computing and technology. Overall, these insights have broader implications for diversifying the representation of undergraduate students in computing and the workforce.

APPENDICES

Appendix A

Overview of Variables and Coding

Variable	Definition/Coding Scheme
Dependent Variable	
—Computing Identity	Three-item factor (follow-up). Based on items from Likert scale ranging from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*) —I see myself as a "computing person." —Computing is a big part of who I am. —I feel like I "belong" in computing.
Pretest Computing Identity	
—Pretest Computing Identity Score	Three-item factor (pretest). Based on items from Likert scale ranging from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*) —I see myself as a "computing person." —Computing is a big part of who I am. —I feel like I "belong" in computing.
Personal Characteristics	
—Gender: Women	Dichotomous: 1 (*Woman*); 0 (*Man*)
Race (White)	
—Race: Asian	Dichotomous: 1 (*Yes*); 0 (*No*)
—Race: Black	Dichotomous: 1 (*Yes*); 0 (*No*)
—Race: Latine	Dichotomous: 1 (*Yes*); 0 (*No*)
—Race: Indigenous, Multiracial, or Other	Dichotomous: 1 (*Yes*); 0 (*No*)
—First-Generation Status	Dichotomous: 1 (*Yes; students' parents/guardians do not have a bachelor's degree*); 0 (*No*)
Background Contexts	
—Class Standing	Dichotomous: 1 (*Yes; first- or second-year student*); 0 (*No; third-year student or beyond*)
—Socioeconomic Status	Categorical: 1 (*Poor*); 2 (*Below Average*); 3 (*Average*); 4 (*Above Average*); 5 (*Wealthy*)
—Transfer Status	Dichotomous: 1 (*Yes; student transferred from another college or university*); 0 (*No; student did not transfer*)
—HS Computing Course	Dichotomous: 1 (*Yes*); 0 (*No*)
—HS GPA	Continuous
Early Computing Identity Experiences	
—Parent Computing Career Status	Dichotomous: 1 (*Computing Career*) 0 (*Non-computing career*)
—Family Computing Support	Four-item factor, based on items from Likert scale ranging from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*) —My family encourages me to pursue a computing degree.

	—My family questions why I would pursue a computing degree. (*Reverse coded*) —My family wonders why I invest so much time and effort into studying computing. (*Reverse Coded*) —My family emphasizes the value of earning a computing degree.
—Amount of Prior Programming Experience	Seven-item aggregate measure: sum of prior programming experiences ranging from 0–7 where responses were summed as each item's response options were 1 (*Yes*) and 0 (*No*). —I did not have programming experience prior to this course. —I took a computer programming course in high school (e.g., Java, Python, HTML, etc.). —I took a computer programming course at computer camp. —I took a computer programming course online. —I took a computer programming course at this college. —I took a computer programming course at another four-year college. —I took a computer programming course at community college. —I did not take a specific course, but I learned to program on my own (e.g., by reading books).
Additional Psychosocial Computing Identity Factors	
—Interest in Computing	Three-item factor (posttest). Based on items from Likert scale ranging from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*) —I am interested in learning more about what I can do with computing —Using computers to solve problems is interesting —I care about doing well in computing.
—Computing Self-efficacy: Professional Competencies	Three-item factor (posttest). Based on items from Likert scale ranging from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*) —I am confident I can—become a leader in the field of computing. —I am confident I can—win a computing-related contest (e.g., programming context, robotics contest, hackathon). —I am confident I can—quickly learn a new programming language on my own.
—Computing Self-efficacy: Professional Milestones	Three-item factor (posttest). Based on items from Likert scale ranging from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*) —I am confident I can—find employment in an area of computing interest. —I am confident I can—get admitted to a graduate computing program. —I am confident I can—complete an undergraduate degree in computing.
College Computing Identity Experiences *Within* the Intro Course	
—Relevant Social/Personal Examples	Four-item factor (posttest), based on items from Likert scale ranging from 1 (*Never*) to 5 (*Always*) -Use of examples involving women -Use of examples involving People of Color -Use of examples involving relevant social issues

	-Discussions addressing misconceptions about the computer science field
—Instructor Inclusivity	Six-item factor (posttest), based on items from Likert scale ranging from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*) —TAs are inclusive and supportive of women —TAs are inclusive and supportive of students of color —TAs are interested in helping me when I come to them with questions —Introductory course instructor(s) are inclusive and supportive of women —Introductory course instructor(s) are inclusive and supportive of students of color —Introductory course instructor(s) are interested in helping me when I come to them with questions
—Feeling Dismissed	Three-item factor (posttest), based on items from Likert scale ranging from 1 (*Never*) to 5 (*All of the Time*) —People tend to attribute your success to special treatment or luck rather than to your competence. —You are "talked down to" by classmates, instructors, or advisors. —Your ideas or opinions are minimized or ignored.
College Computing Identity Experiences *Outside* of the Intro Course	
—Peer Support	Four-item factor (posttest), based on items from Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree) To what extent is each of the following kinds of support available to you from other computing students if you need it? —Someone to hang out with. —Someone to confide in or talk to about your problems. —Someone to get class assignments for you if you were sick. —Someone to help you understand difficult homework problems.
—Mentor Support	Three-item factor (posttest), based on items from Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree) To what extent do you have a mentor who: —Shows compassion for any concerns and feelings you discussed with them. —Shares personal experiences as an alternative perspective to your problems. —Explores career options with you.
—Computing Major/minor	Dichotomous: 1 (*Student indicated they were a computing major or computing minor on the first follow-up survey*); 0 (*Student indicated they were not computing major or computing minor on the first follow-up survey*)
—Undergraduate Research Participation	Dichotomous: 1 (*Yes, student participated in at least one research experience*); 0 (*No, did not participate in undergraduate research*)
—Technical Conference Participation	Dichotomous: 1 (*Yes, attended at least one conference*); 0 (*No, did not attend a conference*)

Appendix B

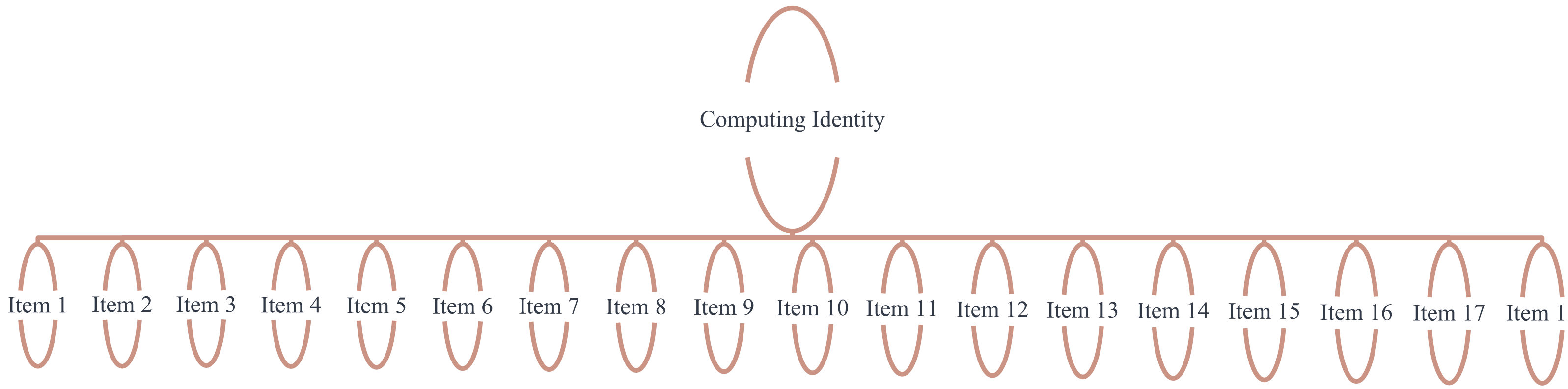
Simple Correlations and Descriptive Statistics of Factor Items

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1 I see myself as a "computing person."	1.00																	
2 I feel like I "belong" in computing.	0.76	1.00																
3 I am interested in learning more about what I can do with computing.	0.40	0.39	1.00															
4 Computing is a big part of who I am.	0.64	0.66	0.36	1.00														
5 Using computers to solve problems is interesting.	0.45	0.45	0.57	0.41	1.00													
6 I care about doing well in computing.	0.35	0.34	0.54	0.28	0.54	1.00												
7 Confidence in finding employment in an area of computing interest.	0.38	0.37	0.32	0.31	0.34	0.32	1.00											
8 Confidence in getting admitted to a graduate computing program.	0.35	0.36	0.25	0.31	0.29	0.24	0.58	1.00										
9 Confidence in completing an undergraduate degree in computing.	0.37	0.37	0.35	0.29	0.39	0.37	0.63	0.57	1.00									
10 Confidence in winning a computing related contest.	0.29	0.33	0.13	0.36	0.18	0.10	0.37	0.44	0.35	1.00								
11 Confidence in becoming a leader in the field of computing.	0.32	0.36	0.15	0.36	0.20	0.15	0.43	0.46	0.38	0.67	1.00							
12 Confidence in quickly learn a new programming language on your own.	0.34	0.34	0.20	0.35	0.27	0.17	0.37	0.39	0.40	0.49	0.52	1.00						
13 Confidence in clearly communicate technical problems and solutions to a range of audiences.	0.34	0.35	0.19	0.33	0.27	0.18	0.43	0.41	0.43	0.46	0.52	0.56	1.00					
14 Self-rated computer skills.	0.46	0.43	0.18	0.46	0.28	0.17	0.28	0.27	0.27	0.27	0.28	0.32	0.32	1.00				
15 My family emphasizes the value of earning a computing degree.	0.17	0.18	0.12	0.18	0.13	0.13	0.18	0.16	0.16	0.16	0.18	0.14	0.18	0.14	1.00			
16 I feel like an outsider in the computing community (R)	0.39	0.43	0.15	0.34	0.21	0.14	0.20	0.19	0.21	0.13	0.13	0.17	0.16	0.33	0.04	1.00		
17 I do not have much in common with other students in my computing classes (R)	0.21	0.26	0.13	0.17	0.15	0.09	0.13	0.11	0.15	0.06	0.04	0.09	0.07	0.17	0.06	0.51	1.00	
18 I feel welcomed in the computing community.	0.43	0.49	0.30	0.48	0.42	0.28	0.32	0.29	0.29	0.32	0.33	0.30	0.29	0.31	0.18	0.32	0.26	1.00
Sample	4,159	4,144	4,154	4,151	4,151	4,156	4,111	4,099	4,094	4,103	4,101	4,106	4,103	4,111	4,066	4,142	4,156	4,147
Mean	4.13	4.02	4.50	3.73	4.51	4.59	4.46	4.19	4.62	3.40	3.58	3.91	3.98	3.76	3.54	3.56	3.30	3.84
Standard Deviation	0.87	0.94	0.70	1.04	0.69	0.69	0.85	1.03	0.81	1.14	1.13	1.08	1.02	0.76	1.05	1.16	1.11	0.93

Notes. + denotes variable is a factor; (R) indicates that item was reverse coded.

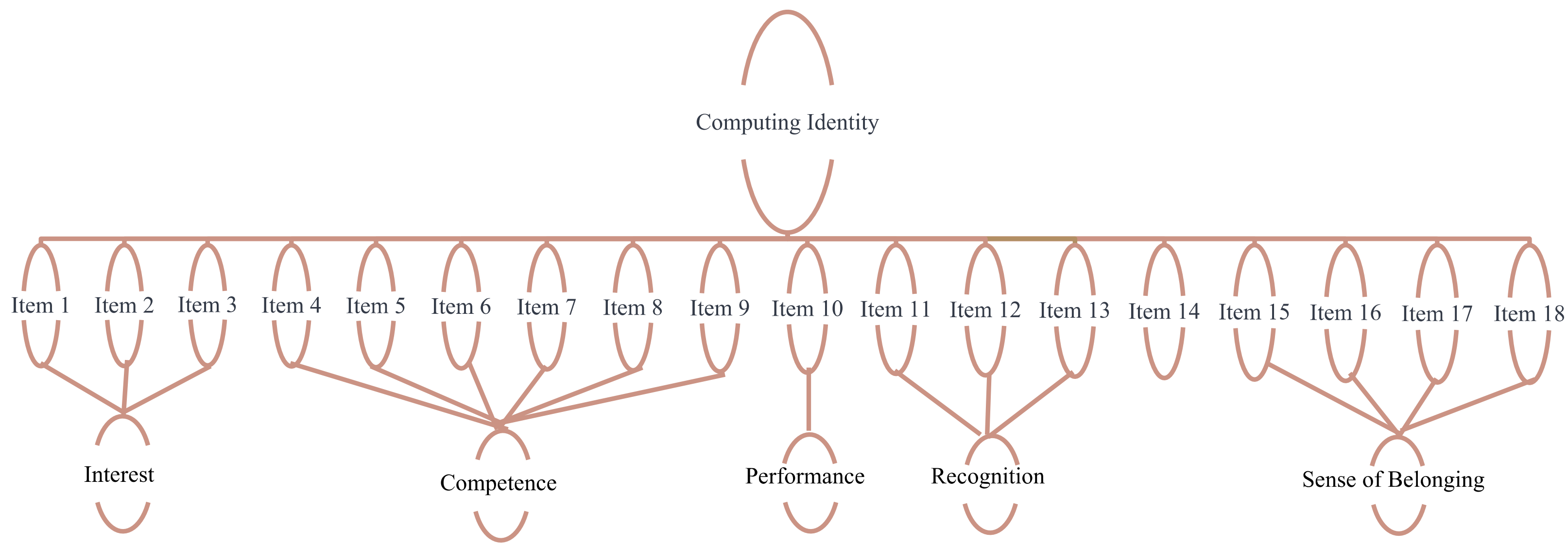
Appendix C

Single Factor (AKA Unidimensional) Model of Computing Identity



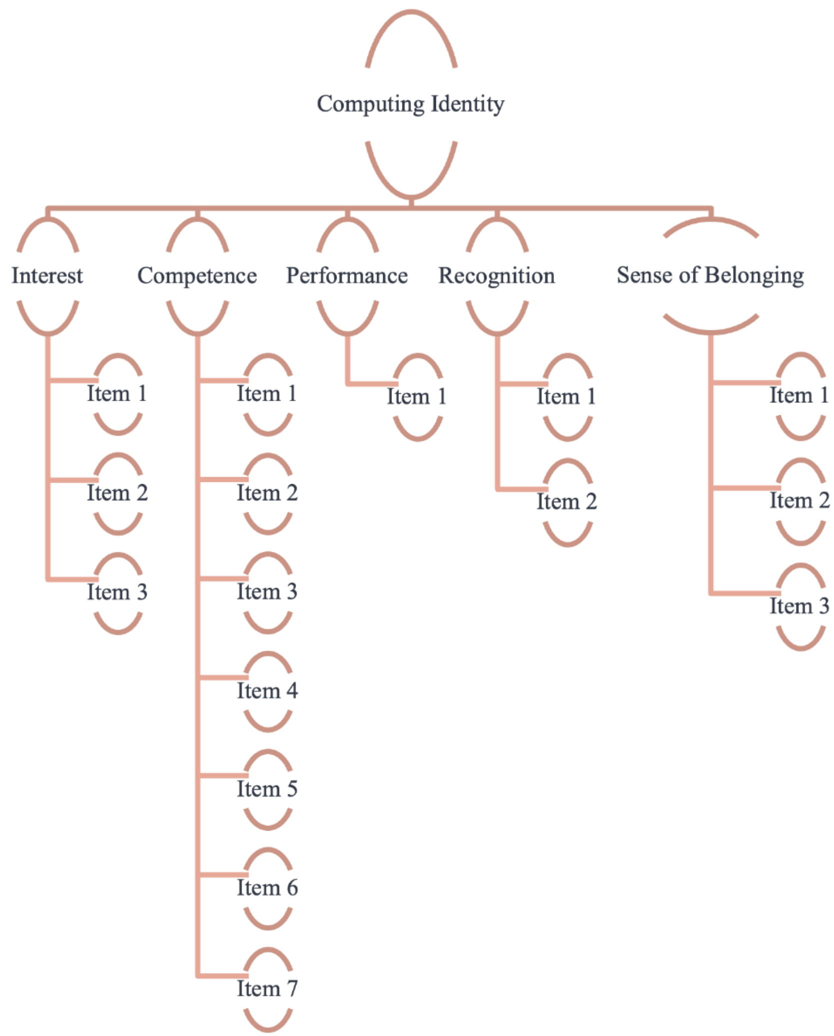
Appendix D

Bifactor Model of Computing Identity



Appendix E

Hierarchical (AKA Second-Order) Model of Computing Identity



Appendix F

Simple Correlations and Descriptive Statistics of Blocked Regression Variables

	DV	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
DV Computing Identity at First Follow-up	1.00																										
1 Pretest Computing Identity	0.62	1.00																									
2 Gender: Women	-0.30	-0.26	1.00																								
3 Race: Asian	-0.08	-0.09	0.06	1.00																							
4 Race: Black	-0.11	-0.06	0.06	-0.18	1.00																						
5 Race: Latine	0.05	0.10	0.02	-0.26	-0.09	1.00																					
6 Race: Other	0.00	0.02	0.03	-0.18	-0.06	-0.09	1.00																				
7 First-Generation Status	0.02	0.06	0.03	0.08	0.01	0.22	0.00	1.00																			
8 Class Standing	-0.02	-0.06	0.00	0.07	0.00	-0.06	-0.06	-0.01	1.00																		
9 Socioeconomic Status	-0.01	-0.08	-0.10	-0.07	-0.04	-0.28	0.05	-0.47	0.01	1.00																	
10 Transfer Status	0.00	0.09	0.01	-0.10	0.07	0.08	0.11	0.10	-0.47	-0.11	1.00																
11 HS CS Course	0.18	0.17	-0.08	0.00	-0.05	-0.08	-0.03	-0.08	0.03	0.12	0.00	1.00															
12 HS GPA	-0.03	-0.13	0.12	0.04	-0.05	-0.06	-0.03	-0.03	0.19	0.08	-0.20	-0.04	1.00														
13 Parent CS Career	0.03	0.00	0.01	0.02	-0.04	-0.08	0.00	-0.22	0.05	0.24	-0.07	0.12	0.01	1.00													
14 Family Computing Support+	0.17	0.26	0.03	-0.14	-0.11	0.00	0.00	-0.14	-0.01	0.18	-0.02	0.09	-0.03	0.20	1.00												
15 Amount of Prior Programming Experience	0.28	0.32	-0.05	-0.02	-0.04	-0.06	0.04	-0.10	-0.12	0.09	0.16	0.42	-0.10	0.05	0.18	1.00											
16 Interest in Computing+	0.38	0.40	-0.02	-0.08	-0.01	0.07	-0.01	-0.02	-0.02	-0.01	-0.03	0.02	0.02	0.00	0.24	0.14	1.00										
17 Computing Self-Efficacy: Professional Competencies+	0.33	0.46	-0.14	-0.04	-0.01	0.05	0.03	0.01	0.04	0.08	0.04	0.12	-0.04	0.15	0.14	0.20	0.22	1.00									
18 Computing Self-Efficacy: Professional Milestones+	0.32	0.47	-0.06	-0.07	0.04	0.03	-0.04	-0.01	0.07	0.00	0.01	0.05	0.01	0.06	0.26	0.12	0.37	0.55	1.00								
19 Relevant Social/Personal Examples+	0.04	0.05	-0.06	-0.03	0.02	0.00	0.06	0.02	0.05	-0.02	-0.03	0.04	-0.04	0.03	-0.03	0.00	-0.07	0.13	0.04	1.00							
20 Perception of Inclusive TAs/Faculty+	0.14	0.15	-0.01	-0.15	0.00	-0.02	0.00	-0.05	0.06	0.07	-0.05	0.02	0.05	0.03	0.25	0.01	0.30	0.04	0.18	0.15	1.00						
21 Feeling Dismissed+	-0.09	-0.10	0.07	0.11	0.10	-0.05	0.06	0.00	0.01	0.03	0.00	0.01	-0.04	0.01	-0.20	-0.02	-0.26	0.01	-0.10	0.16	-0.21	1.00					
22 Peer Support+	0.15	0.12	0.00	0.10	-0.03	-0.03	-0.10	-0.06	0.16	0.09	-0.20	0.06	0.14	0.12	0.10	0.02	0.15	0.12	0.13	0.18	0.21	-0.04	1.00				
23 Mentor Support+	0.09	0.13	0.06	0.07	0.01	-0.07	-0.05	-0.01	-0.01	0.08	-0.04	0.04	0.05	0.08	0.04	0.03	0.04	0.19	0.07	0.20	0.11	0.08	0.37	1.00			
24 Computing Major/minor	0.41	0.27	-0.12	0.11	-0.09	-0.01	0.05	0.02	0.10	-0.03	-0.01	0.11	-0.02	0.01	0.17	0.14	0.25	0.11	0.17	-0.04	0.10	-0.15	0.06	0.00	1.00		
25 Undergraduate Research Participation	0.04	0.00	0.08	-0.02	-0.04	0.05	-0.02	0.00	-0.03	-0.01	0.01	-0.04	0.02	0.01	0.02	-0.01	0.07	0.00	-0.03	-0.03	-0.07	-0.02	-0.06	0.04	0.03	1.00	
26 Technical Conference Participation	-0.03	-0.01	0.08	-0.09	0.06	0.01	-0.01	-0.02	0.01	0.02	0.01	-0.01	0.05	0.01	0.02	-0.02	0.04	-0.15	-0.02	-0.08	0.08	-0.05	-0.06	-0.09	-0.01	-0.11	1.00
Sample	1,024	1,010	1,024	1,030	1,032	1,031	965	1,036	1,036	1,034	1,036	1,005	1,010	1,036	1,004	1,004	1,007	1,005	999	911	796	983	990	989	1,036	1,036	1,036
Mean	9.46	9.76	1.36	0.29	0.07	0.11	0.06	0.32	0.83	3.14	0.18	0.50	7.85	0.22	10.16	1.34	10.61	8.01	10.50	6.10	18.60	4.29	10.79	7.50	0.74	0.09	0.94
Standard Deviation	2.68	2.29	0.48	0.45	0.25	0.32	0.23	0.46	0.37	0.87	0.38	0.50	1.31	0.42	2.01	1.14	1.72	2.32	2.00	2.58	3.44	2.20	3.63	3.45	0.44	0.29	0.24

Note. + denotes variable is a factor

Appendix G

Comparison of Stepwise Linear Regression Findings Predicting Computing Identity One Year after Students' Intro Computing Courses

Dependent Variable —Computing Identity at First Follow-up	*REGRESSION 1* **ALL** (*n* = 768) Final Model (Step 17)		*REGRESSION 2* **MEN** (*n* = 492) Final Model (Step 16)		*REGRESSION 3* **WOMEN** (*n* = 276) Final Model (Step 16)		*REGRESSION 4* **WHITE** (*n* = 375) Final Model (Step 13)		*REGRESSION 5* **ASIAN** (*n* = 268) Final Model (Step 13)		*REGRESSION 6* **USOCC** (*n* = 193) Final Model (Step 13)	
Independent Variables	*b*	Sig	*b*	Sig	*b*	Sig	*b*	Sig	*b*	Sig	*b*	Sig
Pretest Item												
—Pretest Computing Identity Score	0.41	***	0.39	***	0.43	***	0.52	***	0.38	***	0.40	***
Personal Characteristics												
—Gender: Women	−0.16	***	--	--	--	--	−0.08	*	−0.16	***	−0.26	***
Race (White)												
—Race: Asian	−0.08	**	−0.08	*	−0.11	*	--	--	--	--	--	--
—Race: Black	−0.06	*	−0.04		−0.11	*	--	--	--	--	--	--
—Race: Latine	−0.04		−0.02		−0.07		--	--	--	--	--	--
—Race: Other	−0.03		−0.02		−0.07		--	--	--	--	--	--
Background Contexts												
—Class Standing	−0.04		−0.08		0.02		−0.07		0.00		−0.05	
—Transfer Status	−0.04		−0.06		−0.02		−0.04		−0.02		−0.13	
—HS CS Course	0.02		0.01		0.04		0.05		−0.02		−0.01	
—HS GPA	0.03		0.01		0.06		0.03		0.01		−0.01	
Pre-College Computing Identity Experiences												
—Amount of Prior Programming Experience	0.07	*	0.07		0.10		0.04		0.16	***	0.01	
Additional Psychosocial Computing Identity Factors												

—Interest in Computing⁺	0.14	**	0.21	***	0.09		0.16	***	0.10		−0.01	
—Computing Self-Efficacy: Professional Milestones⁺	0.00		−0.01		0.01		−0.12	**	0.01		0.18	**
College Computing Identity Experiences *Within* the Intro Course												
—Instructor Provides Relevant Social/Personal Examples⁺	0.00		0.02		−0.03		0.03		0.04		−0.10	
—Feeling Dismissed⁺	0.04		0.10	**	−0.02		0.02		0.01		−0.04	
College Computing Identity Experiences *Outside* of the Intro Course												
—Peer Support⁺	0.05		0.07		0.01		0.07	*	0.09		0.00	
—Computing Major/minor at First Follow-up	0.23	***	0.21	***	0.27	***	0.24	***	0.24	***	0.16	**
Model R^2	0.48		0.43		0.46		0.55		0.44		0.45	

Note. *b* = unstandardized regression coefficient; Sig = *p*-value; *** = $p < .001$; ** = $p < .01$; * = $p < .05$; -- denotes variable did not enter the model; ⁺ denotes variable is a factor.

Appendix H

Linear Regression Findings Predicting Computing Identity Among Women One Year after Students' Intro Computing Courses (n = 276)

Dependent Variable —Computing Identity at First Follow-up			Pretest Only (Step 1)		Background (Step 10)		Psychosocial (Step 12)		College (Step 16)	
Independent Variables	*r*	Sig	*β*	Sig	*β*	Sig	*β*	Sig	*β*	Sig
Pretest Item										
—Pretest Computing Identity	.57	***	0.57	***	0.54	***		***	0.43	***
Personal Characteristics										
Race (White)										
—Race: Asian	−.04		*−0.01*		−0.08	0.13	−0.07		−0.11	*
—Race: Black	−.13	*	*−0.12*	*	−0.15	***	−0.13	*	−0.11	*
—Race: Latine	.05		*−0.05*		−0.08	0.13	−0.09		−0.07	
—Race: Other	−.02		*−0.03*		−0.07	0.18	−0.06		−0.07	
Background Contexts										
—Class Standing	.02		*0.04*		0.03	0.58	0.03		0.02	
—Transfer Status	.05		*−0.04*		−0.02	0.73	−0.02		−0.02	
—HS CS Course	.22	***	*0.09*		0.04	0.47	0.04		0.04	
—HS GPA	.02		*0.09*		0.08	0.11	0.08		0.06	
Pre-College Computing Identity Experiences										
—Amount of Prior Programming Experience	.30	***	*0.10*		0.10	0.08	0.10		0.10	
Additional Psychosocial Computing Identity Factors										
—Interest in Computing+	.34	***	*0.18*	***	*0.17*	***	0.17	***	0.09	
—Computing Self-Efficacy: Professional Milestones+	.28	***	*0.05*		*0.03*		0.02		0.01	
College Experiences *Within* the Intro Course										
—Relevant Social/Personal Examples+	−.04		*−0.06*		*−0.06*		*−0.05*		−0.03	
—Feeling Dismissed+	−.16	***	*−0.08*		*−0.07*		*−0.05*		−0.02	
College Experiences *Outside* of the Intro Course										
—Peer Support+	.14	*	*0.07*		*0.03*		*0.01*		0.01	
—Computing Major/minor	.44	***	*0.31*	***	*0.30*	***	*0.28*	***	0.27	***
Model R^2				0.321		0.367		0.392		0.457

Note. r = simple correlation; + denotes variable is a factor; Sig = *p*-value; *** = $p < .001$; ** = $p < .01$; * = $p < .05$; *β* = standardized regression coefficient.

Appendix I

Linear Regression Findings Predicting Computing Identity Among Men One Year after Students' Intro Computing Courses (n = 492)

Dependent Variable —Computing Identity at First Follow-up			Pretest Only (Step 1)		Background (Step 10)		Psychosocial (Step 12)		College (Step 16)	
Independent Variables	*r*	Sig	β	Sig	β	Sig	β	Sig	β	Sig
Pretest Item										
—Pretest Computing Identity	.58	***	0.58	***	0.55	***	0.46	***	0.39	***
Personal Characteristics										
Race (White)										
—Race: Asian	−.09	*	*−0.03*		−0.04		−0.03		−0.08	*
—Race: Black	−.04		*−0.02*		−0.01		−0.02		−0.04	
—Race: Latine	.03		*−0.01*		−0.01		−0.01		−0.02	
—Race: Other	.02		*0.01*		0.00		0.00		−0.02	
Background Contexts										
—Class Standing	−.03		*−0.02*		−0.04		−0.04		−0.08	
—Transfer Status	−.02		*−0.04*		−0.07		−0.05		−0.06	
—HS CS Course	.11	*	*0.05*		0.02		0.04		0.01	
—HS GPA	.01		*0.03*		0.03		0.01		0.01	
Pre-College Computing Identity Experiences										
—Amount of Prior Programming Experience	.25	***	*0.10*	*	0.09	*	0.07		0.07	
Additional Psychosocial Computing Identity Factors										
—Interest in Computing+	.43	***	*0.21*	***	*0.20*	***	0.21	***	0.21	***
—Computing Self-Efficacy: Professional Milestones+	.32	***	*0.05*		*0.05*		−0.01		−0.01	
College Experiences *Within* the Intro Course										
—Relevant Social/Personal Examples+	.04		*0.02*		*0.02*		*0.04*		0.02	
—Feeling Dismissed+	−.01		*0.00*		*0.01*		*0.08*	*	0.10	*
College Experiences *Outside* of the Intro Course										
—Peer Support+	.17	***	*0.09*	*	*0.09*	*	*0.07*		0.07	
—Computing Major/minor	.34	***	*0.21*	***	*0.22*	***	*0.20*	***	0.21	***
Model R^2				0.332		0.347		0.379		0.428

Note. r = simple correlation; + denotes variable is a factor; Sig = p-value; *** = $p < .001$; ** = $p < .01$; * = $p < .05$; β = standardized regression coefficient.

Appendix J

Linear Regression Findings Predicting Computing Identity Among USOCC One Year after Students' Intro Computing Courses (n = 193)

Dependent Variable —Computing Identity at First Follow-up			Pretest Only (Step 1)		Background (Step 7)		Psychosocial (Step 9)		College (Step 13)	
Independent Variables	*r*	Sig	*β*	Sig	*β*	Sig	*β*	Sig	*β*	Sig
Pretest Item										
—Pretest Computing Identity	.55	***	0.55	***	0.50	***	0.42	***	0.40	***
Personal Characteristics										
—Gender: Women	−.36	***	*−0.27*	***	−0.27	***	−0.28	***	−0.26	***
Background Contexts										
—Class Standing	.01	0.43	*0.03*		−0.04	0.54	−0.06		−0.05	
—Transfer Status	−.07	0.17	*−0.12*	*	−0.14	0.06	−0.15	*	−0.13	
—HS CS Course	.11	0.07	*0.03*		0.00	0.95	0.00		−0.01	
—HS GPA	−.03	0.36	*−0.01*		0.03	0.57	0.01		−0.01	
Pre-College Computing Identity Experiences										
—Amount of Prior Programming Experience	.13	*	*0.00*		0.02	0.74	0.02		0.01	
Additional Psychosocial Computing Identity Factors										
—Interest in Computing+	.28	***	*0.09*		*0.09*		0.04		−0.01	
—Computing Self-Efficacy: Professional Milestones+	.34	***	*0.14*	*	*0.17*	*	0.15	*	0.18	*
College Experiences *Within* the Intro Course										
—Relevant Social/Personal Examples+	−.06	0.22	*−0.13*	*	*−0.13*	*	*−0.13*	*	−0.10	
—Feeling Dismissed+	−.24	***	*−0.10*		*−0.07*		*−0.07*		−0.04	
College Experiences *Outside* of the Intro Course										
—Peer Support+	.13	*	*0.00*		*−0.03*		*−0.03*		0.00	
—Computing Major/minor	.36	***	*0.22*	***	*0.18*	***	*0.18*	***	0.16	*
Model R^2				0.301		0.386		0.411		0.450

Note. r = simple correlation; + denotes variable is a factor; Sig = p-value; *** = $p < .001$; ** = $p < .01$; * = $p < .05$; β = standardized regression coefficient.

Appendix K

Linear Regression Findings Predicting Computing Identity Among Asian Students One Year after Students' Intro Computing Courses (n = 268)

Dependent Variable —Computing Identity at First Follow-up			Pretest Only (Step 1)		Background (Step 7)		Psychosocial (Step 9)		College (Step 13)	
Independent Variables	*r*	Sig	*β*	Sig	*β*	Sig	*β*	Sig	*β*	Sig
Pretest Item										
—Pretest Computing Identity	.55	***	0.55	***	0.47	***	0.41	***	0.38	***
Personal Characteristics										
—Gender: Women	−.30	***	*−0.17*	***	−0.18	***	−0.19	***	−0.16	***
Background Contexts										
—Class Standing	.05	0.20	*0.05*		0.03	0.52	0.04	0.51	0.00	0.93
—Transfer Status	.01	0.46	*−0.04*		−0.04	0.47	−0.02	0.65	−0.02	0.69
—HS CS Course	.16	***	*0.10*		0.03	0.55	0.04	0.49	−0.02	0.78
—HS GPA	−.02	0.39	*0.01*		0.03	0.56	0.02	0.64	0.01	0.77
Pre-College Computing Identity Experiences										
—Amount of Prior Programming Experience	.27	***	*0.14*	*	0.15	*	0.15	*	0.16	***
Additional Psychosocial Computing Identity Factors										
—Interest in Computing+	.29	***	*0.11*	*	*0.14*	*	0.13	*	0.10	0.07
—Computing Self-Efficacy: Professional Milestones+	.29	***	*0.04*		*0.03*		0.02	0.76	0.01	0.80
College Experiences *Within* the Intro Course										
—Relevant Social/Personal Examples+	.11	*	*0.05*		*0.03*		*0.05*		0.04	0.48
—Feeling Dismissed+	.01	0.41	*−0.01*		*−0.02*		*0.01*		0.01	0.78
College Experiences *Outside* of the Intro Course										
—Peer Support+	.19	***	*0.11*	*	*0.11*	*	*0.09*		0.09	0.09
—Computing Major/minor	.35	***	*0.27*	***	*0.25*	***	*0.24*	***	0.24	***
Model R^2				0.308		0.363		0.379		0.440

Note. r = simple correlation; + denotes variable is a factor; Sig = *p*-value; *** = $p < .001$; ** = $p < .01$; * = $p < .05$; *β* = standardized regression coefficient.

Appendix L

Linear Regression Findings Predicting Computing Identity Among White Students One Year after Students' Intro Computing Courses (n = 375)

Dependent Variable —Computing Identity at First Follow-up			Pretest Only (Step 1)		Background (Step 7)		Psychosocial (Step 9)		College (Step 13)	
Independent Variables	*r*	Sig	*β*	Sig	*β*	Sig	*β*	Sig	*β*	Sig
Pretest Item										
—Pretest Computing Identity	.67	***	0.67	***	0.61	***	0.57	***	0.52	***
Personal Characteristics										
—Gender: Women	−.29	***	*−0.08*	*	−0.09	*	−0.09	*	−0.08	*
Background Contexts										
—Class Standing	−.05		*−0.03*		−0.03		−0.03		−0.07	
—Transfer Status	.04		*0.00*		−0.02		−0.02		−0.04	
—HS CS Course	.24	***	*0.07*		0.05		0.07		0.05	
—HS GPA	−.06		*0.04*		0.05		0.03		0.03	
Pre-College Computing Identity Experiences										
—Amount of Prior Programming Experience	.36	***	*0.09*	*	0.06		0.04		0.04	
Additional Psychosocial Computing Identity Factors										
—Interest in Computing+	.42	***	*0.18*	***	*0.19*	***	0.21	***	0.16	***
—Computing Self-Efficacy: Professional Milestones+	.31	***	*−0.06*		*−0.04*		−0.08		−0.12	**
College Experiences *Within* the Intro Course										
—Relevant Social/Personal Examples+	.01		*0.04*		*0.04*		*0.04*		0.03	
—Feeling Dismissed+	−.09	*	*−0.04*		*−0.04*		*0.00*		0.02	
College Experiences *Outside* of the Intro Course										
—Peer Support+	.22	***	*0.10*	**	*0.10*	*	*0.09*	*	0.07	*
—Computing Major/minor	.46	***	*0.25*	***	*0.25*	***	*0.24*	***	0.24	***
Model R^2			0.451		0.468		0.502		0.553	

Note. r = simple correlation; + denotes variable is a factor; Sig = p-value; *** = $p < .001$; ** = $p < .01$; * = $p < .05$; β = standardized regression coefficient.

REFERENCES

2023 state of CS Report. Code.org. (n.d.). https://advocacy.code.org/stateofcs

Aberson, C. L. (2019). *Applied power analysis for the behavioral sciences*. Routledge.

Abes, E. S., & Jones, S. R. (2004). Meaning-making capacity and the dynamics of lesbian college students' multiple dimensions of identity. *Journal of College Student Development, 45*(6), 612–632.

Abes, E. S., Jones, S. R., & McEwen, M. K. (2007). Reconceptualizing the model of multiple dimensions of identity: The role of meaning-making capacity in the construction of multiple identities. *Journal of College Student Development, 48*(1), 1–22.

Alvarado, C., Dodds, Z., & Libenskind-Hadas, R. (2012). Increasing women's participation in computing at Harvey Mudd College. *ACM Inroads, 3*(4), 55–64.

Anita Borg Institute (2022). 2021 top companies for women technologists [report]. https://anitab.org/research-and-impact/top-companies/2021-results/

Ariella, S. (2022). *25+ telling diversity in high tech statistics [2022]: Tech demographics + trends*. https://www.zippia.com/advice/diversity-in-high-tech-statistics/#:~:text=White%20workers%20make%20up%20a,Black%20employees%20make%20up%207%25

Aschbacher, P. R., Li, E., & Roth, E. J. (2010). Is science me? High school students' identities, participation and aspirations in science, engineering, and medicine. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 47*(5), 564–582.

Ashcraft, C., McLain, B., & Eger, E. (2016). Women in tech: The facts. The National Center for Women and Information Technology (NCWIT). https://ncwit.org/resource/thefacts/

Astin, A. W., & antonio, a. l. (2012). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education (2nd Ed.)*. Lanham, MD: Rowman & Littlefield Publishers, Inc.

Astin, A. W., & Dey, E. L. (1996). [Unpublished report]. Causal analytical modeling via blocked regression analysis (CAMBRA): An introduction with examples. Los Angeles: University of California, Los Angeles. Higher Education Research Institute.

Avraamidou, L. (2020). Science identity as a landscape of becoming: Rethinking recognition and emotions through an intersectionality lens. *Cultural Studies of Science Education, 15*(2), 323–345.

Azmitia, M., Cooper, C. R., & Brown, J. R. (2009). Support and guidance from families, friends, and teachers in Latino early adolescents' math pathways. *The Journal of Early Adolescence, 29*(1), 142–169.

Bailey 2nd, J. A. (2003). Self-image, self-concept, and self-identity revisited. *Journal of the National Medical Association*, *95*(5), 383.

Bandura, A. (1986a). Social foundations of thought and action-A social cognitive theory. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1986b). The explanatory and predictive scope of self-efficacy theory. *Journal of social and clinical psychology, 4*(3), 359.

Baumeister, R. F. (1997). Identity, self-concept, and self-esteem: The self lost and found. In *Handbook of personality psychology* (pp. 681–710). Academic Press.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191.

Barr, V. (2018). Different denominators, different results: reanalyzing CS degrees by gender, race, and ethnicity. *ACM Inroads*, *9*(3), 40–47.

Barker, L. J., Garvin-Doxas, K., & Jackson, M. (2002). The defensive climate in the Computer Science classroom. In *Proceedings of the 33rd Technical Symposium on Computer Science Education*, pp. 43–47. New York, NY: ACM Press.

Barrow, M., Thomas, S., & Alvarado, C. (2016, July). ERSP: A structured CS research program for early-college students. In *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education* (pp. 148–153).

Berk, L. (2015). Child development. Pearson Higher Education AU.

Bertozzi, E. & Lee, S. (2007). Not Just Fun and Games: Digital Play, Gender and Attitudes Towards Technology. *Women's Studies in Communication, 30*(2), 179–204.

Beyer, S. (2014). Why are women underrepresented in Computer Science? Gender differences in stereotypes, self-efficacy, values, and interests and predictors of future CS course-taking and grades. *Computer Science Education, 24*(2–3), 153–192.

Blaney, J. M. (2020, February). Broadening participation in computing: The role of upward transfer. In *Proceedings of the 51st ACM technical symposium on computer science education* (pp. 254–260).

Blaney, J. M., & Barrett, J. (2022). Advancing gender equity and sense of belonging in computing: Are documented best practices representative of upward transfer students? *Community College Journal of Research and Practice, 46*(9), 633–653.

Blaney, J. M., Barrett, J., & Choi, Y. H. (2022). Diversifying STEM pathways: A look into upward transfer students' sense of belonging in computing. *New Directions for Community Colleges, 2022*(198), 63–75.

Blaney, J. M., Hernandez, T. E., Feldon, D. F., & Wofford, A. M. (2024). Transfer Student Receptivity in Patriarchal STEM Contexts: Evidence of Gendered Transfer Student Stigma in Computer Science From a Mixed Methods Study. *Community College Review*, 00915521231218233.

Blaney, J. M., Sax, L. J., & Chang, C. Y. (2019). Incentivizing longitudinal survey research: The impact of mixing guaranteed and non-guaranteed incentives on survey response. *The Review of Higher Education*, *43*(2), 581–601.

Blaney, J. M., & Stout, J. G. (2017, March). 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* (pp. 69–74).

Blum, L., & Frieze, C. (2005). The evolving culture of computing: Similarity is the difference. *Frontiers: A Journal of Women Studies, 26*(1), 110–125.

Bolkan, S. (2017). Development and validation of the clarity indicators scale. *Communication Education*, *66*(1), 19–36.

BPCnet Resource Portal [BPCnet] (n.d.). Diversity drives innovation-Take action to broaden participation in computing (BPC). https://bpcnet.org/

Brickhouse, N. W., & Potter, J. T. (2001). Young women's scientific identity formation in an urban context. Journal of Research in Science Teaching: *The Official Journal of the National Association for Research in Science Teaching, 38*(8), 965–980.

Brickhouse, N. W., Lowery, P., & Schultz, K. (2000). What kind of a girl does science? The construction of school science identities. *Journal of Research in Science Teaching: The*

Official Journal of the National Association for Research in Science Teaching, 37(5), 441–458.

Brown, B. A. (2004). Discursive identity: Assimilation into the culture of science and its implications for minority students. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 41*(8), 810–834.

Carlone, H. B., & Johnson, A. (2007). 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, 44*(8), 1187–1218.

Carter, D. F., & Hurtado, S. (2007). Bridging key research dilemmas: Quantitative research using a critical eye. In *New Directions for Institutional Research*, pp. 25–35.

Chang, M. J., Eagan, M. K., Lin, M. H., & Hurtado, S. (2011). Considering the impact of racial stigmas and science identity: Persistence among biomedical and behavioral science aspirants. *The Journal of Higher Education*, *82*(5), 564–596.

Chang, M. J., Sharkness, J., Hurtado, S., & Newman, C. B. (2014). What matters in college for retaining aspiring scientists and engineers from underrepresented racial groups. *Journal of Research in Science Teaching, 51*(5), 555–580.

Charney, J., Hmelo-Silver, C. E., Sofer, W., Neigeborn, L., Coletta, S., & Nemeroff, M. (2007). Cognitive apprenticeship in science through immersion in laboratory practices. *International Journal of Science Education, 29*(2), 195–213.

Chemers, M. *Unpublished manuscript.* University of California; Santa Cruz: 2006. Science identity and self-efficacy.

Chemers, M. M., Zurbriggen, E. L., Syed, M., Goza, B. K., & Bearman, S. (2011). The role of efficacy and identity in science career commitment among underrepresented minority students. *Journal of Social Issues, 67*(3), 469–491.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling: a multidisciplinary journal, 14*(3), 464–504.

Chen, C., Sonnert, G., & Sadler, P. M. (2020). The effect of first high school science teacher's gender and gender matching on students' science identity in college. *Science Education, 104*(1), 75–99.

Chen, S., Binning, K. R., Manke, K. J., Brady, S. T., McGreevy, E. M., Betancur, L., ... & Kaufmann, N. (2021). Am I a science person? A strong science identity bolsters minority students' sense of belonging and performance in college. *Personality and Social Psychology Bulletin*, *47*(4), 593–606.

Cheryan, S., Master, A., & Meltzoff, A. N. (2015). Cultural stereotypes as gatekeepers: Increasing girls' interest in computer science and engineering by diversifying stereotypes. *Frontiers in psychology*, *6*, 1–8.

Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: how stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology, 97*(6), 1045.

Chickering, A. W. (1969). *Education and identity.* Jossey-Bass Inc.

Chickering, A. W., & Reisser, L. (1993). *Education and Identity. The Jossey-Bass Higher and Adult Education Series.* Jossey-Bass Inc.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohoon, J. M., & Aspray, W. (2006). A critical review of the research on women's participation in postsecondary computing education. The MIT Press.

Collins, P. H. (1986). Learning from the outsider within: The sociological significance of Black feminist thought. *Social Problems, 33*(6), s14–s32.

Compeau, D. R., & Higgins, C. A. (1995). Computer Self-Efficacy: Development of a Measure and Initial Test. *MIS Quarterly, 19*(2), 189–211.

Computing Research Association [CRA] (2017). Generation CS: Computer Science Undergraduate Enrollments Surge Since 2006. https://cra.org/data/Generation-CS/

Corbett, C., & Hill, C. (2015). Solving the equation: The variables for women's success in engineering and computing. American Association of University Women. 1111 Sixteenth Street NW, Washington, DC 20036.

Crenshaw, K. W. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review, 43*(6), 1241–99.

Daily, S., Sadler, C., & Washington, A. (2022, August). Work in Progress: A Novel Professional Development Program for Addressing Systemic Barriers to Computing Participation. In *2022 ASEE Annual Conference & Exposition*.

Daily, S. B., & Eugene, W. (2013). Preparing the future STEM workforce for diverse environments. *Urban Education*, *48*(5), 682–704.

Denning, P. J. (2005). Is computer science science?. *Communications of the ACM*, *48*(4), 27–31.

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. Sage publications.

Diekman, A. B., Brown, E. R., Johnston, A.M., & Clark, E. K. (2010). Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers. *Psychological Science, 21*, pp. 1051–1057.

Diekman, A. B., Clark, E. K., Johnston, A.M., Brown, E. R., & Steinberg, M. (2011). Malleability in communal goals and beliefs influence attraction to STEM careers: Evidence for a goal congruity perspective. *Journal of Personality and Social Psychology,* 101, 902–918.

Eagan, M. K. (2020). Week 3 lab handout: Factor analysis what why how [Class handout]. University of California, Los Angeles, Los Angeles, EDUC 221.

Eagan Jr., M. K., Hurtado, S., Chang, M. J., Garcia, G. A., Herrera, F. A., & Garibay, J. C. (2013). Making a difference in science education: The impact of undergraduate research programs. *American Educational Research Journal, 50*(4), 683–713.

Ensmenger, N. L. (2012). *The computer boys take over: Computers, programmers, and the politics of technical expertise*. MIT Press.

Erikson, E. H. (1968). Identity: Youth and crisis (No. 7). WW Norton & Company.

Erikson, E. H. (Ed.). (1963). Youth: Change and challenge. Basic books.

Erikson, E. H., Paul, I. H., Heider, F., Gardner, R. W., & Klein, G. S. (1959). *Psychological issues: Identity and the life cycle; studies in remembering; on perception and event structure and the psychological environment; cognitive control.* International Universities Press, Incorporated.

Espinosa, L. (2011). Pipelines and pathways: Women of color in undergraduate STEM majors and the college experiences that contribute to persistence. *Harvard Educational Review, 81*(2), 209–241.

Estrada, M., Woodcock, A., Hernandez, P. R., & Schultz, P. W. (2011). Toward a model of social influence that explains minority student integration into the scientific community. *Journal of Educational Psychology, 103*(1), 206.

Fabrigar, L. R., Wegner, D. T., MaCallum, R. C., & Strahan, E. J. (1999). Evaluation the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299.

Festinger, L. (1962). A theory of cognitive dissonance (Vol. 2). Stanford university press.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.

Flynn, J. (2022). 25 trending tech industry statistics [2023]: The state of the U.S. Tech Industry. Zippia. https://www.zippia.com/advice/tech-industry-statistics/

Gallup-Amazon. (2021). Role models matter: How role models influence career awareness and attainment. Gallup. file:///Users/katiestormes/Desktop/Gallup-Amazon-Role-Models-Report.pdf.

Garvey, J. (2020). Quantitative criticalism and social identity items in survey design. Association for the Study of Higher Education Methods Workshop Series #2 (via Zoom).

Gee, J. P. (1999). *An introduction to discourse analysis: Theory and method.* New York: Routledge.

Gee, J. P. (2000). Chapter 3: Identity as an analytic lens for research in education. *Review of Research in Education, 25*(1), 99–125.

Geiser, C. (2022). Multigroup CFA in Mplus: Automated invariance testing [Video]. YouTube. https://www.youtube.com/watch?v=0vVv-TiJp-4&t=107s

Geiser, C. (2023). Automated longitudinal invariance testing in Mplus [Video]. YouTube. https://www.youtube.com/watch?v=VyQf2jn3Hjo

George, K. L., Sax, L. J., Wofford, A. M., & Sundar, S. (2022). The Tech Trajectory: Examining the Role of College Environments in Shaping Students' Interest in Computing Careers. *Research in Higher Education*, 1–28.

Graham, M. J., Frederick, J., Byars-Winston, A., Hunter, A. B., & Handelsman, J. (2013). Increasing persistence of college students in STEM. *Science, 341*(6153), 1455–1456.

Godwin, A., Potvin, G., Hazari, Z., & Lock, R. (2016). Identity, critical agency, and engineering: An affective model for predicting engineering as a career choice. *Journal of engineering Education*, *105*(2), 312–340.

Goodboy, A. K., & Martin, M. M. (2020). Omega over alpha for reliability estimation of unidimensional communication measures. *Annals of the International Communication Association*, *44*(4), 422–439.

Große-Bölting, G., Gerstenberger, D., Gildehaus, L., Mühling, A., & Schulte, C. (2023). Identity in higher computer education research: a systematic literature review. *ACM Transactions on Computing Education*, *23*(3), 1–35.

Hall, R. M., & Sandler, B. R. (1982). The Classroom Climate: A Chilly One for Women?.

Harper, S. R. (2012). Race without racism: How higher education researchers minimize racist institutional norms. *The Review of Higher Education, 36,* pp. 9–29.

Hazari, Z., Sadler, P. M., & Sonnert, G. (2013). The science identity of college students: Exploring the intersection of gender, race, and ethnicity. *Journal of College Science Teaching, 42*(5), 82–91.

Hazari, Z., Sonnert, G., Sadler, P. M., & Shanahan, M. C. (2010). Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study. *Journal of research in science teaching*, *47*(8), 978–1003.

Henn, S. (Host). (2014, October 21). When women stopped coding [Audio podcast episode]. Planet Money: The Economy Explained. National Public Radio. https://www.npr.org/sections/money/2014/10/21/357629765/when-women-stopped-coding

Higher Education Research Institute [HERI]. (2017). CIRP Construct Technical Report: 2016–17 Appendix: Construct Parameters. Higher Education Research Institute. https://www.heri.ucla.edu/PDFs/constructs/Appendix2017.pdf

Hill, C., Corbett, C., & St Rose, A. (2010). Why so few? Women in science, technology, engineering, and mathematics. American Association of University Women. 1111 Sixteenth Street NW, Washington, DC 20036.

Holland, D., & Lave, J. (2001). History in person: Enduring struggles, contentious practice, intimate identities. Santa Fe, NM: School of American Research Press.

Howard, T. C. (2019). Why race and culture matter in schools: Closing the achievement gap in America's classrooms. Teachers College Press.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6, 1–55.*

Hughes, R. (2014). The evolution of the chilly climate for women in science. *Girls and women in STEM: A never ending story*, 71–94.

Hunter, A. B., Laursen, S. L., & Seymour, E. (2007). Becoming a scientist: The role of undergraduate research in students' cognitive, personal, and professional development. *Science education, 91*(1), 36–74.

Jhally, S., & Kilbourne, J. (Eds.). (2010). *Killing us softly 4: Advertising's image of women* [Film]. Media Education Foundation.

Johnson, A. C. (2007). Unintended consequences: How science professors discourage women of color. *Science Education, 91*(5), 805–821.

Jones, S. R., & McEwen, M. K. (2000). A conceptual model of multiple dimensions of identity. *Journal of College Student Development, 41*(4), 405.

Kapor Center (2023). *Culturally responsive-sustaining computer science education: A Framework.* https://www.kaporcenter.org/culturally-responsive-sustaining-computer-science-education-a-framework/

Kapoor, A., & Gardner-McCune, C. (2019, July). Understanding CS undergraduate students' professional identity through the lens of their professional development. In *Proceedings of the 2019 ACM conference on innovation and technology in computer science education* (pp. 9–15).

Kvasny, L., Joshi, K. D., & Trauth, E. (2011). The influence of self-efficacy, gender stereotypes and the importance of IT skills on college students' intentions to pursue IT careers. In *Proceedings of the 2011 iConference* (pp. 508–513).

Kline, R. B. (2023). *Principles and practice of structural equation modeling.* New York, NY, US.

Ladson-Billings, G. (2021, July). Three decades of culturally relevant, responsive, & sustaining pedagogy: What lies ahead?. In *The Educational Forum* (Vol. 85, No. 4, pp. 351–354). Routledge.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for *t*-tests and ANOVAs. *Frontiers in psychology*, *4*, 62627.

Lane, A., Mekonnen, R., Jang, C., Chen, P., & Lewis, C. M. (2021, March). 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* (pp. 816–822).

Lanza, S. T., Bray, B.C., & Collins, L.M. (2013). An introduction to latent class and latent transition analysis. In Schinka, J.A., Velicer, W. F. & Weiner, B. (Eds.), *Handbook of psychology* (2nd ed), 2, 691–716. Hoboken, NJ: Wiley.

Lapan, J. (November 16, 2023). *CS in the "College of Arts and Crafts": Discourses of Gender and Prestige in Computing*. Association for the Study of Higher Education's 2023 General Conference, Minneapolis, MN, United States.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge university press.

Lawrence-Lightfoot, S., & Davis, J. H. (2002). *The art and science of portraiture*. John Wiley & Sons.

Le, P. T., Doughty, L., Thompson, A. N., & Hartley, L. M. (2019). Investigating undergraduate biology students' science identity production. *CBE—Life Sciences Education*, *18*(4), ar50.

Leary, M. R., Kelly, K. M., Cottrell, C. A., & Schreindorfer, L. S. (2013). Construct validity of the need to belong scale: Mapping the nomological network. *Journal of personality assessment*, *95*(6), 610–624.

Lehman, K. J. (2017). Courting the Uncommitted: A Mixed-Methods Study of Undecided Students in Introductory Computer Science Courses (Publication No. 10289323)

[Doctoral dissertation, University of California, Los Angeles]. ProQuest Dissertation and Theses Global.

Lehman, K. J., Newhouse, K. N., Sundar, S., & Sax, L. J. (2022). Nevertheless, They Persisted: Factors that Promote Persistence for Women and Racially/Ethnically Minoritized Students in Undergraduate Computing. *Computer Science Education*, 1–26.

Lehman, K. J., Wofford, A. M., Sendowski, M., Newhouse, K. N., & Sax, L. J. (2020, February). Better Late Than Never: Exploring Students' Pathways to Computing in Later Stages of College. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (pp. 1075–1081).

Lent, R. W., Lopez, F. G., Sheu, H.-B., & Lopez, A. M. (2011). Social cognitive predictors of the interests and choices of computing majors: Applicability to underrepresented students. *Journal of Vocational Behavior, 78*(2), 184–192. https://doi.org/10.1016/j.jvb.2010.10.006

Lewis, C., Bruno, P., Raygoza, J., & Wang, J. (2019, July). Alignment of goals and perceptions of computing predicts students' sense of belonging in computing. In *Proceedings of the 2019 ACM Conference on International Computing Education Research* (pp. 11–19).

Lin, G. Y. (2016). Self-efficacy beliefs and their sources in undergraduate computing disciplines: An examination of gender and persistence. *Journal of Educational Computing Research, 53*(4), 540–561.

Litson, K., & Feldon, D. (2021). Considering Comparability and Change over Time: Measurement Invariance as an Essential Consideration in Theory Development and Testing. In *Theory and Method in Higher Education Research* (pp. 93–111). Emerald Publishing Limited.

Lockard, C. B., & Wolf, M. (2012, January). Employment outlook: 2010–2020. Occupational employment projections to 2020. *Monthly Labor Review*, 84–108.

Lunn, S., Ross, M., Hazari, Z., Weiss, M. A., Georgiopoulos, M., & Christensen, K. (2021a). How do educational experiences predict computing identity? *ACM Transactions on Computing Education (TOCE), 22*(2), 1–28.

Lunn, S., Ross, M., Hazari, Z., Weiss, M. A., Georgiopoulos, M., & Christensen, K. (2021b). The impact of technical interviews, and other professional and cultural experiences on students' computing identity. In *Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education V. 1* (pp. 415–421).

Lunn, S., Zahedi, L., Ross, M., & Ohland, M. (2021c). Exploration of intersectionality and computer science demographics: Understanding the historical context of shifts in participation. *ACM Transactions on Computing Education (TOCE), 21*(2), 1–30.

Mahadeo, J., Hazari, Z., & Potvin, G. (2020). Developing a computing identity framework: Understanding computer science and information technology career choice. *ACM Transactions on Computing Education (TOCE), 20*(1), 1–14.

Marcoulides, K. M., & Yuan, K. (2017). New ways to evaluate goodness of fit: A note on using equivalence testing to assess structural equation models. *Structural Equation Modeling,* 24, 148–153. doi:10.1080/10705511.2016.1225260.

Marakas, G. M., Yi, M. Y., & Johnson, R. D. (1998). The multilevel and multifaceted character of computer self-efficacy: Toward clarification of the construct and an integrative framework for research. *Information Systems Research, 9*(2), 126–163.

Marcia, J. E. (1966). Development and validation of ego-identity status. *Journal of Personality and Social Psychology, 3*(5), 551.

Marcia, J. E. (1980). Identity in adolescence. *Handbook of Adolescent Psychology, 9*(11), 159–187.

Margolis, J., & Fisher, A. (2003). Unlocking the clubhouse: Women in computing. MIT press.

Margolis, J., Fisher, A., & Miller, F. (2000). The anatomy of interest: Women in undergraduate computer science. *Women's Studies Quarterly*, *28*(1/2), 104–127.

Martin, C. D. (2004). Draw a computer scientist. *ACM SIGCSE Bulletin*, *36*(4), 11–12.

Master, A., Meltzoff, A. N., & Cheryan, S. (2021). Gender stereotypes about interests start early and cause gender disparities in computer science and engineering. *Proceedings of the National Academy of Sciences*, *118*(48), e2100030118.

Mayhew, M. J., Rockenbach, A. N., Bowman, N. A., Seifert, T. A., & Wolniak, G. C. (2016). *How college affects students: 21st century evidence that higher education works* (Vol. 1). John Wiley & Sons.

Mayhew, J. M., & Simonoff, J. S. (2015). Effect coding as a mechanism for improving the accuracy of measuring students who self-identify with more than one race, *Research Higher Education, 56*(1), 595–600.

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological methods*, *12*(1), 23.

McGill, M. M., Decker, A., McKlin, T., & Haynie, K. (2019, February). A gap analysis of noncognitive constructs in evaluation instruments designed for computing education. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (pp. 706–712).

McMullen, K., & Waisome, J. (2019-present). *Modern figures* [Audio podcast]. Modern Figures Podcast. https://modernfigurespodcast.com/

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological methods*, *23*(3), 412.

Meade, A. W. (2005, April). Sample size and tests of measurement invariance. In *Annual Conference of the Society for Industrial and Organizational Psychology, Los Angeles, CA*.

Miller, J. E. (2013). The Chicago guide to writing about multivariate analysis. University of Chicago Press.

Miller, D. I., Nolla, K. M., Eagly, A. H., & Uttal, D. H. (2018). The development of children's gender-science stereotypes: A meta-analysis of 5 decades of US Draw-a-Scientist studies. *Child Development*, *89*(6), 1943–1955.

Momentum. (2021). Baseline Computing Education Survey. [Unpublished survey instrument]. University of California, Los Angeles.

Momentum. (2023). Computing Education Survey Follow-up #1. [Unpublished survey instrument]. University of California, Los Angeles.

Muthén, B. O., & Muthén, B. O. (2017). *Mplus user's guide (version 8).* Los Angeles, CA: Muthén & Muthén.

Myers, T. A. (2011). Goodbye, listwise deletion: presenting hot deck imputation as an easy and effective tool for handling missing data. *Communication Methods and Measures, 5*(4), 297–310.

National Center for Education Statistics [NCES] (2018). Digest of Education Statistics: Bachelor's Degrees Conferred by Postsecondary Institutions, by Race/Ethnicity and Sex of Student: Selected Years, 1976–77 through 2018–2019. https://nces.ed.gov/programs/digest/d21/tables/dt21_322.20.asp

National Center for Education Statistics [NCES] (2021). Students' internet access before and during the coronavirus pandemic by household socioeconomic status. https://nces.ed.gov/blogs/nces/post/students-internet-access-before-and-during-the-coronavirus-pandemic-by-household-socioeconomic-status

National Center for Science and Engineering Statistics [NCSES] (2020). Chart Builder: IPEDS Completions Survey from Department of Education 1997–2020. https://ncsesdata.nsf.gov/builder/ipeds_c?type=chart

National Center for Science and Engineering Statistics [NCSES] (2021). Women, Minorities, and Persons with Disabilities in Science and Engineering Report 2021. https://ncses.nsf.gov/pubs/nsf21321/report/field-of-degree-women

NCWIT (n.d.a). *Counselors for computing (C4C) resource booklet: National Center for Women & Information Technology*. ncwit.org. https://ncwit.org/resource/c4cbooklet/

NCWIT (n.d.b). *Guide to inclusive computer science education: How educators can encourage and engage all students in Computer Science: National Center for Women & Information Technology*. ncwit.org. https://ncwit.org/resource/csedguide/

Noble, S. U. (2018). *Algorithms of oppression.* New York University Press.

Ong, M. (2011). The status of women of color in computer science. *Communications of the ACM*, *54*(7), 32–34.

Ong, M., Wright, C., Espinosa, L., & Orfield, G. (2011). 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*, *81*(2), 172–209.

Peckham, J., Harlow, L. L., Stuart, D. A., Silver, B., Mederer, H., & Stephenson, P. D. (2007). Broadening participation in computing: issues and challenges. *ACM SIGCSE Bulletin, 39*(3), 9–13.

Peng, A., Nushi, B., Kıcıman, E., Inkpen, K., Suri, S., & Kamar, E. (2019, October). What you see is what you get? the impact of representation criteria on human bias in hiring. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* (Vol. 7, pp. 125–134).

Peters, A. K. (2014). *The role of students' identity development in higher education in computing* (Doctoral dissertation, Uppsala University).

Peters, A. K. (2018). Students' experience of participation in a discipline—a longitudinal study of computer science and IT engineering students. *ACM Transactions on Computing Education (TOCE)*, *19*(1), 1–28.

Peters, A. K., & Pears, A. (2013, March). Engagement in Computer Science and IT—What! A Matter of Identity?. In *2013 Learning and Teaching in Computing and Engineering* (pp. 114–121). IEEE.

Perez, T., Cromley, J. G., & Kaplan, A. (2014). The role of identity development, values, and costs in college STEM retention. *Journal of educational psychology*, *106*(1), 315.

PremalMatalia. (2020, March 31). *What are the assumptions of linear regression regarding residuals?* https://medium.com/@premal.matalia/what-are-the-assumptions-of-linear-regression-regarding-residuals-19d92dae249a#:~:text=d)%20Independent%20error%20assumption%3A%20It,or%20among%20the%20residuals%20themselves.assumption

Public Broadcasting Service [PBS]. (2021, April 14). Nova: Picture a scientist [Film]. PBS. https://www.pbs.org/video/picture-a-scientist-rlnmdy/

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental review*, *41*, 71–90.

Robnett, R. (2013). The role of peer support for girls and women in the STEM pipeline: Promoting identification with STEM and mitigating the negative effects of sexism. University of California, Santa Cruz.

Robnett, R. D., Chemers, M. M., & Zurbriggen, E. L. (2015). Longitudinal associations among undergraduates' research experience, self-efficacy, and identity. *Journal of Research in Science Teaching, 52*(6), 847–867.

Rodriguez, S. L., & Kaur, A. (November 17, 2023). Understanding Latina Computing Identity Development at a Hispanic-serving Community College. Association for the Study of Higher Education's 2023 General Conference, Minneapolis, MN, United States.

Rodriguez, S. L., & Lehman, K. (2017). Developing the next generation of diverse computer scientists: the need for enhanced, intersectional computing identity theory. *Computer Science Education, 27*(3–4), 229–247.

Rodriguez, S. L., Lu, C., & Ramirez, D. (2020). Creating a conceptual framework for computing identity development for Latina undergraduate students. In *An Asset-Based Approach to Advancing Latina Students in STEM* (pp. 25–39). Routledge.

Rodriguez, S., Ramirez, D., Lehman, K. J., & Sax, L. J. (2023). Utilizing community cultural wealth to explore the experiences of Latina undergraduate students in computing. *Journal* *of Women and Minorities in Science and Engineering, 29*(3), 1–24. 10.1615/JWomenMinorScienEng.2022039365.

Rollins, S. N., Joshi, A., Apedoe, X., Engle, S., Malensek, M., & Bruno, G. (2021, July). Understanding Professional Identity Development Among Computer Science Students. In 2021 ASEE Virtual Annual Conference Content Access.

Rubin, V. (1996). Bright galaxies dark matters (masters of modern physics). New York, NY: Springer-Verlang New York.

Sax, L. J. (2008). *The gender gap in college: Maximizing the developmental potential of women and men.* Jossey-Bass/Wiley.

Sax, L. J., Blaney, J. M., Lehman, K. J., Rodriguez, S. L., George, K. L., & Zavala, C. (2018). Sense of belonging in computing: The role of introductory courses for women and underrepresented minority students. *Social Sciences, 7*(8), 122.

Sax, L. J., & Harper, C. E. (2007). Origins of the gender gap: Pre-college and college influences on differences between men and women. *Research in Higher Education*, *48*, 669–694.

Sax, L. J., Lehman, K. J., Jacobs, J. A., Kanny, M. A., Lim, G., Monje-Paulson, L., & Zimmerman, H. B. (2017). Anatomy of an enduring gender gap: The evolution of women's participation in computer science. *The Journal of Higher Education*, *88*(2), 258–293.

Sax, L. J., & Newhouse, K. N. S. (2019). Disciplinary field specificity and variation in the STEM gender gap. In J. Gaston-Gayles & L. Perez-Felkner (Eds.), Advancing higher education research on undergraduate women in STEM. *New Directions for Institutional Research* (pp. 45–71). Malden, MA: Wiley.

Sax, L. J., Newhouse, K. N., Goode, J., Nakajima, T. M., Skorodinsky, M., & Sendowski, M. (2022). Can computing be diversified on "principles" alone? Exploring the role of AP computer science courses in students' major and career intentions. *ACM Transactions on Computing Education (TOCE), 22*(2), 1–26.

Scott, A., Lopez, A. & Bonilla, S. (2022, May 12). Computer science is everywhere in our world but almost nowhere in California schools. EdSource. https://edsource.org/2022/computer-science-is-everywhere-in-our-world-but-almost-nowhere-in-california-schools/671985#:~:text=Nearly%20two%2Dthirds%20of%20California,needed%20to%20support%20our%20students

Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving* (Vol. 34). Westview Press, Boulder, CO.

Seymour, E., Hunter, A. B., Laursen, S. L., & DeAntoni, T. (2004). Establishing the benefits of research experiences for undergraduates in the sciences: First findings from a three-year study. *Science Education, 88*(4), 493–534.

Shapiro, C. A., & Sax, L. J. (2011). Major selection and persistence for women in STEM. *New Directions for Institutional Research*, *2011*(152), 5–18.

Shetterly, M. L. (2016). Hidden figures. New York: NY: HarperCollins.

Sherman, A. M., & Zurbriggen, E. L. (2014). "Boys can be anything": Effect of Barbie play on girls' career cognitions. *Sex Roles, 70*(5), 195–208.

Simon, R. M., Wagner, A., & Killion, B. (2017). Gender and choosing a STEM major in college: Femininity, masculinity, chilly climate, and occupational values. *Journal of Research in Science Teaching, 54*(3), 299–323.

Smith, D. G., Morrison, D. E., & Wolf, L. E. (1994). College as a Gendered Experience: An Empirical Analysis Using Multiple Lenses. *The Journal of Higher Education, 65*(6), 696–725.

Stage, F. K., & Wells, R. S. (2014). Critical quantitative inquiry in context. In *New directions for Institutional Research, 2013*(158), 1–7.

Strayhorn, T. L. (2012). Exploring the impact of Facebook and Myspace use on first-year students' sense of belonging and persistence decisions. *Journal of College Student Development*, *53*(6), 783–796.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American psychologist, 52*(6), 613.

Stets, J. E., Brenner, P. S., Burke, P. J., & Serpe, R. T. (2017). The science identity and entering a science occupation. *Social Science Research*, *64*, 1–14.

Stewart, D. L., & Nicolazzo, Z. (2018). High impact of [whiteness] on trans* students in postsecondary education. *Equity & Excellence in Education*, *51*(2), 132–145.

Stout, J., Tamer, N., & Alvarado, C. (2018). Formal Research Experiences for First Year Students: A Key to Greater Diversity in Computing? In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education* (pp. 693–698).

Streiner, D. L. (1994). Figuring out factors: the use and misuse of factor analysis. *The Canadian Journal of Psychiatry, 39*(3), 135–140.

Tabachnick, B. G., & Fidel, L. S. (2007). *Using multivariate statistics* (5th ed.) Allyn & Bacon/Pearson Education.

Taheri, M., Ross, M. S., Hazari, Z., Weis, M. A., Georgiopoulos, M., Christensen, K., Solis, T., Chari, D., & Taheri, Z. (2019). Exploring computing identity and persistence across multiple groups using structural equation modeling. In *American Society for Engineering Education (ASEE) Conference Proceedings*. American Society for Engineering Education.

Taheri, M., Ross, M. S., Hazari, Z., Weis, M. A., Georgiopoulos, M., Christensen, K., Solis, T., Garcia, A., & Chari, D. (2018). A structural equation model analysis of computing identity subconstructs and student academic persistence. *In 2018 IEEE Frontiers in Education Conference (FIE)* (pp. 1–7) IEEE.

Tajfel, H. (1979). Individuals and groups in social psychology. *British Journal of Social & Clinical Psychology, 18*(2), 183–190.

Tate, E. D., & Linn, M. C. (2005). How does identity shape the experiences of women of color engineering students?. *Journal of Science Education and Technology, 14*(5), 483–493.

Teo, T., & Koh, J. H. L. (2010). Assessing the dimensionality of computer self-efficacy among pre-service teachers in Singapore: A structural equation modeling approach. *International Journal of Education and Development using ICT*, *6*(3), 7–18.

The Carnegie Classification of Institutions of Higher Education (n.d.) Classification descriptions. Retrieved from: http://carnegieclassifications.iu.edu/descriptions/basic.php

Thompson, C. (2019). The secret history of women in coding: Feature. The New York Times Magazine. https://www.nytimes.com/2019/02/13/magazine/women-coding-computer-programming.html

Tung, L. (2016). Google: This is the scale of race, gender problem in computer science education. https://www.zdnet.com/article/google-this-is-the-scale-of-race-gender-problem-in-computer-science-education/

U.S. Bureau of Labor Statistics [BLS]. (2021). *Labor force characteristics by race and ethnicity, 2021*. U.S. Bureau of Labor Statistics. https://www.bls.gov/opub/reports/race-and-ethnicity/2021/home.htm

U.S. Bureau of Labor Statistics [BLS]. (2022a). Occupational Outlook Handbook*, Field of degree: Computer and Information Technology*. U.S. Bureau of Labor Statistics. https://www.bls.gov/ooh/field-of-degree/computer-and-information/computer-and-information-technology-field-of-degree.htm

U.S. Bureau of Labor Statistics [BLS]. (2022b). Computer and Information Technology Occupations. U.S. Bureau of Labor Statistics. https://www.bls.gov/ooh/computer-and-information-technology/home.htm

Veilleux, N., Bates, R., Jones, D., Allendoerfer, C., & Crawford, J. (2012, February). The role of belonging in engagement, retention and persistence in computer science. *In Proceedings of the 43rd ACM technical symposium on Computer Science Education* (pp. 707–707).

Vincent-Ruz, P., & Schunn, C. D. (2018). The nature of science identity and its role as the driver of student choices. *International journal of STEM education*, *5*, 1–12.

Walker, H. M. (2017). ACM RETENTION COMMITTEE Retention of students in introductory computing courses: curricular issues and approaches. ACM Inroads, 8(4), 14–16.

Washington, N., Barnes, T., Payton, J., Dunton, S., Stukes, F., & Peterfreund, A. (2019). RESPECT 2019: Yes, we still need to talk about diversity in computing. *Computing in Science & Engineering*, *21*(1), 79–83.

Wells, C. S. (2021). *Assessing Measurement Invariance for Applied Research.* New York: Cambridge University Press.

Wells, R. S., & Stage, F. K. (2015). Past, present, and future of critical quantitative research in higher education. *New Directions for Institutional Research*, *2014*(163), 103–112.

Werner, S. M., & Chen, Y. (2024, March). Evaluating Identity and Belonging in Computer Science Students: Instrument Adaptation and Analysis. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1* (pp. 1421–1427). https://dl.acm.org/doi/pdf/10.1145/3626252.3630840?casa_token=oIl8xPFqyykAAAAA:GVNai3_hTOjVa7pbLSe0C69QEjhFMWOYYXpWXdZxUOTLT-tYoatG-XT_XqfL9oXMKSPhfTZ0LWci

Whitney, T. (2021). Women and People of Color still experience difficulty advancing in technology [report]. https://mcusercontent.com/5e8b536fcd530e98bd1cfc24a/files/ed1dc73c-b98a-d51c-3eea-f3f05fba7c4a/TelleWhitney_2021_SurveyReport.01.pdf

Wilson, R. E., & Kittleson, J. (2013). Science as a classed and gendered endeavor: Persistence of two white female first-generation college students within an undergraduate science context. *Journal of Research in Science Teaching, 50*(7), 802–825.

Wofford, A. M. (2021). Modeling the pathways to self-confidence for graduate school in computing. *Research in Higher Education*, *62*(3), 359–391.

Wofford, A. M., Sax, L. J., George, K. L., Ramirez, D., & Nhien, C. (2022). Advancing equity in graduate pathways: *Examining the factors that sustain and develop graduate aspirations. The Journal of Higher Education, 93*(1), 110–136.

Wofford, A. M., & Gutzwa, J. A. (2022). Funds of science identity: Toward an asset-based framework for undergraduate STEM research and praxis. *Journal of Women and Minorities in Science and Engineering, 28*(3), 59–81.

Zweben, S., & Bizot, B. (2015). Continued booming undergraduate CS enrollment; doctoral degree production dips slightly. Computing Research Association 2015 Taulbee Survey. https://cra.org/wp-content/uploads/2016/05/2015-Taulbee-Survey.pdf

Zweben, S., & Bizot, B. (2016). Generation CS continues to produce record undergrad enrollment; graduate degree production rises at both master's and doctoral levels. Computing Research Association 2016 Taulbee Survey. https://cra.org/crn/wp-content/uploads/sites/7/2017/05/2016-Taulbee-Survey.pdf

Zweben, S., & Bizot, B. (2017). Another year of record undergrad enrollment; doctoral degree production steady while master's production rises again. Computing Research Association 2017 Taulbee Survey. https://cra.org/wp-content/uploads/2018/05/2017-Taulbee-Survey-Report.pdf

Zweben, S., & Bizot, B. (2018). Undergrad enrollment continues upward; doctoral degree production declines but doctoral enrollment rises. Computing Research Association 2018 Taulbee Survey. https://cra.org/wp-content/uploads/2019/05/2018_Taulbee_Survey.pdf

Zweben, S., & Bizot, B. (2019). Total undergrad CS enrollment rises again, but with fewer new majors; doctoral degree production recovers from last year's dip. Computing Research Association 2019 Taulbee Survey. https://cra.org/wp-content/uploads/2020/05/2019-Taulbee-Survey.pdf

Zweben, S., & Bizot, B. (2020). Bachelor's and doctoral degree production growth continues but new student enrollment shows declines. Computing Research Association 2020 Taulbee Survey. https://cra.org/wp-content/uploads/2021/05/2020-CRA-Taulbee-Survey.pdf

Zweben, S., & Bizot, B. (2021). CS enrollment grows at all degree levels, with increased gender diversity. Computing Research Association 2021 Taulbee Survey. https://cra.org/wp-content/uploads/2022/05/2021-Taulbee-Survey.pdf