KEY FACTORS INFLUENCING RETENTION RATES AMONG HISTORICALLY UNDERREPRESENTED STUDENT GROUPS IN STEM FIELDS

Divya Premraj

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

December 2018

APPROVED:

Ruthanne Thompson, Major Professor
Pamela Padilla, Committee Member
Lee Hughes, Committee Member
Amanda Wright, Committee Member
Joshua Adams, Committee Member
Jyoti Shah, Chair of the Department of Biological Science
Su Gao, Dean of College of Science
Victor Prybutok, Dean of the Toulouse Graduate School
The aim of the study was to identify the factors that have an influence on the completion rates of undergraduate students in the fields of science, technology, engineering and mathematics (STEM). Using Tinto’s retention rate theory as the theoretical foundation, data were collected from freshman who were enrolled in the years 2005 to 2008. Results showed gender and first-generation status were significant predictors of STEM completion and time taken to complete the degree. Institutional bias played a role in race/ethnicity not being a factor affecting completion rates, as this study was conducted at a Predominantly White Institution. SAT scores and first and second-year college GPA showed to have the most prominent influence on both STEM completion rate and time taken to complete the degree. Females with higher first-year college GPA and higher high school rank finished faster. Similar results were found with first-generation students as well. Students belonging to ethnic minority groups with higher SAT scores and college GPA had greater success in STEM fields as well. The study results can be used to increase completion rates of underrepresented students in the STEM fields, given what we know about the interactions between underrepresented student groups and the most important predictors.
Copyright 2018

By

Divya Premraj
ACKNOWLEDGEMENTS

This work has been possible because of a number of people guiding and supporting me throughout my time in graduate school. I would like to thank my major professor, Dr. Ruthanne Thompson for her valuable guidance and insight throughout this journey. I could not have asked for a better major professor, as her insights into the world of Science Education and the depth of knowledge in the field are par excellence. I am grateful to my committee members Dr. Pamela Padilla, Dr. Lee Hughes, Dr. Amanda Wright and Dr. Joshua Adams for their continued guidance and support throughout this journey.

I would like to extend my gratitude to the members of the Institutional Research and Effectiveness team, who graciously responded to my requests for data.

I am immensely grateful to my parents for raising me to be independent. I am thankful for my wonderful friends who have supported me through the years with their constant support and many words of encouragement- Shabnam Taneja in particular.

I do not know what I would do without the support of my husband, Sailesh Chandran who has not only supported and encouraged me throughout this journey, but has also filled in for me when my two boys needed me the most. Right from the beginning, he has encouraged me to give my best and has constantly pushed me to achieve more than I ever imagined I was capable of. Last, but certainly not the least- my two little boys, Abhinav and Keshav who have put up with a stressed out mom in graduate school. You are my world and I hope I have shown you that persistence, grit and hard work pays off always, no matter what.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .................................................................................................................. iii  
LIST OF TABLES .......................................................................................................................... vii  
LIST OF FIGURES ........................................................................................................................ viii  

CHAPTER 1 INTRODUCTION ......................................................................................................... 1  
  1.1 Retention Problems in STEM Fields .................................................................................... 1  
  1.2 STEM Education and Retention ......................................................................................... 2  
  1.3 Retention Rate Model: Vincent Tinto ................................................................................ 3  
  1.4 Underrepresented Groups in STEM Education ................................................................... 4  
  1.4.1 Race in STEM Education ............................................................................................... 4  
  1.4.2 Women in STEM Education .......................................................................................... 6  
  1.4.3 First-Generation Students ............................................................................................. 7  
  1.5 Pre-College Schooling: High School Rank based on GPA and SAT Scores ...................... 8  
  1.6 Grade Performance: First and Second Year GPA ............................................................... 8  
  1.7 Purpose of the Study and Research Questions .................................................................. 9  
  1.8 Importance of the Study ...................................................................................................... 11  
  1.9 Limitations of the Study ..................................................................................................... 12  
  1.10 Definition of Terms Used in the Study ............................................................................ 12  
  1.11 Organization of the Study ................................................................................................ 13  

CHAPTER 2 REVIEW OF THE LITERATURE ................................................................................ 14  
  2.1 Search Methods and Information Sources ......................................................................... 14  
  2.2 Retention Rate: Models and Theories ............................................................................... 14  
  2.3 Individual Characteristics ................................................................................................... 19  
  2.4 Race and Student Retention ............................................................................................... 19  
  2.5 Gender and Student Retention .......................................................................................... 27  
  2.6 First-Generation Status and Student Retention ................................................................ 32  
  2.7 Pre-College Experiences .................................................................................................... 37  
  2.8 High School Rank, SAT Scores and Student Retention .................................................... 38
<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Frequency and Percentage Summaries of Study Variables (N = 37,739)</td>
<td>54</td>
</tr>
<tr>
<td>4.2</td>
<td>Descriptive Statistics of SAT Scores and GPA</td>
<td>55</td>
</tr>
<tr>
<td>4.3</td>
<td>Frequency and Percentage Summaries of Study Variables (N = 15,774)</td>
<td>55</td>
</tr>
<tr>
<td>4.4</td>
<td>Descriptive Statistics of SAT Scores and GPA</td>
<td>56</td>
</tr>
<tr>
<td>4.5</td>
<td>Comparative Look at Declared vs. Completed STEM and Non-STEM Majors</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>(N for Declared Majors=37,739; N for Completed Majors=15,774)</td>
<td></td>
</tr>
<tr>
<td>4.6</td>
<td>Frequency and Percentage Summaries of Study Variables: Binary Logistic</td>
<td>71</td>
</tr>
<tr>
<td>4.7</td>
<td>Descriptive Statistics of SAT Scores and GPA for Binary Logistic Regression</td>
<td>72</td>
</tr>
<tr>
<td>4.8</td>
<td>Frequency and Percentage Summaries of Study Variables: Linear</td>
<td>72</td>
</tr>
<tr>
<td>4.9</td>
<td>Descriptive Statistics of SAT Scores and GPA for Linear Regression</td>
<td>73</td>
</tr>
<tr>
<td>4.10</td>
<td>Classification Table of Binary Logistic Regression, Step 46</td>
<td>75</td>
</tr>
<tr>
<td>4.11</td>
<td>Binary Logistic Regression Results of Individualized Effects of Variables</td>
<td>76</td>
</tr>
<tr>
<td>4.12</td>
<td>SAT Averages by Degree Completion for Minority and Non-Minority STEM and Non-STEM Students</td>
<td>79</td>
</tr>
<tr>
<td>4.13</td>
<td>SAT Average by Degree Completion, First and Non-First-generation, and STEM and Non-STEM Students</td>
<td>79</td>
</tr>
<tr>
<td>4.14</td>
<td>First Year GPA Average by Degree Completion for Minority and Non-Minority, and STEM and Non-STEM Students</td>
<td>82</td>
</tr>
<tr>
<td>4.15</td>
<td>Linear Regression Results of Individualized Effects of Variables</td>
<td>84</td>
</tr>
<tr>
<td>4.16</td>
<td>Length of Time to Degree (Years) Completion Descriptives for Each High School Ranking</td>
<td>85</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure Number</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>STEM education—United States and beyond (National Science Foundation, 2016).</td>
<td>1</td>
</tr>
<tr>
<td>2.1</td>
<td>Synthesis of literature on dropout from higher education (Tinto, 1975. p. 95).</td>
<td>18</td>
</tr>
<tr>
<td>4.1</td>
<td>Frequency of students who completed versus did not complete their degree, by gender, for non-STEM students.</td>
<td>59</td>
</tr>
<tr>
<td>4.2</td>
<td>Frequency of students who completed versus did not complete their degree, by gender, for STEM students.</td>
<td>59</td>
</tr>
<tr>
<td>4.3</td>
<td>The effect of STEM and SAT score on degree completion.</td>
<td>64</td>
</tr>
<tr>
<td>4.4</td>
<td>The effect of STEM and first year GPA on degree completion.</td>
<td>67</td>
</tr>
<tr>
<td>4.5</td>
<td>The effect of STEM and first year GPA on length of degree completion.</td>
<td>68</td>
</tr>
<tr>
<td>4.6</td>
<td>The effect of STEM and second year GPA on length of degree completion.</td>
<td>70</td>
</tr>
<tr>
<td>4.7</td>
<td>Average SAT score for STEM and non-STEM students.</td>
<td>78</td>
</tr>
<tr>
<td>4.8</td>
<td>Average SAT score for minority and non-minority non-STEM students.</td>
<td>78</td>
</tr>
<tr>
<td>4.9</td>
<td>Average SAT score for minority and non-minority STEM students.</td>
<td>78</td>
</tr>
<tr>
<td>4.10</td>
<td>Average SAT score for first-generation and non-first-generation non-STEM students.</td>
<td>80</td>
</tr>
<tr>
<td>4.11</td>
<td>Average SAT score for first-generation and non-first-generation STEM students.</td>
<td>80</td>
</tr>
<tr>
<td>4.12</td>
<td>Average first year GPA for minority and non-minority non-STEM students.</td>
<td>81</td>
</tr>
<tr>
<td>4.13</td>
<td>Average first year GPA for minority and non-minority STEM students.</td>
<td>81</td>
</tr>
<tr>
<td>4.14</td>
<td>Number of years taken to obtain degree for male and female STEM and non-STEM students.</td>
<td>86</td>
</tr>
<tr>
<td>4.15</td>
<td>Number of years taken to obtain degree for male and female non-STEM students.</td>
<td>87</td>
</tr>
</tbody>
</table>
Figure 4.16. Number of years taken to obtain degree for male and female STEM students. .......................... 87

Figure 4.17. Number of years taken to obtain degree for male and female non-STEM students. .................................................. 87

Figure 4.18. Number of years taken to obtain degree for male and female STEM students. ........................................................... 88

Figure 4.19. Average SAT score by number of years taken to complete degree (by quartile). .......................................................... 88

Figure 4.20. Number of years taken to obtain degree for non-STEM and STEM students. ............................................................. 89

Figure 4.21. Average SAT score by number of years taken to complete degree (by quartile) and GPA for non-STEM students. .................................................. 89

Figure 4.22. Average SAT score by number of years taken to complete degree (by quartile) and GPA for STEM students. ........................................................ 90
CHAPTER 1

INTRODUCTION

1.1 Retention Problems in STEM Fields

Enrollment numbers in STEM fields has been consistently on the rise in the United States (Chen, 2013). Most of this can be attributed to the effectiveness of messaging techniques about the importance of pursuing STEM degrees by schools and other educational institutions. However, studies show that graduation numbers are not keeping up with enrollment, leading to the issue being that of persistence in STEM fields (Kokkelenberg & Sinha, 2010).

Figure 1.1. STEM education—United States and beyond (National Science Foundation, 2016).

The proportion of STEM bachelor’s degrees relative to degrees in all fields awarded remained flat between 2000 and 2013. In addition, during this period,
graduation rates actually declined in much needed fields such as computer sciences, mathematics, physics, engineering, and economics (National Science Foundation, 2016; Carver et al., 2017). A point in fact, according to Le et al., 2014, from 2003 to 2009, only about half of the students who initially enrolled in STEM majors actually persisted in the STEM fields. The others either dropped out of college (20.2%) or switched to a non-STEM major (28.1%). The National Intelligence Council (2008) has stated that as a nation, we no longer have the ability to meet indigenous needs, and thus we must rely on outsourcing to other countries. In response, the President’s Council of Advisors on Science and Technology in 2012 stated that the nation needs to graduate one million more STEM students in the next decade, an additional 34% annually, to meet the needs of current economic projections (Kokkelenberg & Sinha, 2010; Rice et al., 2015; Redmond-Sanogo et al., 2016).

1.2 STEM Education and Retention

There have been several definitions of what retention means in higher education. The Integrated Post-Secondary Education Data System (IPEDS) defines retention as “A measure of the rate at which students persist in their educational program at an institution, expressed as a percentage” (IPEDS Survey, 2017, p. 23). For 4-year institutions, this is the percentage of first-time bachelors (or equivalent) degree-seeking undergraduates from the previous fall who are again enrolled in the current fall, according to IPEDS. Though the language is easy enough, the reality is that student retention, especially in STEM, has been and continues to be a major challenge for all of Higher Education. Researchers such as Seymour (2000) and Tinto (2007, 2012, 2013)
have long reported declining retention rates in STEM and even more recent studies have also highlighted similar declining rates along with the inability of STEM fields to retain their students (Carver et al., 2017; Dagley, Georgiopoulos, Reece, & Young, 2016; Inouye, Ouyang, Couch, & Yeager 2015). Twenty-one years ago, Seymour & Hewitt (1997) stated that the barriers to student success and persistence in STEM were complex and multifaceted. Here, 21 years later, those words still stand true. Thus, the problem is tangible, yet the solution(s) remain elusive, despite the fact that there have been numerous studies conducted which seek to explain and predict student retention in higher education (Mau, 2003; Dekker, Pechenizkiy & Vleeshouwers, 2009; Nora, 2009, Perez et al., 2014; Tinto, 1975, Murphy et al., 2010, Novak et al., 2016). The most notable being the seminal work conducted by Vincent Tinto in 1975. Vincent Tinto’s longitudinal model of college retention is one of the most well-known frameworks for attempting to explain student persistence. The model, which is based on Durkheim’s (1951) theory of suicides, renders the premise that whether a student persists or drops out is strongly predicted by their academic and social integration (Tinto, 1975) (see Figure 2.1).

1.3 Retention Rate Model: Vincent Tinto

As with studies that have come before, this study also uses Tinto's longitudinal model as a baseline. In this case, it is used to identify particular characteristics known to effect retention in general in order to determine how those characteristics affect retention in STEM at a Southwest University in the United States. Thus, the principal focus of this study was to analyze the undergraduate STEM student population in this
large R1 institution to determine if particular previously identified factors affecting retention for college students as a whole also hold true for retention of STEM students. The factors analyzed in this study include the following:

- Individual Attributes of race, gender and first-generation status.
- Pre-College Schooling, consisting of high school rank, and SAT math and reading scores.
- Grade Performance as a part of the Academic System, to include second year GPA.

Of all of the factors in Tinto’s model, these factors in particular were selected based on Griffith, 2010, and Harsh et al., 2011 who stated that primarily these are the deciding factors that play into determining whether a student will drop out or stay in STEM.

1.4 Underrepresented Groups in STEM Education

Although United States has been seeing a decline in STEM graduation among all groups of students, the groups that have been set back the most are racial minorities, women, and first-generation students (Carver et al., 2017).

1.4.1 Race in STEM Education

Sociologist Louis Wirth defined a minority group as “a group of people who, because of their physical or cultural characteristics, are singled out from the others in the society in which they live for differential and unequal treatment, and who therefore regard themselves as objects of collective discrimination” (Wirth, 1945, p.83). The racial minority groups focused on in this study are African-Americans, Hispanic/Latino and Native Americans.
• African-American: They are an ethnic group of Americans with total or partial ancestry from any of the Black racial groups of Africa (Census, 2011).

• Hispanic/Latino: The U.S. Census Bureau defines the term Hispanic or Latino to refer to "a person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin regardless of race" and states that “Hispanics or Latinos can be of any race, any ancestry, any ethnicity” (Passel, Director, & Lopez, 2009, p.24).

• Native American: Native American is someone who is a member of an Indian tribe and an Indian tribe is any tribe, band, nation, or organized Indian community recognized by the United States (U.S. Census Bureau).

Racial groups, such as the ones listed above, though currently in the minority in the U.S., are growing in size and are expected to reach above 40% of the population in the United States by the year 2050 (Pew Research Center, 2008). Despite this fact, they continue to be staggeringly underrepresented in STEM fields nationally. According to 2010 data from the National Science Foundation and the U.S. Census Bureau, Under Represented Minorities (URMs) earned only 18.6% of the total undergraduate degrees from 4-year colleges in 2010. Only 16% of those degrees were in Science fields and less than 13% of those degrees were in Physical Sciences and Engineering (National Science Foundation, 2010). There has been some progress made in recent years but very little. The National Science Foundation reported that as of 2011-2012, African-Americans earned only 13.8% of all STEM degrees, while Hispanic students earned 16.8% (NSF, Science and Engineering Indicators, 2016; Fouad et al., 2016, Perenzadian et al., 2016).

Since 1991, the greatest rise in the share of science and engineering bachelor’s degrees earned by underrepresented minorities has been in psychology, the social sciences, and computer sciences (National Science Foundation, 2012). This study also
stated that since 2000, underrepresented minorities’ shares in engineering and the physical sciences have been flat, and participation in mathematics has dropped.

1.4.2 Women in STEM Education

The underrepresentation of women in STEM is found at multiple levels in education, including high school, college, graduate school, and in academic positions, according to the National Research Council in 2007 (Nimmesgern, 2016). Growing the representation of women in STEM disciplines would increase gender diversity, which in turn could positively affect productivity and innovation according to Woolley et al. (2010). The lack of representation is the result of a number of factors, including such dynamics as gender stereotypes (Bonham & Stefan, 2017; Prentice & Carranza, 2002), discrimination (Christopher & Wojda, 2008; Thomas et al., 2014), stereotypical and social identity threats (Murphy, Steele, & Gross, 2007; Sekaquaptewa & Thompson, 2003; Steele & Aronson 1995), and even unconscious gender-stereotypical cues in the environment (Steele et al., 2009). According to Cundiff et al. (2013), strong gender stereotypes decrease women’s identification with and career aspirations in science.

The 2016 Science and Engineering indicators report by the National Science Foundation showed that only 11.7% of female students chose a STEM major, as opposed to 26% for men. The data from this report also told us that female participants in this same period earned over half of the degrees in non-STEM fields, yet they earned less than a quarter of the STEM degrees during that same academic year. In addition, a closer look into the specific STEM degrees earned by females showed that 84.3% of the earned STEM degrees were in the biological sciences and similar to ethnic minorities in
psychology and the social sciences. Therefore, in the fields of engineering, mathematics, physical sciences, and computer sciences, the gender gap continues to exist (Baiduc et al., 2015).

1.4.3 First-Generation Students

The definition of a first-generation college student varies from one institution to another. For instance, Princeton University defines a first-generation college student as the first person or generation in one’s nuclear family to pursue a bachelor’s degree. While other universities including Brown, Cornell, Harvard, and Yale consider students first generation, if neither parent graduated with a bachelor’s degree, according to data set reports from the universities’ websites. In general, it is accepted that a student who is the first in his/her family to attend college is called a first-generation college student (McConnell, 2000; Williams et al., 2004).

Citing data from the National Center for Education Statistics (BPS: 96/01)a Pell study on college success for low-income, first generation students in 2008 found that most of the 4.5 million, low-income, first-generation students enrolled in post-secondary education (about 24% of the undergraduate population), experienced less success than their peers right from the start (Engle & Tinto 2008). Some additional key findings highlighted in the report included that both low income and first-generation students were four times more likely to leave higher education after their first year as compared to their peers, and that six years after the study began, 43% of first-generation students had left college without earning a degree (Lam. P.C et al., 2005).
1.5 Pre-College Schooling: High School Rank based on GPA and SAT Scores

The two most common variables used for admission to a college or university are GPA and SAT (Burton & Ramist, 2001; Kokkelenberg & Sinha, 2010; Noble & Sawyer, 2004; Rohr, 2012). Due to this fact, there have been a wide variety of studies, which have focused on the predictive nature of the SAT on academic success (Burton & Ramist, 2001; Hull et al., 1995; Reason, 2003, Sedlacek & Adams, 1992; Stearns et al., 2017). Although SAT scores have been found to be a general indicator of academic success, there is little research that has been conducted that specifically looks at SAT scores as an indicator for success or retention in the fields of science, technology, engineering, mathematics, and business (Rohr, 2012). Though there are studies, one by Huang et al. (2000) and then later by Thompson and Bolin (2011), that have found a statistically significant correlation between SAT and high school rank, as it relates to retention in STEM.

1.6 Grade Performance: First and Second Year GPA

Tinto (1975) reported, “with respect to grade performance, many studies have shown it to be the single most important factor in predicting persistence in college” (P. 104). Several other scholars report GPA to be a major and direct predictor of attrition or retention (Cabrera, Nora & Castaneda, 1993; Hachey et al., 2015; Pascarella & Chapman, 1998; Scherer, Talley, & Fife, 2017). However, there have not been many studies that have focused on first and second year GPA, as an important predictor in student retention.
1.7 Purpose of the Study and Research Questions

Post-secondary educational institutions in the United States have responded to the steady decline of student graduation in STEM fields in recent decades, which have proven to be a national crisis (Fairweather, 2008). As a result, enrollment numbers have gone up in STEM fields. About 28% of bachelor's degree and 20% of associate's degree students entered a STEM field at some point within 6 years of entering post-secondary education in 2003-04, according to the U.S. Department of Education in 2013. The problem is, graduation rates are not keeping up with enrollment rates. This is because many of these STEM entrants left STEM several years later by either changing majors or leaving college without completing a degree or certificate. A total of 48% of bachelor's degree students and 69% of associate's degree students who entered STEM fields between 2003 and 2009 had left these fields by 2009, according to the U.S. Department of Education in 2013. This is even more profound in underrepresented groups, such as minorities, women, and first-generation students. The purpose of this study, then, is to first understand the undergraduate student population graduation rates in STEM fields at a Southwest university in the United States. Second, this study also, taking into consideration Vincent Tinto's longitudinal model, involved determining if pre-college characteristics (high school rank and SAT math and reading scores) and Institutional characteristics (first and second year GPA) have an effect on STEM retention rates to determine if any one factor plays a more significant role than another. Third, based on research findings, recommend a retention intervention in the area that is most likely to have the largest impact on retention in STEM. The following research questions guided this study:
RQ1: Are there statistically significant relationships between a STEM student’s completion of a 4-year bachelor’s degree at a postsecondary institution and the student’s gender, race/ethnicity, and first-generation status?

RQ2: Are there statistically significant relationships between a STEM student’s completion of a 4-year bachelor’s degree at a postsecondary institution and the student’s SAT score and high school academic achievement as measured by quartile ranking?

RQ3: Are there statistically significant relationships between a STEM student’s completion of a 4-year bachelor’s degree at a postsecondary institution and the student’s first and second year college GPA?

Additionally, time to completion of the undergraduate degree is a factor that would affect underrepresented students (Chen, 2009; Crisp, Nora, & Taggart, 2009). This factor was addressed by the following questions:

RQ4: For students that completed their degree, are there statistically significant relationships between STEM students’ length of time taken to complete a bachelor’s degree at a postsecondary institution and the student’s gender, race/ethnicity, and first-generation status?

RQ5: For students that completed their degree, are there statistically significant relationships between a STEM length of time to complete a bachelor’s degree at a postsecondary institution and the student’s SAT score and high school academic achievement as measured by quartile ranking?

RQ6: For students that completed their degree, are there statistically significant relationships between a STEM student’s length of time to complete a bachelor’s degree at a postsecondary institution and the student’s first and second year college GPA?

The focus of the research questions is similar to the focus of research conducted through various studies on SAT scores and high school academic achievement, specifically as measured by high school quartile rank, gender, race/ethnicity, and first-generation status. The primary goal for these studies has been to explain a statistically significant amount of the variance in student persistence in STEM fields, as well as the time taken to complete their degrees.
1.8 Importance of the Study

We have made progress in closing the gender gap and the gap between well-represented and underrepresented groups (Goldin et al., 2006; Miyake, 2010). This applies to every field, but the situation in STEM fields seems to be in a crisis (DiPrete et al., 2006). Statistics show, though this trend has been decreasing over the past 20 years, racial minorities continue to have disproportionally higher attrition rates compared with other groups (Braun et al., 2017; Galarza et al., 2007; Maton, Hrabowski, & Ozdemir, 2007; Mendez, Buskirk, Lohr, & Haag 2008; Webb, 2017). As noted previously, these groups tend to enroll in STEM majors and leave in high numbers (Phansalkar, 2007). The importance of this study is that it sheds additional light on the factors that are known to primarily affect undergraduate population of the traditionally underrepresented student groups, particularly at this large Research One University. This population is especially vulnerable and needs to be more represented in STEM fields (Chang, Sharkness, Hurtado, & Newman, 2014; Crisp et al., 2009; Espinosa, 2011; Hurtado et al., 2010). Studies pertaining to underrepresented groups in general are numerous, however this study isolates and compares the variables of SAT scores, high school rank, and college GPA and their relationship in STEM.

This study was an extension of research conducted by Thompson and Bolin (2011), at this same university, which focused on retention rates in STEM fields among the general student population. This study took it further by focusing on specific student groups and looked at two-way interactions between underrepresented minority groups and pre-college characteristics like SAT scores, high school rank, and grade performance, such as first and second year GPA scores.
1.9 Limitations of the Study

The primary limitation of this study is that the findings may not be generalized to institutions of higher education as a whole, as data were collected from a single research university. Trends of student enrollment in this specific Southwest University located in the Dallas-Fort worth Metroplex may not be reflective of trends elsewhere. Data collected for this study were comprised of the review of existing records, including the student enrollment numbers for each department and each category of student from the office of Institutional Research and Effectiveness at UNT. As this is a predominantly White institution there is potential for Institutional bias). Institutions like HBCUs (Historically Black Colleges and Universities) and HSIs (Hispanic Serving Intuitions) along with small 4-year colleges and private intuitions may have different trends. In addition, Data for first-generation students are primarily self-reported, and may not be completely accurate.

1.10 Definition of Terms Used in the Study

- FAFSA: Application used by nearly all colleges and universities to determine eligibility for federal, state, and college-sponsored financial aid, including grants, educational loans, and work-study programs.
- First-year student: A student who has completed less than the equivalent of one full year of undergraduate work (i.e., less than 30 semester hours in a degree program; National Center for Educational Statistics, 2010).
• Persistence: Involves a firm or obstinate continuance in a course of action in spite of difficulty or opposition.

• Retention: An institutional measure of the rate at which students persist in their educational program; measured as a percentage rate of first time degree seeking undergraduates from the previous fall who are enrolled again in the current fall (NCES, 2010).

• STEM: Acronym for Science, Technology, Engineering and Mathematics. For the purposes of this study, educational and workforce disciplines, which come under physics, chemistry, biology, engineering, computer science and math are considered STEM.

• Underrepresented groups: This includes all the racial minorities mentioned above, as well as women.

1.11 Organization of the Study

This chapter involved the presentation of background information for the study, and highlighted its rationale. It also included a discussion on the importance of the study and its research questions. Chapter 2 presents key literature from topics associated with retention, among others, to provide both background and foundation for this study. Chapter 3 outlines the methodology employed for collection and analysis of data. Chapter 4 involves a discussion of the results of the data analysis, and finally Chapter 5 presents conclusions that provide answers to the research questions presented earlier and suggestions that the researcher has for the issue of interest.
CHAPTER 2
REVIEW OF THE LITERATURE

The purpose of this literature review was to familiarize the reader with foundational retention theories and models, which will then be followed by research that focuses on growth and retention in STEM fields, how first-generation students, students from underrepresented minorities, and women perform in STEM fields, and how they compare with the rest of the student population.

2.1 Search Methods and Information Sources

The researcher first narrowed down the voluminous amount of research related to retention. To facilitate that, the following key words were used: retention theories, retention models, student attrition, stem retention, student persistence, retention rates among first-generation students, retention rates among underrepresented minorities, and retention rates among women. In addition, the following sources were used throughout the course of the study: Journal of College Student Retention: Research Theory and Practice, Journal of STEM Education: Innovations and Research, International Journal of STEM Education, International Journal of Science Education, and the Chronicle of Higher Education. The UNT library resources were utilized, including the UNT Digital Library: Theses and Dissertations.

2.2 Retention Rate: Models and Theories

Most of the previously published literature on retention rates in the university setting, whether it is public, private, or community colleges focused on two aspects of
student retention: (1) The big picture view that focuses on the causes of attrition that may in turn help predict retention, and (2) The microscopic view that focuses on specific interventions that could positively affect retention. This study however sought to use 1 to inform 2. In other words, the study used the big picture view (1) to both identify and rank known STEM attrition factors in order to (2) aid the university in prioritization of specific interventions.

When retention of college students appeared on the higher education radar 40 years ago, it was often viewed through the lens of Psychology (Tinto 2007, 2013). A proper definition or understanding of retention and the numerous factors affecting it were essential in order to diagnose potential problems relating to why students drop out in the first place. There are several retention rate theories that have been published and are constantly used in similar studies. Vincent Tinto’s theory, however is the gold standard of retention theories (Stage, 1990) and are based on Durkheim’s theory of suicide (Durkheim, 1951). A lot of literature exists in more recent years which follow his study (Gonzalez & Morrison, 2015; Hlinka, 2017).

Durkheim (1951), in his model addressed the issue of suicide, which, according to his research is more likely to occur when individuals are not sufficiently integrated into society. He goes on to say that the likelihood of suicide in society increases when two types of integration are lacking, which are: (a) Insufficient moral integration and (b) Insufficient collective affiliation. He attributes the lack of integration to prevailing social conditions. In addition, he states that insufficient moral integration is the outcome of one’s holding values highly divergent from those of the social collectivity, and insufficient
collective affiliation is the result of insufficient personal interaction with other members of the society, according to Tinto (1975).

Tinto relates retention to the college environment, in that it is made up of social and academic systems, and therefore, it is essential to distinguish between the normative and structural integration in the academic domain of the college from that in the social domain of the college (Spady, 1970). When college dropout is applied to this model, a direct relationship is found between the person's participation in the academic domain of the college and his future occupational attainment. A student's withdrawal from college can arise either from voluntary withdrawal (like suicide) or from forced withdrawal (dismissal), which arises primarily, from insufficient levels of academic performance (poor grades) and/or from the breaking of established rules concerning proper social and academic behavior (e.g., student strikes, stealing exams, etc.; Tinto, 1975). It is also possible that a student can be integrated in one area and not achieve similar levels of integration in the other. When this happens, dropout can still occur. Not enough integration in any one domain is sufficient for dropout to occur (Tinto, 2013).

Tinto has stated that his theoretical model of dropout can be described as “longitudinal,” (Tinto, 1975, p. 90) because it is a process of interactions between the student and the academic and the social systems of the institution that continue to modify his/her goal and institutional commitments in ways that lead to different forms of dropout (Tinto, 1975). There is certainly an argument to be made for the fact that students enter institutions with a variety of personal attributes (gender, race, and ability), pre-college experiences (GPA, academic and social attainments) and family background (first-generation status, income levels), and each of these factors play a
direct and indirect role on student performance. Background characteristics and individual attributes influence one’s educational expectations and thereby their commitment to staying in college (Tinto, 1975). A student enters the educational institution with a combination of individual characteristics, prior experiences, and commitments, but it is finally the student’s integration into both the academic and social systems of that institution that determines if he/she drops out. The higher the degree of integration of the individual into the college systems, the greater will be his commitment to the specific institution and to the goal of college completion (Eisenberg, Lipson, & Posselt, 2016; Tinto, 1975).

Tinto’s retention theory has reached near paradigmatic status, and it has been used in a number of studies, essentially serving as a baseline for future models (Davidson, 2013).

Nora, Barlow, and Crisp (2005) model is a reformulation of Vincent Tinto’s model of student persistence and degree attainment. It brings more clarity to the academic dimensions of the college environment while maintaining social and academic integration as a central tenet (Chang et al., 2014). This model is primarily based on research on underrepresented groups of students, which are also central to this study. Nora et al. (2005) integration model includes many of the factors that are likely to influence minority, low-income, and non-traditional student populations in important ways, like aspects of pre-college environments (school and home environment), financial assistance, family support, environmental pull factors (family and work responsibilities) and commuting to college (Chang et al., 2014).
Figure 2.1. Synthesis of literature on dropout from higher education (Tinto, 1975. p. 95).
2.3 Individual Characteristics

It has been established that a student's individual characteristics play an important role in determining if he/she decides to stay in college or dropout (Tinto, 1975, 2013). Some of the individual characteristics that this study took into consideration are race, gender and first-generation status.

2.4 Race and Student Retention

As per the Organization for Economic Co-operation and Development (OECD), many of its member countries have seen a significant increase in the number of STEM skilled workers, and in some cases even surpassing the United States (Palmer, Maramba & Dancy, 2011). The Pew Report states that, as of 2015, the United States placed 38th out of 71 countries in Math, and 24th in Science, according to the most recent report released in 2017. The United States is now faced with a challenge of increasing the pool of students who enter STEM fields in colleges and retain them through degree attainment (Fairweather, 2008).

Traditionally underrepresented groups are African American, Hispanic, and Native American students. Studies have shown that despite some progress, minority students continue to experience issues accessing higher education and persisting to graduation, especially in STEM fields (Chen, 2013; Moore, 2006; Museus, 2008; Öztürk, 2007). The U.S. Census Bureau has projected that the minority (African American, Hispanic, Asian American, and Native American) populations are expected to grow rapidly over the next few decades. They will comprise approximately 50% of the total U.S. population by the year 2050 (Palmer et al., 2011; U.S. Census Bureau, 2011).
Given this projection, it is extremely important that the United States enhance its efforts to increase college enrollment, retention, and persistence of minority students in STEM education (Hurtado et al., 2008, Nguyen et al., 2016).

Research has repeatedly shown that students in supportive educational environments during college are positively linked to retention and persistence, and this especially holds true for minority students (Bonous-Hammarth, 2000; Cole & Espinoza 2008; Fries-Britt, Younger, & Hall, 2010; Grandy, 1998; Hurtado et al., 2007). A number of factors have been attributed to minorities performing well and persisting to complete their STEM degree. Some of them are role models of color or similar ethnicity, knowledge, and lesson sharing from advanced students of similar ethnic groups, and relationships with staff of color or similar ethnicity (Bonous-Hammarth, 2000; Fries-Britt et al., 2010; Grandy, 1998; Hurtado et al., 2007; Seymour & Hewitt, 1997; Illumoka et al., 2017). There have been a number of studies using quantitative methods to examine success factors for minority students in STEM (Ampaw & Partlo, 2013; Covington, Chavis, & Perry, 2017; Mckim & Sorenson, 2017; Palmer et al., 2011). There have been fewer qualitative studies on the subject, and one such study examines factors of retention and persistence among college students of color enrolled in STEM education at a predominantly White institution (PWI) (Palmer et al., 2011).

Palmer et al. (2011) conducted a study, which encompasses both STEM and non-STEM fields, at a research-intensive PWI in the Northeastern part of the United States. Forty-five percent of the students enrolled at this institution were White, while their Asian, Black, Hispanic, and their racially/ethnically unknown counterparts comprised 13%, 6%, 7%, and 22%, respectively. Approximately 46% of White students
matriculating at this institution completed a baccalaureate degree within 6 years, while that figure was 13% for Asian, 9% for Black and Hispanic students, and 25% for racially/ethnicity unknown students (Palmer et al., 2011). An interview was conducted among the junior and senior minority students, because they would be more in-depth about factors attributing to their STEM success (Palmer et al., 2011). In this qualitative study, three major success themes emerged: (a) peer group support, (b) involvement in STEM-related activities, and (c) strong high school preparation. Peer group support was helpful in two ways: they served as support for their academic work, and they provided a positive social network. The students viewed this as being extremely important in their decision to stay in STEM. Group studying not only helped them with their courses, but also boosted their morale in more ways than one (Palmer et al., 2011). The second theme that students found to be extremely important is the chance to get involved in STEM-related activities. Activities included involvement in STEM student organizations, being a teaching assistant, participating in STEM summer programs, and interacting with alumni and those currently in STEM professional fields. The final theme that students attributed to their staying in STEM was their strong high school preparation. Participation in advanced science classes and the teachers who invested time to encourage them to pursue science-related fields helped them stick around in STEM (Arcidiacono, Aucejo, & Hotz, 2016; Palmer et al., 2011).

Over the years, concerns over the STEM pipeline have led to additional funding to recruit and retain students in STEM, but somehow these initiatives have not had a significant effect on minority students (Chang et al., 2014; Kendricks & Arment, 2011). National Science Foundation, in 2010 says that STEM graduation rates among minority
students have increased nationally but that increase is not significant (National Science Foundation, 2010). Retaining students who are under-stimulated but well prepared is one big issue (Chang et al., 2014).

Older studies conducted in the year 2000 have found that Underrepresented Minorities (URMs) had lower persistence rates in Science and Engineering (26%) than their White and Asian American counterparts at 46% (Huang, Taddese, & Walter, 2000). Another study conducted by the Higher Education Research Institute in 2010 found that 33% of White students and 42% of Asian American students at a national sample of institutions, completed their bachelor’s degrees in STEM within 5 years of entering college, compared to only 18% of African American students and 22% of Latino students (Chang et al., 2014). Mutegi (2013) pointed out that science education research has been unable to explain how science degree achievement is “racially determined” (Mutegi, 2013, P. 93). Although students’ abilities and pre-college preparation are important, steady progress through the STEM pipeline also depends on the types of opportunities, experiences, and support students receive while in college (Chang et al., 2011; Espinosa, 2011; Wladis, Conway, & Hachey, 2015).

For minorities in STEM fields, combinations of environmental and psychological factors have been shown to affect persistence. It was found that African American students had far less preparation in pre-college sciences, as demonstrated by lower rates of participation in AP Biology, chemistry, physics, and calculus courses (Elliott, Strenta, Adair, Matier, & Scott, 1996). This trend continues today, as more studies have found less participation amongst this student population (Lancaster & Xu, 2017; Means et al., 2017). Other factors include personal contact with faculty, degree of racial
isolation, exposure to racial minority support systems, and an institution’s level of selectivity (Chang et al., 2008; Daempfle, 2003; Espinosa, 2011; Grandy, 1998; Seymour et al., 1997). On taking all of these factors into consideration, these studies seem to suggest that completion of a STEM degree requires not only adequate academic preparation, but also resilience and the ability to negotiate a complex academic context (Chang, 2014; Means et al., 2017).

In 1997, Seymour et al. found that the educational experience and culture of the discipline made a much greater contribution to attrition than the individual inadequacies of students or the appeal of other majors. Regardless, this is not to say that individual attributes do not make much of a difference. On the contrary, there have been many studies that show that they do. Russell and Atwater (2005) noted that a demonstrated competence in Science and Math at the pre-college level is vital to the success of African American students through the science pipeline from high school to college. There are other critical factors that make a difference including receiving family support and teacher encouragement, developing intrinsic motivation, and maintaining perseverance (Ma & Liu, 2015; Russell & Atwater, 2005). Similarly, the presence of family support and guidance from faculty mentors is associated with the development of greater academic self-efficacy and success in the sciences for Latino students as well (Anaya & Cole, 2001; Cole & Espinosa, 2008; Dika & Alvarez, 2016; Torres & Solberg, 2001).

Once URM students are in the university setting, Hurtado et al. (2007) explored the effect of college on two key outcomes in the first year of college: (a) sense of belonging and (b) academic adjustment. Hurtado’s study found that the relevance of
coursework to students’ lives had a positive impact on academic and social adjustment on all students, including the URM students (Hurtado, 2007). This confirmed previous studies that found URM students often leave STEM due to a perceived lack of relevance with their social values and desire to improve conditions for their communities (Bonous-Hammarth, 2000). Offering research opportunities to URM undergraduates has been highly effective in retaining them in the field (Lopatto, 2013). It has the benefit of facilitating students’ learning in the classroom, and also introducing them to the benefits of what science research careers could entail (Kinkead, 2003; Lopatto, 2003). Other advantages of encouraging URM students to participate in research include enhancing their knowledge and comprehension of science (Sabatini, 1997), clarifying graduate school and/or career plans in the sciences (Hurtado et al., 2009; Kardash, 2000; Sabatini, 1997), and obtaining other professional opportunities that develop their scientific self-efficacy (Gándara & Maxwell-Jolly, 1999; Hurtado et al., 2009; Mabrouk & Peters, 2000). The influence of mentors or role models on URM students cannot be discounted. They are generally more confident about their academic abilities; if they feel people like faculty role models (Carlone & Johnson, 2007) recognize them. Additionally, financial constraints are a real issue to underrepresented minorities (McGee, 2016).

Oseguera, Hurtado, Denson, Cerna, and Saenz (2006) found that entering URM undergraduates reported working more hours during high school and were more likely to expect to work full time during college than their White and Asian counterparts. This study also said that having financial concerns and misperceptions about the financial viability of research professions could deter students from choosing a career in STEM fields. Conversely, a similar study found that participating in research-oriented programs
prior to entering college substantially increased entering URM students’ interest in pursuing a scientific research career (Adedokun et al., 2013). Another factor that was found to contribute to persisting in a STEM major is the students’ aspiration to attain a graduate degree (Chang et al., 2008; Dika & Alvarez, 2016).

In fact, studies by Chang et al (2008) have shown that this aspiration to attain a graduate degree has increased the likelihood of persistence in STEM by over 30%. The same study also found that joining a pre-professional or departmental club during a student’s freshman year increased the likelihood of persisting by more than 150%. The selectivity of the institution, where the selectivity is based on the average combined math and verbal SAT score, was also found to have an influence (Chang et al., 2008). If a URM student attended an institution where the average SAT score was 1100 vs. one where the average score was 1000, the student was 30% more likely to drop as a science major (Chang et al., 2008). Another study by Espinosa (2011) also found similarities with the previously mentioned study. It was found that for every 100-point increase in institutional selectivity (average SAT scores), women of color in STEM were almost half as likely as White women to persist in a STEM major—7.6% compared with 14% (Espinosa, 2011). Therefore, higher institutional selectivity, which is measured by student achievement scores negatively, affects all science students, but the effect is stronger for URM students (Espinosa, 2011). It was also observed the opposite effect was true in the case of Historically Black Colleges and Universities (HBCU). In HBCU’s, as the combined SAT score increased, the chances of persisting in STEM for students attending HBCUs also tended to improve (Chang et al., 2008; Gasman, Nguyen, Conrad, Lundberg, & Commodore, 2017).
Chang et al. (2014) adopted key constructs based on a variety of integration models to overturn the link between persistence in STEM to the 4th year of college and student experiences at multiple types of 4-year colleges (Chang et al., 2014). This study tested the hypothesis that STEM persistence is not just the pre-college characteristics that students bring on to the table while entering college, but also the experiences they face while being in the college environment. The study found significant effects for both African Americans and Latinos; students who identified as being from either of these two groups were less likely to persist in STEM majors than their White and Asian American counterparts. Furthermore, in the Chang et al. (2014) study, in order to determine whether the racial differences in STEM persistence were due to different levels of high school achievement and experiences, another set of variables was added to the model: high school academic preparation (grades, SAT scores, and high school course taking patterns). They found that when these variables were added, the race variables were no longer statistically significant, which may suggest that high school preparation moderates the effects of race. Similar effects were observed for college experiences as well. The pre-college variables that increased the likelihood of persisting in a STEM major are the intentions of pursuing an advanced degree and academic and social self-concepts. One of the strongest predictors of persistence is participation in an undergraduate research program. URM students that participated in programs that exposed them to research were 17.4 percentage points more likely to persist in STEM than those who do not (Chang et al., 2014). URM students’ joining a club or organization relating to their major is a positive predictor as well (Chang et al., 2014). Studying more frequently played a positive role (Chang et al., 2014). There were college
experiences that exhibited negative relationships with persistence. Students who worked full time while attending school were 9.7 percentage points less likely to follow through with intentions to major in STEM, compared to those who never worked full time (Buchwitz et al., 2012). At the institutional level, the proportion of student body majoring in STEM fields significantly and positively contributed to the average likelihood of STEM persistence at that institution (Chang et al., 2014).

2.5 Gender and Student Retention

“The under-representation of women in science, engineering and technology threatens, above all, our global competitiveness. It is an issue for society, for organizations, for employers and for the individual” (Greenfield, Peters, Lane, Rees, Samuels, 2002, P. 22). The above quote is testimony to the fact that improving the recruitment, retention, and training of the next generation of women STEM professionals remains an area of concern (Rincon et al., 2016). Its roots are considered to be in the poor quality of primary education in the sciences, high levels of dropouts from STEM courses at the university level, poor pay and career prospects for STEM workers in comparison to other non-STEM fields, and a failure to respond to the growing demands of an increasingly globalized STEM market (Rincon et al., 2016; Smith, 2011). The National Academies Press, in 2010 reported, “While the representation of women among those receiving bachelor’s degrees in all fields from United States universities exceeds 57%, less than 20% of the degrees in engineering are awarded to women, with the most recent trend slightly worsening” (Ramsey et al., 2013; p. 23) found that academic environments could feel very unwelcoming for women. Although there are
examples of blatant discrimination and stereotyping of women in academic settings, it is more often characterized by subtle behavioral biases (Cohn, 2000; Lyness & Heilman, 2006; McGee & Bentley, 2017). As per the National Science Board in 2007, women and minorities are less likely to persist in a STEM major in college than are male and non-minority students.

Over the last few decades, the representation of women in STEM fields has increased, but there are still large gaps that remain that need to be filled (Huang et al., 2000). However, both women and men still tend to choose traditional gender-based fields (Schuster & Martiny, 2017; Staniec, 2004).

Goldman (2012) placed explanations for the gender gap between men and women in STEM fields in three categories from previous research: (a) biological explanations, (b) social and structural explanations, and (c) psychological explanations. Biological explanations simply arise from the explanation that women were not as represented in these fields as their male counterparts, noting biological differences between the sexes explaining the rarity of extremely intelligent women.

As for Social and structural explanations, some researchers have suggested that there is a lack of academic preparation by female students for STEM classes, as women pursue classes in math, science, and engineering less frequently and report higher levels of attrition in these courses than do their male counterparts (Arnold, 1995; Betz, 2004; Ong, Smith, & Ko, 2018). Other research indicates that the STEM gap is linked to the presence, or lack of presence, of role models. Women make up 19.9% of the engineering students who received their degree in the United States in the year 2014 (Yoder, 2016). Other research reveals that higher education may be a “chilly
climate “for women and one in which aspects of the college environment discourages women from reaching success (Stone & McKee, 2000, p. 68). Finally, the presence of sexism or gender discrimination may be perpetuating the gap (Watkins et al., 2006; Smith, 2011). There have been extensive studies on whether the presence of female faculty members in STEM fields encourage female students to persist as compared to when there are not as many female faculties (Felton et al., 2016).

The fraction of female and/or minority faculty in STEM departments is very low, so if role models of the same gender or race are important factors in major choice, this could play an important role in the under-representation of women and minorities in STEM field majors (Page & West, 2010). Results that speak to this theory have been mixed. Some studies suggest that female faculty members do encourage female students to select a major (Ashworth & Evans, 2001; Carrell, Page, & West, 2010; Rask & Bailey, 2002). A paper by Canes and Rosen (1995) found no link between the percentage of female faculty in a department and the percentage of majors that are female. Another study by Bettinger and Long (2005) using data from Ohio public universities found that female faculty members increase the probability of female students taking additional courses in math and geology. However, they found the opposite effect for biology and physics courses. Another study also found that as the percentage of female faculty in STEM departments’ increases, the percentage of 4-year degrees awarded to females in these departments also increases. Sensitivity to grades has also been found to be an important predictor for women in STEM (Qian, Zafar, & Xie, 2010).

There are studies that reiterate the fact that lower grades lead to lower
persistence, and that this effect is stronger for women (Rask & Tiefenthaler, 2008). There has been much literature on how Institutional type affects persistence of women in STEM (Dika & Alvarez, 2016). There have been numerous studies about Historically Black Colleges and Universities (HBCUs) and Hispanic Serving Institutions (HSIs) that have shown higher rates of retention among women and women of color than that of predominantly White institutions (PWIs). Like other racial/ethnic groups, the representation of Black women in most STEM fields is substantially lower than the representation of Black men (Dika & Alvarez, 2016). The largest gender gaps are in engineering, where Black women received only 36% of all bachelor’s degrees awarded to Blacks in engineering in 2001 (Perna et al., 2009). Several statistics have illustrated gaps in the preparation African Americans, in general, but specifically African American women for educational attainment in STEM fields (Sax et al., 2016). While racial and socioeconomic biases may play a role in standardized testing, observed test scores are consistently lower (Freedle, 2003; Sax et al., 2016).

Research has also demonstrated that the relationship between self-efficacy and educational attainment in STEM fields, especially for women and students of color (Colbeck et al., 2001). Zeldin & Pajares (2000) showed the role of other individuals’ encouragement in promoting a woman’s perceptions of her capacity for a math-related career. Using data from interviews with 15 women in math related careers in which women were underrepresented, Zeldin and Pajares found that participants’ self-efficacy for math-related careers was shaped by positive perceptions of family members, teachers, and peers’ math-related skills and careers; support and encouragement for academic success from family members, teachers, and peers; and successful past
experiences with math-related activities (Hill-Cunningham et al., 2018). Self-efficacy beliefs enabled women to continue in these careers despite academic, financial, and other barriers, as well as negative verbal persuasions from peers about the decision to pursue a career in math (Zeldin & Pajares 2000).

Stereotype threat negatively influence performance by shifting an individual’s focus from performing a particular task to worrying that low performance will confirm a negative stereotype about a group to which the individual belongs (Steele & Aronson, 1995). With only a few exceptions (e.g., Cullen, Hardison, & Sackett, 2004), research has shown that stereotype threat contributes to gaps in academic performance between women and men (Brown, Charmsangavej, Keough, Newman, & Rentfrow, 2000; Brown & Josephs 1999; Gonzales, Blanton, & Williams, 2002; Luong & Knobloch-Westerwick, 2017; O’Brien & Crandall 2003; O’Brien et al., 2015).

Regarding the selection of college majors, experiences are different for men and women. Men and women still do not earn equal numbers of STEM undergraduate degrees. According to the U.S Department of Commerce report published in 2017, in 2016, women earned only 30% of all undergraduate degrees awarded in STEM. However, in some STEM areas, women were overrepresented: in 2009, 59.7% of the degrees in biology and 72% of all premedical degrees were awarded to women. In other areas, they were underrepresented: in 2009, women earned only 16.5% of the undergraduate degrees in engineering and 19.3% of the degrees in physics (Heilbronner, 2013). Overall, the data concerning women's enrollment and graduation in STEM majors are conflicting. Chubin, May, & Babco (2005) noted that since 1966, the number of bachelor's degrees awarded to students in STEM majors received annually
by men has remained relatively stable, at around 200,000, but by 2001 the number of bachelor's degrees received by women had increased and was comparable with men in number of degrees awarded. This trend does not include information technology (IT) majors. Chabrow (2007) noted a drop in IT majors since 2000. Similarly Singh, Allen, Scheckler, and Darlington (2007) reported that the enrollment of women in computer-related majors declined from 1994 through 2005. The U.S. Department of Commerce reported that women only received 8% of Computer Science degrees in the year 2015, as opposed to 17% of men with a degree in Computer Science (Economics & Statistics Administration, 2017).

2.6 First-Generation Status and Student Retention

Another underrepresented group in higher education, and specifically in STEM fields, is first-generation college students (FGCS), (Engle & Tinto, 2008). FGCS are also more likely to be members of an underrepresented racial or ethnic group and are more likely to be transfer students (Dika & D'Amico, 2015). First-generation college students comprise at least 25% of the undergraduate population in the United States (Engle & Tinto, 2008). They also comprise much larger proportions in many regional public 4-year institutions. There are quite a few challenges first-generation students experience when seeking to complete a degree. They present lower standardized admission test scores, and are more likely to have jobs in college, have less family support, have lower GPAs in college, and return at lower rates (Carver et al., 2017; Dika & Alvarez, 2016; Hicks et al., 2016; Martinez, Sher, Krull, & Wood, 2009). They are also more likely to begin their studies at a community college with plans of transferring to a 4-year
institution (Chen & Carroll, 2005). All the literature on first-generation college students can be categorized into three major types. The first group is the set of studies that is comparative in nature. It compares the performance of first-generation students with other students in terms of their demographic and pre-college characteristics, their college choice process, and their college experience (Atherton, 2014; Carpenter, Dick, & Clayton, 2014; Jenkins, Belanger, Connally, Boals, & Durón, 2013; Morales, 2012; White, 2012). Studies from this category of research found that first-generation college students, when compared to their colleagues, face additional barriers due to lack of information about the college process (cost and academic requirements), lack of pre-college preparation, and lack of familial support (Uche, 2015).

The second category of research focuses on the experiences of first-generation college students as they move on from secondary to post-secondary school (Harackiewicz et al., 2014; Padgett, Johnson, & Pascarella, 2012; Stuber, 2011; Williams et al., 2012; Woosley & Shepler, 2011). This set of research tells us that first-generation college students are relatively more challenged in their transition from high school to college, as compared to their peers. The literature in this area also indicates that this difficulty can be attributed to dislocation anxiety, social and academic challenges, and normal rigors of the college environment (Katrevich & Arugute, 2017; Pascarella, Pierson, Wolniak, & Terenzini, 2004; Tinto, 2007).

The third part of research focuses on college persistence, degree attainment, and workforce outcomes of those who have first-generation college status (Engle & Tinto, 2008; Gloria, Kanagui-Muñoz, & Rico, 2012; Jehangir, 2010; Soria & Stebleton, 2012; Jehangir, Williams, & Jeske, 2012; Wells & Lynch, 2012; Ward, Siegel, &
These studies tell us that, when compared to the students whose parents have a college education, first-generation college students are more likely to drop out of college after the first year. They are also less likely to graduate from a 4-year institution within 6 years and less likely to pursue a postgraduate or professional degree 5 years after graduation. There is no difference found in career earnings between first-generation students and their peers after accounting for degree attainment (Pascarella et al., 2004). It can however be noted that in the long term, gaps could emerge in occupational status and income due to lower postgraduate school attendance among first-generation college students (Uche, 2015). Whalen and Shelley (2010) found that the strongest predictor of retention/graduation in STEM is college GPA for First-generation students. Others include financial need and availability of financial resources, males and non-minority status, longer time on campus, in-state residency, and higher standardized test scores. The above-mentioned are all academically related factors. Studies have shown that there are a number of social factors at play when it comes to retention in STEM.

Gayles & Ampaw (2011) found that parents’ higher education level and income are positive predictors of STEM degree completion. Students in STEM majors at 4-year institutions with parents having lower levels of education and lower incomes were equally likely to change their programs to a non-STEM major, as compared to students, whose parents had higher educational level and income. They were more likely to depart the STEM major and leave the institution completely (Chen, 2013). Looking at this trend, it can be concluded that first-generation and low-income students were more likely to leave STEM degree programs than their non-first generation and higher income
counterparts. An older study of first-generation students using national datasets by Chen and Carroll (2005) found that they were less likely to select a major in the STEM field. There has been some studies that show why first-generation students were less likely to enroll. A case study by Trenor, Yu, Waight, Zerda, & Sha (2008) showed that first-generation students may have lacked the social capital, such as role models, to gain access to engineering majors, and once in college, all other barriers (e.g., financial, work commitments) may have hindered social capital acquisition once in the major. In many cases, there is also an access issue in that social capital may not exist even when students are enrolled. A qualitative investigation of first-generation students in engineering majors explained that some of the barriers faced by first-generation students in terms of higher education access and persistence in engineering majors include the following: lack of knowledge of the admission process, financial concerns, challenging engineering curriculum, few role models, balancing college with personal commitments, and lack of parental knowledge (Fernandez, Trenor, Zerda, & Cortes, 2008; Garriott & Hultgren, 2017; Hunt et al., 2018). Another qualitative study of socioeconomically disadvantaged students in engineering and physical sciences majors, the majority of whom were first-generation, showed that the selection of a STEM major was because of their passion for research, future societal contributions and job prospects (Conrad, Canetto, MacPhee & Farro, 2009). Regardless, not all studies say that first-generation status is a barrier to STEM enrollment (Engle & Tinto, 2008).

Using logistic regression models, Crisp et al. (2009) found that the likelihood of enrolling in a STEM major was associated with gender, ethnicity, and higher math SAT
score, and high school class rank. First-generation status was neither a positive nor a negative predictor in Crisp’s study. As for success of first-generation students in STEM fields, research has been contradicting. Using data from the college board, Shaw & Barbuti (2010) found that for First-generation students overall, intended STEM majors changed their majors at lower rates than those in other majors, and among STEM majors, first-generation students switched at greater rates than non-first-generation college students. Using data from the Beginning Post-Secondary Students Longitudinal Study (BPS), it was shown that first-generation students were only slightly underrepresented in terms of initial STEM enrollment, but once in STEM majors, they were less likely to complete (Anderson & Kim, 2006). A qualitative study conducted by Wilson & Kittleson (2013), found that first-generation, low income women in STEM majors perceived challenges of not being positioned to compete in these majors due to culturally based priorities.

In the literature concerning first-generation students thus far, the connection between being a first-generation student and other categorizations of underrepresentation, particularly race/ethnicity are quite evident (Dika& D'Amico, 2015). There have been plenty of interventions or supplemental instruction programs that target both underrepresented minorities and first-generation students, which often overlap. One such program is the STRONG-CT program, which is an alliance between four U.S. institutions: the University of Connecticut, Manchester Community College, Quinebaug Valley Community College, and Three Rivers Community College, which look to help diversify and enlarge the STEM communities in Connecticut by increasing enrollment, retention and graduation of racial/ ethnic minority, and first-generation
college students in life science disciplines (McGonagle et al., 2014). There have been studies evaluating the program for its efficacy in recruiting and retaining these groups of students, and the overall results through surveys indicate that STRONG-CT is a promising program that shows benefits for underrepresented minorities and first-generation students academically, and in terms of psychological and behavioral variables (McGonagle et al., 2014).

Many other programs from colleges such as the University of Maryland Baltimore County, University of Akron, Bowling Green State University, and others now offer some form of a pre-engineering or math/science program to promote the pursuit of undergraduate STEM education among under-represented groups, which includes minorities and first-generation students (Yelamarthi & Mawasha, 2008). All these programs are designed for specific groups of students and have the ultimate goal of retaining those students. There are other noteworthy goals of these projects which include the following: (a) reinforcing the self-confidence of underrepresented students, (b) enhancing students’ problem solving skills through hands-on learning approach, (c) increasing awareness of the students to pursue a career in stem disciplines, and (d) providing diagnostic testing in math and find the focus areas that require additional efforts prior to enrollment in college (Dika & Alvarez, 2016; Yelamarthi & Mawasha, 2008).

2.7 Pre-College Experiences

Pre-college experiences play an important role in determining how well students
would do in a college/university setting, and this holds true for STEM specifically (Tinto, 1975, 2008; Gaskins, 2009).

2.8 High School Rank, SAT Scores and Student Retention

There are numerous studies, concerning the importance of high school rank and scores on placement tests, in addition to grades in math, chemistry, and physics on student retention (Mendez et al., 2008). Besterfield-Sacre, Atman & Shuman (1997) have shown that models incorporating cognitive variables such as student high school math and science ability, as well as basic engineering knowledge, allow for better prediction of student retention. Zhang, Anderson, Ohland, and Thorndyke (2004) also found that high school GPA was one of the best predictors of retention and graduation. In the past, at the high school level, students often believe that science is no longer engaging and enjoyable, partly due to the fact that advanced courses are content-heavy but shallow, allowing students little time for investigative experiences (National Research Council, 2002; Perkins-Gough, 2006).

Gaskins (2009) emphasized in his study that student predefined variables such as high school GPA combined with environmental variables such as student living and involvement in first year programs such as a residential living learning community are best predictors of student success. The study was conducted over a 10-year period (fall 1997 through fall 2006), and 35,050 students were involved from all majors.

Over the years, there has been a lot of focus on the predictive nature of SAT scores on academic success (Burton & Ramist, 2001; Kroc, Woodard, Howard, & Hull, 1995; Reasons, 2003; Sedlacek & Adams, 1992). Although SAT scores have typically
been good predictors of academic success, a gap exists in the literature (Rohr, 2012). Sedlacek and Adams (1992) predicted the academic success of student-athletes using SAT scores and non-cognitive variables. The researchers found that standardized tests such as the SAT correlated with freshman grades for White students, but they were not so accurate when it came to predicting the success of student athletes who were non-traditional students. They, however, could be predicted with non-cognitive factors such as self-concept, self-appraisal, leadership, community, and non-traditional knowledge (Sedlacek & Adams, 1992; Rohr, 2012).

There have been some authors, however, who have been critical of standardized tests and have reported that they were biased, placing students of color at a disadvantage (Castenell & Castenell, 1988; Malveaux, 1997; Mills, 2003; Steele, 1997; Zirkel, 2000). Hoffman & Lowitzki (2005) measured the predictive strength of high school GPA and standardized tests of minority and religious minority students. They found that both were not predictors of academic success for minority and religious minority students. Rather, according to the authors, multiple factors played a role in minority student success including race, religion, gender, hours worked on a job, financial need, ability demonstrated on the SAT, housing, and satisfaction with the college’s services (Rohr, 2012). Rohr (2012) conducted a study at a small liberal arts college in Indiana to see if SAT scores would be good predictors for retention among STEM students. The study found that college preparatory GPA and the aggregate SAT score were predictors of retention of STEM and business students. For every point increase in GPA, the odds were more than twice as much that the student would be
retained. For every point increase in SAT, there was 0.3% increase in retention (Rohr, 2012).

2.9 Grade Performances

When students are inside the institutional setting, their experiences are likely to play a large role in whether or not they would graduate or dropout. Tinto’s (1975) model places an emphasis on experiences, such as formal and informal academic interactions with faculty, perceptions of campus climate, validating experiences from faculty, and peers and mentoring relationships (faculty, peer, and advising staff). The model also places an emphasis on academic performance, academic/intellectual development, and non-cognitive gains as intermediate outcomes, which in turn determine persistence in college (Chang et al., 2014).

2.10 College GPA and Student Retention

Way back in 1975, Tinto reported, “with respect to grade performance, many studies have shown it to be the single most important factor in predicting persistence in college” (Tinto, 1975, P. 104). Several other scholars report GPA to be a major and direct predictor of attrition or retention (Cabrera et al., 1993; Fischbach, 1990; Mohammadi, 1994; Pascarella & Chapman, 1983; Wall et al., 1996). Some studies have shown that as students, especially minorities, progress through their academic career, their interest in STEM weakens as their achievements in these classes decline (Crisp et al., 2009; Ma & Liu, 2015; Means et al., 2017). This is the reason why college GPA
matters. It has been consistently shown that higher the GPA, the higher the persistence levels (Rohr, 2012).

2.11 Summary of the Literature Review

The study was based on Vincent’s Tinto model, but the variables that were studied have been the ones that were academic in nature. Individual attributes that were a part of the student’s identity such as Race, Gender and First- generation status have proven to be strong indicators in the past. So have Pre- college schooling such as SAT scores and High school rank. Grade performances once the student gets into the University setting is extremely important as well. These are the variables that were included in this study.
CHAPTER 3

METHODS

3.1 Introduction

Enrollment numbers have gone up in STEM fields (Kokkelenberg & Sinha, 2010). About 28% of bachelor’s degree and 20% of associate’s degree students entered a STEM field at some point within 6 years of entering post-secondary education in 2003-04 (U.S. Department of Education, 2013). Regardless, graduation rates have not kept up with enrollment rates (Kokkelenberg & Sinha, 2010). This is because many of these STEM entrants left STEM several years later by either changing majors or leaving college without completing a degree or certificate (Chen, 2013). A total of 48% of bachelor’s degree students and 69% of associate’s degree students who entered STEM fields between 2003 and 2009 had left these fields by 2009 (U.S. Department of Education, 2013).

The purpose of this study was to identify trends in completion rates of STEM field 4-year degrees, at a single university in Texas, with women, minorities and first-generation students, while considering the student’s academic record in high school and the early years of the degree. This could fill an important knowledge gap in what may contribute to these high rates of lack of completion and act as an extension to a study conducted in 2011 by Thompson and Bolin at the same university. Two-way interactions have not previously been explored as widely as single variables that individually contribute to student retention. This study looked at two-way interactions between underrepresented groups of students, pre-college characteristics, and Institutional experiences as referenced by Vincent Tinto.
This chapter presents a description and justification for the research methods, a description of the sample used in this study, instrumentation and procedures for data collection, the variables and research questions, data analysis plans, and finally a summary of the chapter.

3.2 Description of the Study

Non-experimental, observational research is research that lacks manipulation of the independent variables by the researcher; therefore, the researcher studied what has already occurred and discussed how variables are related. Despite its limitations for studying cause and effect, non-experimental research is very important in education. In this study, schools and their students come as they are, and the variables of this analysis are either impossible to manipulate (such as gender or racial background) or not ethical to manipulate (such as subject of study). As such, a non-experimental, observational design is an appropriate design for this study. As this is an observational, non-experimental dataset, there were no treatment levels or control groups. Control using statistical methods will replace experimental control.

3.3 Subjects and Sample Size

Data have been obtained from a university located in Denton, Dallas-Fort Worth in Texas. The data were collected through a review of existing student records containing the student enrollment numbers for each department and demographic data for each student. Data for first-generation students are primarily self-reported and thus may lack accuracy.
The sample consisted of 37,739 students. Data were obtained from the academic years 2005, 2006, and 2007, which covered enrollment data from fall 2005, spring 2006, fall 2006, spring 2007, fall 2007, spring 2008, fall 2008 and spring 2009. Studies have shown that it takes an average of 6 years to complete a bachelor’s degree. The time taken to degree completion has been shown to be longer in the case of underrepresented students (Wolniak, 2015). The study variables for Research Questions 1 through 3 included the dependent variable of completion of bachelor’s degree, as well as the independent variables of STEM classification, gender, minority, first-generation, first year GPA, second year GPA, high school rank, and SAT score. Mean and standard deviation scores were used to summarize the continuous measured study variables, while frequency and percentage summaries were used to summarize the categorical measured study variables. Multiple logistic regressions were performed to determine how the effect of STEM degree status interacts with student demographics and high school and college achievement measures to predict likelihood of degree completion. The study variables for Research Questions 4 through 6 includes the dependent variable of time taken to complete bachelor’s degree, as well as the independent variables of STEM classification, gender, minority, first-generation, first year GPA, second year GPA, high school rank, and SAT score. Linear regression was used to explore the relationships between STEM degree status, student demographics, high school and college achievement measures, and length of time taken to complete a degree.

In order to figure out two-way interactions between the variables and to determine which two variables played a significant role in contributing to STEM degree
completion, a hierarchical regression model was performed on the entire sample. For this analysis, missing data for the variables of interest were excluded from the sample. The study variables included the dependent variable of completion of bachelor’s degree; independent variables of STEM classification, gender, minority, first-generation, first year GPA, second year GPA, high school rank, and SAT score. Mean and standard deviation scores were used to summarize the continuous measured study variables, while frequency and percentage summaries were used to summarize the categorical measured study variables. A breakdown of the racial and gender distribution for all STEM departments in 2005 is presented in Table 4.1. Overall, there was an even distribution of both genders across all races, and most racial categories had a suitable number of subjects for data analysis. In the event some racial categories do not have enough members for data analysis, racial groups were combined together.

3.4 Instrumentation

There were no instruments used in this study, as all data was collected from student records.

3.5 Procedure for Data Collection

The dataset was obtained from the University Institutional Research and Effectiveness Unit, used in this study after IRB approval (Approval Code: 15-114). Subject anonymity was maintained by use of a student numeric identification, separate from any known student identification numbers.
3.6 Variables and Research Questions

Six main research questions guided this research project:

RQ1: Are there statistically significant relationships between a STEM student’s completion of a 4-year bachelor’s degree at a postsecondary institution and the student’s gender, race/ethnicity, and first-generation status?

RQ2: Are there statistically significant relationships between a STEM student’s completion of a 4-year bachelor’s degree at a postsecondary institution and the student’s SAT score and high school academic achievement as measured by quartile ranking?

RQ3: Are there statistically significant relationships between a STEM student’s completion of a 4-year bachelor’s degree at a postsecondary institution and the student’s first and second year college GPA?

RQ4: For students that completed their degree, are there statistically significant relationships between STEM students’ length of time taken to complete a bachelor’s degree at a postsecondary institution and the student’s gender, race/ethnicity, and first-generation status?

RQ5: For students that completed their degree, are there statistically significant relationships between a STEM length of time to complete a bachelor’s degree at a postsecondary institution and the student’s SAT score and high school academic achievement as measured by quartile ranking?

RQ6: For students that completed their degree, are there statistically significant relationships between a STEM student’s length of time to complete a bachelor’s degree at a postsecondary institution and the student’s first and second year college GPA?

The following is a list of the operationally defined variables for the study:

- **Dependent Variable:** Completion of 4-year degree, classified as a binary variable of whether or not the subject successfully completed a 4-year bachelor.

- **Independent Variable1:** STEM or not, a binary classification determining if the subject had completed study in a STEM field or not. STEM field study is classified as science, technology, engineering or mathematics subjects. STEM classification was used based on the completed field of study for students who complete their 4-year
degree and intended field of study for students who do not complete the 4-year degree. Students who begin their degree as undecided on field were classified as non-STEM.

- Independent Variable 2: Gender, a binary variable with two levels of male or female.
- Independent Variable 3: Minority, a binary variable based on reported race. Ethnicities other than White/Caucasian and Asian were considered minority races for the purposes of this analysis.
- Independent Variable 4: First-generation, a binary variable based on self-reported first-generation status. First-generation was classified in this study, as parents having never enrolled in postsecondary education (NCES, 1998).
- Independent Variable 5: First year GPA, a continuous variable between 0 and 4 representing the subject’s GPA at the end of their first year of study at the university.
- Independent Variable 6: Second year GPA, a continuous variable between 0 and 4 representing the subject’s GPA at the end of their second year of study at the university.
- Independent Variable 7: High school rank, a continuous variable between 0 and 4 representing the subject’s rank at the end of high school.
- Independent Variable 8: SAT score, a continuous variable up to 2,400 representing the SAT score of the subject upon university admission.

In addition to the variables above, one confounding variable was included in this analysis:

- Confounding Variable 1: Year of enrollment, a categorical variable with four levels representing the enrollment year of the student.
3.7 Data Analysis

Summary statistics were generated for the dependent and independent variables. For continuous variables, means and standard deviations were presented, and for categorical variable counts and frequencies were presented. Next, data analysis was undertaken in three phases.

The first phase explored Research Questions 1 through 3, which focused on the dependent variable of STEM degree completion. The study variables for Research Questions 1 through 3 included the dependent variable of completion of bachelor's degree, as well as the independent variables of STEM classification, gender, minority, first-generation, first year GPA, second year GPA, high school rank, and SAT score. Mean and standard deviation scores were used to summarize the continuous measure study variables, while frequency and percentage summaries were used to summarize the categorical measure study variables. Binary logistic regressions were performed to determine how the effect of STEM degree status interacts with student demographics and high school and college achievement measures to predict likelihood of degree completion. The second phase focused on the length of time taken to complete the degree. The study variables for Research Questions 4 through 6 included the dependent variable of time taken to complete bachelor's degree, as well as the independent variables of STEM classification, gender, minority, first-generation, first year GPA, second year GPA, high school rank, and SAT score. Linear regression was used to explore the relationships between STEM degree status, student demographics, high school and college achievement measures, and length of time taken to complete a degree.
In order to figure out two-way interactions between the variables and to determine which two variables played a significant role in contributing to STEM degree completion, a hierarchical regression model was performed on the entire sample. This is the third phase of the analysis. For this analysis, missing data for the variables of interest in the study were excluded from the sample. The study variables included the dependent variable of completion of bachelor’s degree, as well as the independent variables of STEM classification, gender, minority, first-generation, first year GPA, second year GPA, high school rank, and SAT score. Mean and standard deviation scores were used to summarize the continuous measured study variables, while frequency and percentage summaries were used to summarize the categorical measured study variables.

Data analysis was conducted using binary logistic regression with a logit link function. Binary logistic regression is a type of regression analysis that quantifies the probability of a binary outcome (Crawley, 2013), and completion or not of a 4-year degree, as a relationship with the other variables. In this case, it allowed the change in risk of a student not completing their 4-year degree to be quantified for each independent variable. All independent variables were included in the model, and all two-way interactions with the variable for STEM/non-STEM field were included. This was done for several reasons. Firstly, this allowed relationships between the independent variables and completion to be quantified for both STEM and non-STEM subjects, and secondly, it helped determine if those relationships are significantly different between the STEM and non-STEM students. Backwards model selection guided the model
building process using a combination of AIC and BIC as selection criteria (Bozdogan, 1987; Burnham & Anderson, 2004).

Backwards selection can be considered a process of selecting appropriate covariates for a model, allowing conclusions to be drawn about the model under the assumption the model is the truest possible (Buckland, Burnham, & Augustin, 1997). It is a model selection process where all variables are included in an initial model and iterative changes are made removing one variable at each step and assessing model fit to determine if removal of that parameter made a significant change to model fit. AIC and BIC are both measures of model fit that also penalize a model for inclusion of a large number of parameters to minimize overfitting. The end goal is a model that adequately explains the dataset with the minimum number of parameters. Parameter estimations and significant of the parameters included in the final model were used to answer the three research questions posed by this study.

3.8 Design and Ethical Considerations

It was assumed that the data collected during data collection was reliable and valid. As this is an observational, non-experimental study, the researcher was unable to manipulate independent variables, and statistical control took the place of experimental control. Due to the inherent limitations of observational data to identify cause and effect, caution must be exercised when drawing conclusions from the data, particularly when trying to attribute causality. This study also only took place using data from a single school; therefore, applicability of the findings to other schools may be limited unless,
there is a large consensus in the literature for similar findings, indicating the patterns uncovered in this analysis are applicable on a broader scale.

3.9 Summary

The purpose of this study was to identify trends in completion rates of STEM field 4-year degrees, at a single university in Texas, with women, minorities and first-generation students, while considering the student’s academic record in high school and the early years of the degree. Data analysis was conducted using appropriate summary statistics and a binary logistic model to determine the nature and strength of the relationship between dependent and independent variables. Results of this data analysis are discussed in Chapter 4.
CHAPTER 4
RESULTS

4.1 Introduction

The purpose of this quantitative study using non-experimental, observational method was to identify trends in completion rates of STEM field bachelor’s degrees, at a single university in Texas with focus on women, minorities, and first-generation students, while considering the student’s academic record in high school and the early years of the degree. Chapter 4 presents descriptive data and data analysis using binary logistic regression and linear regression. IBM® SPSS® Statistics Version 22 was utilized to conduct the data analysis.

4.2 Research Questions:

The following research questions and hypotheses guided the analysis for this study:

RQ1: Are there statistically significant relationships between a STEM student’s completion of a 4-year bachelor’s degree at a postsecondary institution and the student’s gender, race/ethnicity, and first-generation status?

RQ2: Are there statistically significant relationships between a STEM student’s completion of a 4-year bachelor’s degree at a postsecondary institution and the student’s SAT score and high school academic achievement as measured by quartile ranking?

RQ3: Are there statistically significant relationships between a STEM student’s completion of a 4-year bachelor’s degree at a postsecondary institution and the student’s first and second year college GPA?

RQ4: For students that completed their degree, are there statistically significant relationships between STEM students’ length of time taken to complete a bachelor’s degree at a postsecondary institution and the student’s gender, race/ethnicity, and first-generation status?

RQ5: For students that completed their degree, are there statistically significant relationships between a STEM length of time to complete a bachelor’s degree at
Chapter 4 involves a review of frequencies and descriptives of the study variables. Next, Research Questions 1 through 3 was addressed using binary logistic regression. Research Questions 4 through 6 was addressed using linear regression. Lastly, hierarchical binary logistic regression was used to identify a model using all significant predictors previously explored in Research Questions 1 through 3, and hierarchical linear regression was used to identify a model using all significant predictors previously explored in Research Questions 4 through 6. Results are presented by each variable.

4.3 Summary of Study Variables

The sample consisted of 37,739 students. These students were freshmen enrolled in the academic years 2005, 2006 and 2007. This included fall 2005, spring 2006, fall 2006, spring 2007, fall 2007 and spring 2008. Tables 4.1 and 4.2 summarize the descriptive statistics of the study variables. The study variables included the dependent variable of completion of bachelor's degree, as well as the independent variables of STEM classification, gender, minority, first-generation, first year GPA, second year GPA, high school rank, and SAT score. Mean and standard deviation scores were used to summarize the continuous measured study variables while frequency and percentage summaries were used to summarize the categorical...
measured study variables. In terms of completion of the degree, more than half (58.2%) of the 37,739 students (21,965) did not complete a bachelor’s degree. There were 16.3% of the students in the same study sample-6,156 students who declared a STEM major. For gender, more than half (55.3% or 20,872) students in the sample were women. In terms of ethnicity, 36.7% of the students in the sample (13,854) students were in minority ethnic groups. 32.4% of the students in the sample (12,236) students were first-generation status. Lastly, in terms of high school rank, many of the students in the sample were ranked in the 2nd quarter (11.5%). There were 4,322 students in the 2nd quarter.

Table 4.1

*Frequency and Percentage Summaries of Study Variables (N = 37,739)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>STEM or not</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0 Not STEM</td>
<td>31,583</td>
<td>83.7</td>
</tr>
<tr>
<td>1.0 STEM Field</td>
<td>6,156</td>
<td>16.3</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0 Women</td>
<td>20,872</td>
<td>55.3</td>
</tr>
<tr>
<td>1.0 Men</td>
<td>16,865</td>
<td>44.7</td>
</tr>
<tr>
<td>Missing</td>
<td>2</td>
<td>&lt;.01</td>
</tr>
<tr>
<td><strong>Minority</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0 No</td>
<td>23,885</td>
<td>63.3</td>
</tr>
<tr>
<td>1.0 Yes</td>
<td>13,854</td>
<td>36.7</td>
</tr>
<tr>
<td><strong>First-generation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0 No</td>
<td>15,690</td>
<td>41.6</td>
</tr>
<tr>
<td>1.0 Yes</td>
<td>12,236</td>
<td>32.4</td>
</tr>
<tr>
<td>Missing</td>
<td>9,813</td>
<td>26.0</td>
</tr>
<tr>
<td><strong>High School Rank</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0 Top 10%</td>
<td>1,524</td>
<td>4.0</td>
</tr>
<tr>
<td>2.0 11%-25%</td>
<td>3,310</td>
<td>8.8</td>
</tr>
<tr>
<td>3.0 2nd quarter</td>
<td>4,322</td>
<td>11.5</td>
</tr>
<tr>
<td>4.0 3rd quarter</td>
<td>1,202</td>
<td>3.2</td>
</tr>
<tr>
<td>5.0 4th quarter</td>
<td>105</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Table 4.2

Descriptive Statistics of SAT Scores and GPA

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT Total Score</td>
<td>9,404</td>
<td>500</td>
<td>2260</td>
<td>1074.33</td>
<td>138.64</td>
</tr>
<tr>
<td>First Year GPA</td>
<td>9,809</td>
<td>0</td>
<td>4</td>
<td>2.63</td>
<td>0.92</td>
</tr>
<tr>
<td>Second Year GPA</td>
<td>7,265</td>
<td>0</td>
<td>4</td>
<td>2.89</td>
<td>0.65</td>
</tr>
</tbody>
</table>

For Research Questions 4 through 6, only participants who completed their bachelor's degree were analyzed. The sample size for RQs 4 through 6 is 15,774 which was the number of students who completed their bachelor's degree. To preserve the sample size, cases with incomplete data were included in the analyses as long as the participant had data for the variables used in the present analysis. For hierarchical linear and logistic regression, only cases with complete data were used. Exact statistics on missing values can be seen in Tables 4.3 and 4.4.

Table 4.3

Frequency and Percentage Summaries of Study Variables (N = 15,774)

<table>
<thead>
<tr>
<th>Completion of bachelor's degree</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 Completed Bachelor degree</td>
<td>15,774</td>
<td>100</td>
</tr>
<tr>
<td>-1.0 Not STEM</td>
<td>14,027</td>
<td>88.9</td>
</tr>
<tr>
<td>1.0 STEM Field</td>
<td>1,747</td>
<td>11.1</td>
</tr>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------</td>
<td>---------</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.0 Women</td>
<td>9,758</td>
<td>61.9</td>
</tr>
<tr>
<td>1.0 Men</td>
<td>6,015</td>
<td>38.1</td>
</tr>
<tr>
<td>Missing</td>
<td>1</td>
<td>&lt;.01</td>
</tr>
<tr>
<td><strong>Minority</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.0 No</td>
<td>10,666</td>
<td>67.6</td>
</tr>
<tr>
<td>1.0 Yes</td>
<td>5,108</td>
<td>32.4</td>
</tr>
<tr>
<td><strong>First-generation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.0 No</td>
<td>7,046</td>
<td>44.7</td>
</tr>
<tr>
<td>1.0 Yes</td>
<td>4,476</td>
<td>28.4</td>
</tr>
<tr>
<td>Missing</td>
<td>4,252</td>
<td>27.0</td>
</tr>
<tr>
<td><strong>High School Rank</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0 Top 10%</td>
<td>1,055</td>
<td>6.7</td>
</tr>
<tr>
<td>2.0 11%-25%</td>
<td>1,993</td>
<td>12.6</td>
</tr>
<tr>
<td>3.0 2nd quarter</td>
<td>2,180</td>
<td>13.8</td>
</tr>
<tr>
<td>4.0 3rd quarter</td>
<td>525</td>
<td>3.3</td>
</tr>
<tr>
<td>5.0 4th quarter</td>
<td>44</td>
<td>0.3</td>
</tr>
<tr>
<td>6.0 No Rank</td>
<td>178</td>
<td>1.1</td>
</tr>
<tr>
<td>Missing</td>
<td>9,799</td>
<td>62.1</td>
</tr>
</tbody>
</table>

Table 4.4

*Descriptive Statistics of SAT Scores and GPA*

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SAT Total Score</strong></td>
<td>5,251</td>
<td>500</td>
<td>1540</td>
<td>1082.70</td>
<td>139.64</td>
</tr>
<tr>
<td><strong>First Year GPA</strong></td>
<td>5,876</td>
<td>0</td>
<td>4</td>
<td>2.95</td>
<td>0.70</td>
</tr>
<tr>
<td><strong>Second Year GPA</strong></td>
<td>5,583</td>
<td>0.82</td>
<td>4</td>
<td>3.03</td>
<td>0.56</td>
</tr>
<tr>
<td><strong>Degree completion length</strong></td>
<td>15,773</td>
<td>1</td>
<td>11</td>
<td>4.03</td>
<td>1.06</td>
</tr>
</tbody>
</table>

Table 4.5 shows a comparative look at students who declared STEM and Non-STEM majors, as compared to students who completed their study in these fields.
Comparative Look at Declared vs. Completed STEM and Non-STEM Majors (N for Declared Majors=37,739; N for Completed Majors=15,774)

<table>
<thead>
<tr>
<th>Variable (STEM or Not)</th>
<th>STEM</th>
<th>Non- STEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declared</td>
<td>6,156 (16.3%)</td>
<td>31,583 (83.7%)</td>
</tr>
<tr>
<td>Completed</td>
<td>1,747 (11.1%)</td>
<td>14,027 (88.9%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable (Gender)</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declared</td>
<td>3373 (8.9%)</td>
<td>2449 (6.5%)</td>
</tr>
<tr>
<td>Completed</td>
<td>912 (5.7%)</td>
<td>598 (3.8%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable (Race)</th>
<th>Minority</th>
<th>Non- Minority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declared</td>
<td>2449 (6.5%)</td>
<td>3707 (9.8%)</td>
</tr>
<tr>
<td>Completed</td>
<td>598 (3.8%)</td>
<td>1149 (7.2%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable (First Generation Status)</th>
<th>First Generation</th>
<th>Non- First Generation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declared</td>
<td>2225 (5.8%)</td>
<td>9981 (26.4%)</td>
</tr>
<tr>
<td>Completed</td>
<td>535 (3.4%)</td>
<td>3941 (25%)</td>
</tr>
</tbody>
</table>


4.4 Results

For Research Questions 1 through 3, binary logistic regressions were performed to determine how the effect of STEM degree status interacted with student demographics and high school and college achievement measures to predict likelihood of degree completion. For Research Questions 4 through 6, linear regressions were used to explore the relationships between STEM degree status, student demographics, high school and college achievement measures, and length of time taken to complete a degree. Effects were considered significant at a $p$-value of .05. The results are presented by each independent variable.

4.4.1 Gender: STEM Completion

The first research question examined the relationship between STEM status, gender, minority status, and first-generation status on the likelihood of completing a bachelor’s degree. The sample size for this research question was 37,737. Two values were excluded from the original sample because of two missing values.

In the first model, binary logistic regression was used to regress degree completion (yes, no) on STEM degree status (yes, no), gender (male, female), and their interaction. The logistic regression model was statistically significant ($X^2(3) = 987.316$, $p < 0.00001$), explaining 3.5% (Nagelkerke $R^2$) of the variance in degree completion and correctly classifying 58.2% of cases. There was a significant main effect of STEM status on likelihood of completing a degree, controlling for the effect of gender and their interaction ($Wald (1) = 461.240$, $p < .00001$); Non-STEM students were more likely to finish their degree by a factor of 1.93 compared to STEM students. There was also a significant main effect of gender on likelihood of degree completion ($Wald (1) = 101.239$, $p < .00001$).
controlling for the effect of STEM status and their interaction. Female students were more likely to obtain their Bachelor’s degree compared to male students by a factor of 1.36. There was a significant interaction of STEM status and gender on likelihood of completing a degree (Wald (1) = 27.999, \( p < .000001 \)); women were more likely to complete their degree than men, and this effect was significantly stronger for non-STEM students compared to STEM students.

**Figure 4.1.** Frequency of students who completed versus did not complete their degree, by gender, for non-STEM students.

**Figure 4.2.** Frequency of students who completed versus did not complete their degree, by gender, for STEM students.

4.4.2 Gender: Time Taken to Complete

The fourth research question examined the relationship between STEM status,
gender, minority status, and first-generation status on the length of time taken to complete a bachelor’s degree. The sample size for this research question was 15,772 because the question only took into consideration those students who had completed their degree. This reduces the sample size from 37,737 to 15,772. In the first model, linear regression was used to regress length of time taken to complete a degree (years) on STEM degree status (yes, no), gender (male, female), and their interaction. The model was statistically significant \(F(3, 15768) = 146.417, p< .00001, R^2 = .027\). There was a significant main effect of gender on years to complete degree, controlling for STEM status and their interaction \(F(1, 15768) = 188.048, p< .000001, R^2 = .012\). Female students \((M = 3.90, SD = 0.99)\) took less time to complete their degree on average than male students \((M = 4.27, SD = 1.14)\). There was no significant main effect of STEM status on years to completion degree, controlling for gender and their interaction \(F(1, 15768) = 0.373, p = .541, R^2 < .001\). There was no significant interaction effect of gender and STEM status on years to degree completion \(F(1, 15768) = 0.145, p = .703, R^2 < .001\).

4.4.3 Race/Ethnicity: STEM Completion

For the next model, binary logistic regression was used to regress degree completion (yes, no) on STEM degree status (yes, no), racial minority status (yes, no), and their interaction. The sample size for this question was 37,739. All the values that had a known variable for race were included. The logistic regression model was statistically significant \(X^2(3) = 768.843, p< 0.00001\), explaining 2.7% (Nagelkerke \(R^2\)) of the variance in degree completion and correctly classifying 58.2% of cases. There was a significant main effect of STEM status, controlling for minority status and their
interaction (Wald (1) = 479.840, \( p < .000001 \)). Non-STEM students were 2.0 times more likely to obtain their degree than STEM students. There was also a significant main effect of minority status, controlling for STEM status and their interaction (Wald (1) = 101.574, \( p < .000001 \)). Non-minority students were 1.38 times more likely to obtain their degree compared to minority students. There was no significant interaction effect of STEM status and minority status on likelihood of obtaining a degree (Wald (1) = 0.085, \( p = .770 \)).

4.4.4 Race/Ethnicity: Time Taken to Complete

For the second model, length of time taken to complete a degree was regressed on STEM status, minority status, and their interaction. The sample size for this question included only those who had completed their degree, which was 15,773. The model was statistically significant (\( F (3, 15769) = 3.066, p = .027, R^2 = .001 \)). There was a significant main effect of STEM status on years to complete degree, controlling for minority status and their interaction (\( F (1, 15769) = 7.984, p = .005, R^2 = .001 \)). Non-STEM students (\( M = 4.027, SD = 1.06 \)) took less time to complete their degree on average than STEM students (\( M = 4.108, SD = 1.08 \)). There was no significant main effect of minority status on years to completion degree, controlling for STEM status and their interaction \( F (1, 15769) = 1.741, p = .187, R^2 < .001 \)). There was no significant interaction effect of minority status and STEM status on years to degree completion \( F(1,15769) = 0.698, p = .404, R^2 < .001 \)).

4.4.5 First-Generation Status: STEM Completion

For the last model related to student demographics, degree completion was
regressed on STEM status, first-generation status, and their interaction. The sample size for this question was 27,926, which only included those values for which first-generation status was known. The logistic regression model was statistically significant ($X^2(3) = 634.984, p< 0.00001$), explaining 3.0% (Nagelkerke $R^2$) of the variance in degree completion and correctly classifying 58.7% of cases. There was a significant main effect of STEM status, controlling for first-generation status and their interaction ($Wald (1) = 405.109, p< .000001$). Non-STEM students were 2.04 more likely to obtain their degree compared to STEM students. There was also a significant main effect of first-generation status, controlling for STEM status and their interaction ($Wald (1) = 101.844, p < .000001$). Non-first-generation students were 1.43 times more likely to obtain their degree compared to first-generation students. There was no significant interaction effect of STEM status and first-generation status on likelihood of obtaining a degree ($Wald (1) = 0.697, p = .410$).

4.4.6 First-Generation Status: Time Taken to Complete

For the last model, length of time taken to complete a degree was regressed on STEM status, first-generation status, and their interaction. The sample size for this question was 11,522, which only included students who completed their degree. The model was statistically significant ($F (3, 11518) = 5.377, p = .001, R^2= .001$). There was a significant main effect of STEM status on years to complete degree, controlling for first-generation status and their interaction ($F (1, 11518) = 10.143, p = .001, R^2= .001$). Non-STEM students ($M = 3.908, SD = 1.06$) took less time to complete their degree on average than STEM students ($M = 4.003, SD = 1.08$). There was also a significant main effect of first-generation status on length of degree completion, controlling for STEM
status and their interaction ($F(1, 11518) = 4.510, p = .034, R^2 < .001$). Non-first-generation students ($M = 3.926, SD = 0.943$) took less time to complete their degree on average than first-generation students ($M = 3.986, SD = 0.956$). There was no significant interaction effect of first-generation status and STEM status on years to degree completion $F (1, 11518) = 0.936, p = .333, R^2 < .001$).

4.4.7 SAT Scores: STEM Completion

The second research question examined the relationship between STEM status, high school rank, and SAT score on the likelihood completing a bachelor's degree. The sample size for this question was 9,404, which included only those values for which the SAT scores were known. For the first model, binary logistic regression was used to regress degree completion (yes, no) on STEM degree status (yes, no), SAT score, and their interaction. The logistic regression model was statistically significant ($X^2(3) = 141.135, p < 0.00001$), explaining 2.0% (Nagelkerke $R^2$) of the variance in degree completion and correctly classifying 57.6% of cases.

There was a significant main effect of STEM status on likelihood of obtaining a degree, controlling for SAT score and their interaction ($Wald (1) = 38.048, p < .000001$). Non-STEM students were 24.49 times more likely to obtain their degree compared to STEM students. There was also a significant main effect of SAT score on likelihood of obtaining a degree ($Wald (1) = 71.071, p < .000001$). For every one point increase in SAT score, likelihood of obtaining a degree increased by 1.002. There was also a significant interaction effect of SAT score and STEM status on likelihood of completing a degree ($Wald (1) = 27.053, p < .000001$). SAT scores were higher for those who
completed their degree versus those who did not, and this effect was significantly stronger for STEM compared to non-STEM students (see Figure 4.3).

![Average SAT Score](image)

**Figure 4.3.** The effect of STEM and SAT score on degree completion.

### 4.4.8 SAT Scores: Time Taken to Complete

The fifth research question examined the relationship between STEM status, SAT score, and high school rank on the length of time taken to complete a bachelor’s degree. The sample size for this question was 5,251, which only included students who completed their degree. In the first model, length of time to complete degree was regressed on STEM status and SAT score. The model was statistically significant ($F(3, 5247) = 35.526, p < .000001, R^2 = .013$). There was a significant negative relationship between SAT score and years to complete degree, controlling for STEM status and their interaction ($F(1, 5247) = 47.854, p < .000001, R^2 = .006$). For every one point increase in SAT score, length of time taken to complete degree decreased by 0.001 years, on average. There was no significant main effect of STEM status on years to degree completion, controlling for SAT scores and their interactions ($F(1, 5247) = 2.298, p = .130, R^2 < .001$). There was no significant interaction effect between STEM status and
SAT score on length of time for degree completion \( (F(1, 5247) = 1.078, p = .299, R^2 < .001) \).

4.4.9 High School Rank: STEM Completion

In the next model, degree completion was regressed on STEM, high school rank, and their interaction. The sample size for this question was 10,823, which only included those values for which high school rank was known. The logistic regression model was statistically significant \( (X^2(3) = 304.122, p < 0.00001) \), explaining 3.7\% (Nagelkerke \( R^2 \)) of the variance in degree completion and correctly classifying 58.2\% of cases. There was a significant effect of STEM status on degree completion, controlling for high school rank and their interaction (Wald (1) = 14.340, \( p = .00002 \)). Non-STEM students were 1.751 times more likely to obtain their degree than STEM students. There was also a significant main effect of high school rank on degree completion, controlling for STEM status and their interaction (Wald (1) = 100.986, \( p < .0000001 \)). For every increase in high school rank level, probability of obtaining a degree decreased by 0.76. There was no significant interaction effect between STEM status and high school rank on likelihood of obtaining a degree (Wald (1) = 0.005 \( p = .942 \)).

4.4.10 High School Rank: Time Taken to Complete

In the next model, length of time taken to complete degree was regressed on STEM status and high school rank. The sample for this question was 5,974, which included only those students who completed their degree. The model was statistically significant \( (F(3, 5970) = 69.308, p < .000001, R^2 = .034) \). There was a significant main effect of high school rank on length of time taken to complete degree, controlling for
STEM status and their interaction \((F (1, 5970) = 80.202, p < .000001, R^2 = .013)\). For every increase in high school rank level, length of time to complete degree increased by 0.206 years on average. There was no significant main effect of STEM status on years to degree completion, controlling for the effect of high school rank and their interaction \((F(1, 5970) = 5.196, p = .065, R^2 = .001)\). There was no significant interaction effect between STEM status and high school rank on length of time for degree completion \((F (1, 5970) = 0.007, p = .935, R^2 < .001)\).

4.4.11 First-Year GPA: STEM Completion

The third research question examined the relationship between STEM status, first year GPA, and second year GPA on the likelihood completing a bachelor’s degree. The sample size for this question was 9,809, which included only those values for which the first-year GPA was known. For the first model, degree completion was regressed on STEM status, first year GPA and their interaction. The logistic regression model was statistically significant \((X^2(3) = 304.122, p < 0.000001)\), explaining 25.5% (Nagelkerke R²) of the variance in degree completion and correctly classifying 71.6% of cases. There was a significant main effect of STEM status on likelihood of obtaining a degree, controlling for first year GPA and their interaction \((Wald (1) = 44.779, p < .000005)\). The odds of completing a degree were 7.02 times greater for non-STEM students compared to STEM students. There was also a significant main effect of first year GPA on likelihood of obtaining a degree, controlling for STEM status and their interaction \((Wald (1) = 663.217, p < .000001)\). As first year GPA increased by one point, odds of completing a bachelor’s degree increased by 3.68. There was also a significant interaction effect between STEM status and first year GPA \((Wald (1) = 21.606, p = \)
The likelihood of completing a degree was higher for students with a better first year GPA and this effect was significantly stronger for STEM students compared to Non-STEM students.

Figure 4.4. The effect of STEM and first year GPA on degree completion.

4.4.12 First-Year GPA: Time Taken to Complete

The sixth research question examined the relationship between STEM status, first year GPA, and second year GPA on the length of time taken to complete a bachelor’s degree. The sample size for this question was 5,875, which included only those students who completed their degree. In the first model, length of time taken to complete degree was regressed on STEM status and first year GPA. The model was statistically significant \( F(3, 5871) = 386.277, \ p < .000001, \ R^2 = .003 \). There was a significant main effect of STEM status on length of time taken to complete degree, controlling for first year GPA and their interaction \( F(1, 5871) = 15.276, \ p = .00009, \ R^2 = .003 \). On average, non-STEM students \( (M = 4.290, \ SD = 1.227) \) took less time to complete their degree than STEM students \( (M = 4.640, \ SD = 1.324) \). There was also a significant main effect of first year GPA on length of time taken to complete degree,
controlling for STEM status and their interaction ($F(1, 5871) = 430.392, p< .000001, R^2= .068$). For every one point increase in first year GPA, length of time taken to complete degree decreased by 0.803 years on average. There was also a significant interaction between STEM status and first year GPA on length of time taken to complete degree ($F(1, 5871) = 6.540, p = .011, R^2=.001$). Higher first year GPA was related to shorter degree length, and this effect was significantly stronger for STEM students compared to non-STEM students.

![Average Number of Years Taken to Complete Degree](image)

**Figure 4.5.** The effect of STEM and first year GPA on length of degree completion.

### 4.4.13 Second-Year GPA: STEM Completion

For the last model, degree completion was regressed on STEM status, second year GPA, and their interaction. The sample size for this question was 7,265, which included only those values for which second-year GPA was known. The logistic regression model was statistically significant ($X^2(3) = 1181.333, p< 0.00001$), explaining 22.7% (Nagelkerke $R^2$) of the variance in degree completion and correctly classifying 80.2% of cases. There was a significant main effect of STEM status on likelihood of obtaining a degree, controlling for second year GPA and their interaction ($Wald(1) =$
The odds of completing a degree were 4.66 times greater for non-STEM students compared to STEM students. There was also a significant main effect of second year GPA on likelihood of obtaining a degree, controlling for STEM status and their interaction (Wald (1) = 419.432, \( p < .000001 \)). As second year GPA increased by one point, odds of completing a bachelor’s degree increased by 5.38. There was also no significant interaction effect between STEM status and second year GPA (Wald (1) = 0.145, \( p = .078 \)).

4.4.14 Second-Year GPA: Time Taken to Complete

In the last model, length of time take to complete degree was regressed on STEM status and second year GPA. The sample size for this question was 5,582, which included only those students who completed their degree. The model was statistically significant (\( F(3, 5578) = 437.174, \ p < .000001, \ R^2 = .190 \)). There was a significant main effect of STEM status on length of time taken to complete degree, controlling for second year GPA and their interaction (\( F(1, 5578) = 21.937, \ p = .000003, \ R^2 = .004 \)). On average, non-STEM students (\( M = 4.221, \ SD = 1.160 \)) took less time to complete their degree than STEM students (\( M = 4.534, \ SD = 1.264 \)). There was also a significant main effect of second year GPA on length of time taken to complete degree, controlling for STEM status and their interaction (\( F(1, 5578) = 527.244, \ p < .000001, \ R^2 = .086 \)). For every one point increase in second year GPA, length of time taken to complete degree decreased by 1.037 years on average. There was also a significant interaction between STEM status and second year GPA on length of time taken to complete degree (\( F(1, 5578) = 13.286, \ p = .0003, \ R^2 = .002 \)). Higher second year GPA was related to shorter
degree length, and this effect was significantly stronger for STEM students compared to non-STEM students.

![Average Number of Years Taken to Complete Degree](image)

**Figure 4.6.** The effect of STEM and second year GPA on length of degree completion.

### 4.5 Analysis of Hierarchical Regression Models

The sample consisted of 37,739 students. These were the students who were enrolled in the academic years 2005, 2006 ad 2007. They were from fall 2005, spring 2006, fall 2006, spring 2007, fall 2007, and spring 2008. For these next two analyses, missing data for the variables of interest in the study were excluded from the sample, leaving 2,422 subjects for the analysis. Since this is a hierarchical analysis, all the independent variables had to be in place for the model to predict correctly the outcome. So for this model, every student analyzed had to have data in place for gender, race, first-generation status, SAT scores, high school rank, and first and second year GPA. Tables 4.5 and 4.6 summarize the descriptive statistics of the study variables used in the hierarchical binary logistic regression. The study variables included the dependent variable of completion of bachelor’s degree, as well as the independent variables of STEM classification, gender, minority, first-generation, first year GPA, second year GPA, and second year GPA.
GPA, high school rank, and SAT score. Mean and standard deviation scores were used to summarize the continuous measured study variables, while frequency and percentage summaries were used to summarize the categorical measured study variables. The sample size used for Hierarchical analysis is smaller than that used for Logistic and Linear regression. This sample is not representative of the bigger sample set because it accounts for multiple variables at the same time that was not used for the bigger sample set.

Table 4.6

**Frequency and Percentage Summaries of Study Variables: Binary Logistic**

<table>
<thead>
<tr>
<th>Completion of bachelor’s degree</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.0 Did not complete Bachelor degree</td>
<td>405</td>
<td>16.7</td>
</tr>
<tr>
<td>1.0 Completed Bachelor degree</td>
<td>2017</td>
<td>83.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>STEM or not</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.0 Not STEM</td>
<td>2124</td>
<td>87.7</td>
</tr>
<tr>
<td>1.0 STEM Field</td>
<td>298</td>
<td>12.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.0 Women</td>
<td>1439</td>
<td>59.4</td>
</tr>
<tr>
<td>1.0 Men</td>
<td>983</td>
<td>40.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Minority</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.0 No</td>
<td>1755</td>
<td>72.5</td>
</tr>
<tr>
<td>1.0 Yes</td>
<td>667</td>
<td>27.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>First-generation</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.0 No</td>
<td>1550</td>
<td>64.0</td>
</tr>
<tr>
<td>1.0 Yes</td>
<td>872</td>
<td>36.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>High School Rank</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 Top 10%</td>
<td>439</td>
<td>18.1</td>
</tr>
<tr>
<td>2.0 11%-25%</td>
<td>805</td>
<td>33.2</td>
</tr>
<tr>
<td>3.0 2nd quarter</td>
<td>917</td>
<td>37.9</td>
</tr>
<tr>
<td>4.0 3rd quarter</td>
<td>214</td>
<td>8.8</td>
</tr>
<tr>
<td>5.0 4th quarter</td>
<td>15</td>
<td>0.6</td>
</tr>
<tr>
<td>6.0 No Rank</td>
<td>32</td>
<td>1.3</td>
</tr>
</tbody>
</table>
Table 4.7

*Descriptive Statistics of SAT Scores and GPA for Binary Logistic Regression*

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT Total Score</td>
<td>2,422</td>
<td>560</td>
<td>1530</td>
<td>1084.88</td>
<td>138.73</td>
</tr>
<tr>
<td>First Year GPA</td>
<td>2,422</td>
<td>0</td>
<td>4</td>
<td>2.94</td>
<td>0.69</td>
</tr>
<tr>
<td>Second Year GPA</td>
<td>2,422</td>
<td>0</td>
<td>4</td>
<td>2.98</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Table 4.8

*Frequency and Percentage Summaries of Study Variables: Linear*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 Completed Bachelor degree</td>
<td>2017</td>
<td>100</td>
</tr>
<tr>
<td>STEM or not</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.0 Not STEM</td>
<td>1,792</td>
<td>88.8</td>
</tr>
<tr>
<td>1.0 STEM Field</td>
<td>225</td>
<td>11.2</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.0 Women</td>
<td>1,237</td>
<td>61.3</td>
</tr>
<tr>
<td>1.0 Men</td>
<td>780</td>
<td>38.7</td>
</tr>
<tr>
<td>Minority</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.0 No</td>
<td>1,473</td>
<td>73.0</td>
</tr>
<tr>
<td>1.0 Yes</td>
<td>544</td>
<td>27.0</td>
</tr>
<tr>
<td>First-generation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.0 No</td>
<td>1,314</td>
<td>65.1</td>
</tr>
<tr>
<td>1.0 Yes</td>
<td>703</td>
<td>38.7</td>
</tr>
<tr>
<td>High School Rank</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0 Top 10%</td>
<td>395</td>
<td>19.6</td>
</tr>
<tr>
<td>2.0 11%-25%</td>
<td>681</td>
<td>33.8</td>
</tr>
<tr>
<td>3.0 2nd quarter</td>
<td>746</td>
<td>37.0</td>
</tr>
<tr>
<td>4.0 3rd quarter</td>
<td>157</td>
<td>7.8</td>
</tr>
<tr>
<td>5.0 4th quarter</td>
<td>9</td>
<td>0.4</td>
</tr>
<tr>
<td>6.0 No Rank</td>
<td>29</td>
<td>1.4</td>
</tr>
</tbody>
</table>

The next model used hierarchical linear regression, which was used to analyze the length of time taken to complete their study. Only participants who completed their degree were analyzed. Tables 4.7 and 4.8 summarize the descriptive statistics of the
study variables used in the hierarchical linear regression. The study variables included
the dependent variable of length of time taken to complete a bachelor’s degree, as well
as the independent variables of STEM classification, gender, minority, first-generation,
first year GPA, second year GPA, high school rank, and SAT score. Mean and standard
deviceation scores were used to summarize the continuous variables, while frequency and
percentage summaries were used to summarize the categorical variables.

Table 4.9

Descriptive Statistics of SAT Scores and GPA for Linear Regression

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT Total Score</td>
<td>2,017</td>
<td>560</td>
<td>1530</td>
<td>1088.83</td>
<td>139.42</td>
</tr>
<tr>
<td>First Year GPA</td>
<td>2,017</td>
<td>0</td>
<td>4</td>
<td>3.03</td>
<td>0.65</td>
</tr>
<tr>
<td>Second Year GPA</td>
<td>2,017</td>
<td>1</td>
<td>4</td>
<td>3.08</td>
<td>0.55</td>
</tr>
<tr>
<td>Degree completion length</td>
<td>2,017</td>
<td>1.5</td>
<td>10.75</td>
<td>4.10</td>
<td>1.07</td>
</tr>
</tbody>
</table>

4.6 Results of Hierarchical Binary Logistic Regression

Binary logistic regression was conducted to determine whether the studied
independent variables of STEM classification, gender, minority, first-generation, first
year GPA, second year GPA, high school rank, and SAT score predicted whether the
students completed a bachelor’s degree. In addition, the two-way interaction effects of
each independent predictor variable were also included in the regression model to
determine if the relationships between the independent variables and the dependent
variable were significantly different between the different levels of the predictor
variables. This model identifies all the two-way interactions that are significant to
completion of a STEM degree.
The binary logistic regression used a backwards model selection. To better explore statistical effects of variables that contributed a significant amount of variation to whether or not students completed their bachelor’s degree, a hierarchical model was created. The independent variables of STEM classification, gender, minority status, first-generation status, first year GPA, second year GPA, high school rank, and SAT score, each two-way interaction term, and each three-way interaction term with STEM status, were included. This process ensured determining the effects of the independent variables and the interaction terms on whether the students were able to complete a bachelor’s degree. A level of significance of 0.05 was used in the binary logistic regression analysis. There is a significant effect by any of the independent variables and interaction terms if the $p$-value is less than or equal to the level of significance value of 0.05. The final result of the backward approach was Step 46. This was the final iteration of the binary logistic regression model and contains only variables that contribute a significant amount of independent explained variance to the model.

4.7 Overall Model Description

Of the 37,739 participants, 2,422 cases were observed in the final model. It is necessary to have all the values (gender, race, first-generation status, SAT scores, high school rank, and first and second year GPA) in place for each student for this model to predict accurately the outcome. Of all predictors initially included in the model, STEM status, minority status, and second year GPA remained in the final iteration as significant predictors of the likelihood of finishing a bachelor’s degree. Of all two-way interactions included in the initial model, STEM by SAT score, gender by first-generation
status, gender by SAT score, and minority status by first year GPA were significant in the final iteration. There were also significant three-way interactions between minority status, first year GPA and STEM status; minority status, SAT score and STEM status; first-generation status, SAT score and STEM status; first year GPA, SAT score and STEM status; and also second year GPA, SAT score, and STEM status. This logistic regression model was statistically significant \((X^2(12) = 400.012, p< 0.00001)\). The classification values for the binary logistic regression are presented below in Table 4.9.

Table 4.10

*Classification Table of Binary Logistic Regression, Step 46*

<table>
<thead>
<tr>
<th>Observed</th>
<th>Predicted</th>
<th>% Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1 (Did not complete degree)</td>
<td>96</td>
<td>309</td>
</tr>
<tr>
<td>1 (Did complete degree)</td>
<td>50</td>
<td>1967</td>
</tr>
<tr>
<td>Overall Percentage</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* The cut value is .500, \(N=2,422\).

The model explained 25.6\% (Nagelkerke\(R^2\)) of the variance in bachelor’s degree completion and correctly classified 85.2\% of cases. Twenty-three point seven percent of the 405 participants who did not complete their bachelor’s degree were correctly classified, and 97.5\% of the 2,017 participants who did complete their bachelor’s degree were correctly classified.

4.8 Significant Predictors in the Model

The next analysis involved the beta coefficients and odds ratios to determine the individual effects of STEM status, minority status, second year GPA, STEM by first year
GPA, STEM by SAT score, gender by first-generation status, gender by SAT score, first year GPA by minority status, minority status by first year GPA by STEM status, minority status by SAT score by STEM status, first-generation status by SAT score by STEM status, first year GPA by SAT score by STEM status, and second year GPA by SAT score by STEM status in the prediction of whether the students completed a bachelor’s degree. The last iteration of the backward approach in the binary logistic regression is Step 46. The results for each variable in the final logistic regression equation are presented below in Table 4.10.

Table 4.11

*Binary Logistic Regression Results of Individualized Effects of Variables*

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>STEM or not</td>
<td>-1.858</td>
<td>0.561</td>
<td>10.96</td>
<td>1</td>
<td>.001</td>
<td>0.16</td>
</tr>
<tr>
<td>Minority or not</td>
<td>0.99</td>
<td>0.335</td>
<td>8.76</td>
<td>1</td>
<td>.003</td>
<td>2.69</td>
</tr>
<tr>
<td>Second Year GPA</td>
<td>1.72</td>
<td>0.152</td>
<td>128.41</td>
<td>1</td>
<td>&lt;.001</td>
<td>5.61</td>
</tr>
<tr>
<td>STEM x SAT</td>
<td>0.002</td>
<td>0.001</td>
<td>5.31</td>
<td>1</td>
<td>.021</td>
<td>1.00</td>
</tr>
<tr>
<td>Gender x 1st Generation</td>
<td>-0.136</td>
<td>0.063</td>
<td>4.69</td>
<td>1</td>
<td>.030</td>
<td>0.87</td>
</tr>
<tr>
<td>Gender x SAT</td>
<td>-0.0001</td>
<td>0.00006</td>
<td>4.76</td>
<td>1</td>
<td>.029</td>
<td>1.00</td>
</tr>
<tr>
<td>Minority x 1GPA</td>
<td>-0.33</td>
<td>0.122</td>
<td>7.21</td>
<td>1</td>
<td>.007</td>
<td>0.72</td>
</tr>
<tr>
<td>Minority x 1GPA x STEM</td>
<td>-0.47</td>
<td>0.128</td>
<td>13.32</td>
<td>1</td>
<td>&lt;.001</td>
<td>0.63</td>
</tr>
<tr>
<td>Minority x SAT x STEM</td>
<td>0.001</td>
<td>0.0003</td>
<td>12.75</td>
<td>1</td>
<td>&lt;.001</td>
<td>1.00</td>
</tr>
<tr>
<td>1stGeneration x SAT x STEM</td>
<td>0.0001</td>
<td>0.00006</td>
<td>4.08</td>
<td>1</td>
<td>.044</td>
<td>1.00</td>
</tr>
<tr>
<td>1GPA x SAT x STEM</td>
<td>0.0003</td>
<td>0.0002</td>
<td>4.61</td>
<td>1</td>
<td>.032</td>
<td>1.00</td>
</tr>
<tr>
<td>2GPA x SAT x STEM</td>
<td>-0.0004</td>
<td>0.0002</td>
<td>4.50</td>
<td>1</td>
<td>.034</td>
<td>1.00</td>
</tr>
<tr>
<td>Constant</td>
<td>-3.452</td>
<td>0.423</td>
<td>66.75</td>
<td>1</td>
<td>&lt;.001</td>
<td>.032</td>
</tr>
</tbody>
</table>

*Note.* Significant at level of significance of 0.05, N=2,422.

The final model determined that there were 12 predictors that affected the dependent variable. The main effect of STEM status was significantly related to degree
completion, controlling for all other imputed predictors (Wald(1) = 10.956, \( p = .001 \)); the odds of completing a bachelor’s degree were 0.16 times less likely for STEM students compared to non-STEM students in this sample. The main effect of minority status was significantly related to degree completion, controlling for all other imputed variables (Wald (1) = 8.760, \( p = .003 \)). The odds of completing a bachelor’s degree were 2.69 times more likely for non-minority students compared to minority students. As previously discussed, students who pursued a STEM degree were significantly less likely to complete their bachelor’s degree than those in non-STEM fields. Additionally, the effect of the interaction between SAT score and STEM on likelihood of obtaining a bachelor’s degree was significant, controlling for all other variables imputed in the model (Wald(1) = 5.312, \( p = .021 \)). In general, higher SAT scores increased the likelihood of obtaining a degree; this effect was significantly greater for those pursuing STEM degrees compared to those pursuing a non-STEM degree (Figure 4.7).

The three-way interaction effect of minority status, SAT score and STEM status was statistically significant in this sample, controlling for all other variables imputed in the model (Wald(1) = 12.747, \( p = .0004 \)). In general, higher SAT scores resulted in an increased likelihood of obtaining a degree; this effect was generally stronger for those pursuing a STEM degree. This interaction effect was significant for minority students such that the effect of SAT score on degree completion was largest for minority students pursuing a STEM degree, and weakest for minority students pursuing a non-STEM degree. Average SAT scores for STEM, non-STEM and minority and non-minority students can be observed in Table 4.11. For a visual representation of this interaction, see Figures 4.8 and 4.9.
**Figure 4.7**. Average SAT score for STEM and non-STEM students.

**Figure 4.8**. Average SAT score for minority and non-minority non-STEM students.

**Figure 4.9**. Average SAT score for minority and non-minority STEM students.
A three-way interaction effect of first-generation status, SAT score, and STEM status was statistically significant in this sample, controlling for all other variable imputed in the model (Wald(1) = 4.075, $p = .044$). In general, higher SAT scores resulted in an increased likelihood of obtaining a degree; this effect was generally stronger for those pursuing a STEM degree. This interaction effect was significantly stronger for first-generation students such that the effect of SAT score on degree completion was largest for first generation students pursuing a STEM degree, and weakest for non-first generation students pursuing a non-STEM degree. Average for SAT scores for STEM, non-STEM, and first-generation and non-first-generation students can be observed in Table 4.12.

Table 4.13

SAT Average by Degree Completion, First and Non-First-generation, and STEM and Non-STEM Students
In the final iteration of the hierarchical logistic regression, a significant main effect of first year GPA on likelihood of completing a bachelor's degree was observed, controlling for all other variables imputed in the model (Wald(1) = 128.411, \( p < .000001 \)). As first year GPA increased by one point, the odds of obtaining a bachelor's degree increased by 5.61. A significant three-way interaction between minority status, first year GPA, and STEM status was also observed in this sample, controlling for all other variables imputed in the model (Wald(1) = 13.323, \( p = .0002 \). In general, first year GPA
was associated with a higher probability of completing a bachelor’s degree. This effect was strongest for non-minority students pursuing a STEM degree. For the average first year GPA for those who obtained a degree versus those who did not, for minority, non-minority, STEM, and non-STEM students, see Table 4.12.

![Average First Year GPA for Non-STEM Students](image1)

**Figure 4.12.** Average first year GPA for minority and non-minority non-STEM students.

![Average First Year GPA for STEM Students](image2)

**Figure 4.13.** Average first year GPA for minority and non-minority STEM students.

There was a significant three-way interaction between first year GPA, SAT score, and STEM status, controlling for all other variables imputed in the model (Wald(1) = 4.611, \( p = .032 \)). In general, higher SAT scores increased the likelihood of obtaining a degree; for non-STEM students, this probability decreased slightly as first year GPA
increased, and for STEM students this probability increased as first year GPA increased. Additionally, there was a significant three-way interaction between second year GPA, SAT score, and STEM status, controlling for all other variables imputed in the model (Wald(1) = 4.501, \( p = .034 \)). In general, higher SAT scores increased the likelihood of obtaining a degree, and this probability increased as second year GPA increased. This effect is stronger for STEM students than for non-STEM students.

Table 4.14

*First Year GPA Average by Degree Completion for Minority and Non-Minority, and STEM and Non-STEM Students*

<table>
<thead>
<tr>
<th></th>
<th>STEM</th>
<th>Non-STEM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minority</td>
<td>Non Minority</td>
</tr>
<tr>
<td>Mean</td>
<td>3.00</td>
<td>3.11</td>
</tr>
<tr>
<td>Did not complete degree</td>
<td>2.56</td>
<td>2.41</td>
</tr>
</tbody>
</table>

4.9 Results of Hierarchical Linear Regression

A linear regression was conducted to determine whether the different independent variables of STEM classification, gender, minority, first-generation, first year GPA, second year GPA, high school rank, and SAT score predicted the length of time (in years) it took students to complete their bachelor’s degree. In addition, the two-way interaction effects of each independent predictor variable were also included in the regression model to determine if the relationships between the independent variables and the dependent variable were significantly different between the different levels of the predictor variables. As with the previous logistic regression, the linear regression used a backwards model selection and an alpha of .05. The final iteration of the linear
regression model contains only variables that contribute a significant amount of independent explained variance to the model.

4.10 Overall Model Description

Of the 2,422 complete cases, 2,017 students obtained their bachelor’s degree and were included in the linear regression model. Of all predictors initially included in the model, gender, second year GPA, high school rank, and SAT score remained in the final iteration as significant predictors of length in time in years taken to complete bachelor’s degree. Of all two-way interactions included in the initial model, STEM by gender, STEM by first year GPA, minority status by first year GPA, first year GPA by second year GPA, first year GPA by SAT score, and high school rank by SAT score were significant in the final iteration. There were also significant three-way interactions between gender, first year GPA and STEM status; gender, high school rank, and STEM status; first year GPA, SAT score and STEM status; second year GPA, high school rank, and STEM status; second year GPA, SAT score, and STEM status; and high school rank, SAT score, and STEM status. This linear regression model was statistically significant ($F(18, 1998) = 30.932, p< 0.00001)$. The model explained 21.8% ($R^2$) of the variance in length of time to complete a bachelor’s degree.

4.11 Significant Predictors in the model

There was a statistically significant main effect of gender on the length of time taken to complete a bachelor’s degree, controlling for all other variables imputed in the model ($t(1) = 5.476, p <.001$). Female students ($M = 3.93, SD = 0.97$) completed their
degree in significantly less time than male students ($M = 4.37$, $SD = 1.17$). There was a statistically significant main effect of second year GPA on years taken to complete a bachelor’s degree, controlling for all other variables imputed in the model. For every one point increase in GPA, length of time to complete a bachelor’s degree decreased by 1.68 years on average. The results for each variable in the final linear regression equation are presented in Table 4.14.

There was also a significant relationship between high school rank and length of time taken to complete a bachelor’s degree, controlling for all other variables imputed in the model ($t(1) = -2.71$, $p = .007$). As high school rank increased, years taken to obtain a degree significantly increased. For descriptive statistics related to length of time to complete degree for each high school rank, see Table 4.15.

Table 4.15

*Linear Regression Results of Individualized Effects of Variables*

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E.</th>
<th>T</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>0.198</td>
<td>0.036</td>
<td>5.476</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Second Year GPA</td>
<td>-1.679</td>
<td>0.238</td>
<td>-7.066</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>High school rank</td>
<td>-0.665</td>
<td>0.246</td>
<td>-2.708</td>
<td>.007</td>
</tr>
<tr>
<td>SAT score</td>
<td>0.002</td>
<td>0.001</td>
<td>3.098</td>
<td>.002</td>
</tr>
<tr>
<td>STEM x Gender</td>
<td>0.457</td>
<td>0.146</td>
<td>3.131</td>
<td>.002</td>
</tr>
<tr>
<td>STEM x 1GPA</td>
<td>0.569</td>
<td>0.162</td>
<td>3.511</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>STEM x HSrank</td>
<td>-0.396</td>
<td>0.216</td>
<td>-1.833</td>
<td>.067</td>
</tr>
<tr>
<td>Minority x 1GPA</td>
<td>-0.146</td>
<td>0.073</td>
<td>-1.991</td>
<td>.047</td>
</tr>
<tr>
<td>Minority x 2GPA</td>
<td>0.134</td>
<td>0.072</td>
<td>1.844</td>
<td>.065</td>
</tr>
<tr>
<td>1GPA x 2GPA</td>
<td>0.378</td>
<td>0.061</td>
<td>6.244</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>1GPA x SAT</td>
<td>-0.001</td>
<td>0.0002</td>
<td>-6.551</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>HSrank x SAT</td>
<td>0.001</td>
<td>0.0002</td>
<td>2.890</td>
<td>.004</td>
</tr>
</tbody>
</table>

*(table continues)*
<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E.</th>
<th>T</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender x GPA x STEM</td>
<td>-0.089</td>
<td>0.037</td>
<td>-2.423</td>
<td>.015</td>
</tr>
<tr>
<td>Gender x HSrank x STEM</td>
<td>-0.065</td>
<td>0.024</td>
<td>-2.677</td>
<td>.007</td>
</tr>
<tr>
<td>1GPA x SAT x STEM</td>
<td>-0.001</td>
<td>0.0002</td>
<td>-3.759</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Constant</td>
<td>7.414</td>
<td>0.657</td>
<td>11.29</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

*Note.* Significant at level of significance of 0.05, N=2,017.

Table 4.16

*Length of Time to Degree (Years) Completion Descriptives for Each High School Ranking*

<table>
<thead>
<tr>
<th>HS Rank</th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (top 10%)</td>
<td>395</td>
<td>3.75</td>
<td>0.97</td>
</tr>
<tr>
<td>2 (11-25%)</td>
<td>681</td>
<td>4.05</td>
<td>1.03</td>
</tr>
<tr>
<td>3 (2nd quarter)</td>
<td>746</td>
<td>4.28</td>
<td>1.09</td>
</tr>
<tr>
<td>4 (3rd quarter)</td>
<td>157</td>
<td>4.35</td>
<td>1.12</td>
</tr>
<tr>
<td>5 (4th quarter)</td>
<td>9</td>
<td>5.17</td>
<td>1.66</td>
</tr>
<tr>
<td>6 (no rank)</td>
<td>29</td>
<td>3.88</td>
<td>1.04</td>
</tr>
</tbody>
</table>

There was a significant two-way interaction effect between gender and STEM status on length of time to complete a degree \(t(1) = 3.131, p = .002\). In general, female students took less time to complete their degree than male students, and this effect was significantly larger for students completing a STEM degree. For a visual representation of this interaction, see Figure 4.14.
There was a statistically significant three-way interaction effect between gender, first year GPA and STEM status on length of time to complete a bachelor’s degree, controlling for all other variables imputed in the model ($t(1) = -2.42, p = .015$). In general, higher first year GPA was associated with a shorter length of time to complete a degree. This effect was pronounced for female students pursuing a STEM degree (see Figures 4.15 and 4.16).

There was also a statistically significant three-way interaction between gender, high school rank, and STEM status, controlling for all other variables imputed in the model ($t(1) = 2.677, p = .007$). In general, higher high school rank was associated with a longer length of time to complete a degree. This effect was pronounced for female students pursuing a STEM degree, and male students pursuing a non-STEM degree. For a visual representation of this interaction, see Figures 17 and 18.
Figure 4.15. Number of years taken to obtain degree for male and female non-STEM students.

Figure 4.16. Number of years taken to obtain degree for male and female STEM students.

Figure 4.17. Number of years taken to obtain degree for male and female non-STEM students.
A significant main effect of SAT score on length of time to complete a degree was observed ($t(1) = 3.10, p = .002$), controlling for all other variables imputed into the model. A single point increase in SAT score was actually associated with a 0.002 years increase in length of time taken to obtain a degree, on average (Figure 4.19).

There was a significant two-way interaction between first year GPA and STEM status on length of time to complete degree, controlling for all other variables in the model. In general, a higher GPA was associated with a shorter length of time to
complete a bachelor’s degree, and this effect was stronger for students pursuing a STEM degree (Figure 4.20).

There was a significant three-way interaction effect of first year GPA, SAT score, and STEM status on length of time take to complete a degree, controlling for all other variables imputed in the model ($t(1) = -3.76, p < .001$). As SAT score increased, length of time taken to complete a degree increased. This effect was weaker as first year GPA increased, and this interaction effect was weaker for students pursuing non-STEM degrees. For a visual representation of this three-way interaction, see Figures 4.21 and 4.22.

![Figure 4.20. Number of years taken to obtain degree for non-STEM and STEM students.](image)

![Figure 4.21. Average SAT score by number of years taken to complete degree (by quartile) and GPA for non-STEM students.](image)
Figure 4.22. Average SAT score by number of years taken to complete degree (by quartile) and GPA for STEM students.
CHAPTER 5
CONCLUSION

5.1 Summary

This study, which took place in a Research One university in Denton, TX, aims to identify trends in completion rates of STEM bachelor degrees for underrepresented groups. Generally, in STEM fields, the underrepresented groups include women, African American, Hispanic, Native American, and first-generation students. This study used Vincent Tinto's model of student retention as a basis to determine factors that have been found to affect student retention or completion rates in STEM fields (Engle & Tinto, 2008; Fuchsetal, 2004; Prentice & Carranza, 2002; Scott et al., 2009; Thompson & Bolin, 2011; Dika & Alvarez, 2016). Most retention studies agree that the following aspects are the biggest determinants of persistence:

- Individual attributes: They include gender, race, and first-generation status. These are the attributes that are part of a student’s persona—one that they carry and identify with throughout their lives, and which they bring with them to the college environment. Multiple studies indicate that these attributes play an extremely important role in determining whether the student gets to complete his/her education (Cundiff et al., 2013; Murphy et al. 2007; Sekaquaptewa & Thompson 2002; Steele & Aronson 1995; Woolley et al., 2010).

- Pre-college schooling: The attributes that reflect the student’s academic performance prior to entering college are also of great importance in determining or predicting their course of study at the university level. The pre-college attributes
included in this study were SAT scores and high school rank (Scott et al., 2009; Thompson & Bolin, 2011).

- Grade performance: Once the students are in the college setting, a multitude of factors would help to determine completion rates in the field of STEM (Lessie & Brown, 1996). First and second year college GPA have been known to be excellent predictors of completion rates. The academic experiences of the students in the early years of their bachelor’s degree goes a long way in determining whether or not they stay in college, and this has been especially been found to be true in STEM fields (Cabrera et al., 1993; Fischbach, 1990; Mohammadi, 1994; Pascarella & Chapman, 1983; Wall et al., 1996).

Data for the study were collected from Institutional Effectiveness and Research (IE&R), which tracks student enrollment and graduation in all fields. Data were collected from the years 2005 through 2008, and these students were tracked all the way through graduation. Since research has shown that it takes an average of 6 years for a student to finish their undergraduate degree, data were collected from 2005 through 2013. In reviewing the chart, it could be determined that, of the 37,739 students included in the study, more than half of them had completed their bachelor’s degree (58.2%), which were 21,965 students across all fields. Regardless, considering just the STEM fields, only 16.3% (6,156 students) had actually completed their degree in their respective fields. The rest of the students had either switched to another field or dropped out of the university. Considering the independent variables, for gender, more than half of the students (58.3% or 20,872) in the sample were women. As far as ethnicity was concerned, less than half of the students (36.7%) in the sample belonged to ethnic
minority groups, which results in 13,854 students. When first-generation status of the students was taken into consideration, less than half (32.4% or 12,236) students belong to this category.

This study answered a two-part question about completion rates in STEM fields at a Research One university. The first part was to determine the factors contributing to the completion rate. The factors include gender, race, first-generation status, SAT scores, high school rank, and first and second year college GPA. This is a comparative study, which measured how STEM fields fare at retention or student completion rates as compared to the non-STEM fields. The second part explored the time taken for STEM majors to complete the degree.

The models used binary logistic regression for the questions regarding degree completion status and linear regression for the question on time to completion. The first model used a forward logistic regression model to determine the relative significance of the independent variables with the dependent variable of STEM completion. This means that each independent variable was added to the model one after the other, in order to determine significance for that particular variable. The second part of the model addressed the effect STEM status had on the time taken to complete the degree. For this part of the study, the sample size was considerably different since the focus was on the students who completed their degree. Previous research has shown that the results are highly dependent on the type of university (Flores & Park, 2015).

5.2 STEM Completion: Main Statistical Effects and One-Way Interactions

There were three steps by which this analysis was conducted. Statistical Main
Effects, One-way interactions and Two-way interactions between variables. Main statistical effects for each individual variable were conducted which included all STEM and non-STEM students. The results indicated that, all the variables (Gender, Minority Status, First generation status, SAT Scores, High school Rank, First and Second year GPA) were all found to be statistically significant when main effects were studied. This shows that each of these variables does tend to affect the student population as a whole, when they are not segregated into STEM and non-STEM students.

The second part of the study was to determine all the one-way relationships between STEM status and the independent variables. This told us which of these factors were statistically significant when they were taken alone and without the influence of another variable. As far as STEM completion was concerned, Gender, SAT scores and First year GPA were found to be statistically significant when it came to STEM completion. As far as Gender is concerned, this tells us that women were more likely to complete their degree than men, and this effect was significantly stronger for non-STEM students compared to STEM students. Women tend to perform a lot better than men in non-STEM fields do, but they seem to fall short when it comes to the STEM fields. This has been consistent with research conducted at other institutions or with national studies, which essentially speak along similar lines. When it comes to SAT scores, they were higher for those who completed their degree in STEM fields as compared to non-STEM fields. This tells us that those with higher SAT scores tend to finish with a STEM degree and that their likelihood of completing their degree is higher than if they had a lower SAT score. When it comes to First year GPA, the likelihood of completing a degree was higher for students with a better first year GPA, and this effect
Better the First-year GPA, the more likely it was for the students who were pursuing a STEM degree to finish their degree. This aligns with previous studies, which say that students whose first year GPA are higher than the students who have a lower GPA, tend to persist on in the STEM field. Once the first year GPA goes down, it is more likely that the students with lower GPA gravitate towards non-STEM fields or would drop out altogether.

5.3 Time Taken to Complete Degree: Main Statistical Effects and One-Way Interactions

The results for the main statistical effects for time to complete degree showed that all the variables with the exception of Minority status were found to be significant. Previous research has indicated that race tends to be a very selective variable, as it is highly dependent on the type of institution where the study was conducted. For instance, race would play out differently in a selective institution (such as an HBCU), as opposed to a PWI (Predominantly White Institution), such as the one where this study was conducted. All the other variables (Gender, First generation status, SAT scores, High School Rank, First and Second year GPA) all have statistical main effects which are aligned with previous research conducted.

As for the one-way interactions between time taken to complete STEM degree and the other variables, only first and second year GPA were found to be statistically significant. Higher first year GPA was directly related to shorter degree length, and this effect was significantly stronger for STEM students compared to non-STEM students. This is consistent with research that has shown that First year GPA acts as a deciding
factor when it comes to STEM degrees. It often is thought of as a ‘morale-booster’ for STEM students, when their First year GPA is high, it motivates them to finish quicker. Second year GPA was also found to be statistically significant. A higher second year GPA was related to shorter degree length and this effect was significantly stronger for STEM students as compared to non-STEM students.

5.4 STEM Completion: Two-Way Interactions

Two-way interactions were determined by performing backwards logistic regression analysis where, all the variables were in the model and were removed one by one with every step in order to determine the ones that fit the model very closely. The two-way interactions which were found to be statistically significant with STEM Completion are Minority Status with SAT Scores, Minority Status with First year GPA and First generation status with SAT Scores.

While taking Minority status with SAT Scores, in general it can be interpreted that higher SAT scores resulted in an increased likelihood of obtaining a degree and this effect was generally stronger for pursuing a STEM degree. This effect was significantly stronger for Minority students such that the effect of SAT Scores on degree completion was largest for Minority students pursuing a STEM degree and weakest for Minority students pursuing a non-STEM degree. This gives an indication of the fact that Minority students who came in with higher SAT Scores were more likely to be successful in a college setting while studying STEM. This way, the focus could be on admitting or enrolling Minority students who have higher SAT scores. This would be a short-term approach. The long-term approach could be to equip middle and high school Minority
students with better resources so that their SAT scores could be vastly improved. Minority status with First year GPA was also found to be statistically significant. In general, first year GPA was associated with a higher probability of completing a bachelor’s degree. This effect was strongest for Minority students pursuing a STEM degree. With the knowledge that first year GPA plays an important role in STEM completion, steps could be taken to ensure that resources are in place for Minority students once they are in the college setting in order to improve their chances of getting a higher first year GPA. First generation status with SAT Scores have been to be statistically significant. In general, higher SAT Scores increased the probability of obtaining a degree and this effect was generally stronger for those pursuing a STEM degree. This interaction was significantly stronger for First generation students such that the effect of SAT Scores on degree completion was largest for First generation students pursuing a STEM degree and weakest for First generation students pursuing a non-STEM degree.

5.5 Time to Complete Degree: Two-Way Interactions

There were two interactions, which were statistically significant when it came to time to complete STEM degree. Gender with High School Rank and Gender with First year GPA were found to be significant predictors when it came to time to complete STEM degree.

As for Gender and High School Rank, in general higher High School Rank is associated with a shorter length of time to complete a degree. This effect is more pronounced for female students pursuing a STEM degree, and male students pursuing
a non-STEM degree. Since female students are underrepresented, and need to have more representation especially in the physical sciences, one strategy would be to admit those female students who have a higher High School Rank. Alternatively, a more efficient and long-term strategy would be to tackle the problem at its root and introduce programs at the middle and high school level would increase their High School Rank, thus leading them to be more successful in completing their STEM degree faster. The other two-way interaction that was found to be statistically significant was Gender with First year GPA. In general, higher First year GPA was associated with a shorter length of time to complete a degree. This effect was more pronounced for female students pursuing a STEM degree. Since higher First-year GPA would get female students to finish their degree faster, intervention or supplemental programs could be put in place to ensure maximum amount of success for female students in STEM fields.

5.6 Limitations

This study was conducted in a predominantly White institution (PWI), where enrollment numbers generally tend to be skewed towards Caucasians. The scenario would be different in a Historically Black College or University (HBCU) or a Hispanic Serving Institution (HSI), where enrollment numbers would tend to be skewed towards African Americans and Hispanics respectively. Geographical location of the institution where the study was conducted tends to play an important role as well. This is a university located in the suburbs of a large metroplex, which has seen an increased intake of commuter students as compared to a typical university with a college town
setting. Another limitation is that the data about first-generation status of students is largely self-reported, and so it may not be necessarily accurate.

5.7 Recommendations for Future Research

Science Education has come a long way since the era of the Great Depression, where classrooms, especially in STEM fields, were mostly filled with White, male students. Women have made tremendous progress since 1920, since they were given the right to vote (Watkins, 2016). The same could be said about ethnic minority groups since the passage of the Civil Rights Act in 1964. Regardless, this has been hardly enough. There has been study after study in the last few decades that have highlighted the problems we face because of underrepresentation of the above mentioned groups, including first-generation students who are the first ones in their family to attend college. The world of science neglects talent and tremendous potential simply because of policies and procedures in place that make it difficult for them to succeed once they already are in the college environment (Savaria et al., 2017). Enrollment is not the issue in the present day, but rather it is a combination of factors that come into play once they enter college that prevent some from succeeding (Kokkelenberg & Sinha, 2010).

As noted in the study, SAT scores, high school rank, and first and second year GPA were some of the most prominent predictors of completion rates in college for the underrepresented groups. This study used Vincent Tinto’s model of Student Retention as base for understanding the factors that contribute to degree completion rate and length of study in STEM fields among underrepresented groups of students. The three broad classifications, namely Individual characteristics (gender, race and first-
generation status), pre-college schooling (SAT scores and high school rank), and grade performances (first and second year GPA) of underrepresented groups of students were analyzed. A potential future study that would serve as an extension to this study would involve looking into the social aspects of the underrepresented students’ institutional experiences.

There were variables that have been found to have a profound impact on how successful the underrepresented student groups were at the college level. The institution can best make use of this in order to further strengthen the prospects of STEM completion and shorten the time taken to complete STEM degrees. Since first and second year GPA had been found to be strong predictors, intervention programs could be put in place in order to ensure that the students’ GPA remains high. This proactive measure is better than a remedial one, and it would ensure that chances of the students doing well in the first place remain high. Remedial programs such as foundation courses, tutoring and mentoring, study groups etc. have been found to be successful at various institutions in the past (Goonewardene et al., 2016).
REFERENCES


113


National Science Foundation. (2016). *The United States lags in the percentage of university students earning science and engineering degrees.* https://www.pgpf.org/chart-archive/0212_science_degrees


Prentice, D. A., & Carranza, E. (2002). What women and men should be, shouldn’t be, are allowed to be, and don’t have to be: The contents of prescriptive gender stereotypes. *Psychology of Women Quarterly, 26*(4), 269-281.


