CONTRIBUTING FACTORS PROMOTING SUCCESS FOR FEMALES IN COMPUTING:

A COMPARATIVE STUDY

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Bethlehem Abera Gronneberg

In Partial Fulfillment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

Major Department:
Computer Science

May 2022

Fargo, North Dakota

North Dakota State University
Graduate School

Title

CONTRIBUTING FACTORS PROMOTING SUCCESS FOR FEMALES IN COMPUTING: A COMPARATIVE STUDY

By

Bethlehem Abera Gronneberg

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

DOCTOR OF PHILOSOPHY

SUPERVISORY COMMITTEE:

Dr. Kendall Nygard
Chair

Dr. Kenneth Magel

Dr. Pratap Kotala

Dr. Limin Zhang

Approved:

June 9, 2022	Dr. Simone Ludwig
Date	Department Chair

ABSTRACT

Despite the growing global demand for Computer Science (CS) professionals, their high earning potential, and diversified career paths (U.S. BLS 2021, UNESCO 2017), a critical gap exists between enrollment and graduation rates among female students in computing fields across the world (Raigoza 2017, Hailu 2018, UNESCO 2017, Bennedsen and Caspersen 2007). The largest dropout point occurs during the first two years of their CS studies (Giannakos, et al., 2017). The purpose of this parallelly convergent mixed-methods research was to comparatively investigate, describe and analyze factors correlated to the experiences and perceptions of female undergraduates as it relates to their persistence in CS/Software Engineering (SE) degrees, conducted in two public universities in the U.S. & Ethiopia. Anchored in Tinto's theory of retention, the quantitative part of the study examined three possible predictive factors of success for students who were enrolled in the first two CS/SE courses and evaluated differences between genders and institutions on those factors. Pearson's correlation coefficient tests were applied to test the hypothesis that the perceptions of Degree's Usefulness (DU), Previously Acquired Knowledge (PAK) and Cognitive Attitude (CA) correlate to the decision to persist for the research participants. The results showed a statistically significant positive correlation between perceptions of DU, influence of PAK and the decision to persist. Two sample t-tests revealed gender and institutional differences exhibited in the influence of PAK and CA. The qualitative part of the study reported 12 contributing factors of success for graduating class of females in CS/SE using sentiment analysis and topic modelling from the domain of Natural Language Processing (NLP) through the interpretation of auto transcribed interview responses.

Key words: Females, Persistence, Ethiopia, Success Factors, North Dakota, Computer Science, Software Engineering, Gender Differences.

ACKNOWLEDGMENTS

With deepest gratitude to my longtime advisor and mentor, Professor Kendall Nygard, whose guidance was vital in formulating and refining the research questions and methodologies. Special appreciation to my advisory committee, Professor Kenneth Magel, Professor Pratap Kotala, and Professor Limin Zhang for their invaluable inquiries and insights.

I would like to acknowledge the following people for their incredible support with respect to contextualizing and framing the Ethiopian aspect of this comparative analysis research: Professor Afework Kassu, Professor Frew Tegegn, Dr. Elefelious Belay, Dr. Getasew Tadesse, Dr. Meseret Hailu, Dr. Nancy Hafkin, Ms. Fikir Aweke and Dr. Ayenachew Woldegiyorgis.

I would like to extend my thanks to Ms. Sidney Fisk, Senior Analyst at the Office of Institutional Research and Analysis at North Dakota State University for her continued support in providing the necessary data point for this research. Appreciation to Mr. Curt Doetkott whose expertise guided the statistical data analysis part of this mixed-method research.

I would also like to express my appreciation to Dr. Stacy Duffield, Prof. Damian Lampl, and all the Software Engineering undergraduates at Addis Ababa Institute of Technology, Computer Science undergraduates at North Dakota State University and female Computer Science students at Concordia College for the collaborative efforts during the research data collection and for contributing to the successful completion of this dissertation.

I am grateful to my husband, Ron, for walking beside me on this winding journey, and for being my pillar of support. My love and respect for my parents, Woyzero Asamenech Mammo and Ato Abera Tegegn, for their wise counsel, prayers, and unconditional love.

DEDICATION

I dedicate my dissertation to Ron, Joshua, Gabriel, and Nathaniel.

TABLE OF CONTENTS

ABSTRACT iii

ACKNOWLEDGMENTS iv

DEDICATION v

LIST OF TABLES ix

LIST OF FIGURES x

CHAPTER 1. INTRODUCTION 1

1.1. Females in Computing 1

1.1.1. Longitudinal and International Context 1

1.1.2. Contributing Factors for Attrition 6

1.2. Problem Statement 12

1.3. Significance of the Research 14

1.4. Purpose of the Study 15

1.5. Theoretical and Conceptual Framework 16

1.6. Research Questions 18

1.7. Assumptions and Limitation 18

1.7.1. Assumptions 18

1.7.2. Limitations 19

1.8. Dissertation Structure 19

CHAPTER 2. REVIEW OF LITERATURE 21

2.1. Introduction 21

2.2. The Perceptions of Degree’s Usefulness 23

2.3. Pre-Acquired Knowledge 24

2.4. Cognitive Aptitude 25

2.5. Sense of Belonging 27

2.6. Institutional Support 28
2.7. Involvement in Peer Group Activities 30
CHAPTER 3. RESEARCH METHODOLOGY 31
3.1. Introduction 31
3.2. Purpose of the Study 31
3.3. Research Questions 31
3.4. Research Design 32
3.5. Research Hypothesis 34
3.6. Target Institutions 35
3.6.1. Addis Ababa University (AAU) 36
3.6.2. North Dakota State University (NDSU) 38
3.7. Target Population 40
3.8. Recruitment 41
3.8.1. Quantitative Instrument 41
3.8.2. Qualitative Instrument 42
3.9. Data Collection 43
3.9.1. Quantitative Instrument 43
3.9.2. Qualitative Instrument 44
3.10. Data Analysis 45
3.10.1. Quantitative Analysis 46
3.10.2. Qualitative Analysis 49
3.11. Researcher Positionality 56
CHAPTER 4. RESULTS 57
4.1. Introduction 57
4.2. Quantitative Findings 57

4.2.1. Demographics 57

4.2.2. Research Question I: Pearson's r Correlation 59

4.2.3. Research Question II: Independent t-tests 64

4.3. Qualitative Findings 74

4.3.1. NLP Sentiment Analysis 75

4.3.2. NLP Topic Modelling 84

CHAPTER 5. CONCLUSION 99

5.1. Introduction 99

5.2. Interpretation of the Findings 100

5.2.1. Research Question One: Correlations between DU, PAK and CA with Persistence 100

5.2.2. On Research Question Two: Gender or Institutional Differences 102

5.2.3. On Research Question Three: Contributing Factors Promoting Success 105

5.2.4. On Research Question Three: Replicable Factors of Persistence between Institutions 107

5.3. Recommendation 108

5.3.1. NDSU Computer Science Program 108

5.3.2. Addis Ababa Institute of Technology (AAiT) 111

5.4. Limitations of the Study 114

5.5. Future Work 115

REFERENCES 116

APPENDIX A. SURVEY CONSENT FORM 132

APPENDIX B. INTERVIEW CONSENT FORM 133

APPENDIX C. IRB EXEMPT DETERMINATION 134

APPENDIX D. SURVEY INSTRUMENTS 135

APPENDIX E. SCATTER PLOT AND HISTOGRAM 140

LIST OF TABLES

Table Page

1: Pearson’s correlation table for the perception of Degree’s Usefulness (DU) for courses CS I and CS II measured for NDSU and AAU, respectively. 60

2: Pearson’s correlation table for the perception of the influence of PAK for courses CS and CS II measured for NDSU and AAU, respectively.. 62

3: Pearson’s correlation table for the perception of CA for courses CS I and CS II measured for NDSU and AAU, respectively. 63

4: Two-tailed independent t-tests, gender comparison with respect to Degree's Usefulness (DU) for courses CS I and CS II measured for NDSU and AAU, respectively. 65

5: Two-tailed independent t-tests, gender comparison with respect to PAK for courses CS I and CS II measured for NDSU and AAU, respectively. 67

6: Two-tailed independent t-tests, gender comparison with respect to CA for courses CS I and CS II measured for NDSU and AAU, respectively. 69

7: Two-tailed independent t-tests, institutional comparison using two sample t-test for Degree’s Usefulness (DU). 70

8: Two-tailed independent t-tests, institutional comparison using two sample t-test for Previously Acquired Knowledge (PAK)... 71

9: Two-tailed independent t-tests, institutional comparison using two sample t-test for Cognitive Attitude (CA).. 71

10: Paired t-tests, institutional comparison using two sample t-test for Cognitive Attitude (CA). 72

LIST OF FIGURES

Figure Page

1: Percentage of reported CIS bachelor's degrees conferred to women, 1970-2015............. 5

2: Tinto's Institutional Departure Model (Tinto 1993)... 7

3: Conceptual research framework of female students' persistence in Computer Science or Software Engineering.. 17

4: The parallelly convergent mixed-methods research design process used in this study.. 33

5: Text analysis processes for sentiment analysis and topic modelling............................... 50

6: Pearson's correlation plots distribution for the perception of DU, PAK and CA for course CS II as measured for NDSU. 61

7: Pearson's correlation plots distribution for the perception of DU, PAK and CA for course CS II as measured for AAU... 64

8: (a) Distribution of Degree's Usefulness (DU) by gender for CS II at NDSU; (b) Distribution of Degree's Usefulness (DU) by gender for CS II at AAU. 66

9: Pair t-test for pre-/post- comparisons of PAK for CS I as measured for NDSU. 73

10: Pair t-test for pre-/post- comparisons of PAK for CS I as measured for NDSU. 73

11: Pair t-test for pre-/post- comparisons of PAK for CS I as measured for NDSU. 74

12: Mean sentiment score using Sentimentr package in R. 76

13: Frequency distribution of sentiment polarity by gender and by institution...................... 78

14: Sentiment attributes of polarized words & valence shifters. 79

15: Density plot of average sentiment polarity. 80

16: Sentiment polarity distribution by institution. 81

17: Sentiment attribute distribution by gender, by institution. 82

18: Manual qualitative inductive thematic coding analysis of factors of persistence/success in CS at NDSU... 86

19: Manual qualitative inductive thematic coding analysis of factors of persistence/success in CS at AAU. 87

20: Statistical goodness of fit on semantic coherence in CS at NDSU. 89

21: Beta probability distribution of the top 12-topic modeling results for NDSU. 90

22: Beta probability distribution of the top 12-topic modeling results for AAU. 92

CHAPTER 1. INTRODUCTION

1.1. Females in Computing

The purpose of this parallelly convergent mixed-methods research is to comparatively investigate, describe and analyze factors correlated to the experiences and perceptions of female undergraduates as it relates to their persistence in Computer Science (CS)/Software Engineering (SE) degrees, enrolled in two geographically, culturally, and economically distinct public universities in the North Dakota, the United States & Addis Ababa, Ethiopia. As illustrated in the literature review, the issue of retention of females through graduation in computing fields is multifaceted. Comparatively investigating and enumerating the contributing factors promoting success from the global perspective of undergraduate female students in computing is the central focus of this study.

This chapter introduces this study by discussing the nature and extent of the problem from longitudinal historical and international perspectives, presenting the problem statement, outlining the purpose and the significance of the study as well as pausing the research questions. The theoretical and conceptual framework anchoring the research is detailed. The acknowledged assumptions and limitations are listed. And lastly, the roadmap for the rest of the dissertation structure is provided.

1.1.1. Longitudinal and International Context

Science, Technology, Engineering and Mathematics (STEM) fields drive innovative solutions to current and future challenges of the world. The United Nations Educational, Scientific and Cultural Organization (UNESCO) which collects data on worldwide educational activities, reports female students who embark on STEM disciplines eventually dropout without a degree in disproportionate numbers (UNESCO 2017). 110 countries in UNESCO member states who participated in the data collection reported a critical gap between enrollment and graduation rates for female students in computing fields.

In United States, despite a steady increase in the number of CS bachelor's degrees conferred by postsecondary institutions in the past decade, the share of females has stubbornly been hovering on average around 18% (NAS 2016). CS has also witnessed the lowest numbers of female degree recipients compared to the broader fields of STEM degrees (NSF 2019), down from a peak of 37.1% in 1984 (NCES 2012). In what seems like a paradox, there has been a sharp increase in enrollment interest for undergraduate CS degree since 2011 among first-year full-time students, according to a survey data collected by U.S. Higher Education Research Institute (HERI) (Eagan et al., 2016). However, a considerable number of female students who start on a path to study CS never earn a degree in it, choosing to switch to a different major instead. The largest dropout point occurs during the first two years of their CS studies (Giannakos, et al., 2017). This pipeline problem leads to the reported sparse number of women in the software development and services workforce. According to the National Center for Women & Information Technology (NCWIT), women in the U.S. held 57% of all professional occupations in 2018, but only 26% of computing jobs.

There exists a dearth of literature regarding the retention of women in computing majors in Ethiopia. Women are historically scarcely represented in higher education in general (Selma 2010). In recognition of this problem, the Ethiopian government designed and implemented affirmative action in 1998 aimed to promote gender equality in the public universities (Egne 2014). The affirmative action policy enabled females to enroll at universities with a 0.2 grade point average less than that of males. The policy further included provisions for female-only tutoring, guidance, and counselling support during their first academic year; peer counselling and support from senior female students; academic support from capable senior female students; and overall assertiveness training (Demise et al., 2002). However, the translation of policy into action is one of the critical challenges in Ethiopia (Asfaw 2012). A recent study on female's enrollment in Ethiopian public universities in Engineering and Technology reported to have averaged at 23% from 2007 to 2016 (Melak and Singh, 2021).

With the goal of producing more graduates in STEM to support the country's sustainable development agenda, the Ethiopian Ministry of Education, the governing body of the country's public universities, approved a policy that directs 70% of overall university enrolment to be in science and technology fields, with the remaining 30% in the social sciences in 2008. Consequently, the number of universities in the country increased by many folds to support the development initiatives. This policy was not supported by efforts on preparedness and recruitment of secondary school females to STEM fields (Tsegai 2010). (Egne 2014) cites prominent factor deterring participation includes inadequate preliminary knowledge and academic preparation, lack of self-efficacy, the persistent effects of socio-cultural gender stereotypes, the existence of science educational experiences that do not welcome women, and the absence of adequate support systems in the Ethiopian universities. In spite of increased student enrollment, a rise in the number of institutions, affirmative action policy and an increased focus on science and technology in education policy, gender inequalities in STEM fields remain pronounced (Hailu, 2018).

This research has conducted a longitudinal retrospective aggregate enrollment data analysis of seven years of historical data of three public universities in Ethiopia (Bahir Dar University Institute of Technology, Adama Science and Technology University and Addis Ababa University Institute of Technology). 20% of average enrollment of females in Software

Engineering major from 2013-2019 in all three stated universities was observed. Three public universities in North Dakota (North Dakota State University, University of North Dakota, and Dickenson State University). 8-10% of average enrollment of females in Computer Science undergraduates 2005-2017 was observed. The analysis was confirmed the previous findings of the prevalence of female students' attrition in CS/SE degrees (appendix E).

Noteworthy progress has been made in higher education when it comes to the enrolment of female students in general which almost doubled in numbers between 2000 and 2014 in the United States. Globally, women are graduating from tertiary institutes at a higher rate than men (UNESCO 2017). Women earned 57% of bachelor's degrees overall, and 50% of bachelor's degrees in science and engineering in 2016. However, females accounted for only 18% of bachelor's degrees in the computing sciences. That percentage dips even lower for women of color (National Science Foundation, 2016).

Since the establishment of Computer Science as a distinct academic discipline in the United States in the early 1960s, CS bachelor's degree production is marked by significant peaks and valleys of growth. The U.S. Department of Education's IPEDS, the authority that tracks and reports the number of historical conferring of Computer and Information Sciences (CIS) degrees, reports the number of graduates rose from 89 in 1966 to astounding 59,586 in 2015. The two main periods of high growth were in 1986 (at 42,337) and in 2004 (at 57,488), following a pattern of rising enrollments when there were economic downturns. Conversely, the two decline periods were in 1993 (at 24,519) and in 2009 (at 37,992).

The first and only historical surge of women with CIS graduates occurred in 1985 at 37%, two years before the mid-1980s decline in overall annual CIS bachelor's degree production, as illustrated in Figure 1 (NAS 2016). In 1985, women's share decreased abruptly, levelling off during the mid-1990s, then dropping again precipitously beginning in 2004, one year before the 2005 onset of the decline in overall CIS bachelor's degree production during the dotcom bust. After 2008 the share of CIS bachelor's degrees awarded to women leveled off, remaining steady near 18% overall through 2019. This drop in female CS graduates stands in contrast to the steady growth of women earning degrees in other sciences and in engineering. The "incredible shrinking pipeline" (Camp, 2002, p. 129) is becoming more like a stagnant pipeline.

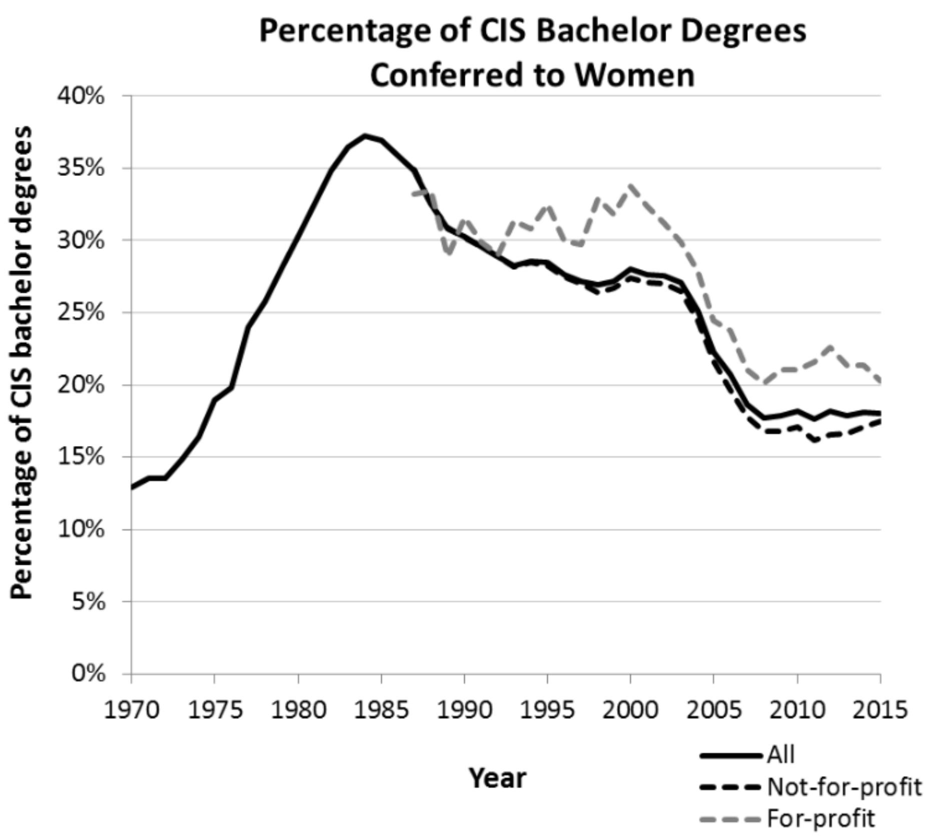


Figure 1: Percentage of reported CIS bachelor's degrees conferred to women, 1970-2015.
Source: Data from IPEDS degree completions survey accessed via the WebCASPAR online database.
Note: Percentage of reported CIS bachelor's degrees conferred to women, 1970-2015, at for-profit (dotted gray line), non-for-profit (dotted black line), and all (solid black line) institutions.

Computing history is full of stories of women who have been pioneering the field of Computer Science for many decades, including but not limited to, a British mathematician named Augusta Ada Lovelace King who published the first set of computer instructions destined for a mechanical general-purpose computing machine (Carlucci Aiello, 2016). Then why didn't the list of women in computing get sizable and expanding over time?

1.1.2. Contributing Factors for Attrition

What transpires during the time when a student enrolls in computer science major and when the same student drops out of the major? One of the central topics in computing education research is the exploration of how a person learns their first programming language in a first course in computer science (Robins 2019). Why do some excel at learning to program while others struggle? What are the strategies employed by effective novice programmers? How is it related to their prior knowledge and their lived-in experiences? Is there a "geek gene" for programming (Lister 2010)? Are the crucial differences of effective novice programmers cognitive, attitudinal, or behavioral, or impossible to separate (Robins 2019)? Researchers have been grappling with this question for quite some time.

The issue of attrition is multifaceted. Researchers have used a variety of theoretical frameworks when exploring the reasons for college dropouts, highlighting an approach on the role of social and academic integration for undergraduate persistence [Tinto 1975; 1993; Spady 1970; Berger 2000]. The well-reputed and influential theoretical framework is Tinto's model of departure, as illustrated in Figure 2, argues that individual departure from institutions can be viewed as arising out of a longitudinal process of interactions between an individual with given attributes, skills, prior educational experiences, dispositions (intentions and commitments) and integration with other members of the academic and social systems of the institution (Tinto .et al 1993, 1997, 2006).

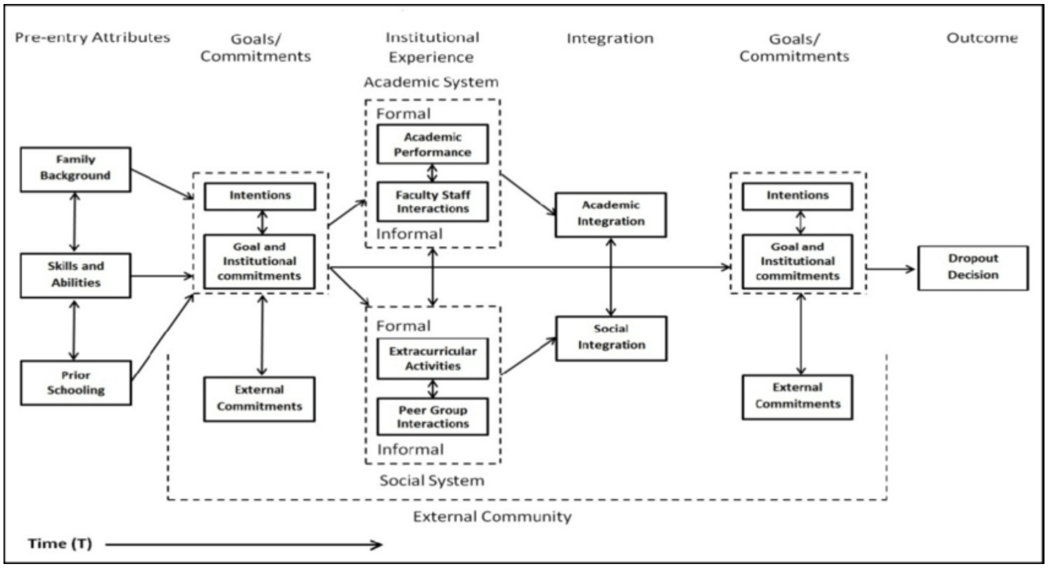


Figure 2: Tinto's Institutional Departure Model (Tinto 1993).
Source: Tinto's institutional departure model (Tinto 1993).

Students bring pre-entry attributes and expectations with them in their first year. Students' associations can then be weakened or strengthened based on the way the student is incorporated into the institutional community. Tinto's model proposes that once the student is emotionally content with the institution, their cognitive and academic performance will inform, and they will perform.

Students enroll into an academic discipline informed by a variety of factors, including planned career path, availability of jobs, salary expectations, intellectual interest, appeal of specific degree programs and courses within an institution, and other social or personal factors. Much research has gone into student retention, and many elements influence whether a student makes it to graduation or not. Here are a few of the most critical factors pertaining to student retention.

- **Misguided Perceptions (Misconceptions)**. The lack of an academic plan prior to joining a university contributes to a feeling of academic displacement for many students. Students who have not done their research might have misconceptions about the field of computer science. This will reflect with students who fail to recognize the connection between their coursework and a future career. Students may come to the university driven mainly by economic or "well-paying employment" necessity. Students may have a poor idea of the true nature of what they are going to study since they have little if any academic experience in the area. Some students are directed into CS by high school guidance counselors, parents, and otherwise influential people in student's life. These people feel that "everything is going to computers" so students "can't go wrong by majoring in computer science (Beaubouef 2005). Often times, these same students were not advised to take college preparatory courses in high school, and therefore, are less than prepared to begin college in CS major.
- **Transition from High School to University**. How students adjust to university and perform, especially in the first year, is a topic of significant interest. Tinto called this a stage of "separation of the individual from past associations," recognizing that they are at their most vulnerable early in their academic career. Students entering full time traditional college are expected to drop their old connections and create new associations in unfamiliar environments. Most students adapt quickly, and their attitudes develop accordingly, but lack of preparedness and unsuitable study skills are seen as primary causes of casualties in this group.
- **Study Habits and Time Management Skills**. Students drop out of the CS major during the first two years. Students may not yet have formed adequate study habits. Students

wait till the last day to work on their programming assignments, not understanding it takes hours to successfully complete it. Better prepared students begin projects earlier and usually complete them on time. The transition from high school way of learning to universities style of knowledge construction through projects or independent activities. Instructors provide course outlines and overviews of the contents; it is up to the students to find relevant reference books, read for details, prepare notes and manage their time. Furthermore, students at the university level are sometimes forced to take subjects which are completely new to them. For instance, in 2008, Ethiopia introduced an intake strategy in universities with a ratio of 70:30 for science and technology to social sciences, meaning that some students would be forced to take science or technology courses with no prior experience.

- **Previously Acquired Knowledge**. Programming is a craft skill that is known to be hard for many to master satisfactorily. Several studies investigating the effect of previous computer experiences on success in computer science courses were conducted in recent years. Lack of prior experience was not only found to negatively affect women's confidence and comfort in the program (Margolis & Fisher, 2003), but it was also found to be a big hindrance to retention (Buzzetto-More et al., 2010) and a big predictor of attrition (Katz et al., 2006) and or failure in CS courses taken (Staehr et al., 2000) Taylor and Mounfield had two studies in 1991 and 1994. The earlier study found that not only did high school computer science have a very positive effect on college computer science, but that the teaching methodology rather than the content of the course seemed to be the major factor of the high school experience that contributed the most to success in college. The later study found that while only certain previous experiences were related to success for males, virtually all previous computer experiences were beneficial for females. The combination of lack of preparation felt by many of the females and the emphasis on computer programming at the entry level of the curriculum is a "lethal combination" (Liu & Blanc 1996)

Students lacking the critical thinking, logic and analysis needed in their courses are also at risk of dropping out. (Robins 2019) Programming is the primary activity of computer science, and therefore most CS programs globally start with an introductory programming course. Programming languages are complex artificial constructs. Learning to program is a difficult process for a good number of individuals. Brooks addressed the practical process of software development and the reasons that so many programming-based projects failed, drawing attention to the need for new ideas about how to manage programmers. (Robins 2019)

"In every introduction to programming course, 20% of the students just get it effortlessly -- you could lock them in a dimly lit closet with a reference manual, and they'd still figure out how to program. 20% of the class never seems to get it." (Guzdial, 2007).

- **Institutional Commitment**. Students may drop out due to the lack of positive expectations, sense of belonging and attraction to a particular institution. All these characteristics of personal college experience correspond to low levels of institutional commitment (Tinto 1993). Students whose choice of institution or major mismatches their initial preferences show lower levels of social integration and are more likely to withdraw (Braxton, Milem, Sullivan 2000). This factor may be especially relevant in the context of Ethiopia, as Ethiopian college students must choose their specialization at the start and have very limited opportunities for horizontal mobility— between majors or institutions—later. Being unsatisfied with their assigned choice of major or college, they have fewer opportunities for a seamless transition as compared to students in education systems where specialization choice occurs at later stages or can switch at any time—and thus face a higher risk of dropping out (Braxton, Milem, Sullivan 2000).

- **Attribution Theory (AT)**. The AT is based on the premise that individuals will assign causal attributions that explain reasons for success and failure (Weiner 2010). AT in CS involves explanations students give for what they perceive to be the contributing factors to their success or struggle in learning to program. (Bernstein 1991) even found that males who were uncomfortable using computers attributed this feeling to "inadequate experience or poor teaching," while females tended to criticize themselves for feeling uncomfortable with the computer. Several studies have suggested that females tend to attribute their successes in computer science to luck and their failures to lack of ability (Bernstein, 1991; Howell, 1993; Moses, 1993; Pearl, Pollack, Riskin, Thomas, Wolf, & Wu, 1990). If these tendencies were substantiated, they would obviously be a barrier to an increase in motivation and self-confidence for female in computer science and certainly could, at least in part, explain the high attrition rates reported in computer science programs.

  Identifying problematic reasoning for causal factors that students attribute to successful and unsuccessful outcomes in an early programming course, may assist our understanding of how to guide or 'retrain' first-year thinking about programming processes and learning processes, in their authentic environments. Such understandings may assist us to prevent students from attributing to external and unchangeable events that ultimately may cause feelings of ‘helplessness’, ‘shame’ or ‘humiliation’ (Weiner 2010) that may result in them dropping out from the degree in the long-term.

- **Social Integration**. Researchers have attributed gender differences in CS to cultural factors such as prevailing gender stereotypes/expectations and the lack of female mentorship. Cheryan et al. (2015) review this literature and provide evidence for how popular stereotypes and typical representations of computer scientists make the field seem inaccessible or inappropriate for females. Given the lack of female representation, the work of Hoh (2009) and Drury et al. (2011) stress the importance of female role models in encouraging women to major in STEM fields such as engineering and computer science.

Factors promoting female students’ departure from computing fields in Ethiopia includes but not limited to inadequate preliminary knowledge and academic preparation, lack of self-efficacy, the persistent effects of socio-cultural gender stereotypes, the existence of science educational experiences that do not welcome women, and the absence of adequate support systems in the Ethiopian universities (Egne 2014).

1.2. Problem Statement

Despite the growing global demand for CS professionals, their high earning potential, and diversified career paths (U.S. BLS 2021, UNESCO 2017), a critical gap exists between enrollment and graduation among female students in computing fields across the world (Raigoza 2017, Hailu 2018, UNESCO 2017, Bennedsen and Caspersen 2007). A breadth of previous studies in the United States highlights that retention of female CS undergraduates through graduation as one of the biggest challenges for higher institutes (Crews & Butterfield, 2003;

Beaubouef & Mason 2005; Bennedsen & Caspersen 2007; Biggers et al. 2008; Barker et al. 2009; Barker et al. 2014; Elared 2016; Raigoza 2017).

Gender imbalance in computing education is problematic for at least three reasons. First, computing jobs provide opportunities for creativity and innovation, are financially rewarding, and secure, and thus, women are missing from jobs that are potentially beneficial for them. Second, computer scientists design tools that shape modern society and diversifying the field can help to ensure that these fields are creating designs appropriate for a broad population (Beyer & Haller, 2006; Cheryan et al., 2015; DuBow, Quinn, Townsend, Robinson, & Barr, 2016). Third, the U.S. and the world are currently not training enough computer scientists to keep up with demand (Caito 2022).

Although past studies provide some understanding of Computer Science retention, the need for further studies comes from the lack of consistent research results, scant literature on enumerated success factors, lack of comparative analysis on the predictive facts from diverse geography and culture perspective, the changing nature of Computer Science fields, the combination of distinct factors identified in the literature as important for CS studies, and the huge current and future demand for Computer Science professionals. Further research in this area is needed to uncover what factors contribute to female students' persistence in Computer Science/Software Engineering fields and whether these factors can be replicated across distinct geographical boundaries. Identifying and enumerating the contributing factors of success from the global perspective of individualized motivations, and experiences of undergraduate female students in CS/SE are largely untapped prior to this study.

A strategic first step is to gain an understanding of students' perceptions of the academic and social experiences related to the subject and investigate why they stay or leave as they move towards their graduation (Biggers et al. 2008).

This research aims to uncover patterns of predicting factors of persistence for the purpose of tracing interventions that can help increase the achievement of female undergraduates in Computer Science, both in Ethiopia and the United States. Both quantitative and qualitative methods will be used to gain in-depth insight into the motivations and perceptions of female undergraduates in CS/SE. This data will be contextualized with a review of recent literature on the and statistical analysis of demographic changes.

1.3. Significance of the Research

Gender disparities in CS/SE are problematic for at least three reasons. First, missed opportunities for the women. According to the U.S. Bureau of Labor Statistics (USBLS), the computer and information technology fields are projected to add 667,600 new jobs and are expected to grow by 13 percent from 2020-2030 (Caito 2022) — faster than the average growth rate of all other occupations. By departing the major before graduating, women are missing out on lucrative, creative and flexible job opportunities that are potentially beneficial for them. (Wodon and Briere 2018) highlight the unrealized potential and economic cost of gender inequality, globally. Women account for only 38 percent of human capital wealth versus 62 percent for men. In low- and lower-middle income countries, women account for a third or less of human capital wealth.

Second, computer scientists and software engineers create and build tools that shape modern society and diversifying the field can help to ensure that these fields are creating and building tools that mirror the broad population (Beyer & Haller, 2006; Cheryan et al., 2015; DuBow, Quinn, Townsend, Robinson, & Barr, 2016).

Third, the U.S. is currently not training enough computer scientists and engineers to keep up with demand (Soper, 2014). The U.S. National Science Foundation and the European Commission (Thilmany 2008) suggested that the lack of women in these skilled professions negatively impacts economic growth, both because of the sheer numbers of STEM professionals required by industry to remain competitive and because of theories that diversity spurs innovation. Helping more women graduate would be an effective way of reducing this gap.

Several stakeholders may benefit from this study. First year and future female CS/SE undergraduates may find the reported experiences of current upperclassmen female students insightful, could draw inspiration and could benefit from the listed factors of persistence. The study will also contribute to the body of knowledge for Ethiopia's literature library on preventative measures of retention and effective strategies for persistence. On a larger scale, universities can benefit from additional insights to better support the needs of first year students, for early interventions mapping out the differences between students who were successful and how their varying characteristics and actions can affect their decisions to stay, and their ability to be successful. A better understand of these populations can assist researchers and stakeholders within the discipline to assist the various student populations, and hopefully begin to increase the diversity in Computer Science as they gain the tools, they need to supplement all student learning, regardless of their backgrounds and risk factors.

1.4. Purpose of the Study

In studying the experiences of undergraduates in Computer Science/Software Engineering education, a strategic first step was to gain an understanding of students' perception of their academic and social experiences related to their major and to investigate why they stay or leave as they move towards their graduation (Biggers et al. 2008). This study seeks to identify what makes a successful CS student. Successful defined as a CS student who has completed the first two required CS courses and has decided to persist through the graduating class of 4th or 5th year. This research aims to understand retention from the perspective of female students who persisted beyond their first two critical years to be in graduating class of their CS/SE majors. The purpose of this parallelly convergent mixed-methods research was to comparatively investigate, describe and analyze factors correlated to the experiences and perceptions of female undergraduates as it relates to their persistence in CS/SE degrees, enrolled in two geographically, culturally and economically distinct public universities in the North Dakota, the United States & Addis Ababa, Ethiopia. The study included students enrolled in the first two CS/SE courses as well as upperclassmen female students in CS and SE major. As illustrated in the literature review, the problem of retention of females through graduation in computing fields is multifaceted. Identifying and enumerating the contributing factors of success from the global perspective of individualized motivations, and experiences of undergraduate female students in CS/SE are largely untapped prior to this study.

1.5. Theoretical and Conceptual Framework

Tinto's theoretical model of student integration is the underpinning of this current research. Tinto proposes that students enter college with pre-existing attributes and experiences, including family background, skills/abilities, and prior schooling, to begin a longitudinal process of interacting with the formal and informal social and academic components of the institution.

According to Tinto's theory of persistence (Tinto .et al 1993, 1997, 2006), students' level of academic and social integration leads to greater commitment to institutions and graduation. In this theory, academic and social integration are complementary but independent processes in a student's life. Guided by this theory and past empirical studies (Seymour and Hewitt 1997, Biggers et al. 2008, Wondimu 2004, Negash 2006, Hailu 2018), this research formulates a conceptual framework, as illustrated in Figure 3, which identified six predictor variables to establish the level of academic and social integration that makes the female students be 'persisters' in CS or SE degree. The six predictor variables included in this study are the student's perception of her/his degree's usefulness, previously acquired knowledge, cognitive attitude, sense of belonging, institutional support, and peer group activities. Anchored on Tinto's comprehensive theory of retention, this mixed-method research will explore the influences of these predicator variables.

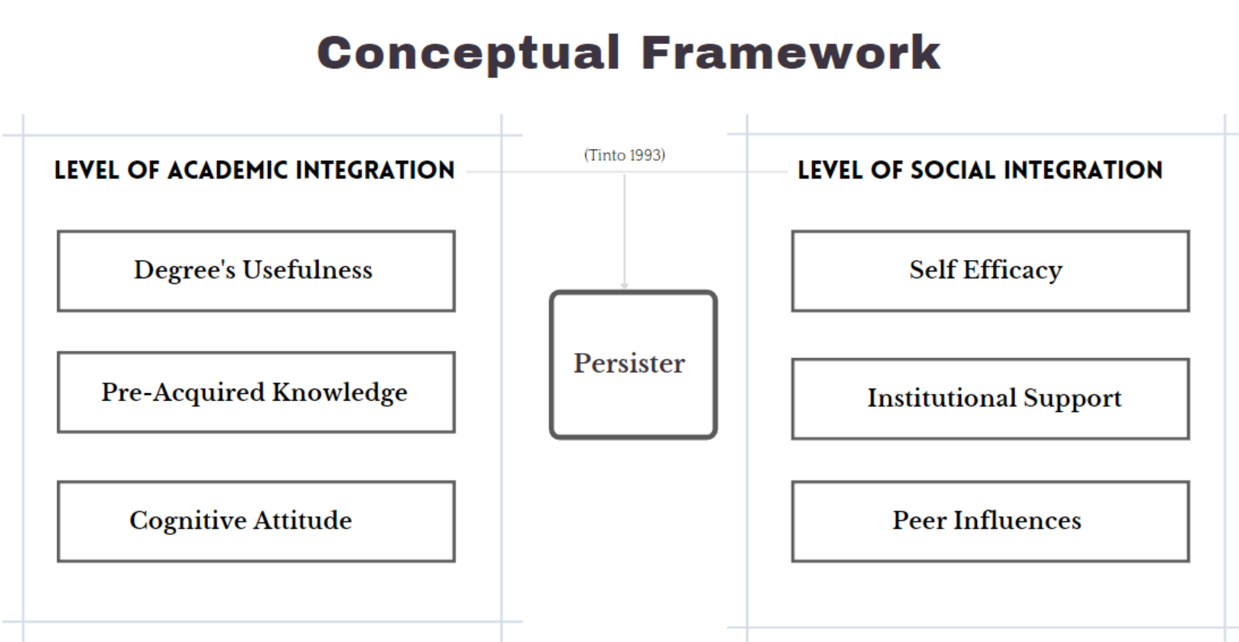


Figure 3: Conceptual research framework of female students' persistence in Computer Science or Software Engineering.

The predictor variables nested within the level of academic and the level of social integration informs the following research questions.

1.6. Research Questions

This study seeks to investigate, evaluate, and enumerate observable factors in answering the following three research questions:

RQ1: Does a relationship exists between students' perceptions of their Degree's Usefulness (DU), their Previous Programming Knowledge (PAK), or their Cognitive Attitude (CA) with their decision to pursue Computer Science (CS)/Software Engineering (SE) degree after completing the first two introductory courses?

RQ2: Is gender difference evident in the perception of DU, PAK, CA for students taking the first two introductory CS/SE courses? Is institutional difference evident in the perception of DU, PAK, CA for students taking the first two introductory CS/SE courses between North Dakota State University (NDSU) and Addis Ababa University (AAU)?

RQ3: Can factors that promote success in pursuit of a degree in Computer Science/Software Engineering for female undergraduates be identified and enumerated? Can persistence factors in pursuit of a Software Engineering degree in Ethiopia be predicted by factors reported to be effective in the U.S. Computer Science degree and vice-versa?

1.7. Assumptions and Limitation

1.7.1. Assumptions

This study assumed the following:

1. The research participants will voluntarily participate and will give honest answers to the questionnaire and interviews.
2. What motivates females to persist in computing fields can be gleaned from coded interviews and NLP topic modelling to be enumerated
3. Previous research can be used as a basis for the design of this project.

4. Many students will drop out before the semester ends. This will offset the pre-/post-survey response results to some extent. (The attrition rates are extremely high in introductory computer science courses.)
5. Surveyed students will self their decision to persist or not beyond the first two computer science courses. A lot of research results show that attitude and motivation are more important than grades in determining which students drop out (Ajzen et al., 1999, Abbott et al, 2002; Bean, 2005; House, 2004).
6. NDSU is the university that houses the College of Engineering where the department of Computer Science is located. AAU is the university that houses the Addis Ababa Institute of Technology where the School of Technology and Engineering is located. To simplify the comparison, this study refers to NDSU and AAU when talking about Computer Science and Software Engineering degrees, respectively.

1.7.2. Limitations

Limitations that affect this study are as follows:

1. This study is limited to the computer science students enrolled in CS I/CS II at NDSU and interview participants as well as students in the pre-engineering and software engineering programs at AAiT.
2. The number of females in college computer science program will be small.

1.8. Dissertation Structure

This dissertation is organized into five chapters. Chapter one introduces the research, presents the problem statement, discusses the research questions, and explains the theoretical and conceptual framework. Chapter two describes the literature review on the six constructs of the conceptual framework. Chapter three discusses the research design, research participants, data collection, analysis methods, and construct validity. Chapter four explains the results of the data collected. Chapter five summarizes the core findings in the conclusion section along with implications, limitations, and suggestions for future research.

CHAPTER 2. REVIEW OF LITERATURE

2.1. Introduction

The central research question guiding this study is “what factors help female students persist in undergraduate Computer Science majors at public universities in Ethiopia and North Dakota?” There is an abundant of published research demonstrating the existence of gender differences in Computer Science in the United States in terms of educational goals, interests, experience with computers, support and encouragement, gender discrimination and attitudes toward the academic environment (Beyer, Rynes, Perrault, Hay & Haller, 2003). Although there is a dearth of literature regarding the participation of women in Computer Science in the Ethiopian context, Gender inequities in STEM education remains pronounced in Ethiopia (Hailu, 2018).

This literature review is organized through the lens of Tinto’s theory of departure constructs and will pause the following questions, why do women choose to pursue Computer Science degree? Why do female Computer Science majors have less computer confidence than the male Computer Science majors? (Beyer 1990) Why do women who succeed in Computer Science degree are often viewed as “exceptional”? (Beyer 1990) Does prerequisite experience (e.g., high school programming courses, enrichment programs, community college transfer students, summer jobs or internships in computing areas) play a role in student retention? How did the female Computer Science freshmen differ from their male counterparts regarding their perceptions of Computer Science degree, and their learning experiences? Do the faculty or teaching demographics of the Computer Science programs impact the retention of women?

Retention literature suggests that students who experience high levels of academic and social integration will, in many cases, have higher levels of retention. Tinto (1993) identifies 3 major sources of student departure in his framework: academic difficulties, the inability of individuals to resolve their educational and occupational goals, and their failure to become or remain incorporated in the intellectual and social life of the institution (Kuh et al., 2006).

Academic integration is defined by students' academic performance, level of intellectual development, and perception of having a positive experience in academic settings. Social integration is defined by individual background, aspirations, involvement in peer group activities, the presence of positive relationships with peers and faculty-student interactions. In this regard, in Tinto's model (Tinto, 1975, 1993), instructors play an important role in students' academic and social integration through the choices they make pertaining to course and program contents, supervision provided, teaching, learning and assessment strategies used, etc. Academic and social integration depend on the appropriateness of course and program contents, teaching and learning, and assessment strategies for students' needs.

In Tinto's theory, academic and social integration are complementary but independent processes in a student's life. Academic and social integration leads to greater commitment to institution and graduation (Bean, 1983). Building up on this model and previous body of literature, this current research has formulated three predicting constructs for high level of academic integration in Computer Science, namely, perceptions of degree's usefulness, pre-acquired knowledge, and cognitive aptitude. This current research also formulates three more predicting constructs for high level of social integration in Computer Science, namely, sense of belonging, institutional support, and involvement in extra-curricular departmental activities.

The issues surrounding why students choose to persist within a given degree or switch to another (or even drop out completely from their studies) are complex. The factors that affect student decisions can be broadly classified into three groups (Hein et al., 2012): (a) academic environment and resources, (b) positive or negative perceptions of the discipline and career, and (c) diverse experiential effects. The academic environments and resources include lectures, recitations, and laboratories; faculty and teaching assistants; university services and so on. Perceptions refer to ideas of the self, including confidence, self-efficacy, and determination to succeed; or to the major and career, including the opportunities and advancements provided by the field with respect to society, and society's perception of the field. Diverse experiential effects involve personal positive or negative experiences, stereotypes, and other input related to students' exposure to the university environment.

The evidenced high rate of student attrition and the reluctance towards studying Computer Science disciplines implies that more research is needed to investigate how students could be retained (Hong, Wang, Ravitz and Fong, 2015). I will examine the female students' perceptions, lived-in experiences and institutional support that helped them succeed in the Computer Science I and II programming courses in their respective universities.

2.2. The Perceptions of Degree's Usefulness

Students' perceptions regarding their future in Computer Science are expected to evolve, and sometimes change, during their studies (Giannakos 2016). The understanding of the student's degree usefulness, and related career prospects are one of the reasons for the student to enroll in a major. Students have reported that the utility of a Computer Science degree influences their decision to complete their studies (Lewis et al. 2011, 2016), where Lewis et al. (2016) used the term "utility" to explain the extent to which Computer Science provides value to society or to the student as an individual. Students' perceived usefulness of their degree should be quite high at enrolment and might increase or decrease depending on the students' experience especially during the first two years (Ohland et al. 2008). Hence, it is important to investigate the relationship between usefulness of the degree and student persistence in Computer Science. The inability of students to resolve their educational and occupational goals as well as their failure to become or remain incorporated in the intellectual and social life of the institution can be a deterrent to their persistence (Tinto 1993). He emphasized that strong intentions or career goals can overpower the effect of negative experiences and poor integration into the culture of the institution.

2.3. Pre-Acquired Knowledge

The pre-entry attributes associated with students provide insight into understanding how they will ultimately respond to their educational environment and persist. There are an abundance of research affirming the correlation between pre-entry attributes such as problem solving, digital literacy and previous programming experiences on one hand and performance in programming on the other (Barlow-Jones G., van der Westhuizen D. 2017). According to Biggers et al. (2008), retention in the Computer Science major depends greatly on having previous experience and knowledge of computing and technology.

Problem solving, computational design thinking and logic-based thinking are some of the foundational knowledge that all students need. According to code.org, Computer Science is marginalized throughout the secondary education system. North Dakota had adopted the K-12 Computer Science and Cybersecurity standards two years ago, but the state does not have a plan on how to make Computer Science education a fundamental part of a state's education system (code.org 2018). Coursework in Computer Science is rarely required of high school students, and the elective offerings are severely limited and even the AP Computer Science curriculum's focus is Computer Programming (Carter 2016). These limitations might inadequately prepare for a degree in Computer Science.

Learning to program during the pursuit of a Computer Science degree has typically been regarded as difficult for novice students, with a high rate of course failure and drop out the first year (Mendes et al., 2012, Watson and Li, 2014). Students who enter with no formal Computer Science and those who enter feeling less than confident that they have made the correct choice of major are more likely to withdraw (Fisher, Margolis & Miller, 1997; Greening, 1999; Morahan-Martin, Olinsky & Schumacher, 1992; Scragg & Smith, 1998). Students who had adequate previous experience before enrolling in computing courses at their institution, such as previous high school computing courses or independently learned skills, were statistically significantly more likely to stay within the major and graduate (Biggers et al., 2008). In this current research, we will investigate if female students that have limited knowledge about Computer Science are likely to be less interested to continue studying in Computer Science.

2.4. Cognitive Aptitude

The ACM/IEEE Computer Engineering Curricula (2016, p. 17) identifies that all Computer Science graduates are expected to: Possess the ability to design computers and computer-based systems that include both hardware and software to solve novel engineering problems, subject to trade-offs involving a set of competing goals and constraints. In this context, "design" refers to a level of ability beyond "assembling" or "configuring" systems.

It is widely accepted that Computer Science competences are important skills for the twenty-first century, since they allow students to "construct" and "create" meaningful artifacts using computers. Skills such as computational thinking, which include problem solving, complex systems design, and evaluation as well as human behavior understanding, are cornerstones of Computer Science education.

Models such as those by Summerskill, and Marks's point to the importance of intellectual attributes in shaping the individual's ability to meet academic demands (Summerskill 1962; Marks 1967; Tinto 1993). Liu and Blanc (1996) states that female students are more familiar with computer applications than computer programming or terminology. The qualitative analysis (Berges 2016) conducted showed them that young men and women do have different perceptions of computer science. “While male students focus on technical aspects like hardware, mathematics, or logical issues, female students are attracted by creativity, communication, or job opportunities.” Survey studies of college students even suggest that students who have an obvious aptitude for CS careers (e.g., STEM majors) have little knowledge about what becoming a computer professional might entail (Carter 2006). Finally, Xenos et al. (2002) identify five categories of dropout reasons in CS studies: professional (mentioned by 62.1 % of the respondents in their study), academic (46 %), family (17.8 %), health-related (9.5 %), and personal (8.9 %).

The challenges faced by a student during the learning phase of computer programming can contribute directly to the development of negative perception on computers. The learning and acquiring of competence in programming is a highly difficult process. As a consequence, students unconsciously reject the task of learning programming (Tan et. all, 2009). Therefore, it is crucial task to determine, before in hand, the students’ perception on programming and take countermeasure to tackle and address the problems associated with this perception. Students might have different type and different level of perceptions: Positive or negative perceptions on programming; a perception on benefit of programming, perceptions related to difficulties of programming process etc. The perception of student on their own competence is defined as self-efficacy and it is stated that “different people with similar skills or the same person under different circumstances may perform poorly, adequately, or extraordinarily depending on fluctuations in their beliefs of personal efficacy" (Bandura, 1997).

In this context, the self-efficacy perception is a self-evaluation of a person, on his/her competence to conduct a task successfully (Korkmaz 2011; Bandura, 1987) and it is considered as a prediction of a person's aptitude on what he/she might accomplish, being in aware of his competence to complete a particular task. A person might have all necessary qualities to accomplish a task, but deficiency in self-belief and lack of motivation may give rise to failure. On the other hand, it is reported in literature that self-efficacy has a direct effect on the process of acquiring new skills and using the acquired skills. Therefore, the level of self-efficacy might be used as a tool and a reliable indicator in order to predict a person's performance (Askar & Davenport, 2009).

This current research will evaluate if the perception of cognitive aptitude in programming is correlated to success in programming courses which leads to persistence in Computer Science degree.

2.5. Sense of Belonging

Cultural stereotypes in computing fields are alive and well. When this stereotype reveals itself in variety of ways, female students' self-efficacy, sense of belonging and identification with computing suffers. Students' sense of belonging is widely recognized as relevant to persistence and engagement (Stout and Tamer 2016). Students are more likely to leave their major when they feel they are being treated differently as a result of belonging to an underrepresented group (Barker et al. 2009). (Marra et al., 2012) found non-academic factors such as sense of belonging and perceived values related to Computer Science programs contribute to detrimental or enabling to their retention in those programs. One of the most important factors that shape students' decisions to major in Computer Science is the fit between their identity and Computer Science; that is, the extent to which their own values and identity align with Computer Science studies and professions (Lewis et al. 2011, 2016). Retention is expected to increase when students feel like they belong to the intellectual and social community of their major (Barker et al. 2009).

(Tinto 1993) also found that a student's sense of academic and social belonging impacts on retention and graduation, and this sense of belonging is increased or decreased through interactions with the academic and social environments of the university. His findings have been extended to include student expectations (Braxton et al. 1995). Moreover, high levels of social support contribute to students' overall sense of belonging in their program, and, ultimately, their likelihood of persistence.

This current research will focus on female student persistence factors with regards to their feelings of belonging and confidence. Persistence efforts focused on students who develop a strong interest in other fields can hardly be a fair target, however, when it is focused on students who leave computer science because of misunderstandings about a negative experience in Computer Science academic major.

2.6. Institutional Support

Walden and Foor (2008) found that a welcoming climate has a significant influence on students' decision to persist or change major. Academic integration reflects a student's experience with the academic systems, classroom environment and academic communities of a college or university. Such experiences find expression in a student's sense of normative congruence and affiliation with these academic systems and communities (Braxton 2019; Braxton & Lien 2004; Tinto, 1975).

The third element of Tinto's model has been expanded to include the formal and informal aspects of institutional experiences and the interaction of the academic and social systems. Academic and non-academic staff are both seen as having the ability to influence the departure decision. Literature in this area has shown that institutional characteristics and culture have both direct and indirect effects on the student's tendency to become involved in both academic and non-academic activities (Braxton et al. 1995).

Bean's model in 1980 established the condition that the background characteristics of students must be considered in order to understand their interactions within the environment of the higher educational institute (Bean 1982). In presenting his model in 1985 Pascarella suggested that quality of student effort, student background/pre-college traits, and interactions with agents of socialization directly influence learning and cognitive development (along with all other variables in the model) (Pascarella 1985). Colleges and universities must integrate students deliberately academically, socially, and intellectually with the culture of the institution. Colleges and universities should create opportunities for peer group activities, informal student interactions, and faculty/student interactions. (Long, 2012).

The lecture format and quality that dominates many Computers Science courses, especially during the first two years, can be detrimental in that it potentially creates a barrier between students and the subject and makes it more likely for them to become disconnected from the Computer Science program. Student-to-faculty interaction is an important aspect of the environment, and ultimately students' success. Typical pedagogical activities in Computer Science (and STEM) degrees may also be incompatible with students' personal approaches and styles with respect to learning course material. Bernold et al. (2007) examine the study path followed by a first-year STEM cohort for three years and identify that different teaching and learning methods are related to learning performance, entering a STEM major, and staying in the major. High-quality teaching and excellent support contribute to students' learning performance, overall assessments in Computer Science programs, and, ultimately, their likelihood of persistence.

2.7. Involvement in Peer Group Activities

Study has shown that collaborative learning activities nullified the negative impact of gender stereotype endorsement on women's self-efficacy, sense of belonging, and identification with computing (Stout and Tamer, 2016). "Simply stated, it is student interactions with peers, advisers, and faculty that increase satisfaction with the institution, create a sense of belonging, and strengthen commitment to the institution's educational goals and standards" (Levin and Levin 1991). The college also has an impact on determining the amount of student involvement and thus the gains and retention (Clagget 1992).

The ongoing social and academic integration stage of retention is defined by the conditions for socialization and mentoring. These factors include access to the proper equipment in upper-level courses and a network of peer and professional role models. (Jama et al., 2008) This current study will further evaluate the correlation between the involvement of female students with their peers and their persistence in undergraduates in Computer Science course.

CHAPTER 3. RESEARCH METHODOLOGY

3.1. Introduction

This chapter introduces the research methodology used in investigating predictive factors of success for female undergraduates in Computer Science and Software Engineering. This parallelly convergent mixed-methods research approach allowed for examining variables independently and interpreting together the contributing agents of persistence for deeper understanding of the experiences of female students. The research questions, the research plan, including description of research design, conceptual framework of hypotheses, data collection methods, research participants, data analysis processes, and assumptions and limitations are also primary components of this chapter.

3.2. Purpose of the Study

The purpose of this parallelly convergent mixed-methods research was to comparatively investigate, describe and analyze factors correlated to the experiences and perceptions of female undergraduates as it relates to their persistence in Computer Science/Software Engineering degrees, conducted in two public universities in the U.S. & Ethiopia. The two selected universities are North Dakota State University (NDSU) and Addis Ababa University (AAU).

3.3. Research Questions

This study sought to investigate, evaluate, and enumerate observable factors in answering the following three research questions:

RQ1: Does a relationship exists between students' perceptions of their Degree's Usefulness (DU), their Previous Programming Knowledge (PAK), or their Cognitive Attitude (CA) with their decision to pursue Computer Science (CS)/Software Engineering (SE) degree after completing the first two introductory courses?

RQ2: Is gender difference evident in the perception of DU, PAK, CA for students taking the first two introductory CS/SE courses? Is institutional difference evident in the perception of DU, PAK, CA for students taking the first two introductory CS/SE courses between North Dakota State University (NDSU) and Addis Ababa University (AAU)?

RQ3: Can factors that promote success in pursuit of a degree in Computer Science/Software Engineering for female undergraduates be identified and enumerated? Can persistence factors in pursuit of a Software Engineering degree in Ethiopia be predicted by factors reported to be effective in the U.S. Computer Science degree?

3.4. Research Design

A parallelly convergent mixed-methods research design approach was chosen to best understand the research questions of this study. A parallelly convergent design entails that the researcher concurrently conducts the quantitative and qualitative elements in the same phase of the research process, weighs the methods equally, analyzes the two components independently, and interprets the results together (Creswell & Pablo-Clark, 2011).

Combining quantitative and qualitative research methodologies allowed the researcher to gain deeper understanding of the contributing agents of success for female students enrolled in undergraduate Computer Science/Software Engineering education at NDSU and AAU. As outlined by Creswell (2007), the correlational quantitative research approach provided a way to elaborate the statistical association of variables as outlined with the defined set of hypotheses focusing on persistence during the first two years of Computer Science/Software Engineering studies. The grounded theory qualitative approach connected data science and transcribed audio interview insights through Natural Language Processing (NLP) sentiment analysis and topic modelling processes to assess the perceptions, feelings, and experiences of female students' majoring in Computer Science/Software Engineering. As described by Stake (2010), the grounded theory qualitative inquiry allowed the researcher to explore the research questions by talking directly with those associated with the phenomenon, allowing them to communicate their lived-in experiences first-hand, through coding the data from interviews, and build a theory based on the interpretation of their shared experiences.

The parallelly convergent mixed-methods research design approach chosen for this study requires the researcher to perform concurrent collection of both qualitative and quantitative data in the same phase of the research process. Figure 4 illustrates the triangulation of the two methods by comparing the quantitative statistical results and qualitative findings, analyzing the two components independently, and interpreting the results together (Creswell & Pablo-Clark, 2011).

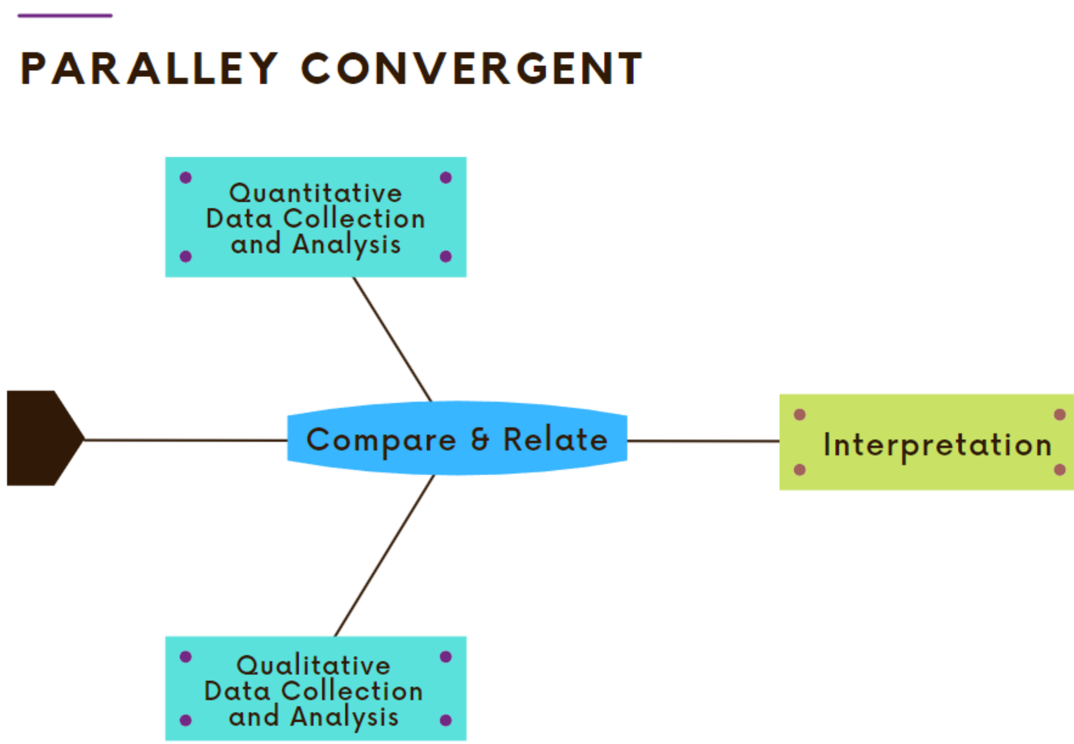


Figure 4: The parallelly convergent mixed-methods research design process used in this study.

The quantitative and qualitative methods were utilized to gather enough supplemental data to understand the holistic characteristics, opinions, and abilities of undergraduate students in introductory Computer Science courses. It was important to utilize these methods in order to reach as many students' population as possible through both the online surveys and virtual meetings. The mixed method also helped to gain multiple forms of data, both in-depth anecdotes from interviews as well as a larger volume of answers from multiple choice and Likert scale questions.

Furthermore, this design methodology was chosen because a combination of quantitative and qualitative methods provides a more detailed representation of the findings that one would receive utilizing one methodology alone (Creswell & Plano Clark, 2007, p.5).

3.5. Research Hypothesis

The mixed-method design used in this study required the use of two instruments to collect the quantitative and qualitative data needed to address the first two research questions and validate the hypotheses. The purpose of the quantitative portion of this study was intended to evaluate the five hypotheses derived from the research questions. Each hypothesis corresponds to the survey subscales namely Degree Usefulness, Pre-Acquired Knowledge, and Cognitive Attitude.

Therefore, the five null hypotheses are: -

Hypothesis 1: Degree’s Usefulness (DU)

H0: There is no relationship between students’ perception of the usefulness of their degree’s and their decision to persist in CS/SE degrees after completing the first two introductory CS courses.

Hypothesis 2: Pre-Acquired Knowledge (PAK)

H0: There is no relationship between students' perception of the influence of their previously acquired knowledge and their decision to persist in CS/SE degrees after completing the first two introductory CS courses.

Hypothesis 3: Cognitive Aptitude (CA)

Ho: There is no relationship between students' perception of their cognitive attitudes and their decision to persist in CS/SE degrees after completing the first two introductory CS courses.

Hypothesis 4: Gender Difference

Ho: There is no gender difference in the perceptions of DU, PAK or CA during the two introductory courses.

Hypothesis 5: Institutional Difference

Ho: There is no institutional difference evident in the perceptions of the three variables identified (DU, PAK, CA).

The qualitative portion of this research included domain knowledge gathering through grounded theory inductive manual coding technique and then unstructured data analysis using Natural Language Processing (NLP) to answer the third research question.

3.6. Target Institutions

North Dakota State University (NDSU) and Addis Ababa University (AAU) are the two public universities selected for the purpose of population sampling for this research. These two institutions are on different spectrums geographically, economically and culturally. The significant cultural contrast between them makes common conclusions of interest, since it is reasonable to extrapolate them to a significant proportion of the sector. A description of each institution is included below.

3.6.1. Addis Ababa University (AAU)

Established in 1950, Addis Ababa University (AAU) is the oldest and largest national public university located in Addis Ababa, Ethiopia. Ranked #11 in Best Universities in Africa (Morse and Castonguay, 2022), AAU has fourteen campuses scattered across the nation. At present the University has 10 colleges, 4 institutes that run both teaching and research, and 6 research institutes that predominantly conduct research (AAU 2022). One of the largest research and teaching institutes is the Addis Ababa Institute of Technology (AAiT). As listed on the institute's website, AAiT's academic faculty comprises of 4 professors, 16 associate professors, 43 assistant professors and 118 lecturers with undergraduate student numbers reaching up to 10,284. The Software Engineering degree is offered at The School of Information Technology and Engineering (SiTE) which is housed at AAiT of AAU. In 2021, 443 undergraduates joined the SiTE, of whom 152 are Software Engineering majors. Approximately 20-30% of the CS majors are female. This research participants were derived from these 152 populations.

Software Engineering bachelor's degree at AAiT takes 5 years to complete. Upon admission to AAiT, students are placed in what is called pre-Engineering cohort and are given orientation. At this point, all admitted students have not yet chosen a major. The first semester of the program is referred to as the *assessment semester* where students are required to examine their interest in various engineering disciplines offered at the institute. "*In this level students will get to know the various fields of engineering, what engineers do, and some preliminary courses before getting to know the core modules; this includes language and writing skills, and other social and humanities courses. This semester is intended to consolidate the students' educational background for higher learning and further help them to adapt to the system*," (SiTE 2021).

Fundamentals of Computer Science (CS I) is one of the courses offered to admitted pre-Engineering students during the *assessment semester*. As CS I is the first course for the Software Engineering degree, it is one of the courses selected for this research. After students complete the *assessment semester*, they are tasked with choosing their top two choices degree. Earning a higher GPA in pre-Engineering will ensure their first choice. Affirmative action is instituted to allow more female students into the Software Engineering major if the female students marked Software Engineering major as their first choice but are shy of .2 points to reach the required GPA to join the department. Women are historically marginalized in education (Hailu 2018). The Ethiopian government designed and implemented affirmative action in 1998, which aimed to promote gender equality, especially in the universities of the country (Egne 2014). This program includes an initial orientation for all first-year female students; female-only tutorials in four subjects chosen by the students during their first academic year; guidance and counselling service led by female professional counsellors on a one-to-one basis, peer counselling and support from senior female students; academic support from capable senior female students; and overall assertiveness training (Demise et al., 2002).

Transfer. Ethiopian college students have to choose their specialization at the very start of their college journey and have very limited opportunities for horizontal mobility. Unlike universities in the United States, the AAU has strict rules and regulations that dictate the transfer of students from one program to another, from one college to another college or even from university to university. In this research, no student reported to have transferred out of Software Engineering or have transferred into the Software Engineering department. The arduous process may have prevented it.

- *"Both inter-college and intra-college transfers shall depend on availability of space and facilities in the receiving college or department and a transfer applicant shall have grade levels that would normally be required of students for enrolment into the department concerned. Inter-college transfer is only applicable in areas of related disciplines. A student dismissed from his previous department shall not be eligible for transfer into another college or department of the University. Granting or denying transfer shall be made by a dean of the receiving college upon the recommendation of the receiving department in case of inter-college transfer and by the head of the receiving department in case of intra-college transfer which shall then be approved by the dean of the receiving college. Accepted transfer requests are subject to final approval by the Registrar." (AAU 2022)*

AAiT is hosts numerous international tech companies for short- or long-term engagements and has a strong network of strong bodies active with hackathon, workshop, etc. The institute also has a few partnerships with universities in United States and Africa which gives great exposure for the students.

3.6.2. North Dakota State University (NDSU)

Founded in 1890, North Dakota State University (NDSU) is a public land-grant research university located in Fargo, North Dakota. Part of the North Dakota University System, it is ranked as an R1 research institution by the Carnegie Classification of Institutions of Higer Education (NDSU 2022). As of 2021, NDSU offers 94 undergraduate majors, 146 undergraduate degree programs, 5 undergraduate certificate programs, 84 undergraduate minors, 87 master's degree programs, 52 doctoral degree programs of study, and 210 graduate certificate programs.

The Computer Science department is part of the College of Engineering at NDSU. NDSU has been offering Computer Science as an academic degree since 1973 with the first Bachelor of Science degree conferred in 1977. Software Engineering major is offered as a graduate degree program. Admitted undergraduate students are assigned an advisor to help them devise their Computer Science plan of study by choosing electives in their particular area of interest. Unlike AAU, NDSU CS department has individualized plan of study for each student depending on their pre-college experiences and credit course. Actual student schedules for each semester will vary depending on start year, education goals, applicable transfer credit, and course availability. Students are encouraged to work with their academic advisor on a regular basis to review degree progress and customize an individual plan of study.

The CS department offers remedial courses for students with no or very limited computer experience. The recommendation is to take college preparatory courses in high school. Advanced undergraduate students may have the opportunity to take graduate courses while completing their undergraduate program. An extensive and varied set of elective courses in every aspect of computer science is available as well. As of 2021, the Department has eight full professors, three assistant professors, two professors of practice, one senior lecturer, and three lecturers. There are approximately 140 graduate students, in both computer science and software engineering, and about 400 undergraduate students. Approximately 8-10% of the CS majors are female. This research participants are derived from this population. The first required Computer Science major course is CSCI 160 Computer Science I, referred to as CS I for the purpose of this research. Students are expected to take MATH 103 (College Algebra) prior to taking CS I. Per Credit by Examination policy of NDSU, students can test out of MATH 103 or get credit for the required score of the Advanced Placement (AP) math exam from College Board. The follow-up course to CS I is CSCI 161 Computer Science II, referred to as CS II for the purpose of this research. Students can skip CS I if they have transfer credit or have the required AP test score in Computer Science from college board.

Students can transfer to another department or college at will at any time during the semester. Consistent with the national statistics, students depart the CS major during their first two years of entering the CS department (Raigoza 2017) which is the interest of this research.

- "*College-level coursework from regionally accredited colleges or universities (or equivalent for international institutions), including Credit by Examination and the Joint Services transcripts, are eligible for acceptance in transfer with receipt of official transcript or score report.*"

There are two student organizations in computing, Association for Computing Machinery and Cybersecurity. The Computer Science program has had a high rate of students who receive a grade of a D, F, withdrew (DFW) from the course, especially in introductory Computer Science courses. In the past, CS I and CS II have produced as high as a 50% DFW rate in recent cohorts, as the students who left the major either switched to another major or left the institution completely. With this percentage of unsuccessful and departing students, there calls for further investigation as to why many students are not staying in the Computer Science major, as it affects NDSU's standing in producing qualified computing professionals.

3.7. Target Population

The target population was Computer Science undergraduates from the College of Engineering, Computer Science department at North Dakota State University (NDSU) and Software Engineering undergraduates from Addis Ababa Institute of Technology and Engineering at Addis Ababa University.

NDSU: Students who were enrolled in the first two introductory Computer Science courses CSCI 160 (Computer Science I) and CSCI 161 (Computer Science II) were invited to participate. All students enrolled in the two courses were invited to voluntarily participate which included non-Computer Science majors and upperclassmen.

AAU: Students who were enrolled in the first two introductory Software Engineering courses SECT-2111 (Fundamentals of Computer Science) in the pre-Engineering program and SECT-3091 (Fundamentals of Data Structures) were invited to participate. All students enrolled in the two courses were invited to participate in the study.

The researcher has re-labeled the courses to CS I and CS II for the purpose of standardizing the nomenclature during the study. The introductory Computer Science curriculum at most universities typically consists of courses which are commonly referred to as the CS1, CS2 and CS3 course series (Raigoza, 2017).

3.8. Recruitment

An informed consent form, as shown in Appendix A, was required for each participant prior to participating. Two instruments were used to collect data from the study participants: a pre-/post- questionnaire for each of the four courses (included in Appendix D) and one-on-one audio-only virtual interviews of upperclassmen (included in Appendix D). The method for recruiting these students for each portion of the study, as well as when each component was administered, are described below.

3.8.1. Quantitative Instrument

Survey. During Sept. - Dec. 2021, students from both institutions were asked to participate in two surveys; one at the beginning of the semester and one at the end of the semester, in order to compare their opinions of the introductory Computer Science course as they persisted throughout the semester. In September 2021, all students enrolled in CS I and CS II were contacted by their professors via email to participate in a 2-part preliminary survey through web based and mobile accessible Qualtrics form. The first part contains demographic questions, and the second part is the 5-point Likert Scale series of statements whose multiple-choice responses ranges from "I strongly Disagree" to "I strongly Agree". In order to ensure a higher participation rate, the students were informed a drawing for a $50 gift certificate (ETB 500 for AAU students) was made available. From the pre-survey, a total of 164 NDSU students responded with CS I (n=66) and CS II (n=98) out of 221 enrolled students; and a total of 210 AAU students responded with CS I (n=148) and CS II (n=62) out of 400 enrolled students.

The post-semester survey, released in early December 2021 during the Fall semester finals week, was administered in the same manner as the preliminary survey; students were invited to participate in a 2-part online Qualtrics form, via email by their course professors and was offered as another $50 gift certificate (ETB 500 for AAU students) as participation incentive. A total of 120 NDSU students responded with CS I (n=55) and CS II (n=65) out of 221 enrolled CS I/CS II students; and a total of 350 AAU students responded with CS I (n=254) and CS II (n=96) out of 400 enrolled students.

3.8.2. Qualitative Instrument

Interview. An email invitation was distributed to the NDSU Computer Science department listserv to invite female junior and senior Computer Science students to participate in an interview during the fall of 2021. In order to increase the number of participants, a drawing for a $50 gift (ETB 200 for AAU students) was announced. At AAU, the department administration sent an open invitation for female upperclassmen in Software Engineering to reach out the researcher’s Telegram account. A total of 20 participants responded from NDSU (n=10) and NDSU (n=10). An informed consent form, as shown in Appendix A, was required for each participant prior to participating.

3.9. Data Collection

Permission for conducting research activities involving human subjects was obtained from the NDSU Institutional Review Board (IRB), as shown in Appendix B.

3.9.1. Quantitative Instrument

Survey. The questionnaire that was handed out to the students was divided into two parts. The first part included questions on the demographics of the sample population (e.g., age, gender, major). The second part included measures of constructs identified in the literature from previous studies as important (1) Degree’s Usefulness, (2) Previously Acquired Knowledge and (3) Cognitive Attitude. Additionally, the questionnaire collected data on the following subscales to help address each of the three research questions and the corresponding hypothesis: Gender, Institution, and Course Reflections (CR) which included statement to measure the student’s decision to persist after competing CS I and CS II. Part of the survey questions were adapted by permission from a Computer Science Attitude Survey (CSAS) authored by Eric Weibe (Wiebe, et al, 2003) that measures Confidence in Learning Computer Science and Programming, Attitude Toward Success in Computer Science, Computer Science as a Male Domain, Usefulness of Computer Science and Effective Motivation in Computer Science and Programming. Additional customized questions were added to cover the hypothesis construct, Pre-acquired Knowledge and Cognitive Aptitude. Below is the description of the way each of the predictor variables was measured:

1. **Degree's Usefulness (DU)** - a continuous variable, which is the summation of the choices made on a 5-point Likert scale measuring the student's perceived usefulness of his/her degree and the related career prospects.
2. **Previously Acquired Knowledge (PAK)** - a continuous variable, which is the summation of the choices made on a 5-point Likert scale measuring the student's perceived influence of their prior programming experiences.
3. **Cognitive Attitude (CA)** - a continuous variable, which is the summation of the choices made on a 5-point Likert scale measuring the student's perceived abilities in programming.

The subscales for comparing predictor variables are:

1. Gender – a dichotomous variable (male or female).
2. Institution – a dichotomous variable (NDSU or AAU).

The survey instrument face validation and data collection period spanning three academic semesters (Summer 2021, Fall 2021, and Spring 2022). A pilot test was given to the AAU students to enable the researcher to find any contextual ambiguities in the survey instrument and revisions were made appropriately. One expert in the field of research and analysis of survey data, two professors of Computer Science and an expert in the field of testing and evaluation were asked to evaluate the face validity of the questionnaire. The questionnaire was found to have high face validity. The same validated instrument was administered to students in the two introductory computer science courses at NDSU and AAU.

3.9.2. Qualitative Instrument

Interview. Once approval was granted, the researcher emailed interview participants within their university contacts to schedule the virtual interview. The interview was conducted over zoom using the audio-only feature. No interview was conducted without confirming the written and verbal informed consent of the participant. The English language was used to conduct the interview since the instructional language of Ethiopian higher education institutes is English. The audio interviews were recorded electronically using zoom desktop recording feature and kept confidential on password protected laptop. Each interview was transcribed automatically using Microsoft Office 365's A.I.-based auto transcribe feature. While auto-transcribe to text does well, a second review of the transcription was needed to edit out any additional noises from the translation of audio to text.

With a goal to answer research question number three, the interview participants were asked to reflect on why they choose to study Computer Science/Software Engineering, what their challenges were and how they overcome those challenges to persist beyond CS I/CS II and advance in their Computer Science major. Grounded theory allows for discovering the phenomenon during the research process (Charmaz, 2006). Because the theory or phenomenon emerges from the data, it is possible that some interview questions may be added, or that the proposed interview questions will be modified during the research study (Birks & Mills, 2011; Charmaz, 2006; Urquhart, 2013). As some initial themes surfaced during the first four interviews, or subsequent interviews, the researcher added clarifying questions or points to subsequent interviews in an effort to explore more on the topic or gap that emerged. Appendix C includes the interview questions.

3.10. Data Analysis

Anchored on Tinto's comprehensive theory of retention and this research's conceptual framework, this section will describe the methods undertaken to analyze the quantitative and qualitative data collected through the online survey and participant interviews. Analysis of all data collected addressed the initial research questions and allowed the researcher to respond to each of the five research hypotheses.

3.10.1. Quantitative Analysis

3.10.1.1. Demographic Data

The demographic data collected from the online surveys was used to ensure the validity and cleanliness of data. Data indicating gender, academic year, and major ensured that the survey data captured reflected the target population of the study; students enrolled in undergraduate Computer Science/software Engineering programs.

Before the survey data analysis, the gathered survey data was prepared. The dataset was checked for missing data and outliers. In addition, for each survey data, the mean and standard deviation was computed for the two sample populations. For this the "outlier labeling rule" was used. All values outside the calculated range were considered outliers (Hoaglin & Iglewicz, 1987). The data was then analyzed using statistical software SAS. All analysis used an alpha level of .05 to determine significance.

3.10.1.2. Pearson's r Correlational Coefficient for Research Question 1

After scrubbing the data, Statistical Analysis Software (SAS) was used to report the descriptive statistics of survey responses for each of the courses (CS I/CS II) per institution. To address the five research hypotheses of this study, the five subscales of this study were analyzed using Pearson's r correlation coefficient and independent t-tests. As the visual investigation of the survey responses with a scatterplot confirm normal distribution of data, the Pearson's r correlation coefficient parametric test was chosen to measure the strength and direction of the linear relationship between each of the three predictor variables (DU, PAK and CA) and the decision to persist. The outcome variable, the decision to persist, is a response from Likert-scale

“This course has helped me decide CS major is not for me.” The bivariate statistical model Pearson’s r correlation was chosen to evaluate the first research question that states the hypothesis of linear relationship existence between each of the predictor variable DU, PAK, CA and decision to persist after completing the first two required CS courses. Pearson's r correlation was chosen for this research question because we are comparing two quantitative continuous variables with normally distributed data which is used to predict and find the strength and direction of a linear relationship between these two independent variables. There are three possible research hypotheses for this statistical model, and they are: 1. There is a positive linear relationship. 2. There is a negative linear relationship. 3. There is no linear relationship between the two variables.

The Pearson's r Correlation Coefficient statistical is calculated by dividing the covariance of the two samples, x and y, by the product of the two variables' standard deviations and by examining the Pearson’s r correlation coefficient and its p-value.

$$r_{xy} = \frac{n\Sigma XY - \Sigma X \Sigma Y}{\sqrt{[n\Sigma X^2 - (\Sigma X)^2][n\Sigma Y^2 - (\Sigma Y)^2]}}$$

- rxy= strength of the correlation between variables x and y
- n = sample size
- ∑ = sum of what follows…
- X = every x-variable value
- Y = every y-variable value
- XY = the product of each x-variable score and the corresponding y-variable score

The steps this research followed

1. State the null hypothesis,
2. Select the probability of type I error level (alpha = 0.05)
3. Calculate correlation coefficient, r
4. Test the statistical significance with p-value to reject/accept the null hypothesis (Prob > |r| under H0: Rho=0)

   $H0$: $\mu 1 = \mu 2$ (No relationship b/n the two variables)

   $H1$: $\mu 1 \neq \mu 2$ (Relationship exists b/n the two variables)

3.10.1.3. Independent T-Test for Research Question 2

Two-tailed independent t-tests, also called two-sample t-test, were applied to evaluate the existence of gender and institutional differences in three predictor variables of DU, PAK and CA perceptions. This inferential statistical test was chosen to evaluate this research question number two to determine the existence of a statistically significant difference between the means in two unrelated groups, male and female as well as NDSU and AAU. In order to run an independent t-tests, we need one independent, categorical variable that has two groups (gender and institution, respectively) and one continuous Likert scale dependent variable (DU, PAK, CA). The null hypothesis for the independent t-test is that the population means from the two unrelated groups are equal. This research evaluates if the null hypothesis can be rejected and the alternative hypothesis, which is that the population means are not equal, can be accepted. An alpha level of .05 was set to determine significance level that allows to either reject or accept the alternative hypothesis.

The following two questions were addressed:

- Is gender difference evident in the perceptions of the three variables identified (DU, PAK, CA)?

- Is institutional difference evident in the perceptions of the three variables identified (DU, PAK, CA)?

Comparing the male and female means, $\mu 1$ and $\mu 2$, and that $\delta = \mu 1 - \mu 2$, the null hypothesis for comparing the two means is $H0$: $\mu 1 = \mu 2$. The alternative hypothesis $H1$: $\mu 1 \neq \mu 2$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

This t-statistic follows a t distribution with $n1 + n2 - 2$ degrees of freedom.

- Ho: There is no gender difference in the perceptions of DU, PAK, CA during the two introductory courses.
- Ho: There is no institutional difference evident in the perceptions of the three variables identified (DU, PAK, CA)

To evaluate the assumption of equal variances in each group the research used the Folded F-test for equality of variances. The null hypothesis for this test is that the variances are equal. The F-value is calculated as a ratio of the greater of the two variances divided by the lesser of the two variances. Thus, if the null hypothesis is true, F will tend to be close to 1.0 and the p-value for F will be statistically nonsignificant ($p > 0.05$).

3.10.2. Qualitative Analysis

The qualitative research is used to unearth the lived experiences of female undergraduates in computing. A unique technique of sentiment analysis and topic modelling from the domain of natural language processing (NLP) was used to bring objectivity to the interpretation of personal interviews. This relatively novel approach of using quantitative analysis on qualitative data is useful in deriving insights by finding patterns and building a simple linear model to explain the variation in sentiment pattern and topic clusters. A dataset of interviews was curated, and an Exploratory Data Analysis (EDA) was performed on it to find patterns and improve understanding. This process mitigated the researcher's unconscious biases and the subjectivity often attached to interview analysis (Pang, 2008).

The audio-recorded and automatically transcribed answers to the questions paused to the upperclassmen of NDSU and AAU undergraduate computer science/software engineering female students (included in Appendix E). The qualitative strategy implemented for this research was a combination of text analysis techniques used to automatically extract valuable insights from unstructured auto-transcribed interview text data through sentiment analysis and topic cluster modelling. As Figure 5 shows, this quantitative analysis of qualitative data involved data acquisition, text preprocessing, text transformation, categorization, visualization, sentiment analysis, and text clustering/modeling looking for interesting patterns in the extracted text.

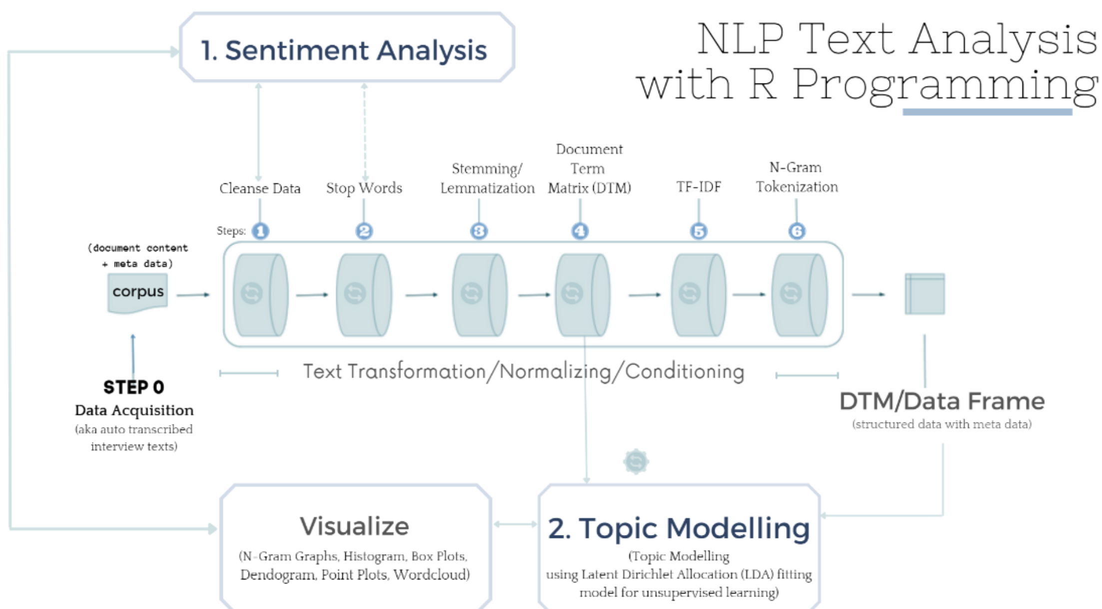


Figure 5: Text analysis processes for sentiment analysis and topic modelling.

3.10.2.1. Sentiment Analysis using Natural Language Processing (NLP)

Sentiment analysis techniques were applied to answers provided by respondents to a set of questionnaires, averaging the value to derive the final sentiment of the interview. Sentiment analysis is an application of NLP focused on analyzing sentiments of datasets where sentiments are quantified with a positive, negative, or neutral value called polarity. The *sentimentr* package of the R programming language was used to compute the sentiment polarity score of the transcribed text. The sentiment polarity score range of –1 to 1 for an element defines the orientation of the expressed sentiment, i.e., it determines if the text expresses the positive, negative or neutral sentiment of the user about the entity in consideration.

20 hours of recorded audio interview were transcribed to text which served as input document for this research sentiment analysis. The dataset was curated by segregating the interview transcripts based on a set of questionnaires related to the analysis. The question paused for the sentiment analysis was, "how was your experience in this course?" Students who completed the course CS I/CS II in the Fall 2021 from the two institutions were asked this question as part of the questionnaire as well as the 20 female interview participants who took the course before. The data was imported into R using excel library function. 422 sentences were extracted as excel responses. The text transformation step involves a series of iterative text pre-processing which involved removal "stopwords" from sentences such as prepositions, pronouns, possessive adjectives, and in general words that doesn't provide meaning, tokenizing, stemming and creating a document term matrix from the corpus.

The average sentiment value was derived for each participant which forms the baseline data for further analysis. The set of average sentiments was further investigated in this dataset using descriptive methods, and visualization to improve this research's understanding. This understanding led to further performing hypothesis-building and validation to unearth the factors influencing the sentiments and establish a statistical relationship. With the proven statistical relationship, independent t-test was run to explain variance in average sentiment which garnered additional observations and insights.

Sentiment analysis is a very powerful tool to gain great insights from extracted data. However, sentiments can be incorrectly interpreted without appropriate context such as polysemy (text containing several meanings), multi-polarity (text containing varying sentiment) and sarcasm and irony. This research evaluated several algorithmic approaches to try and overcome some of these problems. The chosen algorithm for this study is Sentimentr which corrects for inversions, meaning that while a more basic sentiment analysis would judge "*I am not good*" as positive due to the adjective good, Sentimentr recognizes the inversion of good and classifies it as negative. It is an augmented dictionary lookup approach that tries to incorporate weighting for valence shifters (i.e., negators, amplifiers (intensifiers), de-amplifiers (downtoners), and adversative conjunctions). A negator or and adversative conjunctions flips the sign of a polarized word (e.g., "*I do not like it.*").

The equation used by the algorithm to assign value to polarity of each sentence first utilizes a sentiment dictionary (Jockers, 2017) to tag polarized words. The model builds up a representation of whole sentences based on the sentence structure. Each sentence is broken into an ordered bag of words. The words in each sentence are searched and compared to a dictionary of polarized words in the lexicon package. Positive and negative words are tagged with a +1 and −1 respectively. Neutral is 0. These will form a polar cluster which is a subset of the sentence. Neutral words hold no value in the equation but do affect word count. Each polarized word is then weighted based on the weights from the polarity argument and then further weighted by the function and number of the valence shifters directly surrounding the positive or negative word. This sentiment model takes into account valence shifters 1. Negators - A negator flips the sign of a polarized word adjacent to it (e.g., "I do not like it."). 2. Amplifiers (intensifiers) - increases the impact of a polarized word (e.g., "I really like it."), 3. de-amplifiers (downtoners) - reduces the impact of a polarized word (e.g., "I hardly like it.") 4. adversative conjunctions - overrules the previous clause containing a polarized word (e.g., "I like it but it's not worth it.") dictionary lookup. More work is needed in future NLP sentiment analysis algorithm models in understanding of sarcasm in large dataset.

3.10.2.2. Topic Modeling with LDA

The second qualitative analysis strategy implemented answered the third research question to identify and enumerate the contributing factors of success for female Computer Science/Software Engineering undergraduates. A question was paused to twenty female upperclassmen undergraduates of Computer Science and Software Engineering majors who had voluntarily opted to be interviewed for this purpose. The question was "*What would you say helped you persist the first two introductory courses and beyond to become a soon to be graduating class of CS/SE?*" The interview responses were auto transcribed using Microsoft's A.I. powered Office 365 auto transcription tool.

The Latent Dirichlet Allocation (LDA) topic modelling was applied to the auto transcribed documents which was used as input corpus. Two distinct LDA topic modeling scripts were written in R programming language for the topicmodels package and procedures adapted from Grün and Hornik (Grün and Hornik 2011). The procedures and scripts included standard text preprocessing of the corpus: punctuation removal, lemmitizing, stopword removal, and the like as outlined in the methodology section. The LDA topic modelling algorithm was used to build data points, estimate probabilities, cluster similar terms together, and for fitting topic model in discovering abstract "topics" that occur in a corpus for the unsupervised classification of documents.

The LDA topic clustering algorithm doesn't identify the overarching topic for the collections of word tokens. The word tokens provide some hints on what the topic the cluster of terms represents but reading through the interview transcripts was necessary to proximately identify the semantic topics represented by the LDA fitting model. The researcher's experience reflects the observations of Hindle (Hindle et al. 2014), who suggested that domain expertise is necessary to accurately label topics. Defining semantically meaningful topic labels required the non-trivial effort in applying the qualitative method of manually creating inductive coding categories. As a result, the researcher first conducted a manual inductive thematic analysis coding technique to organize, and structure content to aid with the assigning of more distinctive labels to the LDA fitted topics.

The Algorithm to find the latter

- Go through each document and randomly assign each word in the document to one of k topics (k is chosen beforehand).
- For each document d, go through each word w and compute:

1. p (topic t | document d): the proportion of words in document d that are assigned to topic t. Tries to capture how many words belong to the topic t for a given document d. Excluding the current word.

   If a lot of words from d belongs to t, it is more probable that word w belongs to t.

   (#words in d with t +alpha/ #words in d with any topic+ k*alpha)

2. p(word w| topic t): the proportion of assignments to topic t over all documents that come from this word w. Tries to capture how many documents are in topic t because of word w. LDA represents documents as a mixture of topics. Similarly, a topic is a mixture of words. If a word has high probability of being in a topic, all the documents having w will be more strongly associated with t as well. Similarly, if w is not very probable to be in t, the documents which contain the w will be having very low probability of being in t, because rest of the words in d will belong to some other topic and hence d will have a higher probability for those topic. So even if w gets added to t, it won't be bringing many such documents to t.

- Update the probability for the word w belonging to topic t, as
- p(word w with topic t) = p(topic t | document d) * p(word w | topic t)

Topic modelling method "probabilistic models for uncovering the underlying semantic structure of a document collection based on a hierarchical Bayesian analysis of the original texts" [4].

An LDA topic model is defined by two probability distributions:

- $\theta \sim \text{Dirichlet}(\alpha)$: representing the topic distribution over documents
- $\phi \sim \text{Dirichlet}(\beta)$: representing the word distribution over topics

The generated topic model reveals: 1) the latent topics (distributions of words) throughout the corpus, and 2) which topics are most prevalent in each document. Since topics are merely collections of words, it is incumbent upon the analyst to assign semantic meaning to the topics. Thus, the meaningfulness of topics is dependent upon both the semantic cohesion of the topics and the domain knowledge of the analyst.

Sentiment analysis is the computational task of automatically determining what feelings a writer is expressing in text. Sentiment is often framed as a binary distinction (positive vs. negative)

3.11. Researcher Positionality

I am a Software Engineer by profession. I have worked in the technology industry for over 20 years. I have taught CS I – as a part-time adjunct professor. Anchored in the pursuit of gender equity, I founded a non-profit organization called uCodeGirl six years ago to help inspire, equip and empower young women to pursue Computer Science in colleges. I will likely have some unconscious and conscious biases from my workplace experience and advocacy for diversity.

CHAPTER 4. RESULTS

4.1. Introduction

This chapter will present the findings of the parallelly convergent mixed-methods research in investigating predictive factors of success for female undergraduates in Computer Science at NDSU and Software Engineering at AAU.

4.2. Quantitative Findings

This section will present the quantitative findings of this study. The quantitative findings were a direct result of the statistical analysis of data collected through online survey of the sample population for a closer look into the correlation between predictor variables and persistence to answer the following two research questions:

1. **RQ1**: Does a relationship exist between the students' perceptions of their Degree's Usefulness (DU), their Previous Acquired Knowledge (PAK), and their Cognitive Attitude (CA) with their decision to pursue Computer Science/Software Engineering degree after completing the first two introductory courses?
2. **RQ2**: Is gender or institutional difference evident between the three variables identified (DU, PAK, CA) and the decision to pursue in CS/SE major after completing the first two introductory courses?

The demographics of the sample population, and the data analysis are presented here.

4.2.1. Demographics

4.2.1.1. North Dakota State University (NDSU)

Approximately 221 students were enrolled in CS I (Computer Science I) and CS II (Computer Science II) at North Dakota State University during the fall of 2021. A total of 164 students from CS I and from CS II voluntarily participated in the pre-/post- survey. As is the case in most computer science courses, the percentage of students identified as females was incredibly low. Only 16 of the total students who chose to participate in the study identified as females (approximately 11%) and 89% identified as male. In addition to this binary statistic, the sample population also included students who identified as others (these identifiers were grouped together to ensure privacy, as it was a much smaller sample size). CS I is the first programming class required for the Computer Science major and uses java as the programming language. The following percentages represent how the CS I sample was classified by year in school: 71% freshmen, 15% sophomores, 8% juniors, 2% seniors, and 5% others. Of the respondents enrolled in the CS I class, 90% were computer science majors, 4% were computer engineering majors, 2% were mathematics majors, 2% mechanical engineering majors and 2% were from math education. CS II is the second required programming class for computer science major focused on data structure. The instructional language used was java programming however students can use java or C++ to complete the course assignments, projects, or exams. The following percentages represent how the CS II sample was classified by year in school: 16% freshmen, 40% sophomores, 29% juniors, 10% seniors, and 4% others. Of the respondents enrolled in the CS II class, 62% were computer science majors, 34.5% were computer & electrical engineering majors as well as other various majors including but not limited to physics, statistics, Management Information Systems (MIS) 3.5% were undecided.

The discovery that only 11% of NDSU students and 12% of AAU students who had participated in CS I & CS II survey were female was not unexpected. The low percentage is consistent with the low enrollment numbers of females in the major. This also supports the widely discussed concern that there is an acute shortage of women in computer science.

4.2.1.2. Addis Ababa University (AAU)

Approximately 443 students were enrolled in the equivalent courses of NDSU's CS I (Computer Science I) and CS II (Computer Science II) at Addis Ababa University during the fall of 2021. There were 164 students from CS I and 106 students from CS II who voluntarily participated in the pre-/post- survey. As is the global case in most computer science courses, the percentage of students identified as females was incredibly low. Only 20 of the total students who chose to participate in the study identified as females (approximately 12%) and 88% identified as male. CS I is the first programming class required for the Pre-Engineering students and uses C++ as the programming language. 100% of respondents were first-year students as students are not allowed to advance to the next course without completing this course in the same cohort at AAU. Of the respondents enrolled in the CS I class, 100% were in pre-engineering. CS II is the other required programming class for software engineering major focused on data structure. The instructional language used was java programming. Of the respondents enrolled in the CS II class, 100% were software engineering majors. Results of the statistical tests performed on each of the research question hypothesis are presented below.

4.2.2. Research Question I: Pearson's r Correlation

RQ 1: What relationship exists between the students' perceptions of their Degree's Usefulness (DU), their Previous Acquired Knowledge (PAK), and their Cognitive Attitude (CA) with their decision to pursue Computer Science/Software Engineering degree after completing the first two introductory courses?

A correlation matrix was generated to examine how each of the 3 variables correlated with the outcome variable which is the decision to persist after completing each of the first two introductory courses.

4.2.2.1. Degree's Usefulness (DU) and Decision to Persist

- Ho: There is no correlation between students' perception of the usefulness of their degree's and their decision to persist in CS/SE degrees after completing the first two introductory CS courses.

Table 1: Pearson's correlation table for the perception of Degree's Usefulness (DU) for courses CS I and CS II measured for NDSU and AAU, respectively.

Pearson Correlation	DU (DU_mn_a)	
	NDSU	**AAU**
CS I (CR_11b)	r(57) = .11, p = .614	r(98) = -.37, p = .110
CS II (CR_10b)	r(85) = -.65, p <.0001	r(43) = -.18, p = .666

Table 1 shows the Pearson correlation coefficient, denoted by r, computed to assess the linear relationship between students' perception of degree's usefulness (DU) and the decision to persist with the Computer Science for the case of NDSU or Software Engineering degree for the case of AAU for courses CS I and CS II.

There was a strong positive correlation between DU and students taking CS II at NDSU, $r(85) = -.65$, $p < .0001$. Therefore, the null hypothesis was rejected for NDSU CS II course. For all other variables, the null hypothesis is accepted.

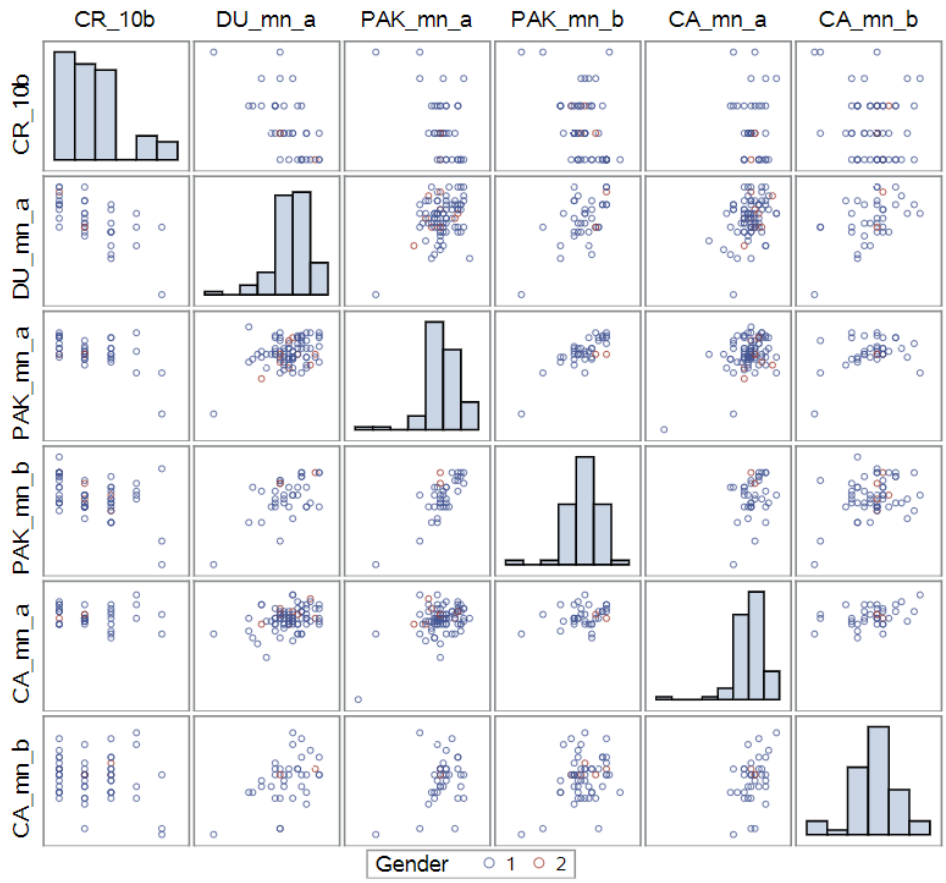


Figure 6: Pearson's correlation plots distribution for the perception of DU, PAK and CA for course CS II as measured for NDSU.

4.2.2.2. Previously Acquired Knowledge (PAK) and Decision to Persist

- Ho: There is no correlation between students' perception of the influence of their previously acquired knowledge and their decision to persist in CS/SE degrees after completing the first two introductory CS courses.

Table 2: Pearson's correlation table for the perception of the influence of PAK for courses CS and CS II measured for NDSU and AAU, respectively.

Pearson Correlation	PAK (PAK_mn_a/PAK_mn_b)			
	NDSU		AAU	
	Pre-	Post-	Pre-	Post-
CS I (CR_11b)	r(56) = .13, p = .550	r(47) = .18, p = .215	r(107) = -.23, p = .288	r(163) = .17, p = .029
CS II (CR_10b)	r(86) = -.55, p = .001	r(55) = -.44, p = .001	r(45) = -.16, p = .679	r(52) = .15, p = .284

Table 2 shows the Pearson's correlation coefficient computed to assess the linear relationship between students' perception of influence of previously acquired knowledge and the decision to persist with the Computer Science for the case of NDSU or Software Engineering degree for the case of AAU for courses CS I and CS II.

- NDSU students exhibited a moderately strong positive correlation between PAK (pre-survey) and CS II CR_10b (decision to persist), $r(86) = -.55$, $p = .001$.
- NDSU students exhibited a moderately strong positive correlation between PAK (post-survey) and CS II CR_10b (decision to persist), $r(55) = -.44$, $p = .001$.
- AAU students did not exhibit a correlation between PAK (Pre/Post-survey) and CS I (CR_11b) or CS II (CR_10b).

Therefore, the null hypothesis was rejected for NDSU CS II course for both Pre-/Post-PAK observations and AAU CS I. For all other variables, the null hypothesis is accepted.

4.2.2.3. Cognitive Attitude (CA) and Decision to Persist

Ho: There is no correlation between students' perception of their cognitive attitudes and their decision to persist in CS/SE degrees after completing the first two introductory CS courses.

Table 3: Pearson's correlation table for the perception of CA for courses CS I and CS II measured for NDSU and AAU, respectively.

Pearson Correlation	CA (CA_mn_a/CA_mn_b)			
	NDSU		AAU	
	Pre-	Post-	Pre-	Post-
CS I (CR_11b)	r(56) = .06, p = .784	r(45) = .14, p = .365	r(101) = .12, p = .618	r(162) = .11, p = .179
CS II (CR_10b)	r(85) = -.15, p = .390	r(55) = -.16, p = .229	r(42) = .17, p = .667	r(46) = .24, p = .103

Table 3 shows the Pearson correlation coefficient computed to assess the linear relationship between students' perception of cognitive attitude (CA) and the decision to persist with the Computer Science for the case of NDSU or Software Engineering degree for the case of AAU for courses CS I and CS II. No statistically significant relationship was evident in all the permutations of tests described in Table 3. Therefore, the null hypothesis was accepted for courses CS I and CS II for both NDSU and AAU.

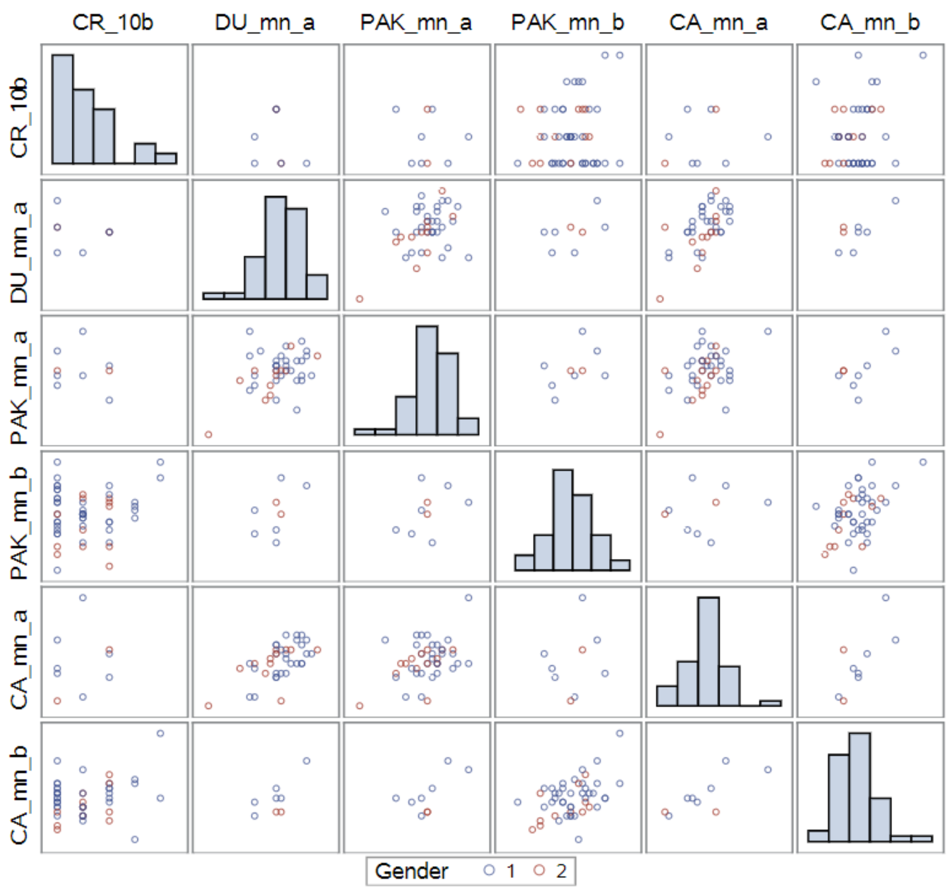


Figure 7: Pearson's correlation plots distribution for the perception of DU, PAK and CA for course CS II as measured for AAU.

4.2.3. Research Question II: Independent t-tests

An independent two-sample t-test was conducted to evaluate if the perceptions of the three variables DU, PAK and CU were significantly different across gender and institutions.

4.2.3.1. Gender Differences in Degree's Usefulness (DU)

There are two parts to the second research question: Gender and Institutional difference.

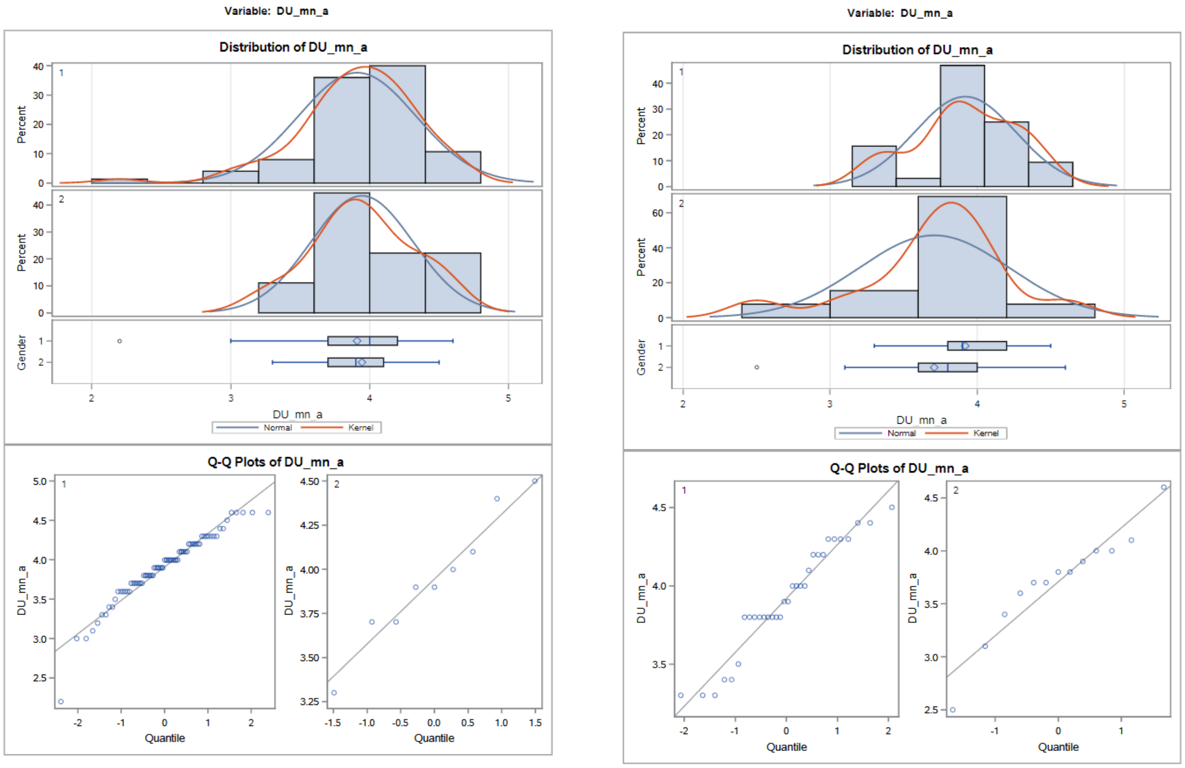
RQ 2: Is gender difference evident in the perceptions of the three variables identified (DU, PAK, CA)?

- Ho: There is no gender difference in the perceptions of DU during the two introductory courses.

Table 4: Two-tailed independent t-tests, gender comparison with respect to Degree's Usefulness (DU) for courses CS I and CS II measured for NDSU and AAU, respectively.

Gender (t-test)	DU (DU_mn_a)			
	NDSU		AAU	
CS I	M (3.87, 0.43) F (3.84, 0.34)	t(56) = .18, p = .536	M (3.93, 0.38) F (3.99, 0.39)	t(98) = .62, p = .826
CS II	M (3.91, 0.42) F (3.94, 0.37)	t(82) = .24, p = .709	M (3.92, 0.34) F (3.71, 0.51)	t(43) = 1.62, p = .079

Table 4 describes the series of two-sample t-tests that was performed to compare perceptions of Degree's Usefulness in male vs female. Examining the CS II course offered at NDSU, there was not a significant difference in perceptions of Degree's Usefulness between male (M = 3.91, SD = 0.42) and female (M = 3.94, SD = 0.37); t(82) = 0.24, p = .709. No statistically significant difference was evident in all the permutations of tests described in Table 4. Therefore, the null hypothesis was accepted for courses CS I and CS II for both NDSU and AAU.



(a) (b)

Figure 8: (a) Distribution of Degree's Usefulness (DU) by gender for CS II at NDSU; (b) Distribution of Degree's Usefulness (DU) by gender for CS II at AAU.

4.2.3.2. Gender Differences in Previously Acquired Knowledge (PAK)

Ho: There is no gender difference in the perceptions of PAK during the two introductory courses.

Table 5: Two-tailed independent t-tests, gender comparison with respect to PAK for courses CS I and CS II measured for NDSU and AAU, respectively.

Gender (t-test)	Previously Acquired Knowledge (PAK)							
	NDSU				AAU			
	Pre-(PAK_mn_a)		Post-(PAK_mn_b)		Pre-(PAK_mn_a)		Post-(PAK_mn_b)	
CS I	M (2.95, .38) F (3.07, .36)	t(55) = .82, p = .890	M (2.82, .30) F (3.02, .14)	**t(46) = 1.67, p = .051**	M (2.92, .45) F (2.55, .40)	t(107) = 3.57, p = .626	M (3.06, .45) F (2.74,.41)	t(163) = 3.03, p = .674
CS II	M (3.83, .59) F (3.69,.48)	t(83) = .73, p = .532	M (3.96,.49) F (4.08,.38)	t(54) = .53, p = .666	M (3.69, .36) F (3.45, .44)	t(54) = 1.98, p = .361	M (3.79,.54) F (3.43,.64)	t(52) = 2.00, p = .406

Table 5 describes the series of two-sample t-tests that was performed to compare perceptions of PAK in male vs female. There was a significant difference in perceptions of the influence of PAK between male ($M = 2.82$, $SD = .30$) and female ($M = 3.02$, $SD = 0.14$); $t(46) = 1.67$, $p = .051$ for course CS I at NDSU. Therefore, the null hypothesis was rejected for NDSU CS I course.

No statistically significant difference was evident in the other permutations of tests described in Table 5. Therefore, the null hypothesis was accepted for courses CS I and CS II for both NDSU and AAU.

4.2.3.3. Gender Differences in Cognitive Attitude (CA)

Ho: There is no gender difference in the perceptions of CA during the two introductory courses.

Table 6: Two-tailed independent t-tests, gender comparison with respect to CA for courses CS I and CS II measured for NDSU and AAU, respectively.

Gender (t-test)	Cognitive Attitude (CA)							
	NDSU				AAU			
	Pre-(CA_mn_a)		Post-(CA_mn_b)		Pre-(CA_mn_a)		Post-(CA_mn_b)	
CS I	M (3.13, .27) F (3.06, .29)	t(55) = .69, p = .633	M (3.19, .35) F (3.15, .33)	t(44) = .28, p = 1.00	M (3.28, .35) F (3.20, .39)	t(101) = .92, p = .410	M (3.47, .41) F (3.26,.47)	t(162) = 2.10, p = .385
CS II	M (3.21, .40) F (3.36,.23)	t(82) = 1.14, p = .089	M (3.22,.05) F (3.33,.08)	t(54) = .70, p = .014	M (3.39, .39) F (3.25, .34)	t(42) = 1.22, p = .584	M (3.39,.36) F (3.18,.34)	t(46) = 1.78, p = .818

Table 6 describes the series of two-sample t-tests that was performed to compare perceptions of CA in male vs female. There was a significant difference in perceptions of CA (post-survey) between male (M = 3.22, SD = .05) and female (M = 3.33, SD = 0.08); t(54) = .70, p = .014 for course CS II at NDSU. Therefore, the null hypothesis was rejected for NDSU CS II post survey course.

No statistically significant difference was evident in the other permutations of tests described in Table 6. Therefore, the null hypothesis was accepted for courses CS I and CS II for both NDSU and AAU.

4.2.3.4. Institutional Differences

- Ho: There is no institutional difference evident in the perceptions of the three variables identified (DU, PAK, CA)

Table 7: Two-tailed independent t-tests, institutional comparison using two sample t-test for Degree's Usefulness (DU).

Institution (t-test)	**DU (DU_mn_a)**	
CS I	AAU (3.94, .39) NDSU (3.86, 0.41)	t(157) = 1.2, p = .596
CS II	AAU (3.79, 0.41) NDSU (3.91, 0.42)	t(185) = 1.97, p = .763

A two-sample t-test was performed to compare perceptions of Degree's Usefulness (DU) in AAU and NDSU. There was no significant difference in perceptions of the influence of DU between male and female for course CS I and CS II at AAU and NDSU. Therefore, the null hypothesis was accepted.

Table 8: Two-tailed independent t-tests, institutional comparison using two sample t-test for Previously Acquired Knowledge (PAK).

Institution (t-test)	Previously Acquired Knowledge (PAK)			
	Pre- (PAK_mn_a)		Post- (PAK_mn_b)	
CS I	AAU(2.84, .46) NDSU (2.97, .38)	t(165) = 1.86, p = .127	AAU (3.02, .46) NDSU (2.85, .29)	t(212) = 2.51, p = .0004
CS II	AAU (3.62, .39) NDSU (3.84,.58)	t(133) = 2.43, p = .005	AAU (3.71,.58) NDSU (3.96,.48)	t(109) = 2.49, p = .175

A two-sample t-test was performed to compare perceptions of Previously Acquired Knowledge (PAK) in AAU and NDSU. There was a significant difference in perceptions of the influence of PAK between AAU (M = 3.62, SD = .39) and NDSU (M = 3.84, SD = 0.58); t(133) = 2.43, p = .005 for course CS II; and PAK and AAU (M = 3.02, SD =.46) and NDSU (M = 2.85, SD = 0.29); t(212) = 2.51, p = .0004 for CS I. Therefore, the null hypothesis was rejected for courses CS I (PAK post-) and CS II (PAK pre-) for both NDSU and AAU.

Table 9: Two-tailed independent t-tests, institutional comparison using two sample t-test for Cognitive Attitude (CA).

Institution (t-test)	Cognitive Attitude (CA)			
	Pre- (CA_mn_a)		Post- (CA_mn_b)	
CS I	AAU(3.27, .36) NDSU (3.13, .28)	t(159) = 2.41, p = .046	AAU (3.44, .42) NDSU (3.19, .33)	t(209) = 3.85, p = .079
CS II	AAU (3.34, .37) NDSU (3.24,.41)	t(129) = 1.33, p = .171	AAU (3.35,.36) NDSU (3.22,.32)	t(103) = 1.87, p = .394

A two-sample t-test was performed to compare perceptions of Cognitive Attitude (CA) in AAU and NDSU. There was a significant difference in perceptions of the influence of CA between AAU (M = 3.27, SD = .36) and NDSU (M = 3.13, SD = 0.28); t(159) = 2.41, p = .046 for CS I pre-survey course. Therefore, the null hypothesis was rejected for courses CS I for CA

(Pre-). No other institutional difference was exhibited for CS II. Therefore, the rest of the null hypothesis was accepted.

4.2.3.5. Cognitive Gains in Previously Acquired Knowledge (PAK)

Table 10: Paired t-tests, institutional comparison using two sample t-test for Cognitive Attitude (CA).

Pre-/Post- (paired t-test)	**PAK & CA Pre-/Post- Comparisons**			
	NDSU		**AAU**	
CS I	PAK (.23, .25)	t(24) = 4.65, p = .0001	PAK(.18,.09)	t(19) = 1.88, p = .075
	CA(.03,.28)	t(22) = .56, p = .584	CA(.07, .51)	t(19) = .58, p = .566
CS II	PAK (.09, .33)	t(36) = 1.74, p = .089	PAK(.15,.52)	t(8) = .90, p = .396
	CA(.11,.34)	t(36) = 1.97, p = .057	CA(.11, .40)	t(8) = .83, p = .429

A paired samples t-test was performed to compare change in the perceptions of Previously Acquired Knowledge (PAK) in pre-survey and post-survey of CS I students at NDSU. There was a significant difference in PAK ($M = .23$, $SD = .25$); $t(24) = 4.65$, $p = .0001$. Therefore, the null hypothesis was rejected for PAK CS I for NDSU. No other variable difference was observed. Therefore, the rest of the null hypothesis was accepted.

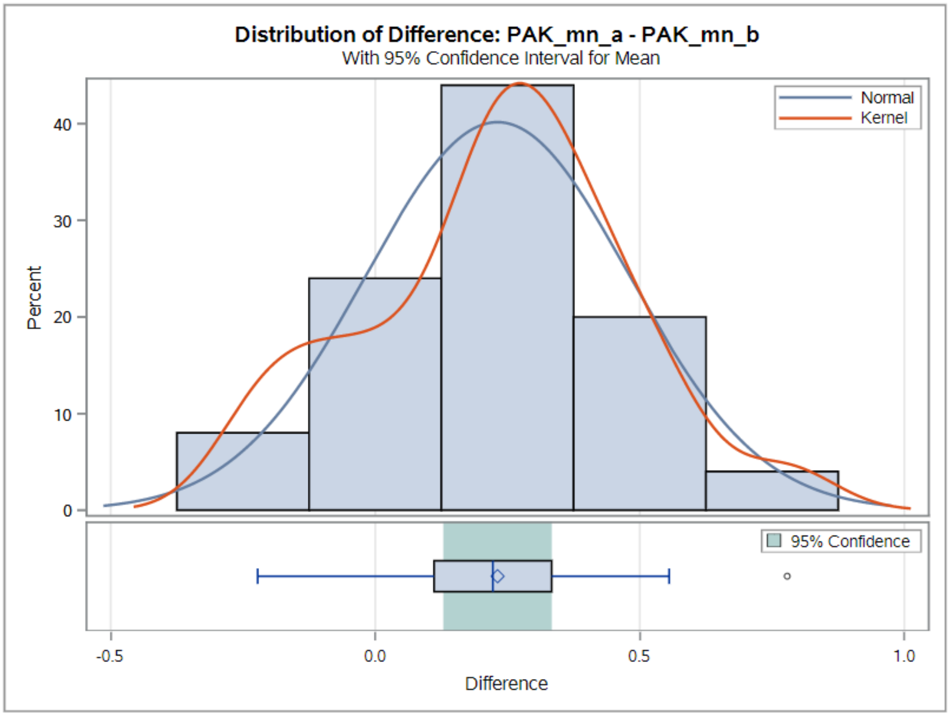


Figure 9: Pair t-test for pre-/post- comparisons of PAK for CS I as measured for NDSU.

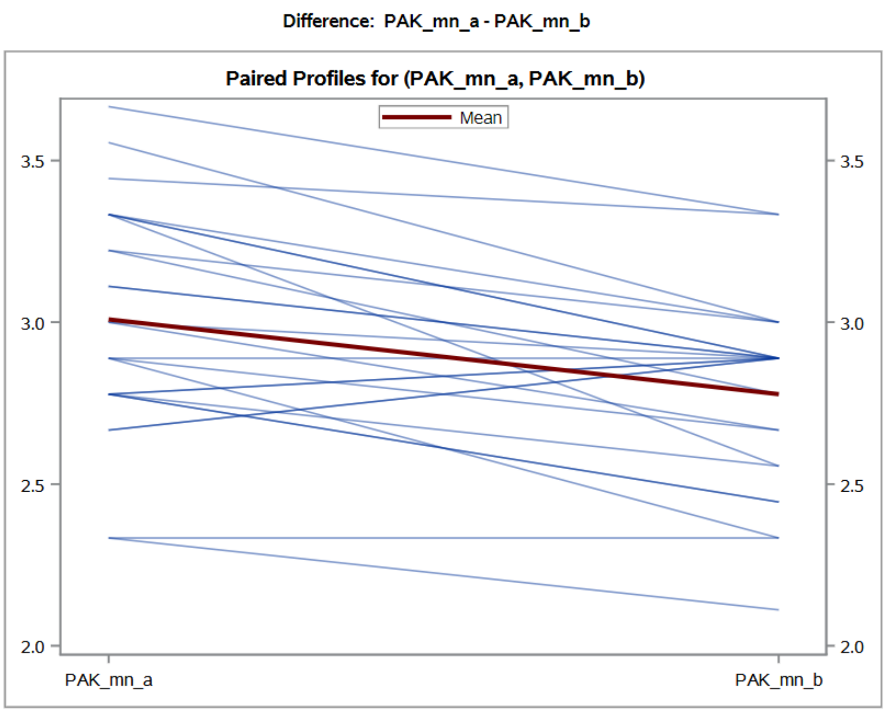


Figure 10: Pair t-test for pre-/post- comparisons of PAK for CS I as measured for NDSU.

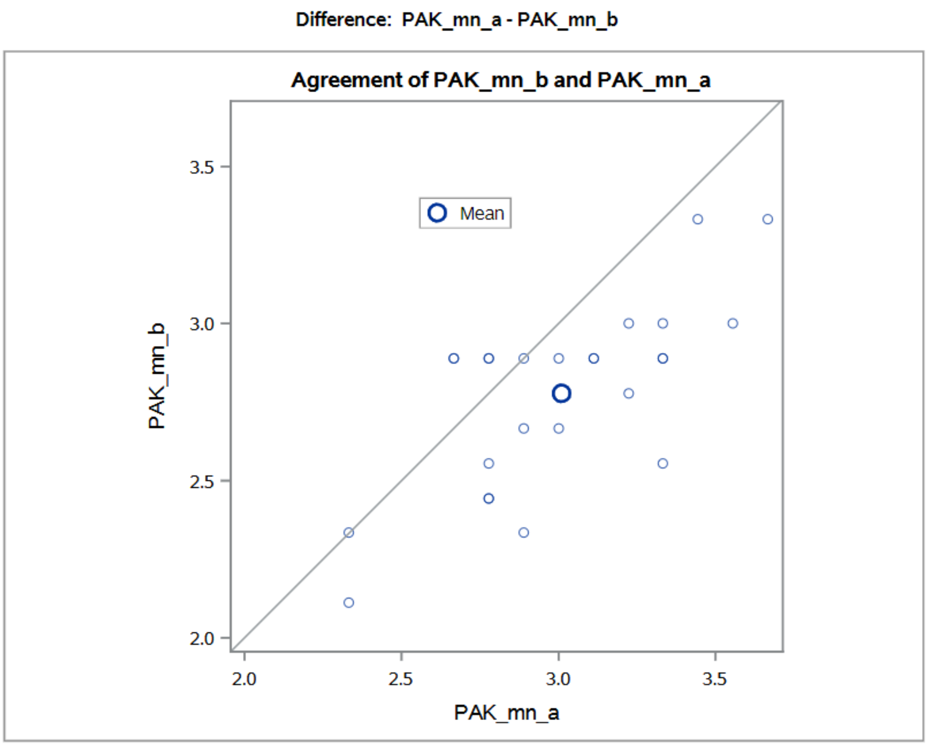


Figure 11: Pair t-test for pre-/post- comparisons of PAK for CS I as measured for NDSU.

4.3. Qualitative Findings

The third and last research question was answered through the qualitative part of this mixed-methods research.

- **RQ3**: Can factors that promote success in pursuit of a degree in Computer Science/Software Engineering for female undergraduates be identified and enumerated? Can persistence factors in pursuit of a Software Engineering degree in Ethiopia be predicted by factors reported to be effective in the U.S. Computer Science degree and vice-versa?

This finding had captured the essence of female perceptions of persistence towards computer science and software engineering degree. The results are organized into two sections: sentiment analysis of course reviews of CS I and CS II using Natural Language Processing

(NLP) and theme-based topic findings of factors that promote success using NLP topic cluster modelling.

4.3.1. NLP Sentiment Analysis

4.3.1.1. Research Question II

The qualitative data was collected through the female participants' interviews and the write-in portion of the online survey. Overall, 422 statements were extracted for analysis. The technique of sentiment analysis from the domain of NLP was used to provide a degree of objectivity to the interpretation of personal interviews and derive significant insights. This unique approach is at the intersection of social research, NLP, and analytics, which had a scant previous precedence. Sentiment analysis algorithm was applied to course reviews provided by the students, averaging the sentiment value to derive the final quantified sentiment of the interview. In total, the study conducted course reviews across two institutions covering 52 females and 376 males with the write-in portion of the online survey as well as curated interviews. Similar to the quantitative study, the vast majority of the respondents (88%) are male. The dataset consists of all CS and SE majors in CS I and CS II. With this analysis, we comprehend the pattern of sentiment about the course review and the demographics. This relationship provides further insights which are difficult to assimilate by simple qualitative analysis. To expand on the findings of the independent two sample t-test quantitative analysis on gender and institutional differences, the sentiment analysis focuses on the reflections of the perceptions and experiences of the students who completed the first two introductory computer science courses. The analysis starts by stating the null hypothesis as follows: -

- Ho: There is no gender difference in the perceptions of cognitive gains and performance from the two introductory courses.

The sentiment was computed using the R language library called Sentimentr. Each sentence is broken into an ordered bag of words. The words in each sentence are searched and compared to a dictionary of polarized words in the lexicon package. Positive and negative words are tagged with a +1 and −1 respectively. Neutral is 0. These will form a polar cluster which is a subset of the sentence. Neutral words hold no value in the equation but do affect word count. Each polarized word is then weighted based on the weights from the polarity argument and then further weighted by the function and number of the valence shifters directly surrounding the positive or negative word. The average sentiment value was derived for each participant which forms the baseline data for further analysis.

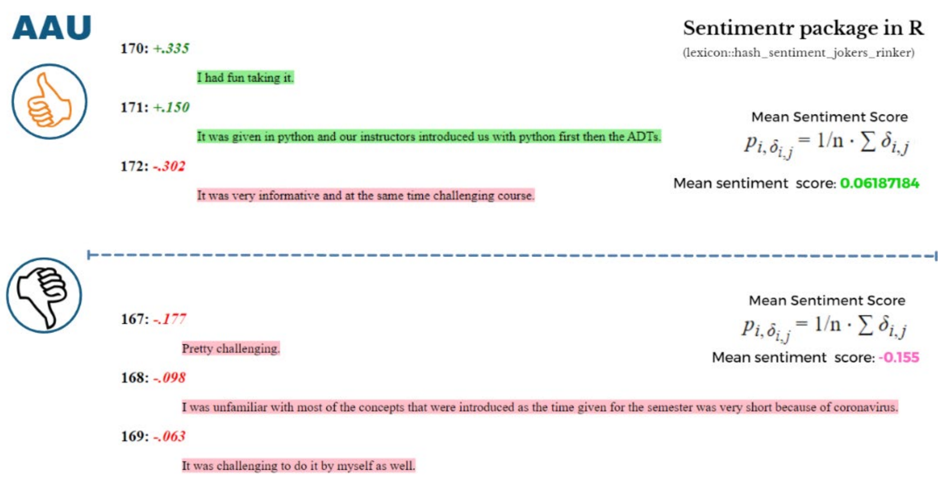


Figure 12: Mean sentiment score using Sentimentr package in R.

Figure 12 analyzes a sentiment of the review of two female students at AAU who took CS II. The first one had a positive sentiment score with a quote that reads, "*It was definitely a challenging class, but it was also rewarding, and I learned a lot of useful things.*" ~female, CS II, NDSU. Each sentence is computed separately. The model of Sentimentr computes the mean of the review response is the average of each sentence divided by the number of sentences.

Polarized word in each sentence was identified by the algorithm and given weight and the valence shifters are identified and amplify or negate the polarized word.

The second sample review response had a negative overall sentiment score, “Pretty challenging. I was unfamiliar with most of the concepts that were introduced, as the time given for the semester was very short because of coronavirus. It was challenging to do it by myself as well.” ~female, CS II, AAU

A comparative negative sentiment score, “It was very difficult. I did not receive a lot of help when I looked for it. I had to do a lot of self-teaching.” ~male, CS II, NDSU

A comparative positive sentiment score: “I had fun taking it. It was given in python and our instructors introduced us with python first then the ADTs. It was a highly informative and at the same time challenging course.” ~female, CS II. AAU

Gender	Institution	Word Count	SD	Average Sentiment
Female	AAU	395 (13%)	0.3819916	0.2215776
Male	AAU	2739 (87%)	0.3616715	0.3404326
Female	NDSU	113 (10%)	0.3617155	0.1475966
Male	NDSU	998 (90%)	0.4464442	0.2250517
other	NDSU	8	0	0.000000

Figure 13: Frequency distribution of sentiment polarity by gender and by institution.

Figure 13 shows a total of 4261 words were recorded on the question, “how was the course?” Traversing the word count column, the male population spoke more words than female population by 87-90% for both institutions. The average sentiment column reveals male tend to have a more positive attitude on course review than females.

```
       Stat    n
1     words 4261
2    tokens 4624
3 sentences  422
4 questions    0

$Attributes
      Attribute Count         Rate
1:      negator   118 0.0276930298
2:    amplifier    54 0.0126730814
3: de-amplifier     4 0.0009387468
4:  adversative    52 0.0122037080
5:     negative    73 0.0171321286
6:     positive   431 0.1011499648
7:    polarized   504 0.1182820934

$Polarized_Cooccurrences
   Valence_Shifter Cooccurrence
1:         negator  0.212962963
2:       amplifier  0.123456790
3:    de-amplifier  0.009259259
4:     adversative  0.132716049

$Cooccurrences
             negator amplifier de-amplifier adversative negative positive polarized
negator          102        22            2          21       25       55        69
amplifier         22        50            0          10        9       37        40
de-amplifier       2         0            4           1        1        2         3
adversative       21        10            1          50       16       39        43
negative          25         9            1          16       64       37        64
positive          55        37            2          39       37      297       297
polarized         69        40            3          43       64      297       324
```

Figure 14: Sentiment attributes of polarized words & valence shifters.

As Figure 14 shows, the extracted sentiment attributes had 504 polarized words and 431 of them were positive. The influence of adjacent pre-occurring words to the polarized words are known as valence-shifters. In this analysis, there were more negators than amplifiers. The occurrence of amplifiers, positive attributes with the presence of positive words makes the review skew to positive as shown below with the density distribution.

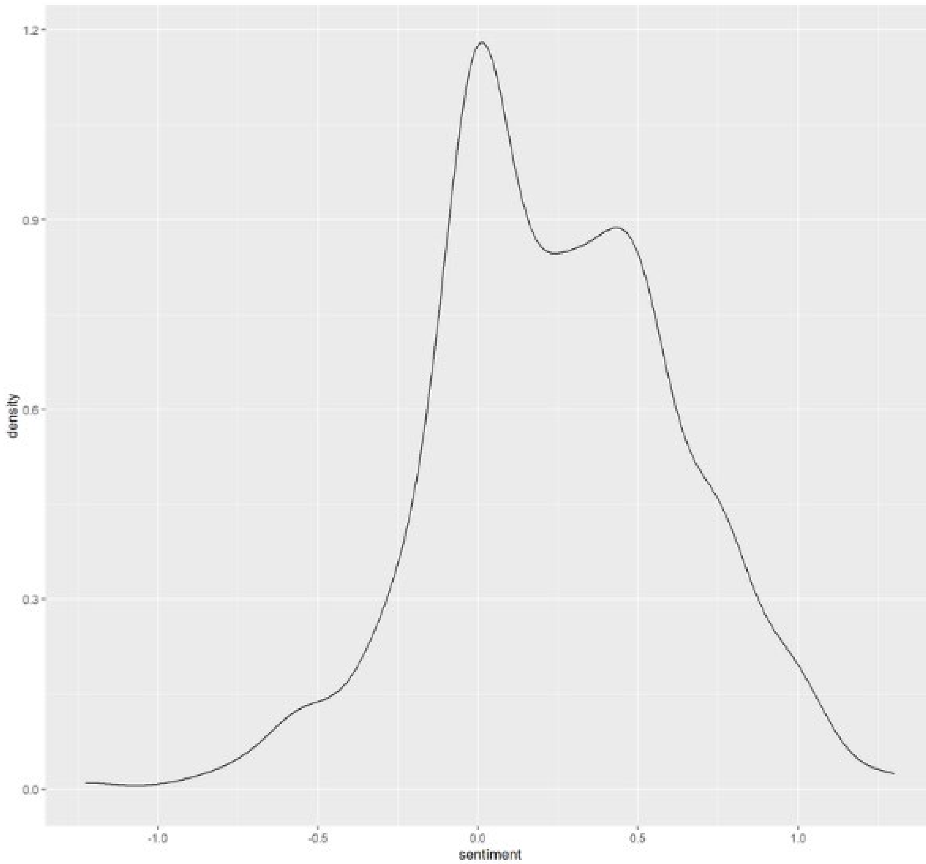


Figure 15: Density plot of average sentiment polarity.

The visualization provides all data points with differentiated markings for gender and institution while plotting the average sentiment over course experience. Figure 15 is the density plot of the average sentiments of each document or sentence of all the course review data. The distribution is normal with a concentration on neutrality with sentiment average of 0.0 and a strong deviation for positive sentiments. It shows a slightly negative but mostly neutral and positive review of the two courses.

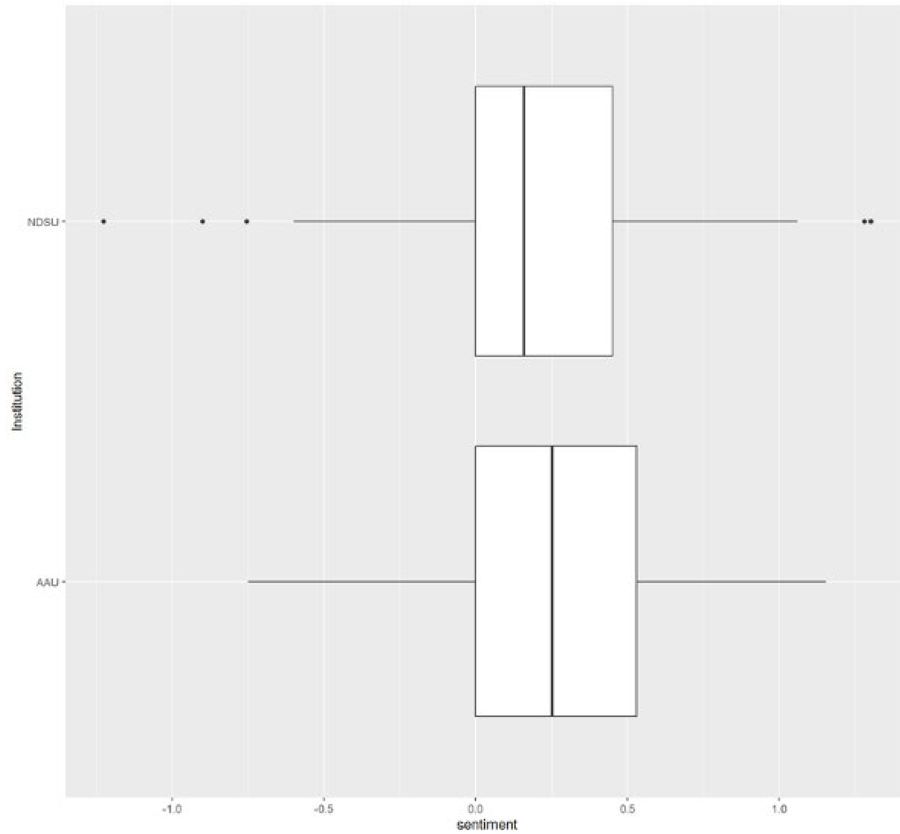


Figure 16: Sentiment polarity distribution by institution.

Figure 16 is the box plot of sentiment distribution as segmented by the two institutions. The number of participants is larger for the AAU students at 264 vs 57 for NDSU students. NDSU students are quite spread on the sentiment score with outliers on both ends with the lowest at -1.45 and the highest at 1.8 average sentiment scores on the courses. 50% of the pupils have an average sentiment score of 0.0 to 0.5. Most of the average sentiment cluster is 0.225 and above.

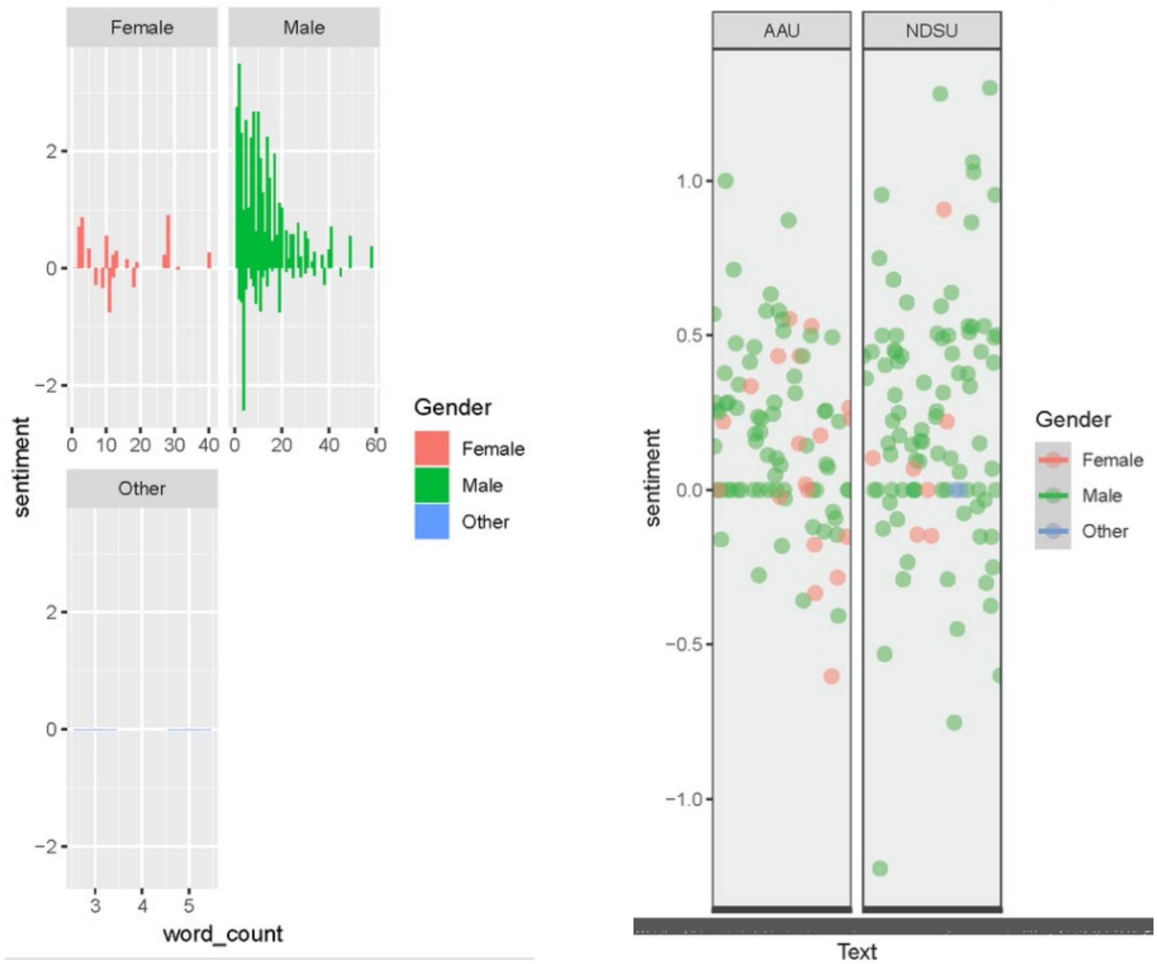


Figure 17: Sentiment attribute distribution by gender, by institution.

From the descriptive analysis and visualization, the comparison of gender is validated by a statistical means using hypothesis testing. Hypothesis validation is critical to converting observations to concrete knowledge with statistical significance.

4.3.1.2. Institutional Difference in Sentiment of Course Review, NDSU vs AAU

A two-sample t-test was performed to compare the institutional sentiment responses of cognitive gains and self-performance in completing course CS I and CS II at NDSU and AAU. There was a significant mean sentiment score difference in the perceptions of cognitive gains between AAU (M = 0.23, VAR= .13) and NDSU (M = .37, VAR = 0.13); $t(376) = -2.48$, $p = .013$. Therefore, the null hypothesis was rejected for course reviews.

4.3.1.3. Course Difference in Sentiment Score of CS I vs CS II at AAU

A two-sample t-test was performed to compare the course sentiment responses of cognitive gains and self-performance in course review of CS II at AAU. There was a significant mean sentiment score difference in the perceptions of cognitive gains between CS I ($M = .26$, $VAR = .14$) and CS II ($M = .17$, $VAR = .08$); $t(330) = 2.30$, $p = .0029$. Therefore, the null hypothesis was rejected for courses.

4.3.1.4. Gender Difference in Sentiment Score of CS II Course Review at AAU

A two-sample t-test was performed to compare the gender sentiment responses of cognitive gains and self-performance in course CS II at AAU. There was no significant gender difference in the sentiment score of perceptions of cognitive gains between Male ($M = 0.18$, $VAR= .07$) and Female ($M = .09$, $VAR = 0.10$); $t(86) = -1.14$, $p = .258$. Therefore, the null hypothesis was accepted for courses.

4.3.1.5. Gender Difference in Sentiment Score of CS II Course Review at NDSU

A two-sample t-test was performed to compare the gender sentiment responses of cognitive gains and self-performance in course CS II at NDSU. There was no significant gender difference in the sentiment score of perceptions of cognitive gains between Male ($M = 0.35$, $VAR = .14$) and Female ($M = .42$, $VAR = 0.05$); $t(96) = -.58$, $p = -.56$. Therefore, the null hypothesis was accepted for courses.

We deployed a unique and innovative approach to apply NLP to derive the average sentiment of interview transcripts (Parmar, et al 2018). We demonstrated the use of quantitative analysis to derive meaningful insights from the sentiment analysis leading to hypothesis validation.

4.3.2. NLP Topic Modelling

4.3.2.1. Research Question III

Addressing the 3rd research question was aimed at identifying and enumerating factors that promote success in pursuit of a degree in Computer Science/Software Engineering for female undergraduates. A question was paused to twenty female upperclassmen undergraduates of Computer Science and Software Engineering majors who had voluntarily opted to be interviewed for this purpose. The question was "*What would you say helped you persist the first two introductory courses and beyond to become a soon to be graduating class of CS/SE?*" The interview responses were auto transcribed using Microsoft's A.I. powered Office 365 auto transcription tool.

The Latent Dirichlet Allocation (LDA) topic modelling was applied to the auto transcribed documents which was used as input corpus. Two distinct LDA topic modeling scripts were written in R programming language for the topicmodels package and procedures adapted from Grün and Hornik (Grün and Hornik 2011). The procedures and scripts included standard text preprocessing of the corpus: punctuation removal, lemmitizing, stopword removal, and the like as outlined in the methodology section. The LDA topic modelling algorithm was used to build data points, estimate probabilities, cluster similar terms together, and for fitting topic model in discovering abstract "topics" that occur in a corpus for the unsupervised classification of documents.

The LDA topic clustering algorithm doesn't identify the overarching topic for the collections of word tokens. The word tokens provide some hints on what the topic it represents but reading through the interview transcripts was necessary to proximately identify the semantic topics represented by the LDA fitting model. The researcher's experience reflects the observations of Hindle (Hindle et al. 2014), who suggested that domain expertise is necessary to accurately label topics. Defining semantically meaningful topic labels required the non-trivial effort in applying the qualitative method of manually creating inductive coding categories. As a result, the researcher first conducted a manual inductive thematic analysis coding technique to organize, and structure content to aid with the assigning of more distinctive labels to the LDA fitted topics. Listed here are some excerpts from the auto transcribed audio responses of the question paused to NDSU & AAU graduating class:-

- "*The professor's teaching and learning approach helped me do better. Besides spending a lot of time, the way the professor taught also made a huge impact on me... I guess I'm the first kid in my family to go to college. I thrive on the challenge. I am not a quitter. I focused on learning and understanding the problem. I also started seeking outside help. I wish I had done it earlier*." ~female, CS Senior, NDSU
- "*Focus on the class you are taking. Things change. Things get better. Don't get hung up on how difficult it is because it gets better. If you are passionate, talk to a senior with experience who had gone through this prior. Realizing that there are not as much intense coding classes after those first introductory classes, helped me persist through this course.*" ~female, CS Senior, NDSU

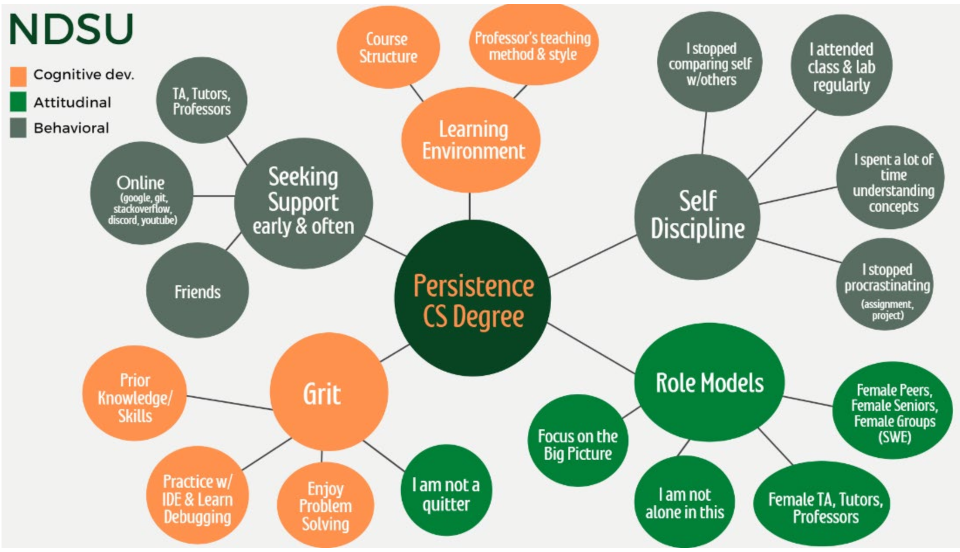


Figure 18: Manual qualitative inductive thematic coding analysis of factors of persistence/success in CS at NDSU.

As Figure 18 shows the thematic coding technique found five main thematic areas with reported correlational value to success or persistence, namely, grit, learning environment, seeking support early and often, self-discipline, and role models. The subcategories related to attitudinal, behavioral and perception of cognitive development. The "*grit*" category, contains a combination of attitudinal perseverance and cognitive growth that comes from acquired knowledge and continued learning. The self-discipline category contains elements of self-efficacy or behavioral changes, when female students decided to stop comparing themselves with other smart students in the class, when they decided to stop procrastinating and when they decided to start spending lots of time to understand concepts and structures instead of focused on completing assignments and tests. The interview transcript also highlighted the importance of attending class & lab sessions regularly. Conducive learning environment was stressed as an impactful aspect that improved the student's cognitive development. Another stated persistence factor was female students' decision to seek institutionally provided support and online resources. Female students had underscored the influence of role models (female professors and female teaching assistants) and mentors (female upperclassmen) had inspired, encouraged and improved their sense of belonging and self-efficacy.

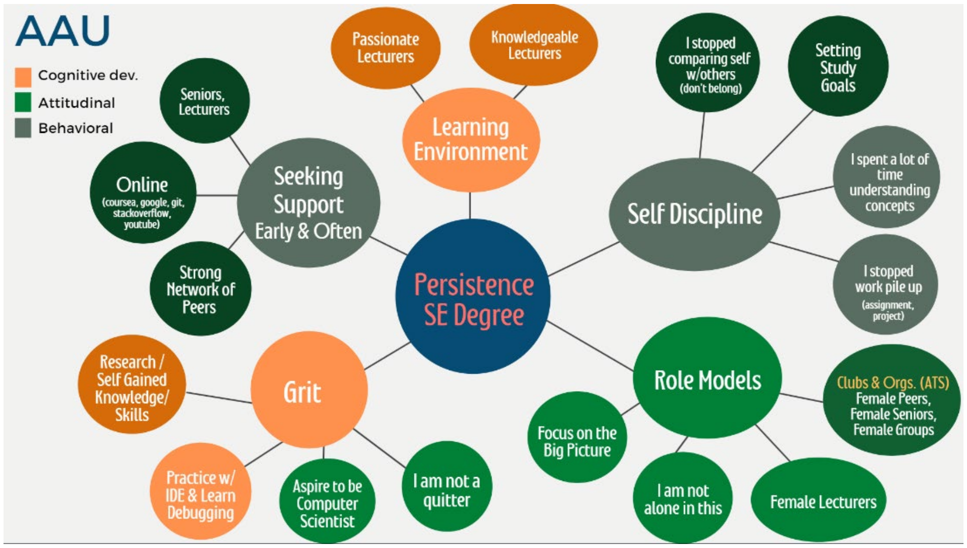


Figure 19: Manual qualitative inductive thematic coding analysis of factors of persistence/success in CS at AAU.

- "*I used to feel like an imposter. I stopped comparing myself to others. Some students are smart because they have experience before joining college. I finally understood this. I came to Computer Science without prior experience. I survived the first semester by talking to my senior friends. I got advice that things are going to get better, and everything is going to improve through practice, and I also saw it for myself, like by practicing and I got it.*" ~female, SE Senior
- "*I suggest students do very good research and find out if they actually do like programming. Do a few programming lessons in high schools just so that students can find out if they actually like programming or not. Getting encouragement from my friends and facing reality that I genuinely had to spend a lot of time practicing if I wanted to be a computer scientist.*" ~female, SE Senior

As Figure 19 depicts the responses of the NDSU female students essentially mirror the responses of the AAU female students to the same paused question. On Seeking Support category, AAU female students had described as having a very strong network of peers which helped them with understanding of computing concepts and stress management. On Self Discipline category, AAU female students work on setting study goals while NDSU work on class attendance and participation. Lastly, in the Role Models category, there were stronger peer level activities, some of them spearheaded by students. Once these categories are identified and labeled, the researcher continued with the process of generating clusters of terms belonging to the same categories.

In order to ascertain the accuracy and reliability of the LDA modelling of persistence factors from the interview response, the researcher has conducted the time consuming qualitative inductive thematic coding technique.

Determining if the collection of words in a topic is semantically coherent is usually a human task, though researchers have designed and validated automated measures of topic coherence. One such measure, statistical goodness of fit in R, is shown in Figure 20. The selection of k, the number of topics to model, significantly impacts the internal validity of trend analysis. Definitive methods and usable tools for evaluating k are needed relative to the size of the corpus. We calculate the statistical fit of models with different K. Semantic Coherence: shows how coherent topics are, i.e., how often features describing a topic co-occur and topics thus appear to be internally coherent.

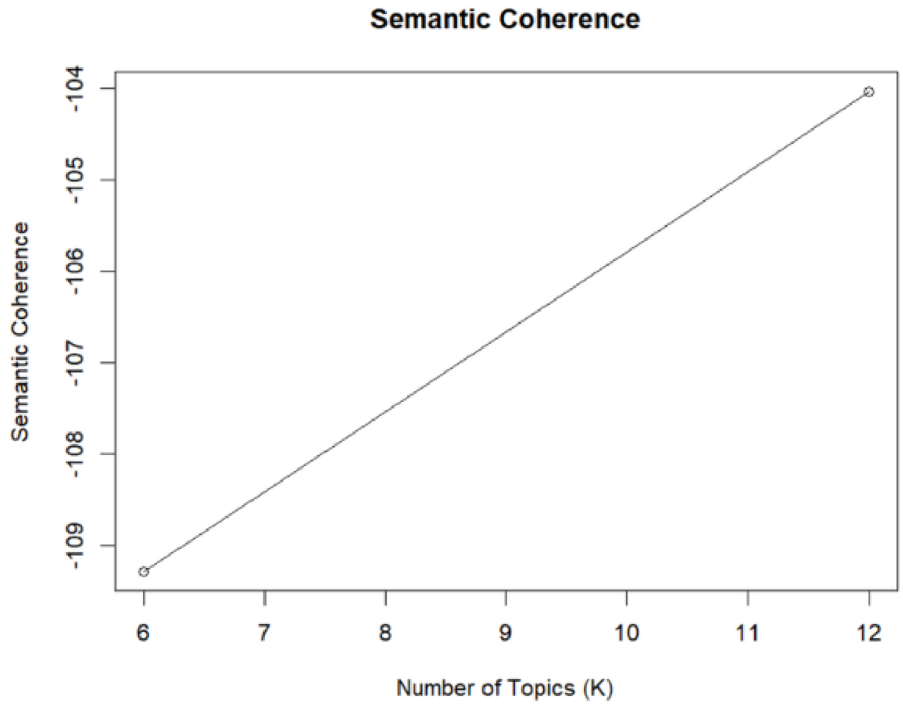


Figure 20: Statistical goodness of fit on semantic coherence in CS at NDSU.

After running a few iterations of the LDA model algorithm for different k values, this research has identified k=12 topics to be a threshold of persistence factor topics from the corpus of interview responses. Figure 21 lists the most frequently occurring terms in each of the 12 topics. Each category contains 10 to 12 terms along with the probability score of the term belonging to that topic, ordered by frequency of occurrence. There was also some degree of overlap between the topics; for example, topic 8 (prob. = 0.09) and topic 9 (prob = 0.08) listed the term professor as the top word. If a word has a high probability of being in a topic, all the documents having the word will be more strongly associated with the topic as well. Performing an inference interpretation of each topic from the manual coding analysis, this research has labelled the topics as Network of Support for topic 1, Learning Environment for topic 2, Class Participation and Practice for topic 2 and so forth, as displayed in Figure 21, for both NDSU and AAU.

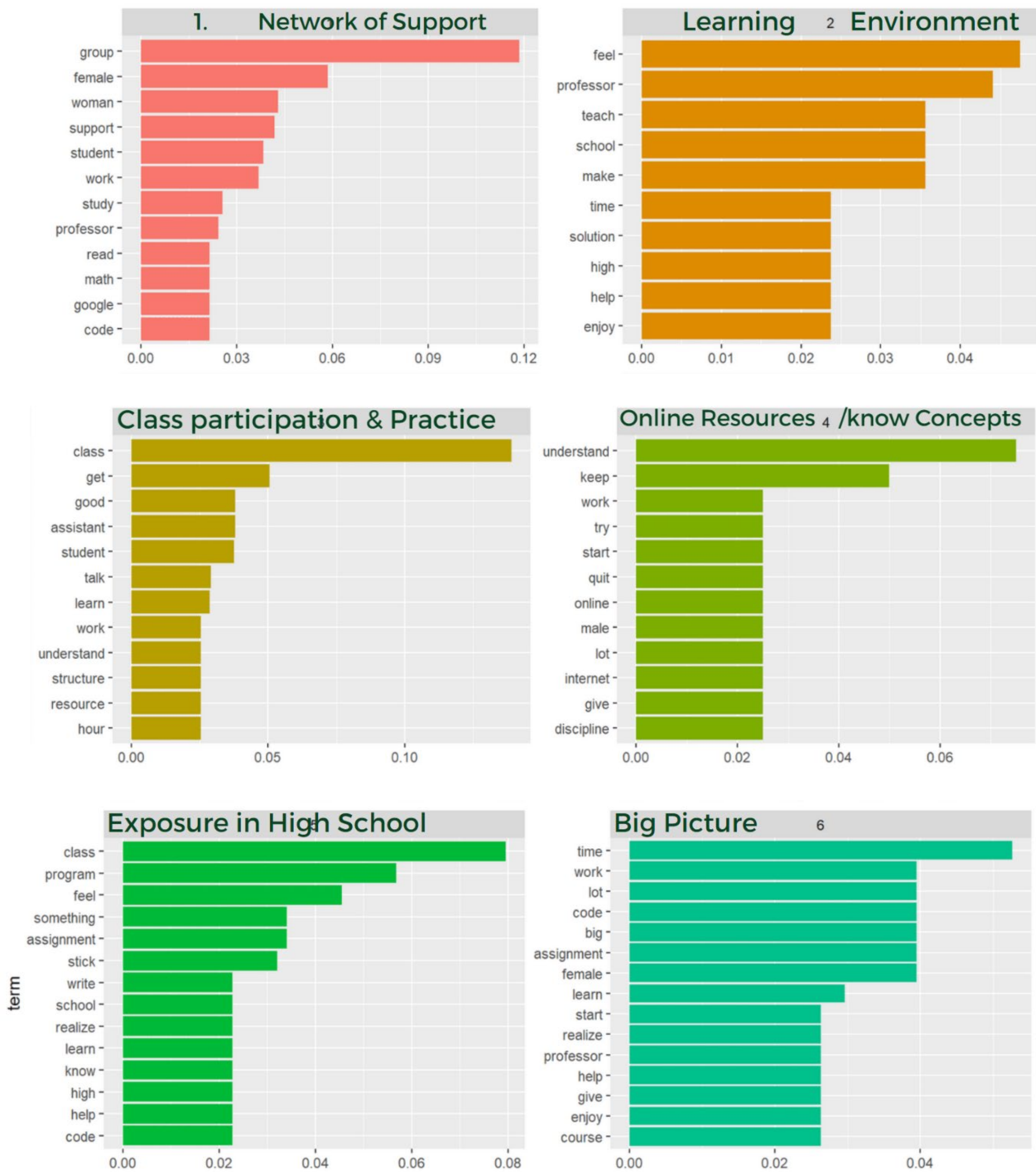


Figure 21: Beta probability distribution of the top 12-topic modeling results for NDSU.

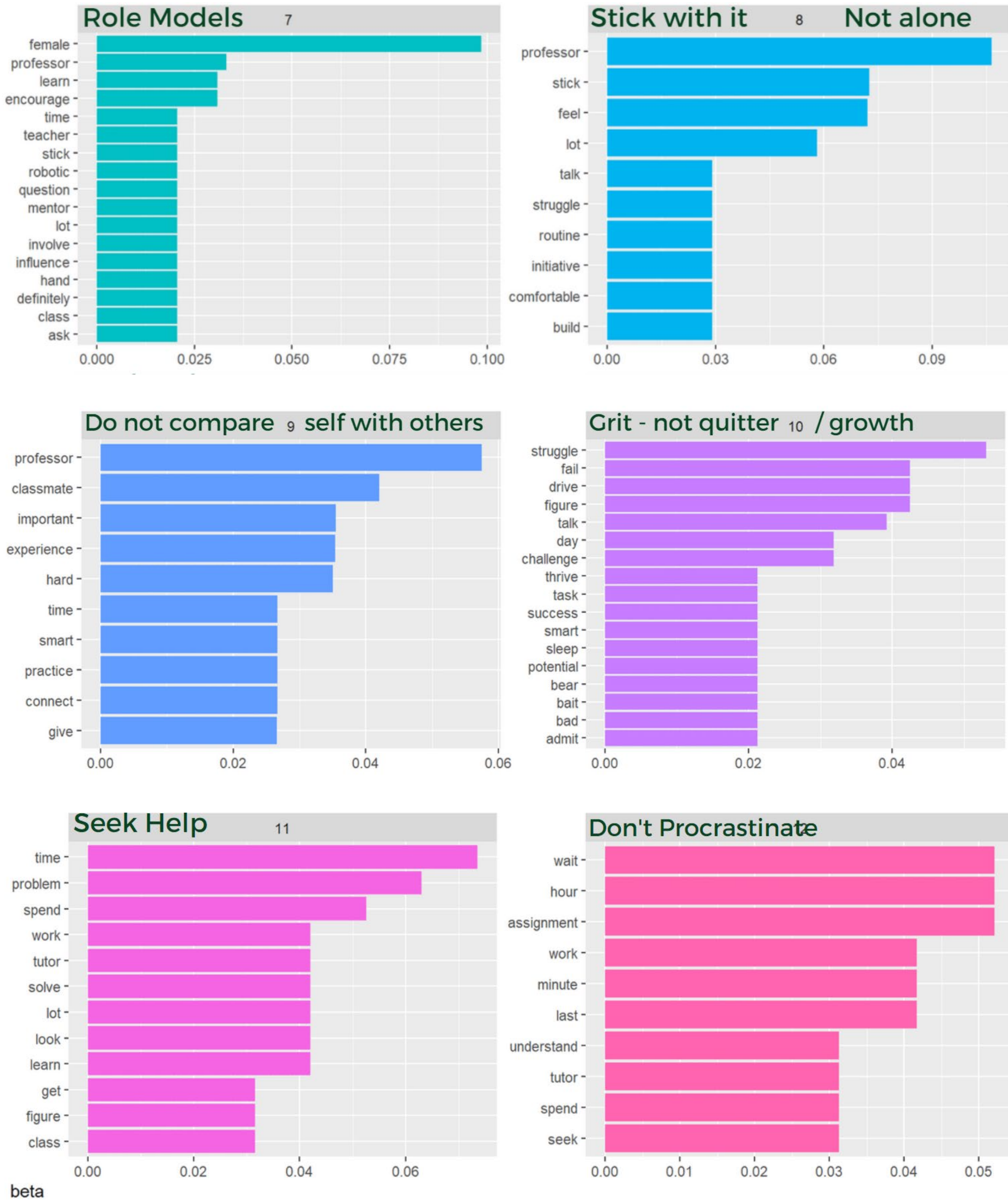


Figure 21: Beta probability distribution of the top 12-topic modeling results for NDSU (continued).

A symmetric distribution would mean that each topic is evenly distributed throughout the document while an asymmetric distribution favors certain topics over others. The β (beta) is a matrix where each row represents a topic, and each column represents a word. Usually, each word is distributed evenly throughout the topic such that no topic is biased towards certain words.

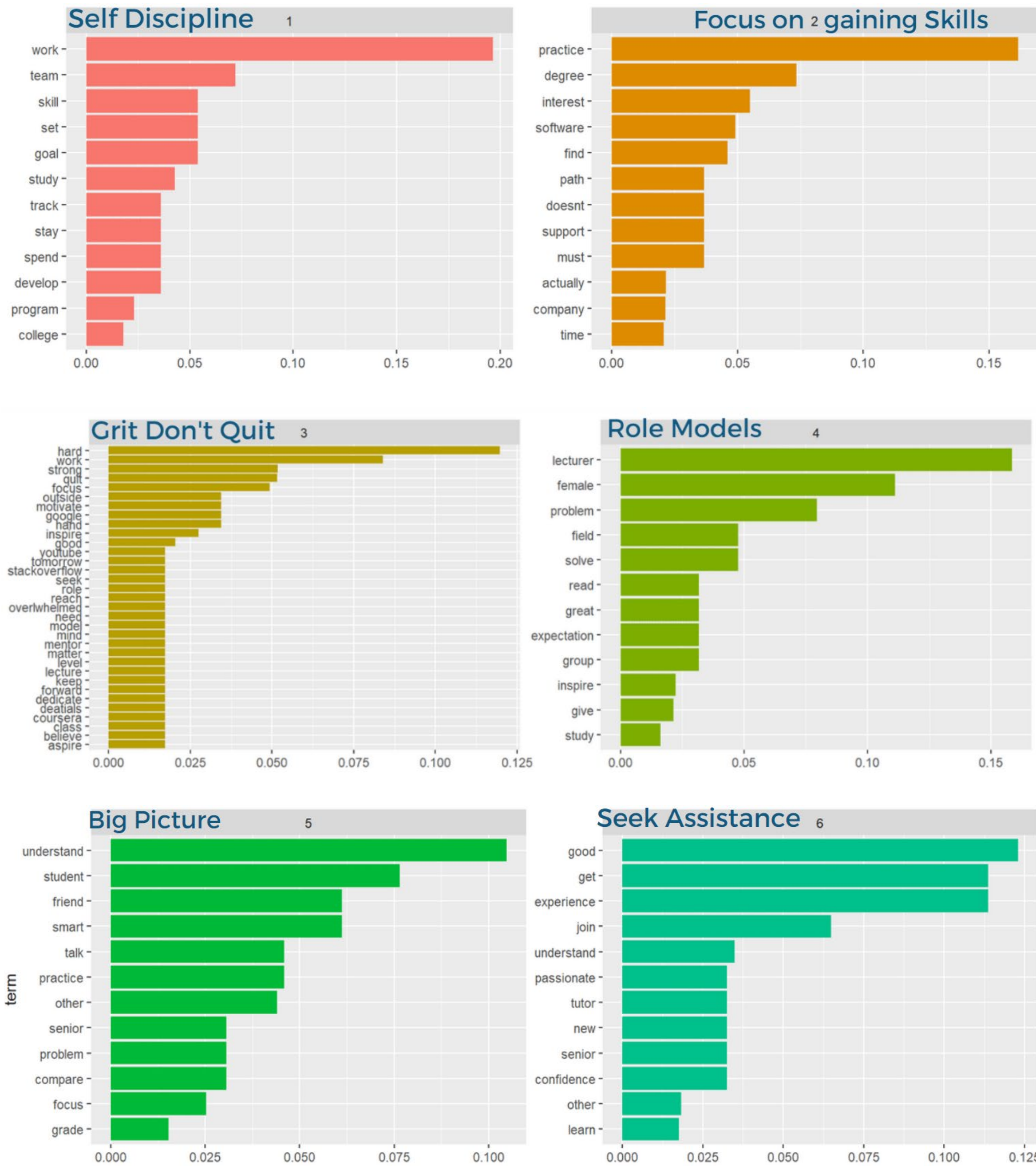


Figure 22: Beta probability distribution of the top 12-topic modeling results for AAU.

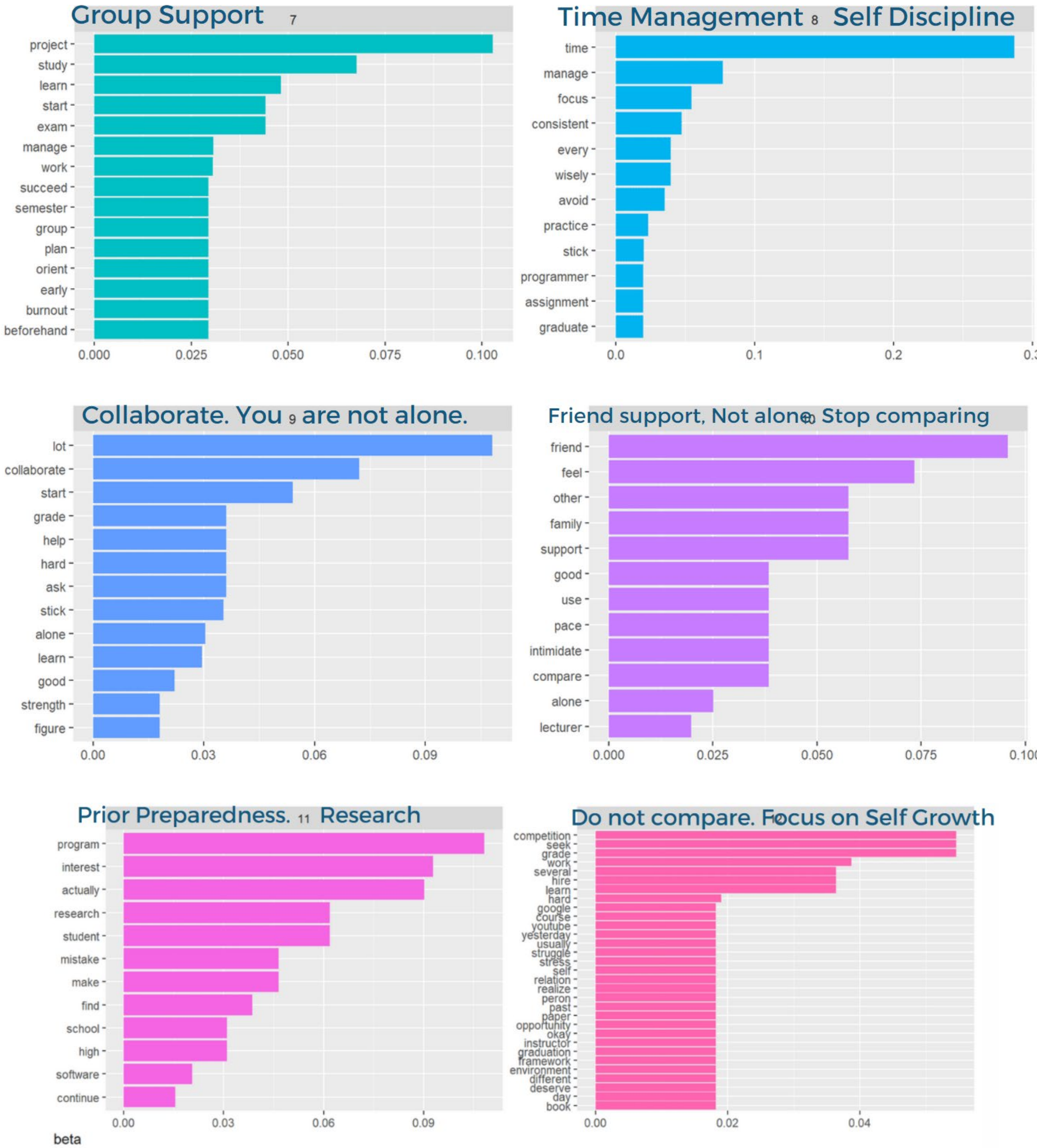


Figure 22: Beta probability distribution of the top 12-topic modeling results for AAU (continued).

The following categories of factors were enumerated.: -

1. Learning Environment & Teaching Style

 - *"The professor's teaching and learning approach helped me do better. Besides spending a lot of time, the way the professor taught also made a huge impact..."*

- *"I think the biggest thing was probably the way the professor taught and how he explained things.*"
- *"I think it's super important to have good connections with your classmates, and your professors and tutors and just to really use those resources."*

2. Role Models and Mentors
   - *"Female professors influence your learning. I wish to have more mentors. She really encouraged you to challenge yourself and figure things out. You ask her for help when you're stuck on something. It's so much easier to ask her for help than not do anything for days. When I got stuck, I want to email the teacher with questions."*
   - *"Empowering to see female professor. You are such an inspiration and I genuinely look up to you. I wanted to be like you"*
   - *"Working in groups with other women on group assignments. I could relate to my female friends more than my male classmates. I'm more reserved with other male students than with female students. So, I think it would be good even just in those classes that don't have many women, to get those women to work together. I do think it would be more helpful to have more female classmates. group projects help you interact with your classmates, build connections."*
3. Self-Efficacy
   - *"Getting encouragement from my friends and facing the reality that I genuinely had to spend a lot of time practicing if I wanted to be a computer scientist."*
   - *"I do ask for help or like go to tutoring if I need to or come ask for a professor or email them. Take more initiative and ask for help when I feel stuck. Good feeling after figuring out something that was so really difficult. Motivated by the rewards of*

getting a program work after such a struggle with it. Stuck more often trying to solve a problem so, but that's the aspect they're chasing that. Challenge and then figuring it out. I really want to be a software engineer."

4. Class Attendance and Participation
   - *"Being present and being involved. What really helped me I know with when my first class is asking questions to my professor and actively learning and not taking the back seat. I feel like you need to get your hands in there so being hands on, it's very helpful to succeed."*
   - *"Attending class regularly and actually paying attention had really helped me as the first step. knowing that I can go ask professors Anything. Use the office hours. they want to help."*
5. Network of Support
   - *"I survived the first semester by talking to my senior friends. I got advice that things are going to get better and everything is going to improve through practice, and I also saw it for myself, like by practicing and I got it."*
   - *"Don't be intimidated by the only few females. We form groups and communicate with Microsoft Teams. collaborate. spend time. ask for help. google for more. professors sponsor study groups or to encourage study groups. upperclassmen provide support. support group. math tutoring. 24 hours online support. Avoid the impostor syndrome You don't have to be more successful than whoever else is in the room. Focus on acquiring more experiences."*

6. Seeking Help Early & Often

    - *"I started to seek out help because I realized, you know I could spend 5 hours on it, or I could go get help you know and spend an hour on it. I did use a typical Google to try to find things and look at other examples. I start to seek help and you know, going to the TA and tutoring and getting help that way. Tutoring worked and helped me understand that it made it easier. stop relying on themselves. Get support when you are stuck."*
    - *"I decided to get tutoring help for computer science project when I got stuck. Tutor encouraged me that I am on the right track while working on debugging. I spent a lot of time and finally figuring out and solving problem."*

7. Stop-Comparing with Others

    - *"I used to feel like an imposter. I stopped comparing myself to others. Some students are smart because they have experience before joining college. I finally understood this."*

8. Online Resource

    - *"I feel like the number one skill for computer science students is Googling. learn how to Google things properly to find the solution. code comprehension, you have to be able to read and fundamentally understand what code does just based on read. I think female and male brains work just the same when problem solving. I would say there was definitely more like you know, group chats, study groups between male students than with the female students. Participating in female centric hackathon stuff, or like the Robotics Club in a less intimidating environment with no prior expectations involved would be really beneficial to learn from each other."*

9. Don’t Procrastinate
   - *“I stopped waiting until the last minute to complete an assignment. Don't wait until the very last day to submit an assignment. Don't be a procrastinator.”*
   - *“Don't wait until deadline to work on assignments and projects. Spending sleepless nights and working code like crazy. Take a break when something is not working and come back to it. Just like coming back to after like 30 minutes can like help you clear mind and then like get you back on track.”*
   - *“If you have a good time management skill, and you spend a lot of time working on it, there's no way why you can't succeed in the courses.”*
10. Grit, Stick with It
   - “*I would say don't get discouraged with it. Realize it is a learning process. Don't get hung up on how difficult it is because it gets better.*”
   - “*I thrive on the challenge. I am not a quitter. I would say focus on learning and understanding the problem.*”
   - “*I have never failed at anything before. And switching a major isn't really necessarily failing. I feel like I chose to this major, and then I feel like dropping out of it and switching majors would just be a failure. And I couldn't handle that.*”
11. Focusing on the Big Picture
   - *“Focus on learning and understanding the problem.”*
   - *“If you are passionate, talk to a senior with experience and gone through this prior. Realizing there is not as much intense coding classes after those first introductory classes, helped me persist through this course.”*

12. Exposure in High School

- "*I had an opportunity to take AP Computer Science my sophomore year. I know programming and robotics in high school. I suggest students do very good research and find out if they actually do like programming. Do a few programming lessons in high schools just so that students can find out if they actually like programming or not.*"
- "*Seeking prior experiences in computer science is super important and super valuable.*"

CHAPTER 5. CONCLUSION

5.1. Introduction

This study used both quantitative and qualitative methods to investing the problem of retention of females through graduation in computing fields which has been researched as fundamental global problem at most tertiary institutions. Why didn't the sharp increase in enrollment interest in the past decade among first-year full-time CS/SE students translate to obtained degrees in CS/SE? What factors if any help females who start out on the path towards a CS degree continue to graduation year after completing the first two years of their CS studies? This research was aimed at comparatively investigating, describing, and analyzing factors correlated to persistence for female undergraduates in Computer Science/Software Engineering degrees. Conducted in two distinct public universities in the U.S. & Ethiopia, the study sought to test the stated hypothesis, and enumerate results for the sample population in answering the following research three questions:

RQ1: Does a relationship exist between the students' perceptions of their Degree's Usefulness (DU), their Previous Acquired Knowledge (PAK), and their Cognitive Attitude (CA) with their decision to pursue Computer Science/Software Engineering degree after completing the first two introductory courses?

RQ2: Is gender or institutional difference evident between the three variables identified (DU, PAK, CA) and the decision to pursue in CS/SE major after completing the first two introductory courses?

RQ3: Can factors that promote success in pursuit of a degree in Computer Science/Software Engineering for female undergraduates be identified and enumerated? Can persistence factors in pursuit of a Software Engineering degree in Ethiopia be predicted by factors reported to be effective in the U.S. Computer Science degree and vice-versa?

The purpose of this study was to add to the understanding of retention from the perspective of female students who persisted in the major. The study's conceptual research framework was constructed from the well-established theory that students' level of academic and social integration leads to greater commitment to institutions and graduation (Tinto .et al 1993, 2006).

The following sections present the summary of findings by research question in comparison to existing literature. These will be followed by sections on the limitations of the study, suggestions for practitioners, and future work.

5.2. Interpretation of the Findings

This study examined the correlation of predictor variables and the decision to persist, the effect of gender and institutions as well as the influential factors CS students ranked as significant predictors of their decision to complete their CS/SE degree.

5.2.1. Research Question One: Correlations between DU, PAK and CA with Persistence

Data collected for each of the three predictor variables (DU, PAK and CA) outlined in the proposed conceptual framework were analyzed with the parametric Pearson's correlation coefficient equation using SAS (Statistical Analysis Software), denoting an alpha level of .05 for the probability of type I error.

This research's findings indicate a strong positive linear correlation between NDSU students' understandings of their degree's usefulness (DU) with their decision to continue to persist with their computer science major beyond their CS II course completion, $r(85) = -.65$, $p < .0001$. The scatter plot graph showed a more noticeable negative trend from upper left to lower right confirming the +ve linear relationship. Therefore, the null hypothesis was rejected for NDSU CS II course. This relates well to Giannakos's study that states student's perceptions regarding their future in computer science and career fit evolves during the course of their studies. It explains why the DU is stronger in the second computer science course (CS II data structure) than the first as strong intentions on career goals can overpower the effect of negative experiences (Tinto 1993). Students have reported that the utility of a Computer Science degree influences their decision to complete their studies (Lewis et al. 2011, 2016).

No statistically significant relationship between DU and the decision to persist exhibited for AAU students ($r(98)=-.37$, $p=.110$ and $r(43)=-.18$, $p=.666$). This can be attributed to the fact that students go through rigorous processes to get accepted into the Software Engineering major and once accepted, the option to switch to another major is an uphill road that most students are discouraged from trying it. The frequency distribution of the DU scale indicated that 92% of the AAU respondents do believe they need a firm mastery of programming for their future work (Appendix E). Previous research shows the lack of an academic plan contributes to a feeling of academic displacement for many students who fail to recognize the connection between their coursework and a future career (Beaubouef, et. al. 2001).

Previously Acquired Knowledge (PAK) was also exhibited to have moderately strong positive linear relationship with their decision to continue to persist with their Computer Science major beyond their CS II course completion for NDSU students ($r(86) = -.55$, $p =.001$ and $r(55) = -.44$, $p =.001$) as well as CS I students of AAU ($r(163) = .17$, $p = .029$). Therefore, the null hypothesis of no correlation was rejected for NDSU CS II course for both Pre-/Post- PAK observations and AAU CS I. For all other variables, the null hypothesis is accepted. This finding supports Biggers et al study that confirms retention in Computer Science major depends greatly on having previous experience in computing and knowledge of problem-solving skills. 53% of NDSU students had reported to have taken one or more formal computer programming courses before attending CS I and 43% have reported to have learned how to code on their own before attending CS I, both of which contributed to the affirming correlation between PAK and persistence in this study. For AAU pre-Engineering students who completed the CS I coursework and who responded with a positive relationship of PAK and persistence matches with the timeline where students choose from Pre-Engineering to Software Engineering major. 59% of respondents have self-taught programming before attending CS I.

This study did not find any statistically significant correlation between cognitive attitude (CA) with students' decision to continue to pursue their major or not for both at CS I ($r(56) = .06$, $p = .784$) & CS II ($r(55) = -.16$, $p = .229$) courses at both NDSU and AAU. It is the most interesting finding. According to this result, the students' perception of learning and acquiring competence in programming has no merit to their decision to persist or leave the major. The result of the frequency distribution confirms this no relationship factor. A whopping 70% of students do not believe CS II courses were challenging but 61% report that "this course has helped me decide computer science major is not for me". Competence on computer and computer literacy are not just related to the level of knowledge, constraints, applications, and the effect of computer but, it is directly related to individual's attitudes towards computers (Levine & Donitsa-Schmith, 1998).

5.2.2. On Research Question Two: Gender or Institutional Differences

This research weaves together the qualitative and quantitative results to interpret the influence of gender and institution on DU, PAK and CA. A two-sample independent t-test was performed on data collected using SAS as well as exploratory text analysis using NLP. Although the reason for choosing to pursue Computer Science degree differs by gender according to the results of the qualitative study, no statistically significant gender differences were exhibited in this quantitative study in the students' perceived usefulness of their degree for both NDSU ($t(56) = .18$, $p = .536$ and $t(82) = .24$, $p = .709$) and AAU ($t(98) = .62$, $p = .826$ and $t(43) = 1.62$, $p = .079$) introduction to Computer Science (CS I & CS II) course.

When asked why students choose to major in Computer Science in the qualitative study, male students have reported to be motivated by specific aspects of Computer Science such as the pay, the prospect of creating video games, job security, diversified career opportunities such as Cyber Security and Artificial Intelligence. Laser sharp focus on career fit. Whereas female come to the major with aspirations of making an impact in the real-world and interested in problem-solving endeavors.

No gender difference is revealed in the cognitive attitude of students who took the course CS I at both NDSU and AAU. However, after taking data structure course (CS II) at NDSU, difference in cognitive attitude was shown between male and female students. There was a significant difference in perceptions of CA (post-survey) between male ($M = 3.22$, $SD = .05$) and female ($M = 3.33$, $SD = 0.08$); $t(54) = .70$, $p = .014$ for course CS II at NDSU. Therefore, the null hypothesis of no correlation was rejected for NDSU CS II post survey course. The qualitative study reveals male students are more likely to seek outside help (including but not limited to Teaching Assistants, professors, friends, or the internet) when stuck with challenging programming task sooner than female students. Female students reported to have spent a lot of time on being stuck on challenging programming tasks alone. Female students from both institutions had said they have cried themselves to sleep out of frustration. As a result, they become more introspective about their abilities, loose self-confidence, and become less comfortable asking for help sooner. Smith's study on the problem women faces with low self-confidence in computer science performance compared to male self-confidence confirm this research's findings.

Gender differences have also been revealed with the influence of Previously Acquired Knowledge for CS I course at NDSU. A two-sample t-test was performed to compare perceptions of the influence of Previously Acquired Knowledge (PAK) in female and male. There was a significant difference in perceptions of the influence of PAK between male ($M = 2.82$, $SD = .30$) and female ($M = 3.02$, $SD = 0.14$); $t(46) = 1.67$, $p = .051$ for course CS I at NDSU.

The largest difference found was in previous programming experience where the female mean was considerably lower than the male mean but not at the significance level in this. Previous literature is full of cases of struggling students who didn't have previous experience in programming (Barker et al., 2009; Biggers et al., 2008; Ott et al., 2018). 27% of CS I respondents wished they had more programming knowledge and skills before taking this course, while only 19% wished more knowledge in programming would have helped them in CS II at NDSU. Qualitative study found male students are more likely to have self-taught programming or acquired formal programming trainings before joining college. Whereas some female students are more likely to have participated in exploratory STEM summer camps. This is supported by Denner who states women are more likely than men to enter college with no prior experience in programming and or computing (Denner et al., 2014). With 95% confidence interval for mean, there exists a statistically significant difference between the pre- and post- perception of the influence of PAK for NDSU introduction to Computer Science (~CS I) course. Students changed their perspective by 50% from the start of the semester to the end of the semester for course CS I regarding their perception on the impact of their acquired programming knowledge.

Institutional difference was revealed in the perception of PAK between NDSU and AAU for CS I course. With no formal Computer Science education in post-secondary education in Ethiopia, 50% of the respondents said they have no prior experience in computer programming before taking the first introduction to CS I course whereas 65% of NDSU said they have some level of exposure before taking CS I. 48% of both NDSU and AAU student respondents either disagree or strongly disagree on the phrase, “programming is enjoyable and stimulating to me”. A stark difference in attitude is observed between NDSU and AAU student respondents who said, "For some reason even though I work hard at it, programming seems unusually hard for me," referring to the course CS II. With no formal institutional out-of-classroom support at AAU, students have developed an effective and strong network of peer groups with clubs, workshops & recurring hackathons. AAU students exhibited more grit despite limited resources.

5.2.3. On Research Question Three: Contributing Factors Promoting Success

Students’ characteristics, aspirations, and abilities all play a significant role in determining their potential pathway through the Computer Science degree program (Tinto 2006). A wealth of research results show that attitude and motivation are more important than grades in determining which students drop out or persist (Ajzen et al., 1999, Abbott et al, 2002; Bean, 2005; House, 2004). In this research, the following questions were explored, “what are effective novice programmers doing that worked so well? Can it be identified, and can it be used to support future novice programmers? Are the crucial differences cognitive, attitudinal, or behavioral, or is it impossible to separate?” (Robin 2019) Through this manual qualitative interview thematic coding techniques and automated NLP Topic Modelling analysis of the interview responses, this study has enumerated 12 factors that CS/SE female research participants cataloged as significant influencers of their decision to complete their degree.

This research had ascertained those successful students in Computer Science at the NDSU and AAU found themselves having a better idea of their intrinsic and extrinsic motivation to continue their degree, join the major with more previously acquired knowledge than their peers, and were more likely to have in-person connections with teaching assistants, faculty, and peers (Margolis et al., 2000; Margolis and Fisher, 2003). Moreover, students who persisted had learned to manage their time better and stay on track with their assignments and projects. "*Persisters*" possess the ability to integrate their pre-existing knowledge in understanding algorithms and problem-solving aspects, and the semantics of a particular programming language than just focus on completing the assignments for grade value.

A combination of attitudinal and behavioral factors had a higher probability to be more influential than cognitive factors for female students to persist through graduation. Female students' attribute their success and persistence on how well they understood their career fit, the efforts they put into their work, how much and how often they seek help from their institutionally provided support system such as tutors, lab assistants and instructors' office hours, the level of support and encouragement they receive from their professors, and female role models. Furthermore, the qualitative study identified and enumerated twelve distinct categories promoting and/or influencing success for female undergraduates pursuing Computer Science or Software Engineering degrees. The attitudinal shift the students reported to exhibit was a crucial factor in the spectrum of success. It is when the female students at NDSU and AAU stopped doing self-limiting activities such as procrastinating, not seeking help for fear of being perceived "dumb" and comparing themselves with other peers they consider "smart" that their selfactualization happened, and they start doing better. Their self-growth was transformational when they started reaching out for institutionally provided help regularly, opened themselves up to working with peers, focused on the bigger picture of understanding concepts and fundamentals, and augmenting their knowledge with online resources.

While these findings should be almost assumed, it is important to map out the differences between students who were successful and how their varying characteristics and actions can affect their decisions to stay, and their ability to be successful. A better understand of these populations can assist researchers and stakeholders within the discipline to assist the various student populations, and hopefully begin to increase the diversity in Computer Science as they gain the tools, they need to supplement all student learning, regardless of their backgrounds and risk factors.

5.2.4. On Research Question Three: Replicable Factors of Persistence between Institutions

The challenging experiences and the described self-efficacy of the successful female students in CS interviewed at NDSU are the mirror image of their peers in SE at AAU. The enumerated success factors are uncannily similar to one another. Despite the contextual differences between the two institutions under study, the research's cross-sectional comparative investigation reveals that persistence and success factors overlap between the two geographically, economically, and culturally distinct institutions. The statistical analysis unearthed similar findings that answered the research number three question that persistence factors in pursuit of a Software Engineering degree in Ethiopia can be predicted by factors reported to be effective in the U.S. Computer Science degree and vice-versa.

5.3. Recommendation

According to Tinto's theory, social and academic integration are equally important to the retention and success of students in the chosen university. Our findings confirm the key role of activities that support academic and social integration in promoting persistence of students to degree completion. Based on these findings, practitioners could consider the following suggestions:

5.3.1. NDSU Computer Science Program

The themes that emerged from the qualitative analysis, coupled with the quantitative findings were the basis for the researcher to suggest the following recommendations for future practice.

1. **Recommendation 1: Overcoming the Previously Acquired Knowledge Gap.**

The first required course for Computer Science majors at NDSU is CSCI 160 (CS I). At the end of the CS I course, 49% of surveyed students at NDSU and 68% of students at AAU said they wished they had more programming knowledge and problem-solving skills before taking this course. 78% of the same group of students responded negatively to this statement, *"I am confident that I will do well in this programming class."* Problem solving is an indispensable part of Computer Science. The ability to solve problems is often very weak in first year computer science students (Beaubouef, et. al. 2001). This problem solving is a skill that can be learned. Failing to teach these skills might add to more students dropping from the course (Beaubouef, et. al. 2001). MATH 103 (college algebra) is a requirement for taking CS I at NDSU. NDSU does offer such a problem solving and algorithm development course titled CSCI 159 (Computer Science Problem Solving) but 91% of the CS I (CSCI 160) students have not taken CSCI 159. 100% of the students interviewed said "*nothing about CSCI 159 was mentioned to me during the* *first-year academic advising*." Rudimentary problem solving should be a prerequisite for CS I (CSCI 160). Moreover, the researcher had experience teaching CS I at the same institution and have witnessed how students struggle with problem-solving. This study suggests that requiring students to take a course in elementary problem solving and algorithm design would result in better CS1 experience and by extension better grade as well as lay the foundation for future courses.

2. Recommendation 2: More Female Mentors and Role Models.

Several studies have suggested that interactions with female peer-role models had a positive effect on retention (Biggers et al. 2008; Cohoon 2002 2006; Cuny and Aspray 2002; Frieze et al. 2012). The research also underscores that the availability and accessibility of role models positively and significantly affect female students' choice of learning engineering and computing education. Interviewed female students have highlighted how they are inspired and encouraged by female professors, female teaching assistants and tutors. Comfort level were found to have a positive influence on success in Computer Science (Wilson 2000).

- *"Female professors influence your learning. I wish to have more mentors. She really encouraged you to challenge yourself and figure things out. You ask her for help when you're stuck on something. It's so much easier to ask her for help than not do anything for days. When I got stuck, I want to email the teacher with questions."*
- *"Empowering to see female professor that looked like me. You are such an inspiration and I genuinely look up to you. I wanted to be like you."*
- *"I enjoy working in groups with other women on group assignments. I could relate to my female friends more than my male classmates. I'm more reserved with other male students than with female students. So, I think it would be good even just in those*

classes that don't have many women, to get those women to work together. I do think it would be more helpful to have more female classmates. group projects help you interact with your classmates, build connections."

The lack of female engineering faculty role models and the failure to account for women's different cognitive styles are cited as contributing factors to the shortages of women students pursuing engineering degrees. A conscious effort to hire female upperclassmen and graduate students as teaching assistants improves the first-year female students getting comfortable to seek assistance and help for a positive experience which leads to persistence.

3. Recommendation 3: More Networking Experiences for Females.

CS department should be a place where students who are passionate about CS come to build their knowledge and expertise in CS by interacting with their faculty, tutors and peers on ongoing basis. (Porter, et. al 2013) suggests that by pairing students which is commonly used as an agile method in industry, the students' retention rates improve without negatively impacting the students' final exam scores. Cooperative group learning may foster theme works whereby students learn from each other through reflection and experience sharing (Taylor et al. 2001). Students in the interview discussed how they could benefit from connecting with NDSU CS alumni, allowing them to learn more about the different types of careers they could have with a Computer Science degree, and what types of steps they should take to be successful in a tech career. In addition, the female students suggested more real-world applications to the work done in current courses, and how it would help them shape their understanding of applied Computer Science, as well as balance theory and practice during the introductory courses.

Suggestions including hosting speaker series, women in CS clubs where students learn peer counselling and academic support from senior female students (Demise et al., 2002).

5.3.2. Addis Ababa Institute of Technology (AAiT)

The themes that emerged from the qualitative analysis, coupled with the quantitative findings were the basis for the researcher to suggest the following recommendations for future practice.

1. **Recommendation 1: Preparedness. Pre-College Curriculum.**

49% of the survey respondents in CS I (Fundamentals of Programming) course said they had no previous programming experience before taking the course. 43% said they have taught themselves to code before taking the course. Programming is a craft skill that is known to be hard for many to master satisfactorily. Lack of prior experience was not only found to negatively affect women's confidence and comfort in the program (Margolis & Fisher, 2003), but it was also found to be a big hindrance to retention (Buzzetto-More et al., 2010) and a big predictor of attrition (Katz et al., 2006) and or failure in CS courses taken (Staehr et al., 2000) The combination of lack of preparation felt by many of the females and the emphasis on computer programming at the entry level of the curriculum is a "lethal combination" (Liu & Blanc 1996) Several studies investigating the effect of previous computer experiences on success in computer science courses were conducted in recent years. Students lacking the critical thinking, logic and analysis needed in their courses are also at risk of dropping out. (Robins 2019) Programming is the primary activity of computer science, and therefore most CS programs globally start with an introductory programming course. Programming languages are complex artificial constructs.

2. **Recommendation 2: More Opportunities for Practice in Labs.**

89% of the surveyed students from AAU enrolled in Software Engineering courses stressed the importance of being given opportunities to practice what they learnt in lectures with practical real-world projects. "*It focuses mainly on theory rather than experiments and* *laboratory sessions, projects. I think it should be focused on developing solutions for real world problems and knowing new technologies. I have acquired most of my practical knowledge from paid courses like coursera."* (Walker 2002) said "applying active learning exercises to supplement passive learning activities such as lecture achieves positive educational results." Lab sessions are a critical part of computer science courses (Beaubouef 2005) because it gives students hands-on experiences with immediate expert assistance, help students develop and hone programming skills, and encourage the freedom to experiment with ideas. Not offering enough lab sessions for students to practice can be detrimental. (Denning, et al 1989) cements the importance of the pedagogic as "Computing sits at the crossroads among the central processes of applied mathematics, science, and engineering. The three processes are of equal-and fundamental-importance in the discipline, which is a unique blend of interaction among theory, abstraction, and design. The binding forces are a common interest in experimentation …"

3. Recommendation 3: More Female Mentors and Role Models.

Several studies have suggested that interactions with female peer-role models had a positive effect on retention (Biggers et al. 2008; Cohoon 2002 2006; Cuny and Aspray 2002; Frieze et al. 2012). The research also underscores that the availability and accessibility of role models positively and significantly affect female students' choice of learning engineering and computing education. Interviewed female students have highlighted how they are inspired and encouraged by female professors, female teaching assistants and tutors. Comfort level were found to have a positive influence on success in Computer Science (Wilson 2000).

- *"Female professors influence your learning. I wish to have more mentors. She really encouraged you to challenge yourself and figure things out. You ask her for help* *when you're stuck on something. It's so much easier to ask her for help than not do anything for days. When I got stuck, I want to email the teacher with questions."*

- *"I enjoy working in groups with other women on group assignments. I could relate to my female friends more than my male classmates. I'm more reserved with other male students than with female students. So, I think it would be good even just in those classes that don't have many women, to get those women to work together. I do think it would be more helpful to have more female classmates. group projects help you interact with your classmates, build connections."*

The lack of female engineering faculty role models and the failure to account for women's different cognitive styles are cited as contributing factors to the shortages of women students pursuing engineering degrees. A conscious effort to hire female upperclassmen and graduate students as teaching assistants improves the first-year female students getting comfortable to seek assistance and help for a positive experience which leads to persistence.

4. Recommendation 4: Formal Institutional Support.

AAiT interviewed participants reported to have a strong network of student peer support group and extracurricular activities which plays a crucial role in the social integration with the institution. Tinto's model of integration suggested that once the student is emotionally content with the institution, their cognitive and academic performance will inform, and they will perform. However, the same interviewed students reported to wish to have extra academic support from the institution such as after hour tutors and extra Teaching Assistants (TA). Drawn from either the graduate or undergraduate student bodies, TAs are crucial for the successful delivery of education in many institutions. TAs lead small lab sections and discussion sessions, prepare projects, grade projects and assignments, and provide individual instruction (NAS 2018).

5.4. Limitations of the Study

The following were limitations for the findings of this study's results:

1. This study focused on undergraduate females in computer science programs. The low numbers reported by the literature could result in a low number of females eligible to participate in this study.
2. Since this study utilized a questionnaire, participants may not respond truthfully to the statements/questions.
3. Participation in the various components of this study were skewed according to the necessity of its completion. Statistically significantly more Computer Science/Software Engineering students participated at the beginning of the semester, while a number of the students did not take the postsurvey. Reasons could be student drop out/withdrawal from the course, or survey fatigue.
4. Data Sparsity in NLP. Small Sample Size (~400-900 terms vs ~20,000 terms). Large dataset may yield more cohesive insights for inferences on predictive model algorithm. While sample size limits the broader generalizability of results, the approach provides new insights into quantitative analysis of qualitative data for better objectivity.

NLP topic modelling algorithm requires domain knowledge augmentation. Defining semantically-meaningful topic labels requires non-trivial effort and domain expertise. The process of selecting the topic modeling parameters lacks definitive guidance.

Lack of Diversity in the U.S. data population. In addition to the gender gap, there is also a lack of diversity in terms of ethnicity and race, which resulted in a lack of feedback from diverse populations. It would be interesting to do similar research at institutions in different national locations to understand the differences of responses from students.

5.5. Future Work

Further research in the future is needed to expand the sample population in NLP text analysis with supervised data to include different geographic representation of institutions to make inference with greater degrees of confidence. An automated way to measure cognitive aptitude, and previously acquired knowledge predictors could avoid the introduction of respondents' bias in the study. Future studies could also include a mathematics background, and the influence of AP computer science courses taken in high school.

REFERENCES

AAU (2022). http://www.aau.edu.et/aau-at-a-glance/

Alvarado, C., Cao, Y., & Minnes, M. (2017). Gender Differences in Students' Behaviors in CS Classes throughout the CS Major. Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education, 27–32.

Appianing, J. & Eck, R.N.V. (2015). Gender Differences in College Students' Perceptions of Technology-Related Jobs in Computer Science. In International Journal of Gender, Science and Technology, 7(1), 29–56.

Asfaw, A. (2012). Gender inequalities in tertiary education in Ethiopia: Mediating the transition to university through the development of adaptive competencies. Center for Universal Education at Brookings.

Ashcraft, C., McLain, B., & Eger, E. (2016). Women in Tech: The facts (2016 update). https://www.ncwit.org/sites/default/files/resources/womenintech_facts_fullreport_05132016.pdf.

Aspray, W. & Cohoon, J.M. (2017). Women in IT: Reasons on the Reasons of Under-Representation. Cambridge, MA: MIT Press. Google Inc. The world of women and mobile gaming. http://services.google.com/fh/files/misc/changethegame_white_paper.pdf.

Awis (2017). Inspiring Stories from Real STEM Heroes. Making the world we share a better place. The Association for Women in Science. https: www.awis.org/member-spotlights/.

Barker, L. J., & Aspray, W. (2006). The state of research on girls and IT. In J. M. Cohoon & W. Aspray (Eds.), Women and information technology: Research on underrepresentation

(pp. 2–54). Cambridge, MA: MIT Press.
http://dx.doi.org/10.7551/mitpress/9780262033459.003.0001

Basow, S. (2004). The hidden curriculum: Gender in the classroom. In M. A. Paludi (Ed.), Praeger guide to the psychology of gender (pp. 117 – 131). Westport, CT: Praeger/ Greenwood.

Beaubouef, T., & Mason, J. (2005). Why the high attrition rate for computer science students: some thoughts and observations. ACM SIGCSE Bulletin, 37(2), 103-106.

Beaubouef, T., Lucas, R., & Howatt, J. (2001). The UNLOCK system: Enhancing problem solving skills in CS-1 students. ACM SIGCSE Bulletin, 33(2), 43-46.

Beckwith, L., Burnett, M., Grigorenu, V., & Wiedenbeck, S. (2006). Gender HCI: What about the software? IEEE Computer Society Press.

Beekhuyzen, J. & Dorries, R. (2008). Tech girls are chic! (not just geek). tech2morrow, Brisbane.

Bennedsen, J. & Caspersen, M. E. (2007). Failure rates in introductory programming. AcM SIGCSE Bulletin, 39(2), 32-36.

Bers, M., Flannery, L., Kazakoff, E., & Sullivan, A. (2014). Computational thinking and tinkering: Exploration of an early childhood robotics curriculum. Computers and Education 72, 145–157.

Beyer, S. (2016). Women in CS: Deterrents. Encyclopedia of Computer Science and Technology, (2), CRC .

Biggers, M., Brauer, A., & Yilmaz, T. (2008). Student perceptions of computer science: a retention study comparing graduating seniors with cs leavers. In ACM SIGCSE Bulletin (Vol. 40, No. 1, pp. 402-406). ACM.

Box, S. (2018). Digital transformation must focus on women and girls. Policy Options Politiques. https://policyoptions.irpp.org/magazines/november-2018/digital-transformation-must-focus-women-girls/

Brackmann, C.P., Romàn-Gonzàlez, M., Robles, G., Moren-Leòn, J., Casali, A., & Barone, D. (2017). Development of Computational Thinking Skills through Unplugged Activities in Primary School. Proceedings of the 12th Workshop on Primary and Secondary Computing Education, Erik Barendsen and Peter Hubwieser (Eds.), 65–72.

Braun, F. (2008). Mehr Frauen in der Sprache. Leitfaden zur geschlechteren Formulierung. Ministerium für Justiz, Frauen, Jugend und Familie des Landes Schleswig Holstein.

Brewer, K. (2017). How the tech industry wrote women out of history. The Guardian. https://www.theguardian.com/careers/2017/aug/10/how-the-tech-industry-wrote-women-out-of-history.

CAITO (2020). Computer and Information Technology Occupations. https://www.bls.gov/ooh/computer-and-information-technology/home.htm.

Carter, L. (2006). Why students with an apparent aptitude for computer science don't choose to major in computer science. Proceedings of the 37th SIGCSE technical symposium on Computer science education, 27–31.

Cheryan, S., Plaut, V.C., Handron, C., & Hudso, L. (2013). The stereotypical computer scientist: Gendered media representations as a barrier to inclusion for women. Springer Science & Business Media, 69(1-2), 58.

Cheryan, S., Siy, J.O., Vichayapai, M., Drury, B.J., & Kim, S. (2011). Do female and male role models who embody STEM stereotypes hinder women's anticipated success in STEM? Journal of Personality and Social Psychology 2, 656–664. 22

Cheryan, S., Plaut, V., Davies, P., & Steele, C. (2009). Ambient belonging: How stereotypical cues impact gender participation in computer science. Journal of Personality and Social Psychology, 97, 1045–1060.

Clarke-Midura, J., Allan, V., & Close, K. (2016). Investigating the Role of Being a Mentor as a Way of Increasing Interest in CS. Proceedings of the 47th ACM Technical Symposium on Computing Science Education, 297–302.

Cohoon, J. M., & Aspray, W. (2006). Women and information technology: Research on underrepresentation. The MIT Press.

Craig, A., Coldwell-Neilson, J., & Beekhuyzen, J. (2013). Are IT interventions for Girls a Special Case? Proceeding of the 44th ACM technical symposium on Computer science education, 451–456.

Creswell, J. W., & Clark, V. L. P. (2017). Designing and conducting mixed methods research. Sage publications.

Dasgupta, N. & Stout, J.G. (2014). Girls and Women in Science, Technology, Engineering, and Mathematics. Policy Insights from the Behavioral and Brain Sciences 1(1), 21–29.

Denning, P. J., Comer, D. E., Gries, D., Mulder, M. C., Tucker, A., Turner, A. J., & Young, P. R. (1989). Computing as a discipline. Computer, 32(1), 202-210.

Eagan, M. K., Stolzenberg, E. B., Zimmerman, H. B., Aragon, M. C., Sayson, H. W., & Rios-Aguilar, C. (2017). The American freshman: national norms fall 2016. Los Angeles: Higher Education Research Institute, UCLA; 2017.

Egne, R. M. (2014). Gender Equality in Public Higher Education Institutions of Ethiopia: The Case of Science, Technology, Engineering, and Mathematics.

Elarde, J. (2016). Toward improving introductory programming student course success rates: experiences with a modified cohort model to student success sessions. Journal of Computing Sciences in Colleges, 32(2), 113-119.

El-Nasr, M.S., Yucel, I., Zopko, J., Andrea, B., & Smith, T.B. (2007). Middle-to-High School Girls as Game Designers - What are the Implications? Academic Days '07. Eurostat (2017) ICT specialists in employment. http://ec.europa.eu/eurostat/statistics-explained/index.php/ICT_specialists_in_employment.

Engler, S. & Wieland H.F. (1995). [German]: Ent-Dramatisierung der Differenzen. Studentinnen und Studenten der Technikwissenschaften (Ent-dramatization of differences. Students of the engineering sciences). Bielefeld: Kleine Verlag.

European Statistics (2018). Women in science and technology. https://ec.europa.eu/eurostat/web/products-eurostat-news/-/EDN-20180210-1?inheritRedirect=true.

Ford, J.L. (2009). Scratch programming for Teens. Computer Science Books.

Formanowicz, M., Cislak, A., Horvath, L., & Scesny, S. (2015). Capturing Socially Motivated Linguistic Change. Different effects of gender-fair language on support for social initiatives in Austria and Poland. Frontiers in Psychology 6, 1617.

Frieze, C. & Quesberry, J. (2015). Kicking Butt in Computer Science: Women in Computing at Carnegie Mellon University. Dog Ear Publishing.

Froyd, J. E., Shryock, K. J., Tripathy, M., Srinivasa, A. R., & Simon, R. C. (2015, June). Patterns of Students' Success: How Engineering Students Progress through a Course Sequence. In 2015 ASEE Annual Conference & Exposition (pp. 26-1215).

Jokers, M. (2017). https://github.com/mjockers/syuzhet

Gabay-Egozi, L., Shavit, Y., & Yaish, M. (2015). Gender Differences in Fields of Study: The Role of Significant Others and Rational Choice Motivations. European Sociological Review. 31(3), 284–297. German Federal Statistical Office (2017) [German] Zahl der Studierenden steigt im Wintersemester 2017/2018 erneut an (Number of students increases again in the winter semester 2017/2018). https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2017/11/PD17_427_213.html. 23

Galdi, S., Cadinu, M., & Tomasetto, C. (2014). The roots of stereotype threat: When automatic associations disrupt girls' math performance. Child development, 85(1), 250-263.

Giannakos, M. N., Jaccheri, L., & Leftheriotis, I. (2014). Happy girls engaging with technology: Assessing emotions and engagement related to programming activities. In International Conference on Learning and Collaboration Technologies (pp. 398-409). Springer, Cham.

Giannakos, M. N., Pappas, I. O., Jaccheri, L., & Sampson, D. G. (2017). Understanding student retention in computer science education: The role of environment, gains, barriers and usefulness. Education and Information Technologies, 22(5), 2365-2382.

Giltemeister, R. & Robert, G. (2008). [German] Geschlechterdifferenzierung in lebenszeitlicher Perspektive (Gender differentiation in life-time perspective). Interaktionen - Institute - Biografie. Wiesbaden, VS Verlag für Sozialwissenschaften.

Goode, J., Estrella, R., & Margolis, J. (2006). Lost in translation: Gender and high school computer science.

Grandl, M. & Ebner, M. (2017). [German] Informatische Grundbildung - ein Ländervergleich (Informatic basic education - a country comparison.

http://www.medienimpulse.at/pdf/Medienimpulse_Informatische_Grundbildung_____ein _La endervergleich_Grandl_20170514.pdf.

Grigoreanu, V., Cao, J., Kulesza, T., Bogart, C., Rector, K., Burnett, M., & Wiedenbeck, S. (2008). Can feature design reduce the gender gap in end-user software development environments? IEEE Symp. Visual Lang. Human-Centric Computing, 149–156.

Habte, D.T. (2017). Dialogue System for Advising Ethiopian Public Universities Students., Master's thesis, Addis Ababa University. http://etd.aau.edu.et/handle/123456789/19153

Harris, E., Winterbottom, M., Xanthoudaki, M., Calcagnini, S., & de Pijer, I. (2016). Tinkering: A Practitioner Guide for Developing and Implementing Tinkering Activities, (2014-1-IT02- KA200-003510), support from the Erasmus+ Programme of the European Union.

Hercy, N., Winston, M., Calvin, C., & Tak-Wai, C. (2009). Equal opportunity tactic: Redesigning and applying competition games in classrooms. In Computers & Education, 53(3), 866–876.

Hindle, A., Bird, C., Zimmermann, T., & Nagappan, N. (2015). Do topics make sense to managers and developers?. Empirical Software Engineering, 20(2), 479-515.

Hornik, K., & Grün, B. (2011). topicmodels: An R package for fitting topic models. Journal of statistical software, 40(13), 1-30.

Jayanth, M. (2014). 52% of gamers are women–but the industry doesn't know it. The Guardian, 18, 2014.

Johnson, C., McGrill, M., Bouchard, D., Bradshow, M.K., Bucheli, V.A., Merkle, L.D., Scott, M.J., Sweedyk, Z., Angel, J., Xiao, Z., & Zhang, M. (2016). Game Development for Computer Science Education. Proceedings of the 2016 ITiCSE Working Group Reports. 23–44.

Kahn, K. (2017). A half-century perspective on Computational Thinking. Tecnologias, sociedade e conhecimento, 4(1), 23-42.

Khaleel, F. L., Ashaari, N. S., Meriam, T. S., Wook, T., & Ismail, A. (2015). The study of gamification application architecture for programming language course. In Proceedings of the 9th international conference on ubiquitous information management and communication (pp. 1-5).

Khazan, O. (2018). The More Gender Equality, the Fewer Women in STEM. The Atlantic. https://www.theatlantic.com/science/archive/2018/02/themore-gender-equality-the-fewer-women-in-stem/553592/. Kimmel, M.S. 2004. The Gendered Society. Oxford University Press, Oxford.

Ko, A. J., & Davis, K. (2017). Computing mentorship in a software boomtown: Relationships to adolescent interest and beliefs. In Proceedings of the 2017 ACM conference on international computing education research (pp. 236-244).

Krieger, S., Rown, A., (2015). Are Females Disinclined to Tinker in Computer Science? Proceedings of the SIGCSE '15.

Kumar, D. (2011). Ready for a third peak? In ACM Inroads 2(3), 10–11.

Ladner, R. E., & Israel, M. (2016). For all" in" computer science for all. Communications of the ACM, 59(9), 26-28.

Lamborelle, A. & Fernandez, L. (2016). Woman in ICT - How do EU member states measure up? http://www.euractiv.com/section/digital/infographic/women-in-ict-how-do-eu-member-statesmeasure-up/.

Layman, L., Nikora, A. P., Meek, J., & Menzies, T. (2016). Topic modeling of nasa space system problem reports: Research in practice. In Proceedings of the 13th International Conference on Mining Software Repositories (pp. 303-314).

LeBlanc, C.A. (2020). First-Year Computer Science Students: Pathways and Perceptions in Introductory Computer Science Courses. Electronic Theses and Dissertations. 3160 https://digitalcommons.library.umaine.edu/etd/3160

Lewis, C.M., Anderson, R.E. & Yasuhara K. (2016). I Don't Code All Day: Fitting in Computer Science When the Stereotypes Don't Fit. ACM Conference on International Computing Education Research. ACM, New York.

Lister, R. (2011). Computing education research geek genes and bimodal grades. ACM Inroads, 1(3), 16-17.

Lockwood, P. (2006). Someone like me can be successful: do college students need same-gender role models? Psychology of Women Quarterly.

Lynch, T., Tompkins, J.E., Van Driel, I.I., & Fritz, N. (2016). Sexy, Strong, and Secondary: A Content Analysis of Female Characters in Video Games across 31 Years. Journal of Communication, 66(4), 564–584.

Mannila, L., Dagiene, V., Demo, B., Grgurina, N., Mirolo, C., Rolandsson, L., & Settle, A. (2014). Computational thinking in K-9 education. Proceedings of the Working Group Reports of the 2014 on Innovation & Technology in Computer Science Education Conference, 1–29.

Margolis J., Fisher A., Miller F. (2000). The anatomy of interest: women in undergraduate computer science. Women's Stud. Q. 28, 104–127.

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

Master, A., Sapna, C., & Meltzoff, A.N. (2016). Computing Whether She Belongs: Stereotypes Undermine Girls' Interest and Sense of Belonging in Computer Science. Journal of Educational Psychology, 108(3), 424–437.

Matlin, M. (1999). Bimbos and Rambos: The Cognitve Basis of Gender Stereotypes. Eye on PsiChi published by the national Honor Society in Psychology, 3(2), 13–14. 25

McLean, M. & Harlow, D. (2017). Designing Inclusive STEM Activities: A Comparison of Playful Interactive Experiences Across Gender. Proceedings of the 2017 Conference on Interaction Design and Children, 567–574.

Medel, P. & Pournagshband, V. (2017). Eliminating Gender Bias in Computer Science Education Materials. Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education, 411–416.

Melak, A., Singh, S. (2021). Factors Affecting Women's Choice of Learning Engineering and Technology Education.

Melak, A., & Singh, S. (2021). Women's participation and factors affecting their academic performance in engineering and technology education: A study of Ethiopia. Sustainability, 13(4), 2246.

Microsoft (2017). [German] Microsoft-Studie: Kreativität ist Schlüssel für mehr Frauen in MINT-Berufen (Microsoft study: Creativity is the key to get more women in STEM-jobs). https://news.microsoft.com/de-de/microsoft-studie-mehrfrauen-mint-berufen/.

Morse, R. and Castonguay, A. (2022). 2022 Best Global Universities in Africa. https://www.usnews.com/education/best-global-universities/africa

National Academies of Sciences, Engineering and Medicine (2018). Assessing and Responding to the Growth of Computer Science Undergraduate Enrollments. Washington, DC: The National Academies Press. https://doi.org/10.17226/24926.

National Science Foundation, National Center for Science and Engineering Statistics (2019). *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2019.* Special Report NSF 19-304. Alexandria, VA. https://www.nsf.gov/statistics/wmpd.

National Science Foundation (2015). Computer Sciences Degrees Awarded, by Degree Level and Sex of Recipient: 1966 – 2012. http://www.nsf.gov/statistics/2015/nsf15326/?WT.mc_id=USNSF_178#chp2.

NCES (2012). Bachelor's Degrees Conferred by Degree-Granting Institutions, by Field of Study: Selected Years, 1970-71 Through 2010-11. National Center for Education Statistics. http://nces.ed.gov/programs/digest/d12/tables/dt12_313.asp

NCWIT (2017). NCWIT's Women in IT: By the Numbers presents the most compelling statistics on women's participation in IT on a single page. www.ncwit.org/bythenumbers. National Center for Women and Information Technology.

NDSU (2022). https://www.ndsu.edu/about

NewZoo (2017). Male and Female Gamers: How Their Similarities and Differences Shape the Games Market. https://newzoo.com/insights/articles/maleand-female-gamers-how-their-similarities-and-differences-shape-the-games-market/.

Outlook (2020). What is the Employment Outlook for Computer Science Careers? https://www.computersciencedegreehub.com/faq/what-is-the-employment-outlook-for-computer-science-careers/

Parmar, Manojkumar, Maturi Bhanurekha, Dutt Jhuma Malik, Phate, Hrushikesh (2018). Sentiment Analysis on Interview Transcripts: An application of NLP for Quantitative Analysis.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in information retrieval, 2(1–2), 1-135.

Pedro-Croft (2015). TOMB RAIDER 1996-2015. http://fav.me/d95eqwd.

Porter, L., Guzdial, M., McDowell, C., & Simon, B. (2013). Success in introductory programming: What works? Communications of the ACM, 56(8), 34-36.

Raigoza, J. (2017). A study of students' progress through introductory computer science programming courses. In 2017 IEEE Frontiers in Education Conference (FIE) (pp. 1-7). IEEE.

Ramalingam, V., LaBelle, D., & Wiedenbeck, S. (2004). Self-efficacy and mental models in learning to program. In ITICSE, Leeds, UK, 171- 175.

Resnick, M., Maloney, J., Monroy-Hernàndez, A., Rusk, N., Eastmond, E., Brenan, K., Millner, A., Rosenbaum, E., Silver, J., Silverman B., & Kafai., Y. (2009). Scratch: programming for all. In Commun. ACM, 52(11), 60–67.

Robins, A. V. (2019). 12 Novice Programmers and Introductory Programming. The Cambridge handbook of computing education research, 327.

Sadler, P.M., Sonnert, G., Hazari, Z., & Tai, R. (2012). Stability and volatility of STEM career interest in high school: A gender study. Science Education, 96(3), 411–427.

Schwartz, K. (2013). Giving Good Praise to Girls: What Messages Stick. https://ww2.kqed.org/mindshift/2013/04/24/giving-good-praise-to-girls-whatmessages-stick/. KQED Inc.

Semela, T. (2010). Who Is Joining Physics and Why? Factors Influencing the Choice of Physics among Ethiopian University Students. International Journal of Environmental and Science Education, 5(3), 319-340.

Shillabeer, A. & Jackson, K. (2013). Gender imbalance in undergraduate IT programs — A Vietnamese perspective. In Innovation in Teaching and Learning in Information and Computer Sciences 12(1), 70–83.

Siobhan, A. (2018). A Man's World: Women in Computer Science. The Saint. http://www.thesaint-online.com/2018/11/a-mans-world-women-incomputer-science/.

SiTE (2021). School of Information Technology and Engineering (SiTE) Modular Curriculum for Degree of Bachelor of Science in Software Engineering and Computing Technology (Software Engineering, Cyber Security, Information Technology, Artificial Intelligence Streams).

Slany, W., Luhana, K. Müller, M., Schindler, C., & Spieler, B. (2018). Rock Bottom, the World, the Sky: Catrobat, an Extremely Large-scale and Long-term Visual Coding Project Relying Purely on Smartphones, Proceedings of Constructionism 2018. 20-25.

Spieler, B., & Slany, W. (2018). Female Teenagers and Coding: Create Gender Sensitive and Creative Learning Environments. arXiv preprint arXiv:1805.04366.

Spieler, B. (2018). Reinforcing gender equality by analyzing female teenagers' performances in coding activities: a lesson learned. In Proceedings of the 4th Conference on Gender & IT (pp. 209-219).

Spieler, B., Schindler, C., Slany, W., & Mashkina, O. (2017a). App Creation in Schools for different Curricula Subjects - Lessons Learned. Proceedings of the 9th International Conference on Education and New Learning Technologies, 5814–5823.

Spieler, B., Schindler, C., Slany, W., & Mashkina, O., Beltràn, M.E., Boulton, H., & Brown, D. (2017b). Evaluation of Game Templates to support Programming Activities in Schools. Proceedings of the 11th European Conference on Games Based Learning, 600–609.

Spieler, B., Petri, A., Slany, W., Schindler, C., Beltràn, M.E., & Boulton, H. (2016). Pocket Code: A Mobile App for Game Jams to facilitate Classroom Learning through Game Creation. Proceedings of the Irish Conference on Game-Based Learning, 61–79.

Steele, C. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. American Psychologist, 52(6), 613-629.

Stoet, G. & Geary, D.C. (2018). The Gender-Equality Paradox in Science, Technology, Engineering, and Mathematics Education. In Psychological Science. Unesco (2015) STEM and Gender Advancement. http://uis.unesco.org/en/news/stem-and-gender-advancement.

Misa J.T. (2010). Gender Codes: Why Women Are Leaving Computing. IEEE Computer Society Press.

Unfried, A., Da Faber, M., Stanhope, D.S., & Wiebe, E. (2015). The Development and Validation of a Measure of Student Attitudes Toward Science, Technology, Engineering, and Math (S-STEM). Journal of Psychoeducational Assessment, 33(7), 622–639.

U.S. Bureau of Labor Statistics (2021). Computer and Information Technology Occupations. The Atlantic. https://www.bls.gov/ooh/computer-and-information-technology/home.htm

Veilleux, N., Bates, R., Allendoerfer, C., Jones, D., Crawford, J., & Smith, T.F. (2013). The relationship between belonging and ability in computer science. Proceeding of the 44th ACM technical symposium on Computer science education, 65–70.

Vervecken, D. & Hannover, B. (2012). Ambassadors of gender equality? How use of pair forms versus masculines as generics impacts percept ion of the speaker. European Journal of Social Psychology, 42, 754–762.

Vilnius, Lithuania Slany, W. (2014). Tinkering with Pocket Code, a Scratch-like programming app for your smartphone. In Proceedings of Constructionism 2014.

Walker, G. (2004). Experimentation in the Computer Programming Lab. Inroads, (36:4), 69-72.

Weibert, A., von Rekowski, T., & Festl, L. (2012). Accessing IT: a curricular approach for girls. Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design, 785–786. WES (2017) The Women's Engineering Society (WES). http://www.wes.org.uk/statistics.

Weiner, B. (2010). The development of an attribution-based theory of motivation: a history of ideas. Educational Psychologist 45(1), 28- 36.

West, C., & Zimmerman, D. H. (1987). Doing gender. Gender & society, 1(2), 125-151.

Wisniewski, T. (2017) The Future Tech Workforce: Breaking Gender Barriers. http://www.isaca.org/SiteCollectionDocuments/Breaking-GenderBarriers_res_eng_0317.PDF.

Wodon, Q.T., and B. de la Brière (2018). Unrealized Potential: The High Cost of Gender Inequality in Earnings. Briefing report 126579. Washington, DC: World Bank.

Wong, B. & Kemp, P.E.J. (2017). Technical boys and creative girls: the career aspirations of digitally skilled youths. The Cambridge Journal of Education.

Wu, B. & Wang, A.I. (2012). A Guideline for Game Development-Based Learning: A Literature Review. International Journal of Computer Games Technology.

Yee, N. (2017). Just How Important Are Female Protagonists? https://quanticfoundry.com/2017/08/29/just-important-female-protagonists/.

Young, D.M., Rudman, L.A., Buettner, H.M., & McLean, M.C. (2013). The Influence of Female Role Models on Women's Implicit Science Cognitions. Psychology of Women Quarterly, 37(3), 283–292.

Zagami, J., Boden, M., Keane, T., Moreton, B., & Schulz, K. (2015). Girls and computing: Female participation in computing in schools. Australian Educational Computing, 30(2), 1–14.

APPENDIX A. SURVEY CONSENT FORM

Informed Consent: Adult Consent
for participation in a PhD Research Survey

Research Title:

Persistence Factors in Computer Science: A Comparative Study

Researcher:

Bethlehem Gronneberg, PhD candidate in Software Engineering,
Department of Computer Science,
North Dakota State University
Contact: bethlehem.gronneberg@ndsu.edu

Principal Researcher:

Kendall Nygard, Professor Emeritus
Department of Computer Science,
North Dakota State University
Contact: kendall.nygard@ndsu.edu

Welcome. You are being invited to participate in PhD research to understand attrition and persistence factors for Computer Science undergraduates. The purpose of this research is to investigate and conduct comparative analysis of underlying factors of persistence by studying the holistic learning experiences of Computer Science undergraduates during their first two years. This research is being conducted by a PhD candidate in Software Engineering, Ms. Bethlehem Gronneberg. Her research advisor Prof. Kendall Nygard is the principal investigator of this research.

Your participation in this survey is confidential and your responses will not identify you by name. Your instructors will not see your responses. Nothing you say on the questionnaire will, in any way, influence your present or future course grades. There are no known risks if you decide to participate in this research, nor are there any costs for participating in the study. The information you provide will contribute to an important research area that may shed some insights and lessons on effective persistence methods to support future students of Computer Science.

Your participation in this research is voluntary. You have the right to withdraw at any point during the study. You will be entered in a drawing for a chance to win a $50 Amazon gift certificate for your participation. There is only one drawing for this survey participants.

The Principal Investigator of this research, Prof. Kendall Nygard, can be contacted at kendall.nygard@ndus.edu. If you have questions about your rights or complaints about this research, you may talk to the researcher, Bethlehem Gronneberg at bethlehem.gronneberg@ndsu.edu or contact the NDSU Human Research Protection Program at 701.231.8995, toll-free at 1-855-800-6717, by email at ndsu.irb@ndsu.edu.

By clicking the button below, you acknowledge that your participation in the study is voluntary.

- I consent, begin the research survey
- I do not consent, I do not wish to participate

APPENDIX B. INTERVIEW CONSENT FORM

Informed Consent: Adult Consent
for Participation in a PhD Research

Research Title:
Persistence Factors in Computer Science: A Comparative Study

Researcher:
Bethlehem Gronneberg, PhD candidate in Software Engineering, Department of Computer Science, North Dakota State University
Contact: bethlehem.gronneberg@ndsu.edu

Principal Researcher:
Kendall Nygard, Professor Emeritus
Department of Computer Science, North Dakota State University
Contact: kendall.nygard@ndsu.edu

Welcome. You are being invited to participate in PhD research to understand attrition and persistence factors for Computer Science undergraduates. The purpose of this research is to investigate and conduct comparative analysis of underlying factors of persistence by studying the holistic learning experiences of Computer Science undergraduates during their first two years. This research is being conducted by a PhD candidate in Software Engineering, Ms. Bethlehem G Gronneberg. Her research advisor Prof. Kendall Nygard is the principal investigator of this research.

This interview should only take you around 30 minutes to 45 minutes of your time. You will get to choose how to conduct the interview, either in-person or in a virtual setting using Blackboard Collaborate, Zoom, WebEx or Microsoft Teams. The interview will be audio recorded with your permission.

Your participation in this interview is confidential and your responses will be anonymous. Your instructors will not see your responses. Nothing you say during the interview will, in any way, influence your present or future course grades. Only the primary investigator and a professional transcriber will have access to the recordings of your audio interviews.

The information you provide will contribute to an important research area that may shed some insights and lessons on effective persistence methods to support future students of Computer Science. There are no known risks if you decide to participate in this research, nor are there any costs for participating in the study.

Your participation in this research is voluntary. You have the right to withdraw at any point during the study. You will be entered in a drawing for a chance to win a $50 Amazon gift certificate for your interview participation. There is only one drawing for all Computer Science major participants for your university.

The Principal Investigator of this research, Prof. Kendall Nygard, can be contacted at kendall.nygard@ndus.edu. If you have questions about your rights or complaints about this research, you may talk to the researcher, Bethlehem Gronneberg at bethlehem.gronneberg@ndsu.edu or contact the NDSU Human Research Protection Program at 701.231.8995, toll-free at 1-855-800-6717, by email at ndsu.irb@ndsu.edu.

By selecting the options below, you acknowledge that your participation in the research is voluntary, you give permission to be recorded and that you are a Computer Science major.

○ I consent to be interviewed and audio recorded, begin the research interview

○ I do not consent to be neither interview nor audio recorded, I do not wish to participate

____
Full Name

____
Signature

APPENDIX C. IRB EXEMPT DETERMINATION

NDSU NORTH DAKOTA STATE UNIVERSITY

09/13/2021

Dr. Kendall E Nygard
Computer Science

Re: IRB Determination of Exempt Human Subjects Research:
Protocol #IRB0003563, "The First-Two Critical Years of Female Computer Science Undergraduates: A Comparative Study in Ethiopia & the United States"

NDSU Co-investigator(s) and research team:

- Kendall E Nygard
- Bethlehem A Gronneberg

Approval Date: 09/13/2021
Expiration Date: 09/12/2024
Study site(s): Method I: Survey - Online The course instructors of NDSU, UND, and Ethiopian Universities will distribute anonymous link to the Qualtrics based online written survey for volunteer participants. Method II: Interview - online or in-person The volunteer participants at NDSU will choose how to answer the interview, either online, or in-person. The in-person interviews will be conducted at the Computer Science department computer labs and the online one will be conducted using the medium chosen by the participants such as Blackboard Collaborate, WebEx, Zoom or Microsoft Teams. Interviews will be conducted online for the Ethiopian universities using Zoom or Microsoft Teams. We will not conduct interview at UND.
Funding Agency:
The above referenced human subjects research project has been determined exempt (category 2,3) in accordance with federal regulations (Code of Federal Regulations, Title 45, Part 46, *Protection of Human Subjects*).

Please also note the following:

- The study must be conducted as described in the approved protocol.
- Changes to this protocol must be approved prior to initiating, unless the changes are necessary to eliminate an immediate hazard to subjects.
- Promptly report adverse events, unanticipated problems involving risks to subjects or others, or protocol deviations related to this project.

Thank you for your cooperation with NDSU IRB procedures. Best wishes for a successful study.

NDSU has an approved FederalWide Assurance with the Department of Health and Human Services: FWA00002439.

RESEARCH INTEGRITY AND COMPLIANCE
NDSU Dept 4000 | PO Box 6050 | Fargo ND 58108-6050 | ndsu.research@ndsu.edu

Shipping Address: Research 1, 1735 NDSU Research Park Drive, Fargo ND 58102

NDSU is an EO/AA university.

APPENDIX D. SURVEY INSTRUMENTS

Interview Questions

1. Why did you choose to pursue Computer Science/Software Engineering as a college degree?
2. What was your experience like during your first CS course in college? What prepared you in high school?
3. What was the challenge?
4. What made you persist?
5. How was the department support?
6. How was your interaction with your peers?
7. What aspects of Computer Science do you enjoy?
8. What is your advice for future aspiring Computer Science/Software Engineering majors?

CS I Pre-/POST- Survey

Section II:

Statements on Your Degree's Usefulness:

	Strongly Disagree	Disagree	Neither Disagree Nor Agree	Agree	Strongly Agree
I study Computer Science because it helps me secure a strong financial future.	○	○	○	○	○
I study Computer Science because I hope to make a meaningful difference in the world by helping solve real-world problems.	○	○	○	○	○
I am taking this course because my advisor said it is necessary for my future.	○	○	○	○	○
I believe a Computer Science qualification gives me access to a more diverse range of opportunities and industry than almost any other qualification.	○	○	○	○	○
I believe my future career will require the use of Computer Science concepts.	○	○	○	○	○
I am taking this course because it is required for my non-Computer Science major.10	○	○	○	○	○
I believe the tools and techniques I learn from programming can be useful in the study of other disciplines.	○	○	○	○	○
I believe analytical thinking and problem solving skills used in programming can be helpful in my everyday personal and professional life.	○	○	○	○	○
I believe a degree in Computer Science allows me to choose where I want to live and who I want to work for because the skill is in-demand everywhere in the United States.	○	○	○	○	○
I believe I need a firm mastery of programming for my future work.	○	○	○	○	○

Section III:

Statements on Your Previously Acquired Knowledge & Your Expectations for CSCI 160:

	Strongly Disagree	Disagree	Neither Disagree Nor Agree	Agree	Strongly Agree
I have no previous experience in computer programming before attending this class.	○	○	○	○	○
I have taken one or more formal computer programming courses before attending this class.	○	○	○	○	○
I have learned how to code on my own before attending this class.	○	○	○	○	○
I have taken the course CSCI 159 (Computer Science Problem Solving) before this class.	○	○	○	○	○
I believe the programming courses I took before college has prepared me to take this class.	○	○	○	○	○
I believe the advanced math courses I took before this class, such as College Algebra, Calculus, Statistics, etc, will help me tackle the programming challenges.	○	○	○	○	○
I expect learning how to code in CSCI 160 to be difficult.	○	○	○	○	○
I don't think I will do well in programming.	○	○	○	○	○
I know how to write syntactically correct statements in one or more of the high level programming languages.	○	○	○	○	○
I understand the Object Oriented Programming (OOP) paradigm.	○	○	○	○	○

Section IV:

Statements on Your Knowledge and Attitude on Programming:

	Strongly Disagree	Disagree	Neither Disagree Nor Agree	Agree	Strongly Agree
Programming is enjoyable and stimulating to me.	○	○	○	○	○
I am confident I will do well in this programming class.	○	○	○	○	○
For some reason even though I work hard at it, programming seems unusually hard for me.	○	○	○	○	○
If I get stuck on a programming problem, there is no chance I will figure it out on my own.12	○	○	○	○	○
I cannot complete a programming task unless I have a lot of time to complete the program.	○	○	○	○	○
I can find a way to concentrate on my program, even when there are many distractors around me.	○	○	○	○	○
I know how to manage my time efficiently if I have a pressing deadline on a programming project.	○	○	○	○	○
I can find ways of motivating myself to program, even if the problem area is of no interest to me.	○	○	○	○	○
I can come up with a suitable strategy for a given programming project in a short time.	○	○	○	○	○
Understanding programming means being able to recall something I have read or been shown.	○	○	○	○	○
I have no confidence that I can handle more difficult programming problems.	○	○	○	○	○
When working on a programming problem, I find it useful to brainstorm about solution strategies before writing code.	○	○	○	○	○

Section II:

Your Reflection on CSCI 160 course:

	Strongly Disagree	Disagree	Neither Disagree Nor Agree	Agree	Strongly Agree
This course is challenging for me.	○	○	○	○	○
After taking this course, I am less confident I will do well in future programming classes.	○	○	○	○	○
I wish I had more programming knowledge and skills before taking this course.	○	○	○	○	○
The professor was instrumental in helping me understand the concepts of this course,	○	○	○	○	○
The book was instrumental in helping me understand the concepts of this course.	○	○	○	○	○
Spending a lot of time on the lab assignments was instrumental in helping me understand the concepts of this course.	○	○	○	○	○
Support from lab assistants and tutors was available when I needed it.	○	○	○	○	○
I believe my programming skills have increased from taking this course.	○	○	○	○	○
To do well in this course, I had to seek outside help (google or friends) on a regular basis.	○	○	○	○	○
To do well in this course, it was critical to not put off my course assignments until tomorrow or the next day.	○	○	○	○	○
This course has helped me decide Software Engineering/Computer Science major is not for me.	○	○	○	○	○

CS II Pre-/POST- Survey

Section II: Instruction: Read the statements below and select the choice that best fits your feelings.

Statements on Your Degree's Usefulness:

	Strongly Disagree	Disagree	Neither Disagree Nor Agree	Agree	Strongly Agree
I study Computer Science because it helps me secure a strong financial future.	○	○	○	○	○
I believe a Computer Science qualification gives me access to a more diverse range of opportunities than almost any other qualification.	○	○	○	○	○
I believe I need a firm mastery of programming for my future work.	○	○	○	○	○
I have no interest in Computer Science. I am taking this course because it is required for my non-Computer Science major.	○	○	○	○	○
I believe the tools and techniques I learn from programming can be useful in the study of other disciplines.	○	○	○	○	○
I believe a Computer Science degree equips me with a problem-solving frame of mind if I want to build a tech start-up in the future.	○	○	○	○	○
I believe the analytical thinking and problem solving skills used in programming can be helpful in my professional life.	○	○	○	○	○
I believe an advanced degree in Computer Science allows me to work on innovative research projects at research institutes or big tech companies.	○	○	○	○	○
A degree in Computer Science allows me to choose where I want to live or who I want to work for because the skill is in-demand everywhere in the United States.	○	○	○	○	○
The Computer Science academic advising I received was helpful to me.	○	○	○	○	○

Section III: Regarding Your Previously Acquired Knowledge or Skills on Programming: Instruction: Read the

Statements on Your Previously Acquired Knowledge or Skills on Programming:

NOTE: In the statements below, the phrase "programming language" refers to your knowledge of any one of a high level programming languages, e.g. Java, C#, Python or C++.

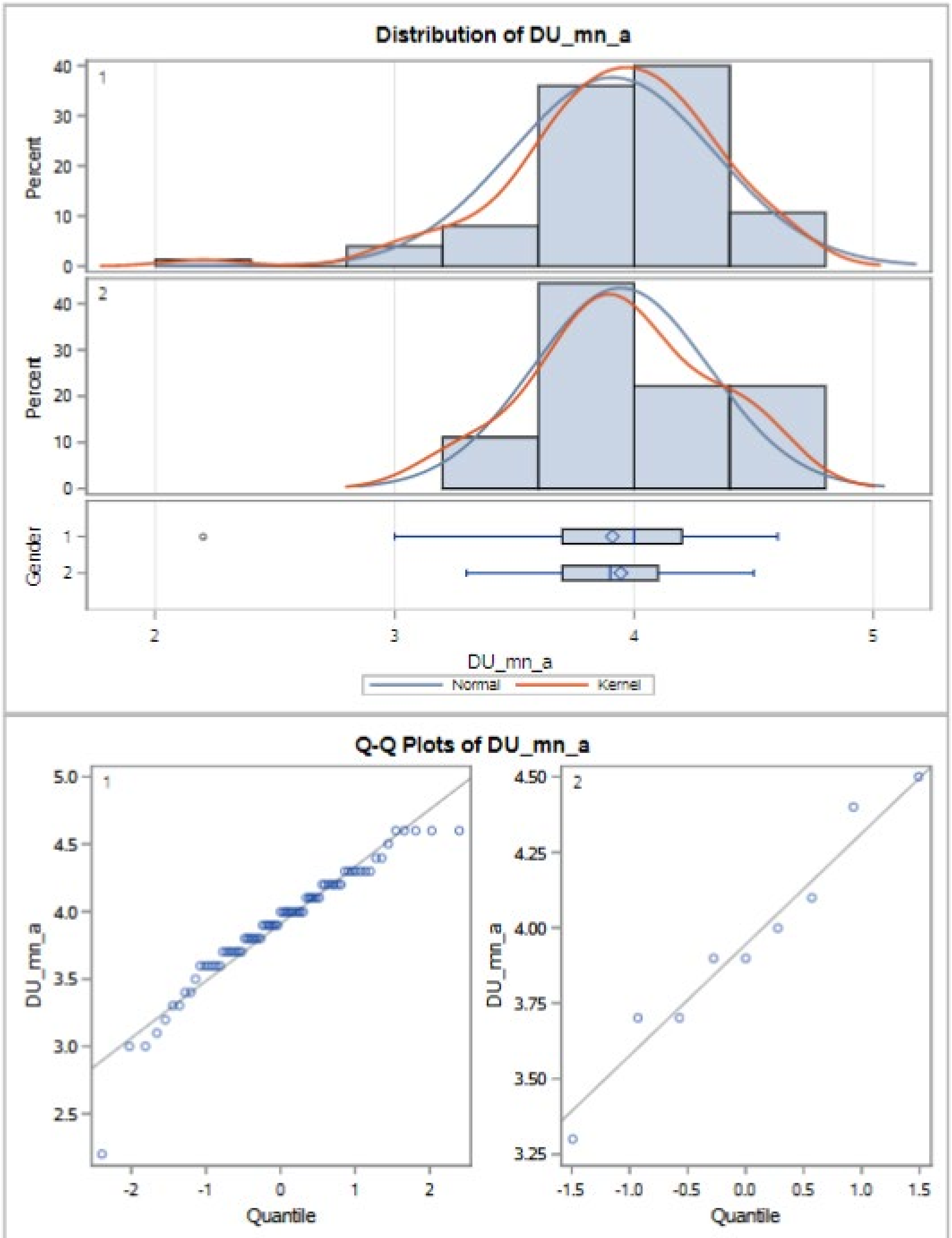
	Strongly Disagree	Disagree	Neither Disagree Nor Agree	Agree	Strongly Agree
I know how to write syntactically correct statements in one or more of the programming languages.	○	○	○	○	○
I understand the Object Oriented Programming (OOP) paradigm.	○	○	○	○	○
I know how to identify, declare, define, and use objects in the problem domain.	○	○	○	○	○
I know how to write a simple program given a problem domain I understand.	○	○	○	○	○
I know how to debug a program that I have written and make it work.	○	○	○	○	○
I know how to make use of a pre-written method, given a clearly labeled declaration of the method.	○	○	○	○	○
I know how to design my program in a modular manner to be reusable, extended and adapted.	○	○	○	○	○
I can comprehend a long, complex program.	○	○	○	○	○
I cannot complete a programming task unless I can call someone for help when I get stuck.	○	○	○	○	○
If I get stuck while working on a programming task, I can find ways of overcoming the problem.	○	○	○	○	○

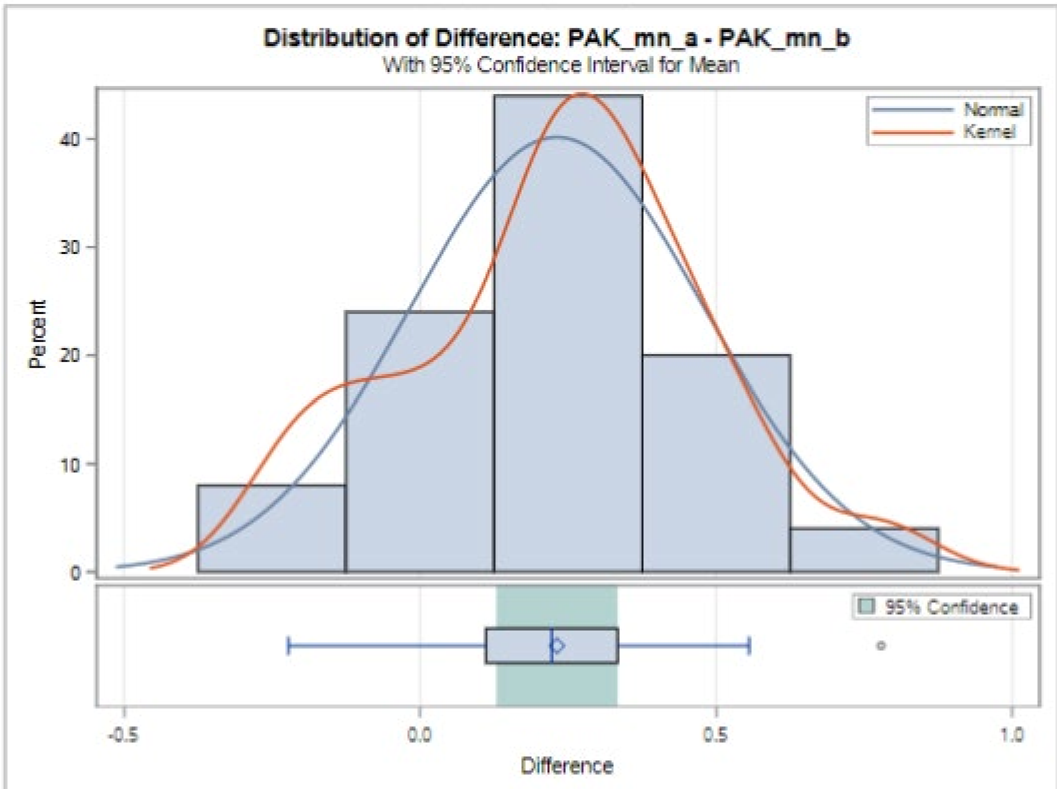
Section IV:

Statements on Your Knowledge and Attitude on Programming:

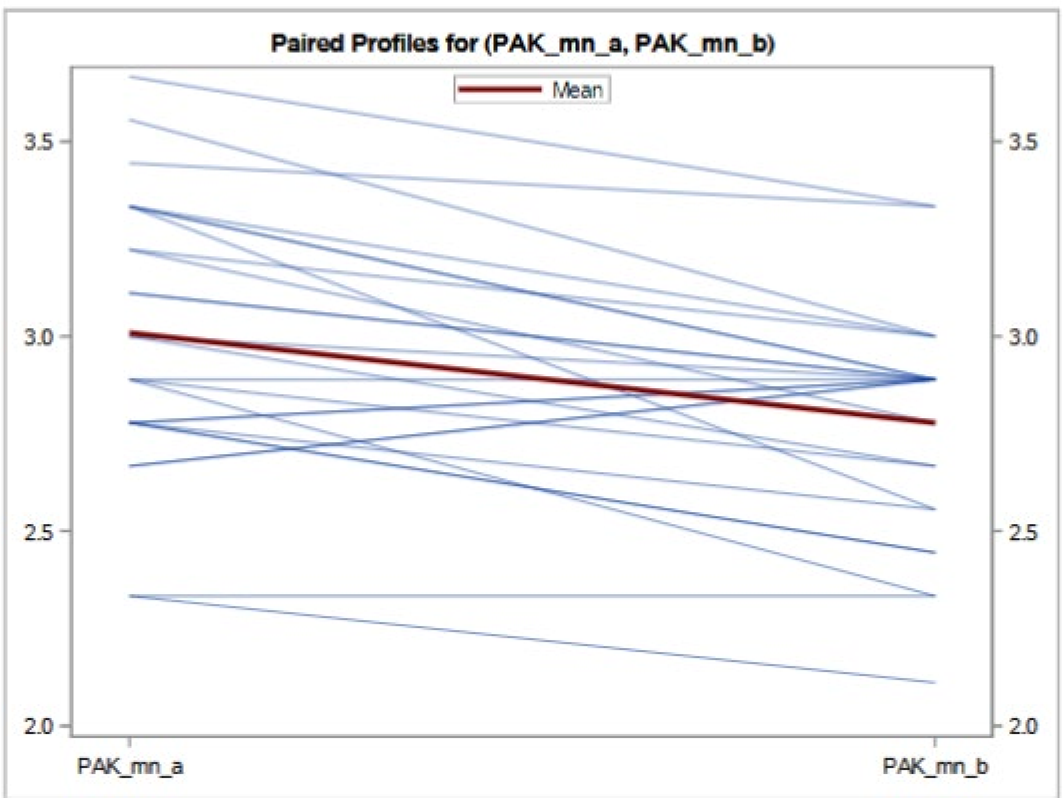
	Strongly Disagree	Disagree	Neither Disagree Nor Agree	Agree	Strongly Agree
I have considered dropping out during this course, this semester.	○	○	○	○	○
I am unable to keep up with the workload.	○	○	○	○	○
I am no longer enjoying the programming course.	○	○	○	○	○
I feel like I lack the practical skill development to pursue in Computer Science major.	○	○	○	○	○
Click to write Statement 18	○	○	○	○	○
Programming is enjoyable and stimulating to me.	○	○	○	○	○
I have decided to drop out of Computer Science major.	○	○	○	○	○
I am confident I will do well in programming class.	○	○	○	○	○
For some reason even though I work hard at it, programming seems unusually hard for me.	○	○	○	○	○
When I feel stuck on a programming problem, I give up easily and seek help from someone else.	○	○	○	○	○
I cannot complete a programming task unless I have a lot of time to complete the program.	○	○	○	○	○
I know how to manage my time efficiently if I have a pressing deadline on a programming project.	○	○	○	○	○
I can find a way to concentrate on my program, even when there are many distractors around me.	○	○	○	○	○
I can find ways of motivating myself to program, even if the problem area is of no interest to me.18	○	○	○	○	○
I can come up with a suitable strategy for a given programming project in a short time.	○	○	○	○	○
I can write a program that someone else can comprehend at a later date.	○	○	○	○	○
I look forward to writing a complex program that can solve a complex problem.	○	○	○	○	○
I have no confidence that I can handle more difficult programming problems.	○	○	○	○	○

APPENDIX E. SCATTER PLOT AND HISTOGRAM

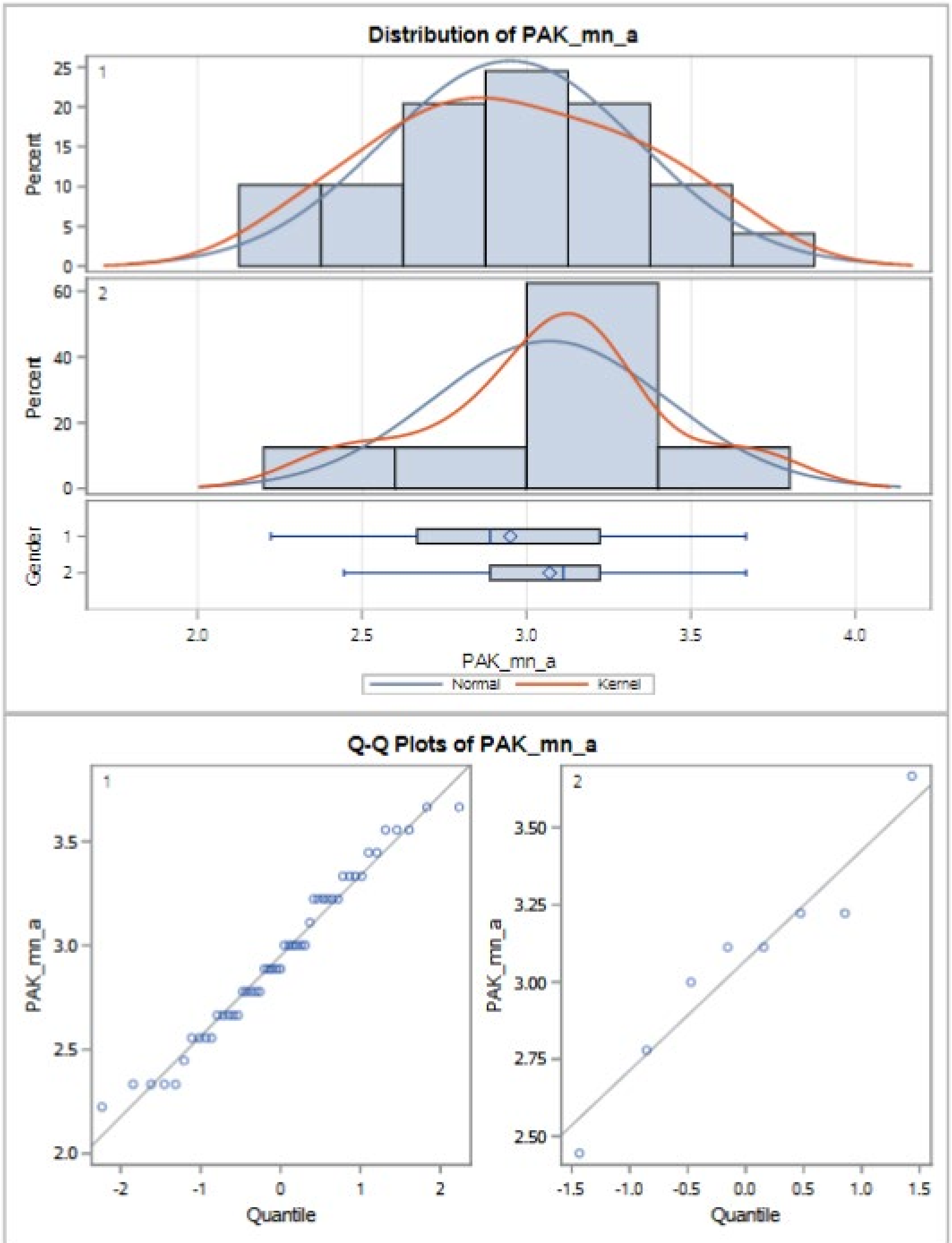


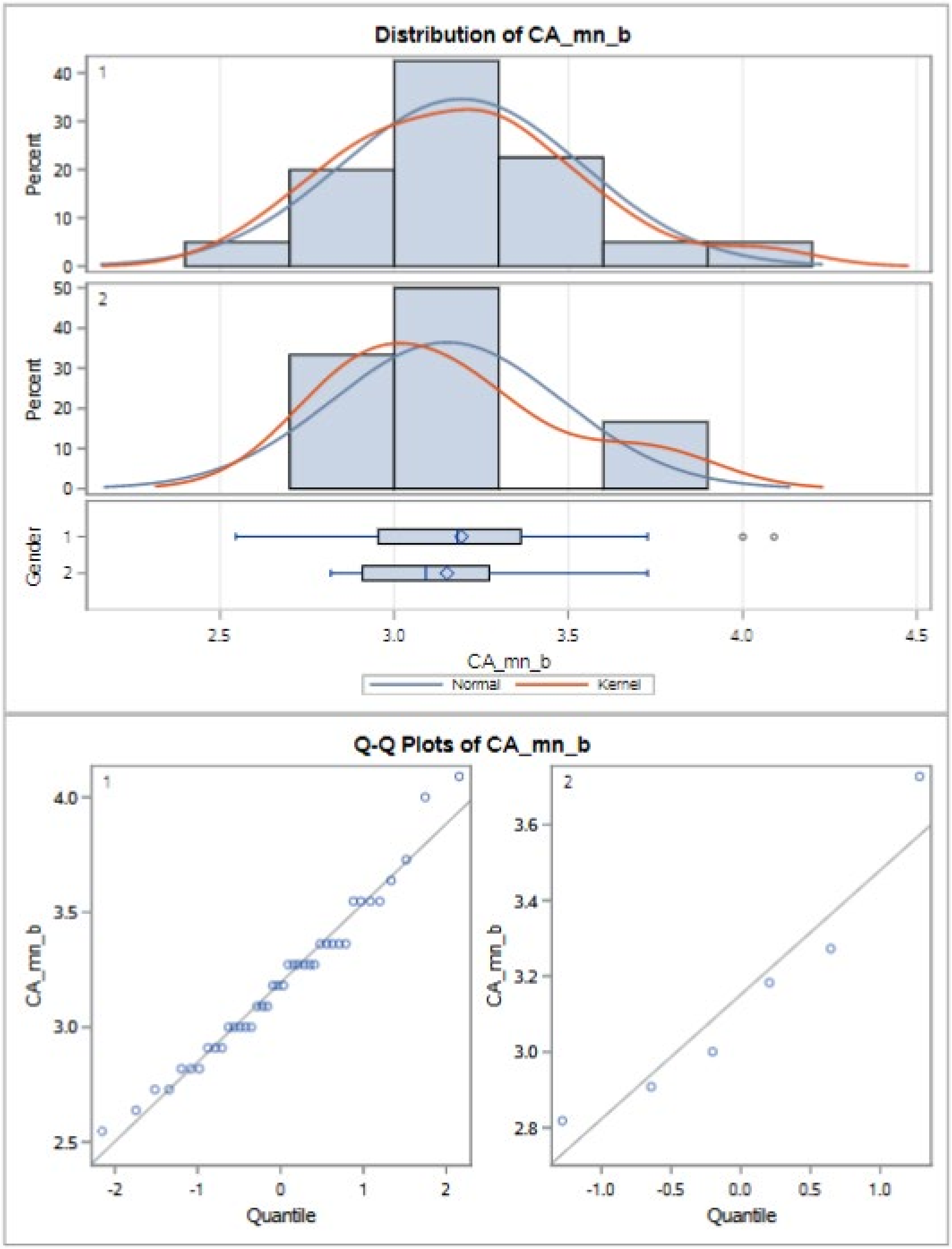


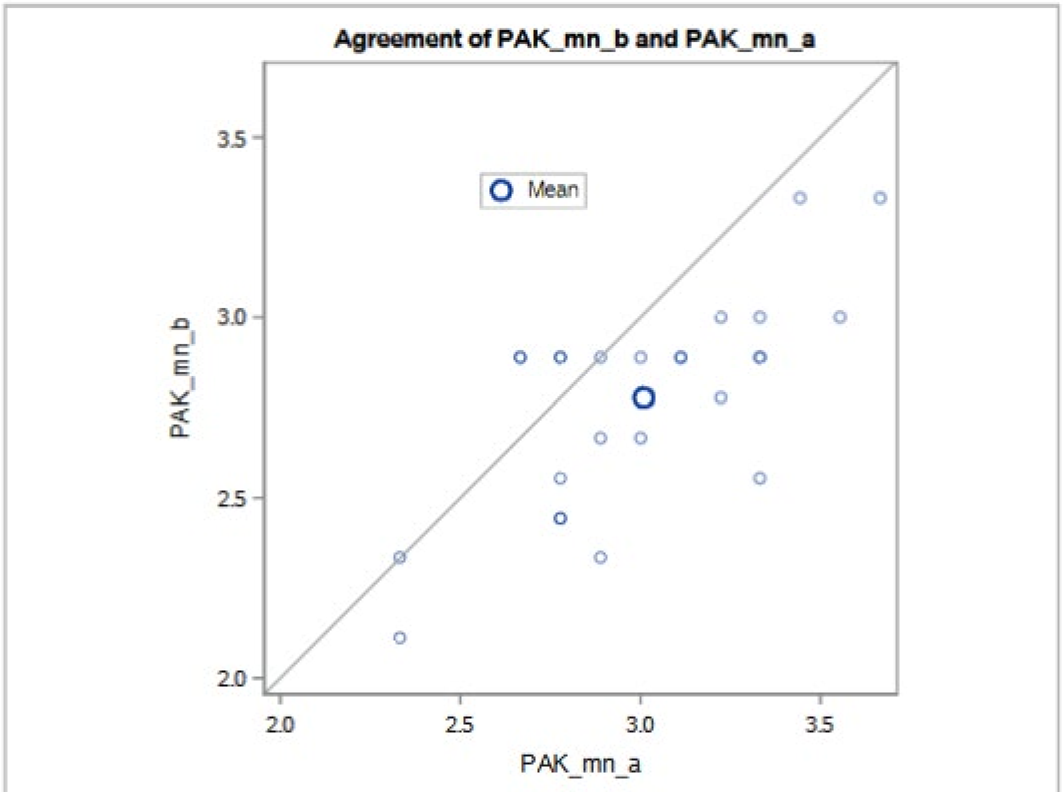
Distribution of Difference: PAK_mn_a - PAK_mn_b
With 95% Confidence Interval for Mean
Normal
Kernel
Percent
0
10
20
30
40
95% Confidence
-0.5
0.0
0.5
1.0
Difference

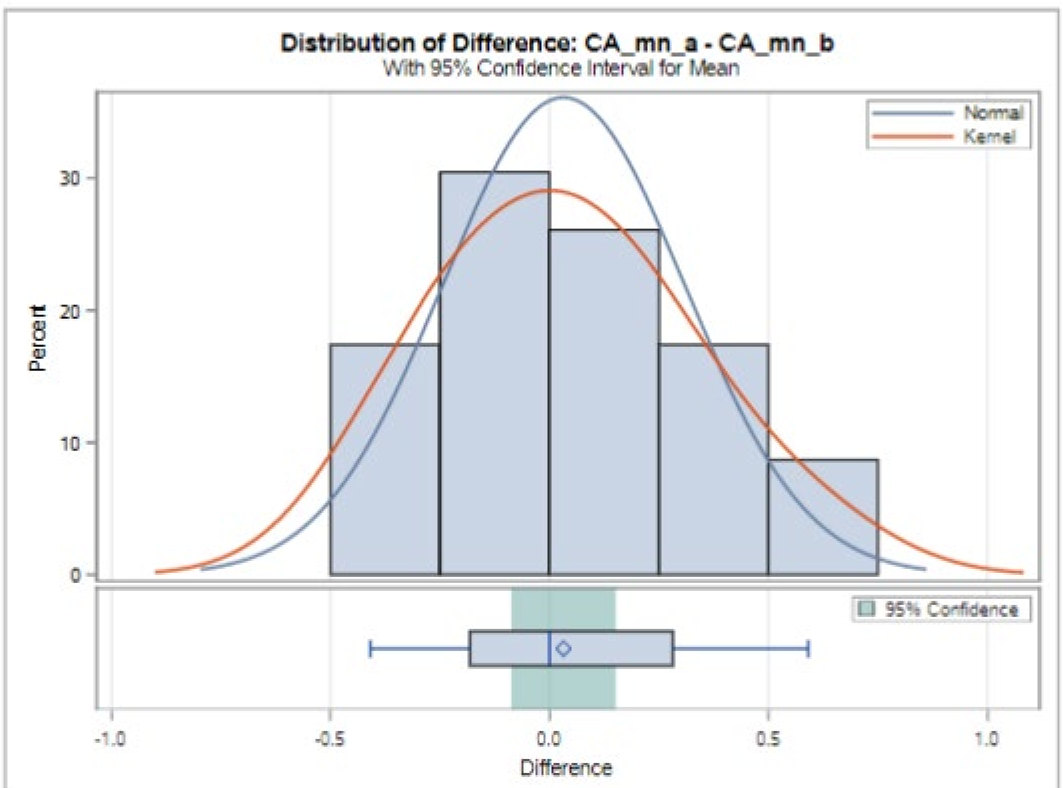


Paired Profiles for (PAK_mn_a, PAK_mn_b)
Mean
2.0
2.5
3.0
3.5
PAK_mn_a
PAK_mn_b


Distribution of PAK_mn_a
1
2
Percent
Gender
PAK_mn_a
Normal
Kernel
Q-Q Plots of PAK_mn_a
PAK_mn_a
Quantile


Distribution of CA_mn_b
Percent
Gender
CA_mn_b
Normal
Kernel
Q-Q Plots of CA_mn_b
CA_mn_b
Quantile


Agreement of PAK_mn_b and PAK_mn_a
Mean
PAK_mn_b
3.5
3.0
2.5
2.0
2.0
2.5
3.0
3.5
PAK_mn_a


Distribution of Difference: CA_mn_a - CA_mn_b
With 95% Confidence Interval for Mean
Normal
Kernel
Percent
30
20
10
0
95% Confidence
-1.0
-0.5
0.0
0.5
1.0
Difference