12-2016

The Numbers Game: What an Un-Predictive Ability of Success in First-Year Mathematics Courses and Subgroup Bias Mean to Students at Hispanic-Serving Institutions

Jayme Gonzales Agozie
University of the Incarnate Word, jgonza30@uiwtx.edu

THE NUMBERS GAME: WHAT AN UN-PREDICTIVE ABILITY OF SUCCESS IN FIRST-YEAR MATHEMATICS COURSES AND SUBGROUP BIAS MEAN TO STUDENTS AT HISPANIC-SERVING INSTITUTIONS

by

JAYME GONZALES AGOZIE

A DISSERTATION

Presented to the Faculty of the University of the Incarnate Word
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF THE INCARNATE WORD

May 2016

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my Lord and Savior for blessing me with the courage, tenacity, and patience to get through the past few years. I would like to thank my mother, father, and husband for their constant encouragement, tolerance, and willingness to put up with me through all of this. To my amazing daughter, Elisheva…Mommy has been far too busy taking care of this, which has taken time away from you. I promise I will make up the last few months to you – soon we will have nothing but nosey nosey, patty cake, monster face, and play time. To my committee members, Dr. Glenn E. James, Dr. Joleen Beltrami, and Dr. Elizabeth Kreston, I could not have completed this process without your guidance, support, and confidence in my ability to complete what seemed like such a monumental task – thank you, from the bottom of my heart. To my editor, Joyce Deuley, thank you for doing such an amazing job and allowing me to pay you in the form of friendship, food, laughter, pool, movie nights, and bowling – you have been an amazing friend.

DEDICATION

For David, Dot, and Elisheva.

THE NUMBERS GAME: WHAT AN UN-PREDICTIVE ABILITY OF SUCCESS IN FIRST-YEAR MATHEMATICS COURSES AND SUBGROUP BIAS MEAN TO STUDENTS AT HISPANIC-SERVING INSTITUTIONS

Jayme Gonzales Agozie, PhD

University of the Incarnate Word, 2016

The purpose of this quantitative study was to examine the ability of standardized test scores to predict the performance of students in first-year mathematics courses, and the extent to which these tests displayed differential validity among various subgroups. Using discriminant analysis, it was established that the following percentages of students were correctly classified into passing and not passing groups, using the listed independent variables: a) SAT mathematics scores - 58.6%, b) SAT verbal scores - 50.6%, c) ACT mathematics scores - 54.7%, and d) ACT verbal scores - 56.3%. New predictive models were created using standardized test scores in combination with students high school GPA, and high school rank to increase correct classification into passing and not passing groups to 67.1% using the SAT, 67.4% using the ACT, and 68.5% using a combination of the SAT and ACT.

Using three-way ANOVA, it was determined that there was a significant three-way interaction between gender, ethnicity, and socioeconomic status (income), and a significant two-way interaction between gender and socioeconomic status (income) for both the SAT and ACT. An analysis of main effects determined that ethnicity and socioeconomic status (income) displayed statistically significant differences in the mean scores of students on the SAT and ACT.

TABLE OF CONTENTS

Table of Contents—Continued

Table of Contents—Continued

Table of Contents—Continued

LIST OF TABLES

List of Tables—Continued

List of Tables—Continued

LIST OF FIGURES

List of Figures—Continued

Chapter 1: Standardized Tests

Introduction to the Problem

For decades, standardized test scores have been used as one of the top three factors considered for a student's admission and placement in higher education institutions. According to Breland, Maxey, Gernand, Cumming, and Trapani (2002), there was a steady increase from 50% to 67% of students taking these exams for admission purposes between the years 1979 and 2000. Now, over 90% of four-year colleges require either the Scholastic Aptitude Test (SAT) or the American College Test (ACT) for admission (Breland et al., 2002). Researchers responsible for analyzing data for The State of College Admissions, a report based on the National Association for College Admission Counseling's (NACAC) Admission Trends Survey of Colleges, stated "the importance assigned to test scores increased with college size, with 55 percent of institutions with less than 3,000 students attributing considerable importance to tests and 92 percent of institutions with 20,000 or more students doing so" (Hawkins & Lautz, 2005, p. 42).

Despite reforms in education at the kindergarten through 12th grade (K-12) levels and an increased use of standardized tests in admission and placement processes, "long-standing concerns with standardized tests have persisted, and the role of the ACT and SAT in determining who gains entry into the nation's colleges and universities continues to be a hotly debated topic" (Zwick, 2007, pg. 2). In his book, *Choosing Elites* (1985), Harvard professor Robert Klitgaard posed the thought provoking question of selectivity,

> The first question to ask about selective admissions is why it should be selective at all? It has unpleasant connotations of elitism, unfairness, snobbishness, and uniformity. On the other hand, we laud excellence, recognize its scarcity and utility, and endorse admissions on the basis of merit. (p. 51)

The use of standardized tests for admissions purposes have not only brought about a sense of elitism and unnecessary selectivity, but has also proven to be biased on the basis of race and ethnicity, gender, and socioeconomic status (Breland, 1998; Breland et al., 2002; Crouse & Trusheim, 1988; Geiser & Santelices, 2007; Lohman, 2004; Zwick, 2007, 2012; Zwick & Himelfarb, 2011). Research has also shown an increase in scoring errors on the SAT, leading many to question the validity and role of standardized tests in the admission and placement processes. This has prompted a handful of higher education institutions to adopt "test-optional" admission policies, (Geiser & Santelices, 2007; Zwick, 2007).

In his writings, William Shakespeare once asked, "What's in a name?" In a nutshell, the question to ask regarding the use of standardized tests is, "What's in a number?" Although the intentions of many institutions may be to use standardized tests to identify students who are most capable of handling the academic rigor of higher education, the question most often asked is what do the scores of these tests actually tell us about a student? Can they accurately predict student success in a specific course, success in their first-year, and/or graduation rates? This brings about another question: are standardized tests – and their use for admission and/or placement procedures – biased against certain subgroups of a population? If one score cannot answer any of these questions, or if standardized tests are biased, then why are they continually used to determine a student's fate in higher education?

Statement of the Problem

Many institutions rely heavily on cognitive assessments for admissions decisions and placement of students into appropriate courses. Yet, research has shown that the capability and ability of a student to successfully complete coursework should not be determined by one cut score (Geiser & Santelices, 2007; Geiser & Studley, 2002). Based on one score alone, students

can be misplaced into courses, which can lead to a number of challenges and frustrations. For instance, students may find themselves repeating courses for which they were incorrectly labeled ready, while others waste time and money on developmental course sequences they did not need, impacting institutional resources for student success and remediation. As a result, one of several scenarios may happen: a) a student may continually drop a course and lengthen their time of study; b) a student may run out of financial resources to fund their education; or c) a student may give up on their education all together. These scenarios have an impact on both students and institutions, most specifically on funding and resources, retention, and matriculation rates. These consequences might be reduced with a more accurate approach to predicting student success and placement, such as the analysis and use of a student's high school performance for placement, developing new cut scores for placement, and implementing the use of a departmental, content specific exam for placement.

Purpose of the Study

The purpose of this quantitative study was to examine the predictive ability of standardized test scores on the performance of students in their first mathematics course, and the extent to which these tests displayed differential validity among various subgroups at a specific private institution with an interesting mission and demographics. According to its mission statement, the institution aims to provide higher education opportunities to populations underserved by higher education. The total global enrollment of the institution is currently 10,984 students, which includes main campus, satellite campuses for extended studies, virtual university, and international campuses. Segmenting this number by degree category and enrollment, there are 8,713 undergraduate students, 1,213 master's students, 189 doctoral students, and the remaining 869 are professional students. The ethnic and racial breakdown of

the student population is uniquely diverse: 56.45% Hispanic, 12.82% White, 7.91% Other Minority/Unknown, 1.72% Nonresident Alien, and 21.1% International. The preliminary data set received from the institution contained information for 6,642 students; a target sample from this population yielded 1,037 students.

Rationale and Significance of the Study

Although studies have been conducted on the predictive ability and differential validity of standardized test scores, many of these studies are thought to be biased due to funding by the College Board and ACT, they do not consider the predictive ability of verbal scores in mathematics course performance, and they only examine one level of subgroup bias. The majority of these studies have also been conducted on data obtained from testing entities in conjunction with first-year student information, neglecting to focus on institutions that are recognized by the Department of Education as Hispanic-Serving Institutions (HSIs) or Historically Black Colleges and Universities (HBCUs). Researchers both for and outside of the College Board have stated the need for predictive ability and differential validity studies to be performed at such institutions to research the possible impact that recent demographic shifts in the United States may have on student performance, as well as to address an increase in the number of students who are English language learners (ELL) and are now entering higher education. A study performed by Mattern, Patterson, and Kobrin (2012) specifically stated,

> Future research should examine issues related to the predictive validity of SAT scores for English as a second language students, who have increasingly comprised a larger percentage of the college-going population recently…and examine course-taking patterns for Hispanic students in order to understand the differential validity results for mathematics courses. (pp. 24-25)

The institution of interest is an HSI, showing a noticeable increase in: 1) the number of first-year students enrolled in developmental courses; and 2) the number of ELL students granted

admission. In the Fall of 2014, 74% of incoming freshmen were placed in developmental courses. A comparable increase has also been seen in the number of students required to repeat both developmental and college-level courses in order to obtain a successful passing grade of “C” or higher. On a local level, the significance of this research was to determine whether or not there is a need to adopt new admission and placement procedures to address possible misplacement of students, as well as potential subgroup bias. The significance of this research on a nation-wide level was to contribute to the body of knowledge in the role that standardized tests play in the admission and placement processes used to admit and place students in HSIs, or those with significant ELL populations.

Research Questions

1. What is the extent to which mathematics and verbal standardized test scores predict the performance of students in their first-year mathematics courses?
2. What is the extent to which standardized test scores display differential validity among subgroups?

Assumptions and Limitations

There are several limitations that the researcher considered prior to the start of the study. First, a study of this nature could be conducted to follow all incoming, first-year students throughout their undergraduate studies in order to thoroughly examine their attrition, retention, and graduation rates. The institution of interest exhibits low retention rates, which would make a valid longitudinal study very difficult to complete. However, the duration of the research was limited. Second, it is possible that not all freshmen will enroll in a mathematics course their first year or stay enrolled throughout the timeframe of a semester. This limited the data to a timeframe that did not include student information for Fall 2015. Third, it was not wise to treat

students who have withdrawn from a course the same as those who have not passed a course. To address these limitations, the researcher only included students in the study who enrolled in and received grades in their first mathematics course.

Fourth, the extended period of time over which the data was collected is problematic. There have been changes made to the mathematics curriculum and departmental placement procedures, which may have an impact on the consistency of the results. In Fall 2014, the changes included moving away from solely using a student's standardized test scores on the SAT, ACT, and/or ACCUPLACER to considering the last high school course a student has taken. The ACCUPLACER was designed by the College Board to identify a student's strengths and needs in the content areas of mathematics, reading, and writing. Prior to Fall 2014, the Texas Assessment of Knowledge and Skills (TAKS) test scores of students were used as the primary means of placement for students who attended public high schools in Texas. For students who attended private schools, out-of-state high schools, or were home-schooled, the SAT, ACT, and/or ACCUPLACER was used as an alternate means of placement. To address these limitations, the researcher only included participants who have taken both the SAT and ACT, and examined their scores as the sole basis for placement. This provided a study geared more towards current policies and procedures, as well as insight into past cohorts and their successes and/or failures. Lastly, to determine an alternate placement procedure, students' high school transcripts could be used to find trends in success rates; however, they are not readily accessible for bulk analysis.

Chapter 2: Literature Review

The significant rise in the population of students underprepared for higher education is alarming. A great concern is with the simple principle that, while the caliber of students has changed, the processes and procedures for admissions and placement into courses has not. In particular, student scores on standardized tests are still weighted very heavily in the decision processes that have an impact on their academic career in higher education, but these tests have been modified only slightly since their creation and initial use. The United States government has no control over the admission and education of students in higher education. Certain agencies do control accreditation of these institutions; however, they do not control education in the same respect as the United States government over K-12 education. Based on this discrepancy, it may be presumed that students might be put at an unfair disadvantage during the admissions process if institutions and testing agencies refuse to accommodate changes within the student population.

Despite the positive leap toward equality, the one-size-fits-all stance on education has had a significant negative impact – most recently with the No Child Left Behind Act (NCLB). Critics of NCLB have raised concerns about accountability and its detrimental effects on the school systems of individual states, as well as on the United States as a whole (Guthrie, 2001; National Commission of Excellence in Education [NCEE], 1983; Whilden, 2010). Students who have been educated from the initiation of national education reform to today are entering institutions of higher education less prepared for college-level coursework (Thiel, Peterman, & Brown, 2008). More specifically, research studies evaluating elementary and secondary curricula in the United States have shown that schools are not sufficiently preparing students in the content areas of science, technology, engineering and mathematics (STEM) (Newcombe, Ambady,

Eccles, Gomez, Klahr, Linn, & Mix, 2009). With such a large, growing population, higher education institutions are continuously seeking ways to expand access to higher education for students by creating developmental courses as well as success and peer-mentoring programs.

National Education Reform and the Decline of the Education System

Prior to 1965, the federal government had little presence in the educational system of the United States (Guthrie, 2001; Whilden, 2010). On April 11, 1965, Congress passed the Elementary and Secondary Education Act (ESEA). This act was a product of President Lyndon B. Johnson's War on Poverty, and signaled the federal government's interest in improving public K-12 education (Whilden, 2010). Prior to this, low-income children had legal access to public schooling; however, there were few resources available for their education (Guthrie, 2001; Whilden, 2010). The ESEA explicitly forbade the establishment of a national curriculum, and still is the largest source of funding for elementary and secondary education (Guthrie, 2001; Whilden, 2010). As such, national education reform laid the foundation for significant changes in the education of the nation's children by expanding access to education. The curricula for K-12 are implemented by state-regulated education agencies, which determine what subject content should be covered by a teacher and mastered by a student at each grade level.

Since its inception, the ESEA has been updated and reauthorized for educational reform numerous times. These reforms include 1994's Improving America's Schools Act (IASA) and most recently, 2001's NCLB. The driving force for its extensive revision in 1994 was the widespread controversy and panic caused by the publication of *A Nation at Risk* by the NCEE. Secretary of Education T. H. Bell created the NCEE on August 26, 1981, which was established to conduct research on the quality of education in America (NCEE, 1983). Some of the findings reported by the NCEE included a loss of sight by educational institutions of the basic purpose of

schooling, a significant decline in American education when compared to other industrialized nations, a continuous decline in the scores of the SAT for students between 1963-1980, a 72% increase of remedial courses in public 4-year colleges between 1975-1980, and lower testing achievement scores of students graduating from college (NCEE, 1983). According to NCEE (1983),

> It is important, of course, to recognize that the average citizen today is better educated and more knowledgeable than the average citizen a generation ago…nevertheless, the average graduate of our schools and colleges today is not as well-educated as the average graduate of 25 or 35 years ago, when a much smaller proportion of our population completed high school and college. (p. 2)

There has been talk of the need for additional education reform. With the positive changes brought about by the ESEA, problems with NCLB have perplexed many in the field of education. “Evidence indicates that despite higher grade point averages, these students’ skills and competencies are at the lowest in American history” (NCEE, 1983, p. 7). Fast-forward to 2010, and results indicate that more than 90% of students entering higher education are deemed insufficiently prepared to start college-level coursework (Cullinane & Treisman, 2010; Kerrigan & Slater, 2010; Saxon, Sullivan, Boylan, & Forrest, 2005). It appears that the fundamental basis of education – teaching and making students proficient in basic competencies and skills – has taken a backseat to the pressure of accountability and the funding it brings.

Accountability of NCLB and its Impact

Although the ESEA and IASA created equal access to education, social and economic changes during the new millennium ultimately forced lawmakers to redefine general education since access to a general education was no longer sufficient (Guthrie, 2001). In January 2002, President George W. Bush reauthorized ESEA, now known as NCLB. It provided additional funding to school districts, but in order for schools to be eligible for funding, states and their

school districts were required to implement and assess learning standards (Guthrie, 2001). The goal of these learning standards was to provide a foundation of assessment for student performance in achieving content specific standards. Students were now expected to learn and excel in content areas with the new learning and assessment guidelines the federal government required school districts to impose; however, there have been concerns about the abilities of all states to meet the national expected proficiency rate of 100% (Irons & Harris, 2007; McCombs, Kirby, Barney, Darilek, & Magee, 2004; Singh & Al-Fadhli, 2011). Due to strict regulations implemented by NCLB and its amendments, state agencies are required to also implement standards of accountability that must be met by each school. In a push to drive all schools toward acceptable levels of accountability, schools receive additional government funding for programs, and in some cases, teacher bonuses are based on the passing rates of their students (Posner, 2004).

On the other side of the spectrum, “schools that continue to fail to improve may be closed, and districts that continue to fail may be subject to state takeover” (Posner, 2004, p. 749). Due to the new accountability regulations under the NCLB, teachers, schools and school districts face penalties for unsatisfactory school performance. In the most severe cases, reform and restructuring can be mandated within a school, throughout a school district or statewide. Singh and Al-Fadhli (2011) noted, “low-performing schools are subject to sanctions if achievement goals are not met for two consecutive years” (p. 752). Research conducted by Irons and Harris (2007) noted that such measures were imposed on the entire Philadelphia school district system. In this particular case, the school district was taken over by the Commonwealth of Pennsylvania for being academically and financially distressed. Operations of the school

district were delegated to several private sector organizations, including the University of Pennsylvania (Irons & Harris, 2007).

It can be argued that testing results do not accurately reflect student performance or instruction and should not be used for accountability purposes. However, teachers and school districts are being held accountable, and at times penalized for student performance. In order to prevent take over and/or restructuring of schools and districts, administrators within school districts and teachers are trying to increase their proficiency rates by whatever means necessary in order to be identified as successful by the state (Irons & Harris, 2007; Menken, 2006; Singh & Al-Fadhli, 2011). As a consequence, some teachers are providing classroom instruction that in many cases is solely devoted to preparing students for standardized tests, known more commonly as "teaching to the test" (Popham, 2001; Posner, 2004). In the majority of school districts, administrators condone this method of teaching by mandating that teachers adhere to strict content delivery programs for the academic year. In most local school districts, these programs are known as scope-and-sequence guidelines. "In these programs, teaching behavior is regimented down to the exact material, timing, and wording of the instruction" (Posner, 2004, p. 750). When "teaching to the test" is used as the sole means for curriculum instruction, the results of standardized high-stakes tests do not truly reflect a student's mastery of a specific content area (Popham, 2001). As this vicious cycle continues in the K-12 education system, curriculum instruction may become less effective over a student's K-12 education and specific content knowledge, its retention, and applicability may be reduced (Posner, 2004; Popham, 2001).

"Supporters claim the NCLB is an effective national education reform initiative requiring intrusion into state and local education to ensure high quality, uniformity, and success, while

opponents argue it is politically motivated and a detriment to public schools" (Irons & Harris, 2007, p. 122). Despite a difference in opinions, one significant fact remains apparent – changes made to meet accountability standards have not been and are not currently equal – "across the [Nation], within states, or among all students" (Singh & Al-Fadhli, 2011, p. 752). Although school districts receive funding to finance K-12 education, only six to seven percent comes from federal sources. "All fifty states, with the exception of Hawaii, create public systems organized into local school districts that rely on financing from local property taxes" (Irons & Harris, 2007, p. 111). Property taxes are based on property values, thus unequal distribution of funding for K-12 education is highly presumed (Irons & Harris, 2007; Singh & Al Fadhli, 2011). With such an unequal distribution of financial resources in the education system, not only multicultural, but socioeconomic aspects of an educational system should be taken into consideration when measuring accountability based on student performance.

Historical Roots of Intelligence/Standardized Testing

"Although standardized testing was used by the Chinese Imperial Civil Service at least two thousand years ago, university admissions tests did not make their debut until centuries later" (Zwick, 2007, pg. 4). Admissions testing is believed to have started with entrance examinations in France in the 20th century; however, several researchers have provided conflicting dates. According to Stewart (1998), admissions testing started in the late 16th century, while Webber (1989) suggests that admissions testing began in the mid-to-late 18th century. Despite conflicting dates, all researchers agree that the birthplace of standardized testing was in Europe, as "in most countries, the use of tests to get out of universities preceded the use of tests to get in" (Zwick, 2007, pg. 4).

It is believed that standardized tests have historical roots in intelligence testing (Zwick, 2004). Alfred Binet administered the first intelligence test, the Binet/IQ Test, in 1905. Lewis Terman of Stanford University used an adapted version of Binet's intelligence test, the Stanford-Binet IQ Test, as an aid in placing and classifying children in schools who were perceived to have physical or mental disabilities that would hinder cognitive development (Lemann, 2004). According to Lemann (2004),

> Terman, not Binet, is responsible for the notion that every person has an innate, numerically expressible "intelligence quotient" that a test can discern. His primary interest was in identifying the very highest scorers and then making sure they were given special educational opportunities. (p. 6)

The goal of using this standardized intelligence test was to provide access to educational opportunities to the highest scorers of the test. It was believed that access to these opportunities would allow the United States to benefit from the talents of such scholars. For the posterity of psychometric research, several scientists persuaded the United States Army to administer an IQ test (the Alpha Army Test) to all recruits during the First World War (1914-1918) for assignments (Lemann, 2004; Zwick, 2007). According to Lemann (2004),

> This was the first mass administration of an IQ test, and the results were used, in that era when eugenicist ideas were conventional wisdom, to demonstrate the danger that unrestricted immigration posed to the quality of our national intellectual stock. (p. 6)

Carl Brigham, a Princeton University psychologist, made further adaptations to the Alpha Army Test for use in college admissions (Lemann, 2004). The Alpha Army Test is believed to be the first appearance of the IQ test in the United States (Lemann, 2004; Zwick, 2004, 2007).

The College Entrance Examination Board – The College Board, SAT, and ACT

Prior to 1900, pre-requisite course requirements and entrance exams used by universities for admissions lacked common criteria for content and standards. The leaders of the top 12 northeastern universities founded the College Entrance Examination Board (CEEB) in 1900

(Zwick, 2004, 2007). This organization is now known simply as the College Board. “Its purpose, then as now, was to act as an interface between high schools and colleges, which was something both sides wanted, for somewhat different reasons” (Lemann, 2004, p. 6). High schools were eager for the development of a uniform admissions process, and universities wanted to ensure that students applying for admission were held to a uniform standard for college preparation. The outcomes of instituting rigor would be: a) a common admissions process that would allow students to apply for admission to more than one university at a time; and b) a platform for improvement that would have a significant impact on the curricula of high schools (Lemann, 2004).

Not all universities in the United States were member institutions of the College Board. According to Zwick (2007), for nearly two decades after its inception, only the top 12 northeastern universities in the United States were member institutions. Member institutions would administer essay exams in all nine subject areas and send them back to the College Board for hand grading. As the hand grading of exams became laborious, the College Board made the decision to look for an alternative to replace essay exams with predominantly multiple-choice questions. It later developed the Scholastic Aptitude Test. This precursor to today’s SAT was first administered in 1926 to about 8,000 candidates. The first SAT consisted of questions similar to those included in the Army Alpha Test.

The passing of the Higher Education Act (HEA) of 1965 provided more access to higher education to those considered “underrepresented” in higher education institutions. The goals of HEA were to provide access to higher education and training programs. It allowed the federal government to provide substantial sums to higher education institutions for “financial aid, special services, and minority recruitment programs” (Boylan, 1988, p. 2). Universities were required to

help an ever-expanding enrollment of underprepared students by providing college preparatory services. During this period, colleges and universities were eager to “rid themselves” of college preparatory programs on their campuses and find alternative means to serve their students’ needs (Richardson, Martens, Fisk, Okun, & Thomas, 1981). The birth of the junior/community college served as the solution to the “underprepared student” problem. “The junior colleges provided an alternative to the college preparatory program by offering the equivalent of the first two years of college courses combined with a large menu of preparatory or remedial courses” (Boylan, 1988, p. 6). As students were encouraged to attend junior/community colleges as means of college preparation, “the community and junior colleges expanded their efforts to provide remedial and developmental services while four-year institutions reduced them” (Boylan, 1988, p. 11).

The philosophy of “extreme selectivity” in the 1960’s gave way to the philosophy of “open admissions” in the late 1970’s (Boylan, 1988). Although the name of these services has changed frequently from college preparatory to learning assistance to remedial, and now to developmental, they have all attempted to accomplish the same goal – to help underprepared students successfully adjust to the college atmosphere. In 1999, Poiani did an extensive study on developmental mathematics programs and noted that Saint Peters College offered such educational opportunities in mathematics as far back as 1968. By 1977, over 80% of higher education institutions in the United States offered assistance in one form or another to underprepared students in order to promote their success in college (Roueche & Snow, 1977). By 1985, several studies estimated that this number increased to over 80%, including the same institutions that offered such services in 1889 (Boylan, 1988; Canfield, 1889; Roueche & Snow, 1977; Wright & Cahalan, 1985). According to Boylan (1988),

> The fact that a large number of students enter college underprepared for success in college-level studies is not a new phenomenon. Instead, it simply represents the

> continuation of a situation that has existed since the very earliest days of American postsecondary education. (p. 13)

SAT. "In 1926 – by which time, to his [Carl Brigham's] immense credit, he had loudly renounced his commitment to eugenics – the College Board experimentally administered Brigham's Scholastic Aptitude Test for the first time" (Lemann, 2004, p. 7). The Scholastic Aptitude Test was administered to approximately 8,000 incoming freshmen during its initial trial. This exam later became the precursor to the today's SAT (Zwick, 2004). The driving force behind the decision for a permanent change to the College Board's entry exam procedures came from persuasion by Henry Chauncey and James Bryant Conant. Their primary goal was to redefine the merit of the Ivy League student body – an effort to expand the admissions pool for the Harvard National Scholarship Program. "In 1938, Conant and Chauncey persuaded all College Board schools to use the SAT as the main admissions test for scholarship applicants" (Lemann, 2004, p. 8).

In a move to accept a larger student body from the Midwest for admissions, Chauncey and Conant wanted to move from the traditions of character and legacy towards intellectualism. They felt the SAT was the answer and hoped to expand the use of the test on a national level for more than scholarship purposes. Carl Brigham remained an obstacle in the expansion of using the SAT until his death in 1943 (Lemann, 2004). After the death of Brigham, Chauncey persuaded the United States Army and Navy to use a version of the SAT (Army-Navy College Qualifying Test) as an officer candidate test. It was administered to approximately 316,000 officer candidates across the United States. (Lemann, 2004). Chauncey used this pilot testing to demonstrate that the SAT could be used to test a large population simultaneously, while maintaining security and the accuracy of the scoring. According to Lemann (2004),

> This made it clear that the SAT could be used as a screen for the entire American high school cohort (only during the war did high school become the majority experience for American adolescents), rather than a handful of private-school kids – that it could be the basis for what one College Board official called the "great sorting" of the national population. (p. 8)

When World War II ended, "a commission on higher education appointed by President Harry Truman issued a clarion call for the expansion of public universities" Lemann, 2004, p. 10). These institutions mainly served in-state students and relied heavily on high school transcripts for admission purposes. They were known to be minimally selective, the curriculum was similar to that of local area high schools, and the attrition rates were typically very high (Lemann, 2004). "As faculties became more ambitious, they began to see admission by SAT as a way of nationalizing, academicizing, and reducing student bodies, which would free them to concentrate on their research" (Lemann, 2004, p. 10). It was at this time that Chauncey and Conant sought to build a monopoly on the industry of postsecondary educational testing (Lemann, 2004). In 1942, the old College Boards were suspended, "for the duration," and never resumed, so the SAT became the admissions test for all applicants to College Board schools, not just scholarship applicants.

In 1947, the Educational Testing Services (ETS) was founded in Princeton, New Jersey. It came about via a merger of the testing services provided by the College Board, the Carnegie Foundation for the Advancement of Teaching (funding Chauncey and Conant), and the American Council on Education (Lemann, 2004). Chauncey was to become the President of ETS and Conant was elected Chairman of the Board. Prior to being chartered, the ETS opened a second office in Berkeley, CA. By the opening date of the second branch office, the SAT had become "the" national admissions test. This version, now called the SAT I: Reasoning Test, consisted mainly of multiple-choice questions, with very few mathematics questions requiring

"student-produced" answers (Zwick, 2004). The timing of Chauncey and Conant's decision to monopolize the post-secondary testing industry came at a perfect time as, "there was a strong fit between ETS's ambitions and faculties' ambitions that wound up linking the SAT to the growth of public universities" (Lemann, 2004, p. 10).

ACT. In 1953, the University of Iowa provided funding to the statistician E. F. Lindquist to develop testing programs and the Measurement Research Corporation (MRC) was founded. Lindquist became director of what is known today as the Iowa Testing Programs, which are composed of the Iowa Tests of Basic Skills and the Iowa Tests of Education Development. According to Zwick (2004), the Iowa testing programs were the first statewide systems to be used in high schools. As with the original SAT, all Iowa tests were hand-graded. Not satisfied with the process of administration or the speed of grading, Lindquist and Phillip Dalon of Harvard invented the Iowa scoring machine (Zwick, 2004; Zwick, 2007). It was the first device to use electronic scanning techniques to score answer sheets and was considered a "marvel of blinking panels…that could emit a record of achievement from the brief encounter of small black marks on paper and the photocells in a reading head" (Peterson, 1983, p. 114). In 1959, shortly after the introduction of the Iowa scoring machine, ACT, Inc. was founded by Lindquist in Iowa City. It may be all too coincidental that the invention of the scoring machine came with such perfect timing with the founding of ACT, Inc., but it did halt the monopolization of admissions testing.

The SAT was well established at the time ACT, Inc. was founded; however, test developers viewed the SAT as an elitist testing tool that lacked close ties to instructional objectives taught in high schools (Zwick, 2004). With long standing success rates since 1942, the original version of the ACT consisted of test items from the subject areas of English,

mathematics, social studies, reading and natural sciences of the Iowa Tests of Educational Development (Zwick, 2004). The version of the ACT now administered to students consists of the analysis of instructional objectives taught from grades seven through 12. The information is obtained from surveys of secondary school teachers and curriculum experts. “As well as being more strongly linked to instructional goals than the SAT, the ACT also places a greater emphasis on facilitating course placement and academic planning” (Zwick, 2007, p. 9).

Purpose and Use of Standardized Tests

There are unfounded theories on how students choose an undergraduate school: popularity, sports teams, Greek life organizations, being a "legacy", the choice of parents or a significant other, and so on. The process by which institutions ultimately choose an applicant for admission is a bit more objective. The sorting process of college admissions begins with the consideration of academic ability. The question widely asked by those applying for admissions is “What is the importance of test scores in consideration for admission?”

As we now know, standardized admissions tests were developed to provide a uniform, rigorous, and reliable process of identifying applicants who would perform well in institutions of higher education. According to Breland et al. (2002), a survey performed between 1979 and 2000 showed that over 90% of four-year colleges required a SAT or ACT score as part of the application process for admissions. The joint survey of SAT/ACT and the NACAC Admission Trends Survey found that standardized test scores are the second most important factor, after high school grades, showing an increase of importance from 1979 to 2000 (Zwick, 2007). “The third most important factor in all surveys was pattern of high school course work” (Breland et al., 2002, p. 67). More specifically, Geiser and Santelices (2007) claim that grades in college-preparatory subjects are the “best” indicator of student performance.

"While conceding the importance of high school record as an admissions criterion, advocates of standardized admissions tests nevertheless state that, used as a supplement to high school record, tests provide additional information that can aid admissions officers and improve decision making" (Geiser & Santelices, 2007, p. 24). The focus of Geiser and Santelice's research has been on this importance. After controlling for high school grades, Bridgeman, Pollack and Burton (2004) found that students who scored higher on the SAT earned higher grades in college. Researchers agree that high school grades are the best predictors of college performance; however, standardized admissions test scores do add significantly to performance prediction (Burton & Ramist, 2001; Camara & Echternact, 2000). Thus, high school course grades, high school GPA and standardized test scores should all be considered to predict college performance and success (Breland et al., 2002; Geiser & Santelices, 2007; Zwick, 2004). It appears that, although high school grades play an important factor, approximately 70 percent of four-year institutions use standardized test scores as the primary tool for admissibility (Breland et al., 2002). Of these "roughly 40 percent of four-year schools reported that they had minimum test score requirements for admission; 57 percent had minimum requirements for high school GPA (Breland et al., 2002, p. 81). There are only a few "open-door" colleges in which standardized admissions tests play no role in the admissions process. In a 2000 survey conducted by ACT, Inc., the Association for Institutional Research, the College Board, ETS, and the NACAC, only eight percent of four-year institutions have an "open door" policy (Breland et al., 2002).

Concerns with Using SAT/ACT for Placement and Admissions

With regard to minority students, researchers suggest that minority students suffer from psychological and sociocultural stress when entering institutions of higher education (Smedley,

Hector, & Harrell, 1993). Studies by Pascarella and Terenzini (1991) indicated that this undue stress was caused by what the students perceived to be the presence of a hostile racial climate in one or several classrooms, leading students to withdraw. In his study, Olivas (1979) noted that a shortage in minority instructors, specifically of African-American and Hispanic ethnic background, posed a particular problem for students feeling accepted and welcomed in higher education institutions. Although this study was performed more than thirty years ago, a more recent study performed by the National Center for Education Statistics (NCES) (2010), detailed the breakdown of higher education institutions' faculty as follows: 80% Caucasian, 7% African American, 6% Asian/Pacific Islander, 4% Hispanic, and 1% American Indian/Alaska Native. Such a small number of minority faculties may contribute to perceived racial hostility.

Researchers suggest that there may be a possible relationship between a student's high school GPA and his or her success in college (Astin, 1985, 1991; Pascarella & Terenzini, 1991). Hagedorn, Siadat, Fogel, Nora, & Pascarella (1999) stated, "students in remedial math placements were more likely to come from families with lower incomes and lower educational levels, and were more likely to receive less encouragement to enroll in college" (p. 278). Researchers also suggest that "students from higher economic backgrounds may be receiving a better grounding in mathematics than their less affluent counterparts" (Hagedorn et al., 1999, p. 280). These statements point to a possible relationship between high school quality, minority status and socioeconomic background. Demographic studies have shown that schools in low socioeconomic and/or predominantly minority areas may not be delivering the same quality instruction as schools in affluent areas. Although these situations are external factors that higher education institutions cannot control, they pose a concern to the admissions process. In terms of

other minority groups, a study by Hoyt (1999) suggested that older, non-traditional students have higher dropout rates than traditional students.

Demographic Shift in the U.S. and ELL

Demographic shift. The United States has been a melting pot of cultures, religions and growing ethnic and racial diversity. Our multicultural nation was created by "immigration and subsequent births to the new arrivals during the last few decades of the century [which] played a major role in changing the racial and ethnic composition of the U.S. population" (Hobbs & Stoop, 2002, p. 80). How much of a role? "At the beginning of the [20th] century, 1 out of 8 Americans was of a race other than White" (Hobbs & Stoop, 2002, p. 71). It is now estimated that 20% of Americans are of a race other than White (Hobbs & Stoop, 2002). According to Hobbs and Stoop (2002), the composition of the immigrant population has nearly doubled and is due to,

> Large-scale immigration, primarily from Latin America and Asia, [which] underlies both increased racial and ethnic diversity. In just the last two decades of the [20th] century, the Asian and Pacific Islander population tripled, and the Hispanic population more than doubled…the increasing racial and ethnic diversity of the U.S. population in the 20th century has largely been a post-1970 development, with regional patterns generally reflecting the trend of the United States as a whole. (pp. 73 & 88)

Hobbs and Stoop (2002) have noted that the population in the United States more than tripled in the 20th century. According to their research, "the growth of 32.7 million people in the 1990's represented the largest numerical increase of any decade in the U.S. history" (Hobbs & Stoop, 2002, p. 1). The population growth has changed across the United States and is not restricted to one region, with Blacks, Asians and Pacific Islanders, American Indians and Alaska Natives, and Hispanics representing increasing shares of the national population, and of each region's population. "From 1900 to 2000, the number of non-Southern states with populations of

at least 10 percent of races other than White increased from 2 to 26, reflecting the spread of diversity across the country" (Hobbs & Stoop, 2002, p. 73).

Fastest growing populations and the switch to a "majority-minority" nation. Based on census data analyzed from the last two decades of the 20th century, researchers found that the "aggregated Minority population [people of races other than White or of Hispanic origin], increased by 88 percent between 1980 and 2000, while the White non-Hispanic population grew by only 7.9 percent during the 20-year period" (Hobbs & Stoop, 2002, p. 71). This data showed a trend that has continued in recent years. Colby and Ortman (2015) have projected the top three fastest-growing populations in the United States to be: a) populations claiming one or more races, b) the Asian population, and c) the Hispanic population. The Asian population alone is projected to increase by 128% between 2014 and 2060. Colby and Ortman (2015) suggest a large increase is prevalent in this population due to a small population at the beginning of the 20th century, along with a considerable increase in immigration from Asian countries. Taking a comparative look at the census data from 1900 to 2000, "the Hispanic population (of any race) more than doubled…growing by 20.7 million people from 1980 to 2000" (Hobbs & Stoop, 2002, p. 71 & 78). According to Colby and Ortman (2015), "the Hispanic population is projected to increase from 55 million in 2014 to 119 million in 2060, an increase of 115 percent" (p. 9). With these projections, Colby and Ortman (2015), estimate that Hispanics will account for more than one-quarter of the total population in the United States by 2060. The increases of both the Asian and Hispanic populations are thought to be the results of a combination of increased immigration between 1980 and 2000, as well as high fertility rates (Hobbs & Stoops, 2002).

The non-Hispanic White population continues to be the largest racial and ethnic group, the majority, accounting for more than 50% of the United States' total population. However, it is

projected that by 2044, more than half of all Americans will belong to a minority group, a phenomenon described as “the point at which we become a “majority-minority” nation” (Colby & Ortman, 2015, p. 9). “The child population within the United States is even more diverse and is projected to experience the majority-minority crossover in 2020, just 6 years into the future” (Colby & Ortman, 2015, p. 13).

Foreign-born and ELL. “By 2060, nearly one in five of the nation’s total population is projected to be foreign born” (Colby & Ortman, 2015, p. 1). Due to its large growth rate, researchers predict the growth rate of the foreign-born population to surpass that of the native born population. It is expected to account for “an increasing share of the total population, reaching 19 percent in 2060, up from 13 percent in 2014” (Colby & Ortman, 2015, p. 2). This population is commonly referred to as ELL or nonnative English speakers. “ELL refers to a students whose first language is not English and encompasses both students who are just beginning to learn English and those who have already developed considerable proficiency” (Case, 2003, p. 2).

This subgroup’s rapid growth in the United States should be taken into consideration when researching the impact a student’s reading ability may have on the use of standardized tests in the admissions process. In turn, its impact on placement and mathematics performance should also be considered as the bulk of this population is concentrated in the adult age range, “with fewer than 10 percent of its population ages 17 and under, as compared with nearly a quarter of the native population.” (Colby & Ortman, 2015, p. 5). The National Clearinghouse for English Language Acquisition (NCELA) estimated that “more than 4.8 million ELL students were enrolled in public schools (Pre-K through Grade 12) for the 2004-2005 school year…approximately 9.9% of total school enrollment, and a 47.6% increase over the reported

1994-1995 total public school enrollment" (Das, 2008, p. 2). Among states, California enrolled the largest number of ELL students, followed by Texas with an enrollment of 616,466 in the 2004-2005 school year, an increase of over 45% between 1994-2004 (Das, 2008). Samway and McKeon (1999) predict that by 2050, it is highly likely that every teacher in the United States will have ELL students, suggesting a crucial need to adequately prepare teachers for such students, and demand that higher education institutions revisit their admissions and placement procedures.

Correlation Between Mathematics and Reading

The connection and problems with the relationship. The relationship between mathematics and reading has been well documented over the last six decades (e.g. Breen, Lehman & Carlson, 1984; Fuchs, Fuchs, Eaton, Hamlet, & Karns, 2000; Helwig, Rozek-Tedesco, Tindal, Heath, & Almond, 1999; Jerman & Mirman, 1974; Ní Ríordáin & O'Donoghue, 2009; Pitts, 1952; Reikerås, 2006; Thompson, 1967; Thurber, Shinn, & Smolkowski, 2002; Walker, Zhang & Surber, 2008). Research indicates that student proficiency in reading can be a strong indicator of success in mathematics (Jiban & Deno, 2007). Boero, Douek, and Ferrari (2008) suggest the correlation between mathematics and reading skills is founded on the need of mastery of one's "natural language" in order to understand the context in which that same language is used in mathematics. In essence, the way in which mathematical assessment items are written can have a negative impact on the achievement of a poor reader. Grimm (2008) reported that reading comprehension skills also play a role in relating that conceptual understanding of mathematics and its application. This relation may be due to general reading deficiencies and/or the deficiencies of learning the language, as is the case with ELL

populations (Beal, Adams, & Cohen, 2010; Kyttälä, 2008; Lee, 2004; Shaw, Bunch, & Geaney, 2010; Stoddard, Pinal, Latzke, & Canaday, 2002; Turkan & Liu, 2012).

There is growing concern that test items in the physical sciences are geared toward assessing reading comprehension rather than content understanding (Flick & Lederman, 2002; Korpershoek et al., 2015; Rutherford-Becker & Vanderwood, 2009; Walker et al., 2008). According Korpershoek et al. (2015), while the general relationship between previous and future school achievement has been researched prolifically, the connection between reading and mathematics ability on academic achievement in the physical sciences remain scarce. The bulk of research available concerning the relationship between reading ability and mathematics achievement has been done in primary education (e.g. Grimm, 2008; Kintsch & Greeno, 1985; Lee, Deaktor, Hart, Barnett, & Enders, 2005; Powell, Fuchs, Fuchs, Cirino, & Fletcher, 2009; Stern, 1993; Vilenius-Tuohimaa, Aunola & Nurmi, 2008; Walker et al., 2008). Walker et al. (2008) determined that "mathematics items designed to measure higher order cognitive skills, such as problem solving and mathematical reasoning, are two-dimensional in that they measure both reading ability and mathematics skills" (p. 163-164). This is what Bruner (1986) refers to as the difference between the two modes of learning - paradigmatic and narrative knowing. Narrative knowing is the sufficient grasp of reading skills in the social context; however, paradigmatic knowing is the use of focused and context-free mathematical models. Rutherford-Becker and Vanderwood (2009) note that these two components are conceptually distinct.

The growing ELL population is of concern as "these growing numbers suggest the crucial need for adequate preparation of teachers to serve these students" (Das, 2008, p. 2). In 2008, the states with the highest public school enrollment of ELL students, in order, were: California, Texas, Florida, New York, Illinois, and Arizona (Das, 2008). In Texas alone, the population of

students identified as ELL grew by 45.1% from 1994-1995, and again from 2003-2004 (Das, 2008). According to Samway and McKennon (1999), it is probable that by 2050, every teacher in the United States will have ELLs students in their classroom. Colby and Ortman (2015) note that such a drastic change in the diversity of the student population suggests a need to remedy general reading deficiencies in this population, especially since the majority-minority population crossover in children will occur in 2020.

Potential solutions. In general, students with deficiencies in both mathematics and reading tend to perform lower on mathematics assessments than students with just one deficiency (Fuchs, Fuchs, & Prentice, 2004; Rutherford-Becker & Vanderwood, 2009). Several studies have suggested that the correlation of reading comprehension and mathematics can be improved through interventions such as computer-based tutors, as well as transactional reading strategies to improve vocabulary (Borasi, Siegel, Fonzi, & Smith, 1998; Brown & Ryoo, 2008; Carter & Dean, 2006; Helwig et al., 1999; Nathan, Kintsch, & Young, 1992; Rutherford-Becker & Vanderwood, 2009). Although there does seem to be a correlation, Rutherford-Becker & Vanderwood (2009) suggest, “math abilities do not seem to have a significant influence on reading growth” (p. 3). On the other hand, an increase in reading ability does have a positive effect on mathematics growth, especially when solving word problems (Jordan, Kaplan, & Hanich, 2002; Jordan, Hanich, & Kaplan, 2003). “Thus, when intervening with children that demonstrate deficits in both reading and math, reading intervention may warrant consideration as the first step” (Rutherford-Becker & Vanderwood, 2009, p. 32).

According to Das (2008), the ELL student population in his study was the worst performing group out of their peers. In order to address the rapidly changing diversity of the student population in Texas, the Texas State University System and the Texas Education Agency

(TEA) have started an initiative to develop “instructional resources designed to increase the effectiveness of mathematics instruction for ELL students [in preparation for the TAKS test]” (Das, 2008, p. 2).

Several measures of readability that increase the overall reading levels of assessments include higher word character, sentence, and syllabus count, as well as word and sentence length. As these measures of readability increase, they provide complexity and confusion for a reader, which, in turn, has an overall negative impact on performance (Bolt & Thurlow, 2007; Jerman & Mirman, 1974; Powell et. al, 2009; Thompson, 1967; Walker et al., 2008). “Ideally, assessment items should minimize reading difficulty without jeopardizing mathematical complexity. Therefore, investigating ways of writing mathematics assessment items that require students to read and synthesize text without going beyond the students’ reading grade level is imperative” (Lamb, 2010, p. 32).

The Institution of Interest

The university of interest for this study is a small, liberal education private institution that was founded in 1881. From its inception until 1969, the university was an all-female higher education institution. In 1970, its status was changed to co-educational. It holds accreditations with several local and national associations, and under federal guidelines, it is recognized as an HSI. According to its mission statement, the institution aims to provide higher education opportunities to populations underserved by higher education. One of its goals is to provide students with a solid foundation of guidance and personal attention, which is supported by small class sizes. It is the university’s hope that through its inclusive culture and smaller class sizes, students will become enlightened individuals that are concerned with the world around them.

During his “State of the University Address” in Fall 2009, the president of the university provided the audience with statistics on the demographics of the student population. The total enrollment had doubled since 2000, with enrollment at 7,166 students in Fall 2009. Segmenting this number by degree category and enrollment during this time, there were 5,628 undergraduate students, 1,121 graduate students, and 417 professional students. Also in Fall 2009, the institution welcomed its largest incoming freshman class, which consisted of 724 students. Of the 724 newly admitted students, 254 graduated from a high school in Bexar County (Texas Higher Education Coordinating Board [THECB], 2010; Texas Education Agency [TEA], 2010). The students enrolled represented the socioeconomic spectrum of the area the institution serves, and the 2009 ethnic/racial breakdown of the student population was: 0.32 % American Indian/Alaskan Native, 3.32% Asian/Pacific Islander, 6.99% Black, 52.79% Hispanic, 12.63% International, and 23.95% White.

Admissions criteria and placement procedures at the institution. Developmental mathematics has existed at this institution since the early 1990s (E. Kreston, personal communication, June 7, 2010). Students applying for admission are placed in developmental mathematics courses based on the college readiness score provided by the results of their TAKS Test, SAT, ACT or ACCUPLACER. Placement requirements include a minimum college readiness score of 2202 for the TAKS test, 520 for the SAT, 22 for the ACT, and 63 for the ACCUPLACER to indicate a student’s passing of 70% of the mathematics objectives. These scores set the university’s requirements for students to be labeled as college-level mathematics ready, and then, upon admission and registration they are placed into their degree-required mathematics course. Students who score below these requirements are labeled as not college-level mathematics ready and are placed in developmental mathematics.

Chapter 3: Methodology

This is a quantitative multivariate analysis, which examined: a) the ability of standardized test scores to predict the performance of students in their first mathematics courses, and b) the differential validity of standardized test scores among subgroups.

Research Design

According to Creswell (2009), "quantitative research is a means for testing objective theories by examining the relationship among variables. These variables in turn, can be measured…so that numbered data can be analyzed using statistical procedures" (Creswell, 2009, p. 4). It is well known that quantitative research is founded on the post-positivist paradigm.

> Postpositivists hold a deterministic philosophy in which causes probably determine effects or outcomes. Thus, the problems studied by postpositivists reflect the need to identify and assess the causes that influence outcomes, such as found in experiments. It is also reductionistic in that the intent is to reduce the ideas into a small, discrete set of ideas to test, such as the variables that comprise hypotheses or research questions. (Creswell, 2009, p. 7)

The researcher engaged in this form of inquiry using multivariate analysis, to include direct discriminant analysis and three-way ANOVA.

Grimm and Yarnold (2000) explain that "multivariate statistics [analysis] provide simultaneous analysis of multiple independent and dependent variables" (p. 5). In focusing on test "significance", the results will show whether or not two groups differ with respect to a composite variable (Grimm & Yarnold, 2000; Pallant, 2007; Szafran, 2012). Discriminant analysis is a statistical technique used to examine the linear combination of continuous variables that predict group membership on one categorical dependent measure (Norušis, 2009; Pallant, 2007; Tabachnick & Fidell, 2006). Due to its predictive nature, the researcher used discriminant analysis to examine the ability of standardized test scores to predict the performance of students in their first mathematics course. Typically, the dependent variable is a dichotomous criterion,

such as pass or fail, as is the case of this study. Due to its ability to examine the interaction effect between three independent variables concomitantly, the researcher used three-way between-groups ANOVA to examine the differential validity of standardized test scores among three subgroups – ethnicity, gender, and socioeconomic status, labeled as income in this study.

Discriminant analysis. According to Tabachnick and Fidell (2006), "discriminant analysis is MANOVA turned around" (p. 375). Instead of independent variables being the grouping criterion and the dependent variables the predictors, in discriminant analysis, the independent variables are the predictors and the grouping criterion is the dependent variable. Mathematically, they are similar, but the emphasis differs in that unlike multivariate analysis of variance (MANOVA), discriminant analysis answers the question whether or not predictors can be combined in a way that reliably predict group membership, as well as "interpret the pattern of differences among predictors in an attempt to understand the dimensions along which groups differ" (Tabachnick & Fidell, 2006, p. 376). The dichotomous dependent variable for this study is a grade of passing or not passing. The label of a passing grade was determined using the university's grade point system on letter grades, as well as the letter grade that is required for all prerequisite courses in the mathematics department. The standard for satisfactory completion is a letter grade of "C"; all other students were labeled as not passing. The independent variables, which the university uses to place students, are standardized test scores in specific content areas, and include SAT mathematics and verbal scores, and ACT mathematics and verbal scores.

There are three types of discriminant analysis, all analogous to the three types of multiple regression: standard (direct), sequential, and statistical (stepwise) (Pallant, 2007; Tabachnick & Fidell, 2006). For the purpose of this study, the researcher interpreted the results of direct discriminant analysis to find the best linear combinations of standardized test mathematics and

verbal predictors. The goal was to determine which combination of predictors would create the best predictive model for classifying and separating students into passing and not passing groups. "In discriminant analysis, the first discriminant function provides the best separation among groups" (Tabachnick & Fidell, 2006, p. 378). The adequacy of this function to predict group membership was checked via cross-validation. Cross-validation is the process of testing a model on more than one sample. This technique is often undertaken to assess the reliability and generalizability of the findings, which is particularly crucial in discriminant analysis because the solutions are often unreliable (Hair, Anderson, Tatham, & Black, 1995).

The assumptions and limitations to discriminant analysis are similar to those of MANOVA.

- The dependent variable must be categorical.
- The data should exhibit multivariate normal distribution.
- The sample size of the smallest group needs to exceed the number of predictor variables.
- Homoscedasticity must be present for the predictor variables.
- Non-multicollinearity must be present among all pairs of predictor variables.
- Correlations must exist between means and variances.
- The variables used to discriminate between groups should not be completely redundant.
- The data set should be void of outliers (Norušis, 2009; Tabachnick & Fidell, 2006).

It is important to note "discriminant analysis is robust to failures of normality and homoscedasticity if the violation is caused by skewness rather than outliers" (Tabachnick & Fidell, 2006, p. 382). This means that the variance around the regression line should be the same for all values of the predictor variables.

In terms of limitations, unequal sample sizes do not hinder analysis, but missing data can create problems with normality and homoscedasticity. According to Tabachnick and Fidell (2006), the pattern of missing data is more important than the amount of data that is missing – “missing values scattered randomly through a data matrix pose a less serious problems” (p. 62). In the case of a large data set, such as this study, a good alternative is “deletion of cases with missing values [as this is] the default option for most programs in the SPSS and SAS packages” (Tabachnick & Fidell, 2006, p. 63). Missing data in this study was thoroughly investigated before proceeding with analysis.

As discriminant analysis predicts group membership in naturally occurring groups rather than groups formed by random assignment, questions regarding the reliability of prediction of group membership and/or differential group membership are often not examined. The predefined groups for this analysis included the grouping of students into passing and not passing groups based on grades received in their first mathematics course. The label of a passing grade is determined by the university’s standard for satisfactory completion (letter grade of “C”); all other students are labeled as not passing. Therefore, the reliability of group membership prediction was not examined beyond cross-validation. Completion of this portion of the study allowed the researcher to proceed with the second research design used for analysis.

Three-way ANOVA. Analysis of variance (ANOVA) and regression are the two most common forms of multivariate analysis used in the social sciences; however, regression is designed for use with continuous independent variables, and ANOVA is designed for use with categorical variables (Szafran, 2012; Tabachnick & Fidell, 2006). ANOVA is used to “compare two or more means to see if there are any statistically significant differences among them” (Tabachnick & Fidell, 2006, p. 37). It provides both descriptive and inferential statistics;

however, one-way ANOVA is limited to examining the effect of one independent variable on one dependent variable (Szafran, 2012). An ANOVA that examines the effect of more than one independent variable on a dependent variable is classified as a factorial ANOVA, being named by the number of independent variable being examined (Cohen, 2002; Grim & Yarnold, 2000; Pallant, 2007; Szafran, 2012; Tabachnick & Fidell, 2006).

According to Pallant (2007), "two-way analysis of variance (ANOVA) allows you to test the impact of two independent variables on one dependent variable" (p. 104). The use of two independent variables allows for a researcher to test for an interaction effect, which examines the influence of one independent variable by another (Pallant, 2007; Szafran, 2012; Tabachnick & Fidell, 2006). "Two-way ANOVA can be understood as a series of hypothesis tests…the result of each hypothesis test determines whether additional hypothesis tests are performed" (Szafran, 2012, p. 350). According to Cohen (2002),

> Just as there are many patterns of cell means that lead to two-way interactions, there are even more distinct patterns in a three-way design…the simplest is when all of the means are about the same, except for one, which is distinctly different. …The three-way interaction can be defined as the difference between two simple interaction effects. (p. 3)

Two separate three-way ANOVAs were performed on the two test scores available in the data set – total SAT score, and total ACT score. The independent variables for which differential validity was examined include ethnicity, gender, and socioeconomic status (derived from parental income).

As with any statistical procedure, there are some assumptions when using one-way ANOVA for analysis, which extend to factorial ANOVA. The assumptions for ANOVA include:

- The dependent variable must be continuous.
- Random sampling of data must be present.
- Independence of observations must exist.

- Data should exhibit multivariate normal distribution.
- Homogeneity of variances must be present (Pallant, 2007).

It is important to note that with a sample size over 30, factorial ANOVA is robust to departures from normality and unequal variance (Pallant, 2007, Tabachnick & Fidell, 2006).

Population and Sample

The participants of this study were from a target sample of the student population at a private, liberal education institution who have completed their first mathematics course in their first year of enrollment. It was assumed that students in this study were placed in either developmental or college-level mathematics courses based on their performance on the mathematics portion of one of four assessments considered for placement during the timeframe selected for the study:

1) 2202 on the TAKS
2) 520 on the SAT
3) 22 on the ACT
4) 63 on the ACCUPLACER.

The two tests not described in detail in the literature review include the TAKS test and ACCUPLACER. These tests are not widely used for placement at the institution of interest, and since these scores are not available for all students, they were not considered in the analysis. The TAKS test was the fourth Texas state standardized test administered by the Texas Education Agency in an attempt to assess students' attainment of reading, writing, mathematics, science, and social studies skills in grades three through eight and nine through 11. It was administered from 2003 to 2011. The ACCUPLACER is administered by the College Board. It is an

assessment that claims to identify student strengths and needs in mathematics, reading, and writing.

The demographic data collected about the participants provided the researcher with information pertaining to age, gender, ethnicity, and financial aid information, to include parent and student income, total expected family contribution, gross need, and total aid awarded. The population used for this study consisted of 1,037 undergraduate students who enrolled on the main campus at the university of interest and completed their first mathematics course between Fall 2010 and Spring 2015. The researcher chose the population and setting of this study based on personal interest, questions raised by several faculty members and administrators, extensive interaction with the student population in mathematics courses, as well as the availability, and access to the data required to perform such an analysis.

Setting

The setting of this experimental study was a private, liberal education institution that holds accreditations with several local and national associations. Under federal guidelines, it is recognized as an HSI. According to its mission statement, the institution aims to provide higher education opportunities to populations underserved by higher education. As of Fall 2015, the total global enrollment of the institution, including main campus, satellite campuses for extended studies, virtual university, and international campuses was 10,984 students. Segmenting this number by degree category, enrollment consisted of 8,713 undergraduate students, 1,213 graduate-master's students, 189 graduate-doctoral students, and 869 professional students. The ethnic/racial breakdown of the total student population is: 56.45% Hispanic, 12.82% White, 7.91% Other Minority/Unknown, 1.72% Nonresident Alien, and 21.1% International.

Data Collection

Consent was obtained from the Institutional Review Board of the university of interest. Data obtained were provided to the researcher via information that resides in data collection and reporting software. The two data fields that were not available upon request included ELL student status, and high school transcripts and/or records of high school coursework and grades.

Data Clean-Up and Analysis

The researcher received data from the Office of Institutional Research in the form of Microsoft Excel files. The researcher used Microsoft Excel to organize and compile the data received into one complete data set. The complete data set was transferred to Statistical Package for the Social Sciences (SPSS) for analysis. The researcher ultimately used the discriminant analysis and three-way ANOVA functions in SPSS to analyze the data collected.

The initial data set received consisted of 6,642 participants enrolled from Fall 2009 through Fall 2015, on main campus, satellite campuses for extended studies (adult education) and the virtual university. The first review of the data exhibited extreme outliers. These outliers represented students who had much higher than average standardized test scores, and those who were missing test scores. The researcher went through four processes to eliminate participants from the study. First and foremost, differing from main campus, the extended study satellite campuses and the virtual university have different criteria for admission requirements, placement procedures, course offerings, and prerequisites. As such, only students who were enrolled on main campus were included in the data set, reducing the participant number to 4,502. Secondly, not all students in the data set ultimately completed their first mathematics course, and withdrew at different times within their semester of enrollment. It was not statistically sound to treat these individuals as non-passing, and one cannot assume these students withdrew solely for academic

reasons. These participants were removed from the data set, reducing the participant number to 4,104. Thirdly, several students were granted admission with successful completion of dual credit courses, as well as AP exam credits. Although these individuals do have standardized test scores, they are not part of the target population desired for analysis since they did not complete their mathematics course at the study institution. These participants were removed from the data set, reducing the participant number to 3,450. Lastly, several students were enrolled at the university through dual credit partnerships with local high schools. Standardized test scores were not available for these students. Thus, the researcher only included participants who reported both SAT mathematics and verbal scores, as well as ACT mathematics and verbal scores, reducing the final participant number for analysis to 1,037.

Upon removal of extreme outliers, the researcher began the process of coding letter grades into grade points by applying the grade point scale used by the university. The researcher then ran preliminary descriptive statistics and normality plots to ensure that no assumptions were violated. The researcher then proceeded with running discriminant analysis, followed by three-way ANOVA. For the purpose of three-way ANOVA, the researcher recoded gender, ethnicity, and socioeconomic status (income) into categorical variables for analysis.

Ethical Considerations

Ethical considerations and practices were put in place to provide the utmost security of the data collected on students' academic and demographic information used during this analysis. In order to protect student anonymity, the Office of Institutional Research was asked to create a participant number for each student whose academic and demographic information was queried. The participant number was used for the sole purpose of analysis. In order to protect participant privacy and confidentiality, only the researcher and Office of Institutional Research had direct

access to the academic and demographic information collected. All data and its analysis were kept in encrypted files within a folder on the principal investigator's personal computer. The computer is in a secure location and requires a passcode for access. A backup copy of the data and its analysis was kept in the possession of the principal investigator in the form of an encrypted flash drive. Data, in the form of an SPSS file, were provided to the dissertation committee statistician upon request, and only in review of progress towards completion of the research study. For the purpose of answering questions and concerns that may arise, and for possible future research, any and all information collected during the study will be retained for two years following the completion of the study.

Chapter 4: Results

The purpose of this multivariate analysis study was to examine the extent to which standardized test scores predicted the performance of students in their first mathematics course, and displayed differential validity among various subgroups. This chapter will begin descriptive analysis for the data, and conclude with a detailed analysis for each research question.

Descriptive Statistics

Demographics. The preliminary data set received from the institution contained information for 6,642 students; a target sample from this population yielded 1,037 students. The demographic information for the target population is displayed in Table 1. The top three majors are Nursing (15.7%), Biology (10.4%), and Pre-Pharmacy (6.5%). It is worth noting that 62.5% of the study population is Hispanic, mirroring the demographics of the community the institution serves—of that Hispanic population, 69.3% are female. The typical age of traditional freshmen students is prevalent, with 88.3% of students 18 years old, and 8.3% of students 19 years old.

Standardized test scores. Descriptive statistics were run for the total standardized test scores of the SAT and ACT. SAT total scores ranged from 590-1420. The range of scores for the mathematics section was 260-750, and 270-790 for the verbal section. ACT total scores ranged from nine through 34. The range of scores for the mathematics section was seven through 35, and 13 through 34 for the verbal section. Figures 1-6 show the normal distribution for the total scores of each standardized test. The normality of these scores is important in the application of statistical analysis to build predictive models. Descriptives and tests for normality for mathematics and verbal components of each test will be discussed separately.

Table 1

Frequencies of Demographic Information

	Frequency	Percent of Participants
Gender		
Female	712	68.7
Male	325	31.3
Age		
17	29	2.8
18	916	88.3
19	86	8.3
20	4	0.4
26	1	0.1
27	1	0.1
Ethnicity		
American Indian or Alaska Native	4	0.4
Asian	28	2.7
Black	62	6.0
Hispanic	648	62.5
Native Hawaiian or Other Pacific Islander	1	0.1
Nonresident Alien	12	1.1
Two or more races	10	1.0
Unknown	79	7.6
White	193	18.6
First Generation Student		
Yes	463	44.6
No	574	55.4
Total	1037	100

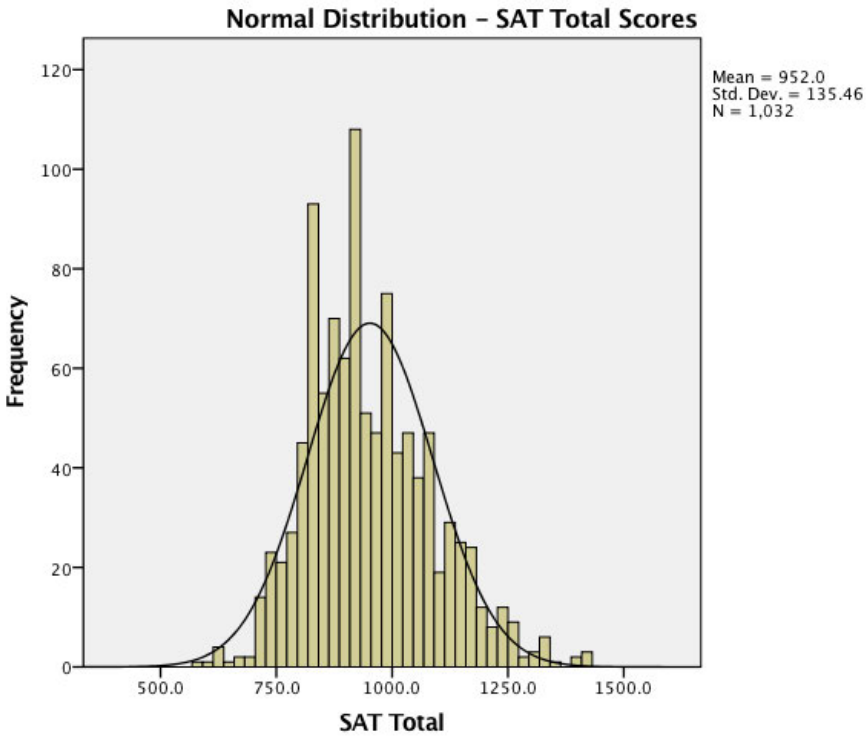

Figure 1. Normal Distribution of SAT Total Scores.

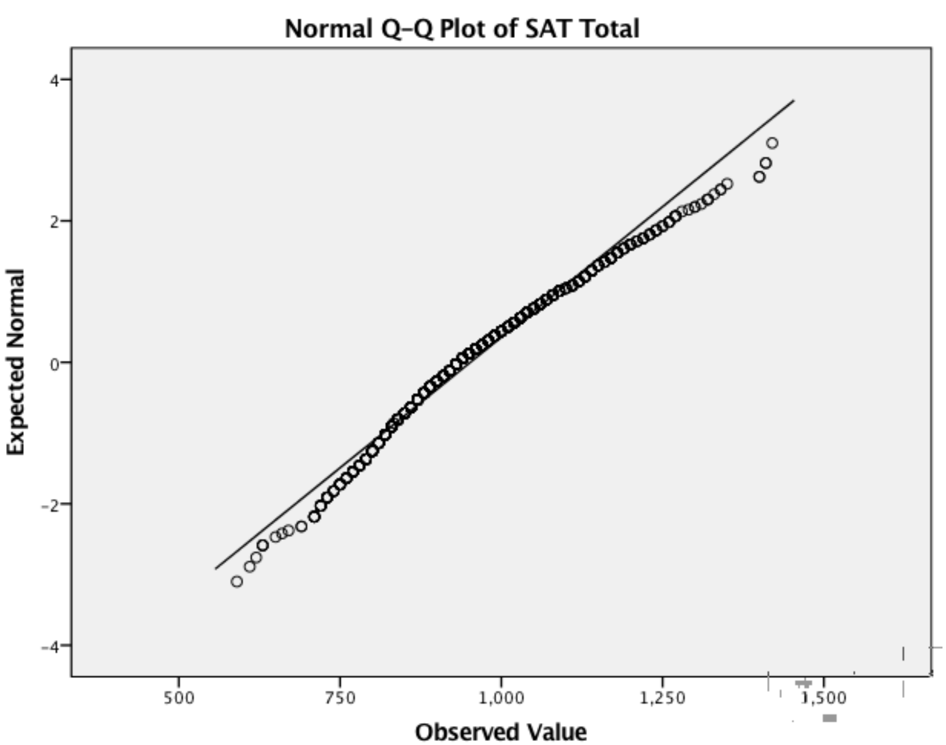

Figure 2. Q-Q Plot of SAT Total Scores.

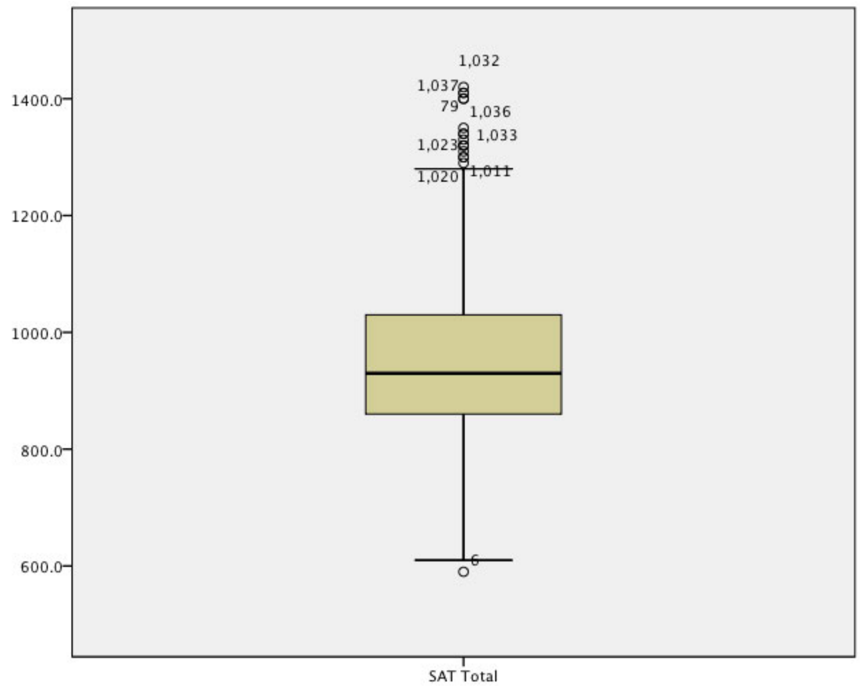

Figure 3. Boxplot of SAT Total Scores.

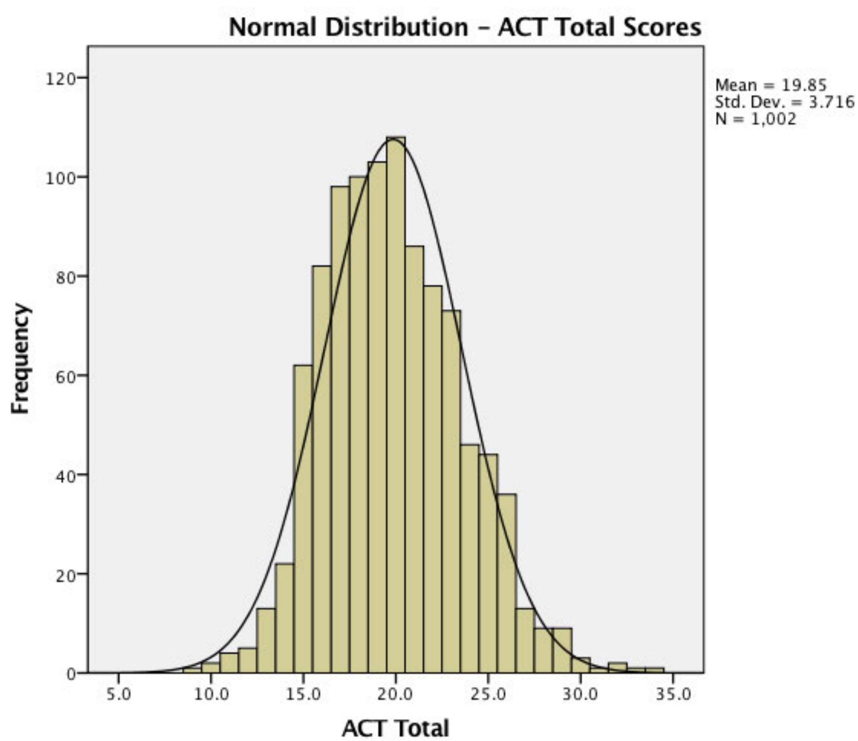

Figure 4. Normal Distribution of ACT Total Scores.

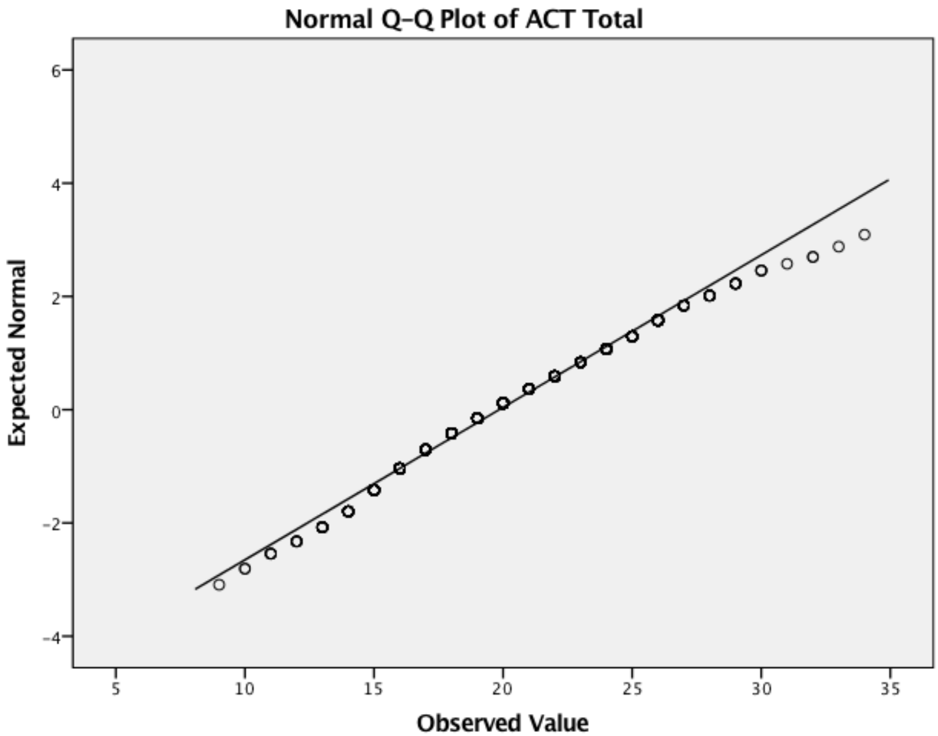

Figure 5. Q-Q Plot of ACT Total Scores.

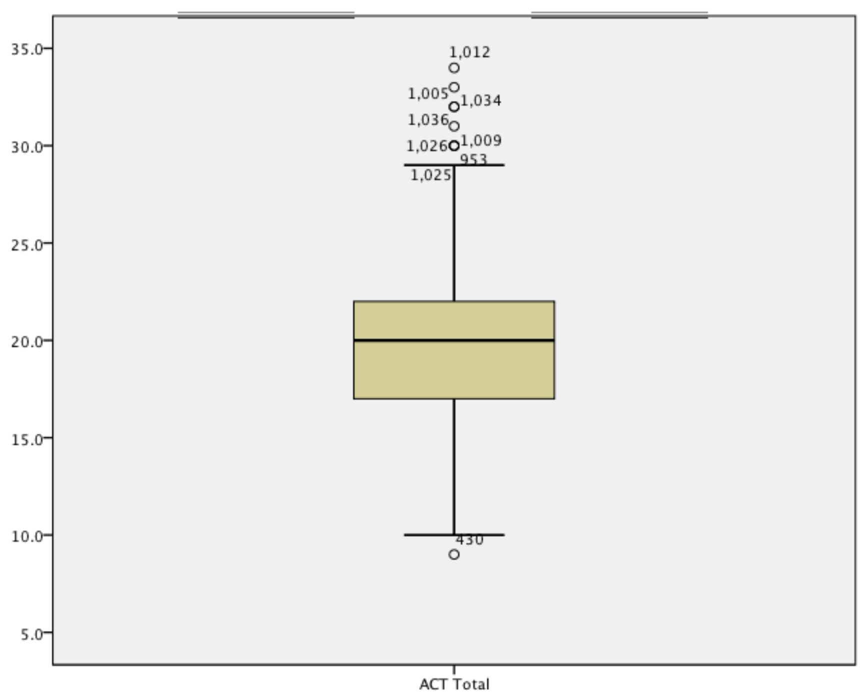

Figure 6. Boxplot of ACT Total Scores.

First semester of enrollment, courses taken, and grades received. Regarding the breakdown of course enrollment for first mathematics courses, it is refreshing to know that the majority of students take and pass their first mathematics course in the fall semester. Previous research suggests "students whose first academic experience in college is positive and successful are more likely to remain in school…[and are] likely to persist towards their goals" (Driscoll, 2007, p. 2). More importantly, it has been suggested that student procrastination in taking college courses leads to dramatically lower chances for success (Beswick, Rothblum, & Mann, 1988; Onwuegbuzie, 2004). Based on the data set, 34.3% of students were enrolled in developmental mathematics courses, and 65.7% were enrolled in college-level mathematics. The breakdown of each course and the grades received are shown in Tables 2-6.

Table 2

First Mathematics Course Grades for MATH 0318

	Frequency	Percent of Participants
A	19	20.9
A-	10	11.0
B+	16	17.6
B	11	12.1
B-	7	7.7
C+	5	5.5
C	13	14.3
D+	0	0
D	3	3.3
D-	4	4.4
F	3	3.3
Total	91	100

Table 3

First Mathematics Course Grades for MATH 0319

	Frequency	Percent of Participants
A	48	18.1
A-	27	10.2
B+	18	6.8
B	24	9.1
B-	31	11.7
C+	15	5.7
C	43	16.2
D+	6	2.3
D	19	7.2
D-	7	2.6
F	27	10.2
Total	265	100

Table 4

First Mathematics Course Grades for MATH 1304

	Frequency	Percent of Participants
A	132	32.5
A-	37	9.1
B+	32	7.9
B	43	10.6
B-	34	8.4
C+	22	5.4
C	48	11.8
D+	11	2.7
D	13	3.2
D-	6	1.5
F	28	6.9
Total	406	100

Table 5

First Mathematics Course Grades for MATH 1306

	Frequency	Percent of Participants
A	11	24.4
A-	4	8.9
B+	2	4.4
B	7	15.6
B-	9	20
C+	1	2.2
C	6	13.3
D+	2	4.4
D	0	0
D-	0	0
F	3	6.7
Total	45	100

Table 6

First Mathematics Course Grades for MATH 2303

	Frequency	Percent of Participants
A	86	37.4
A-	24	10.4
B+	29	12.6
B	21	9.1
B-	13	5.7
C+	9	3.9
C	22	9.6
D+	5	2.2
D	6	2.6
D-	0	0
F	15	6.5
Total	230	100

Assessing Normality of the Independent Variables

According to West, Finch, and Curran (1995), tests for normality for sample sizes larger than 300 should depend on histograms and absolute values of skewness and kurtosis. Absolute values of skewness larger than two, and absolute values of kurtosis larger than seven suggest substantial nonnormality. Due to the large sample size (n > 50), the researcher focused on the results of the Kolmogorov-Smirnov test for significance over the Shapiro-Wilk test for normality. It must be noted, as previously discussed in Chapter 3, "discriminant analysis is robust to failures of normality and homoscedasticity if the violation is caused by skewness rather than outliers" (Tabachnick & Fidell, 2006, p. 382). It is also important to note that with a sample size larger than 30, three-way ANOVA is robust to departures from normality and unequal variance (Pallant, 2007, Tabachnick & Fidell, 2006). The researcher used these guidelines to determine normality of each independent variable.

SAT mathematics scores. The descriptives of the independent variable SAT mathematics scores (SATM), shown in Table 7, presents a calculated mean of 478.05, and a standard deviation of 74.04. The histogram (Figure 7) is close in resemblance to the normal curve. In terms of skewness, the QQ-plot (Figure 8) displays fairly linear data, with slight sagging and the first and last segments, possibly attributed to outliers. The results of the Kolmogorov-Smirnov test presents a p-value less than 0.05, suggesting the data have violated the assumption of normality; however, this is quite common in larger samples. The boxplot (Figure 9) gives visual identification of close symmetry in both halves of the box and whiskers, identifying six outliers in the data, but they are not extreme. To further examine skewness and kurtosis, their absolute values are examined. The absolute value of skewness is 0.38, and the absolute value of kurtosis is 0.26, which is not significant enough to discredit normality.

Table 7

Descriptive Statistics for SAT Mathematics Scores

		Statistic	Std. Error
Mean		478.05	2.29
95% Confidence	Lower Bound	473.53	
Interval for Mean	Upper Bound	482.56	
5% Trimmed Mean		476.42	
Median		470.00	
Variance		5482.50	
Std. Deviation		74.04	
Minimum		260.00	
Maximum		750.00	
Range		4900	
Interquartile Range		100.00	
Skewness		0.38	0.08
Kurtosis		0.26	0.15

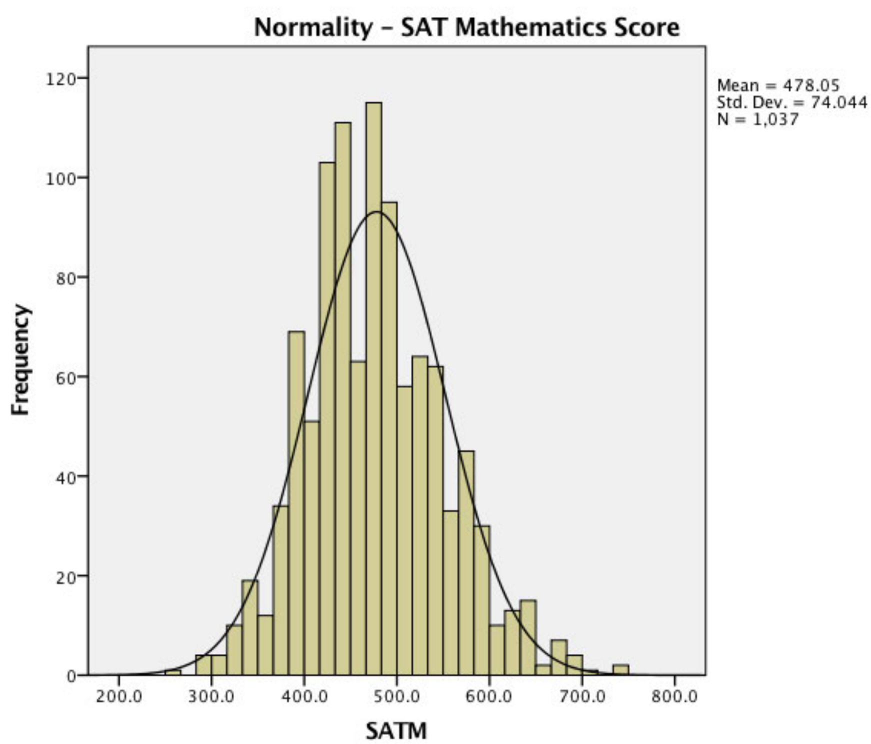

Figure 7. Normal Distribution of SAT Mathematics Scores.

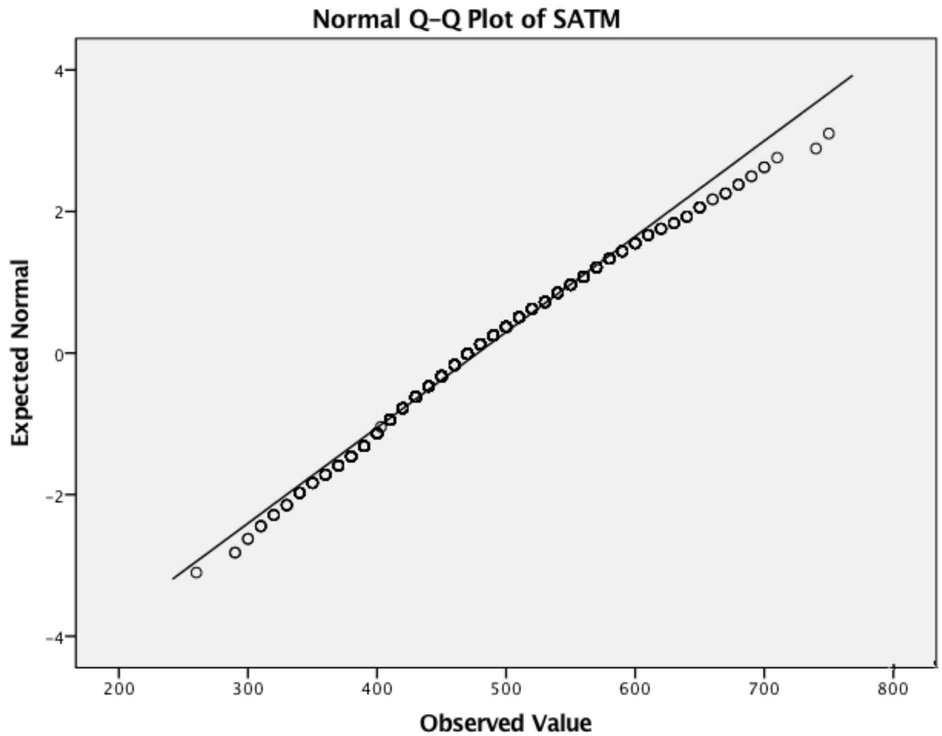

Figure 8. Q-Q Plot of SAT Mathematics Scores.

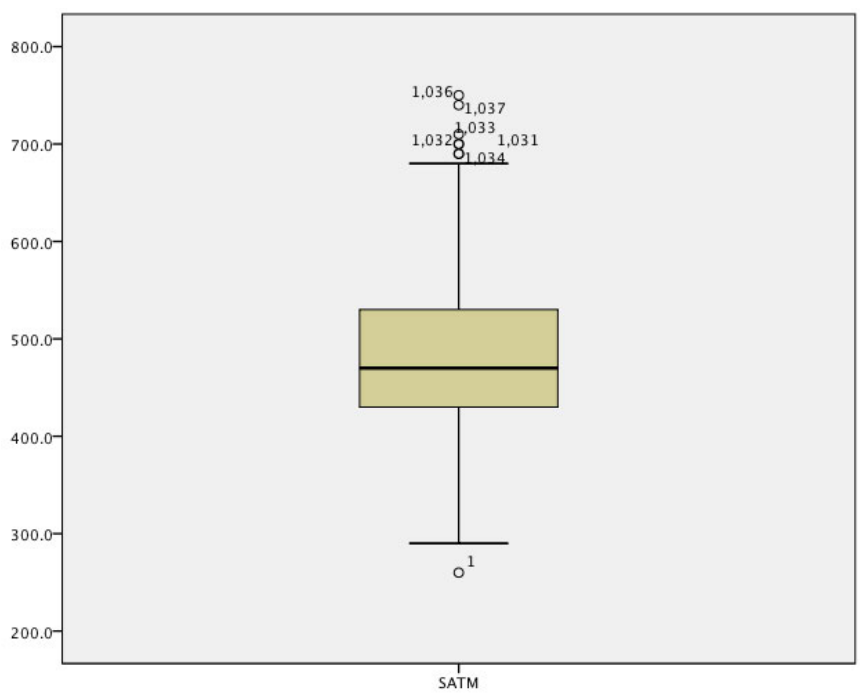

Figure 9. Boxplot of SAT Mathematics Scores.

SAT verbal scores. The descriptives of the independent variable SAT verbal scores (SATV), shown in Table 8, presents a calculated mean of 469.72, and a standard deviation of 75.55. The histogram (Figure 10) is close in resemblance to the normal curve. In terms of skewness, the QQ-plot (Figure 11) displays fairly linear data, with slight sagging and the first and last segments, possibly attributed to outliers. The results of the Kolmogorov-Smirnov test presents a p-value less than 0.05, suggesting the data have violated the assumption of normality; however, this is quite common in larger samples. The boxplot (Figure 12) gives visual identification of close symmetry in both halves of the box and whiskers, identifying nine outliers in the data, but they are not extreme. To further examine skewness and kurtosis, their absolute values are examined. The absolute value of skewness is 0.59, and the absolute value of kurtosis is 0.56, which is not significant enough to discredit normality.

Table 8

Descriptive Statistics for SAT Verbal Scores

		Statistic	Std. Error
Mean		469.72	2.35
95% Confidence	Lower Bound	465.12	
Interval for Mean	Upper Bound	474.32	
5% Trimmed Mean		467.20	
Median		460.00	
Variance		5707.16	
Std. Deviation		75.55	
Minimum		270.00	
Maximum		790.00	
Range		520.00	
Interquartile Range		100.00	
Skewness		0.59	0.08
Kurtosis		0.56	0.15

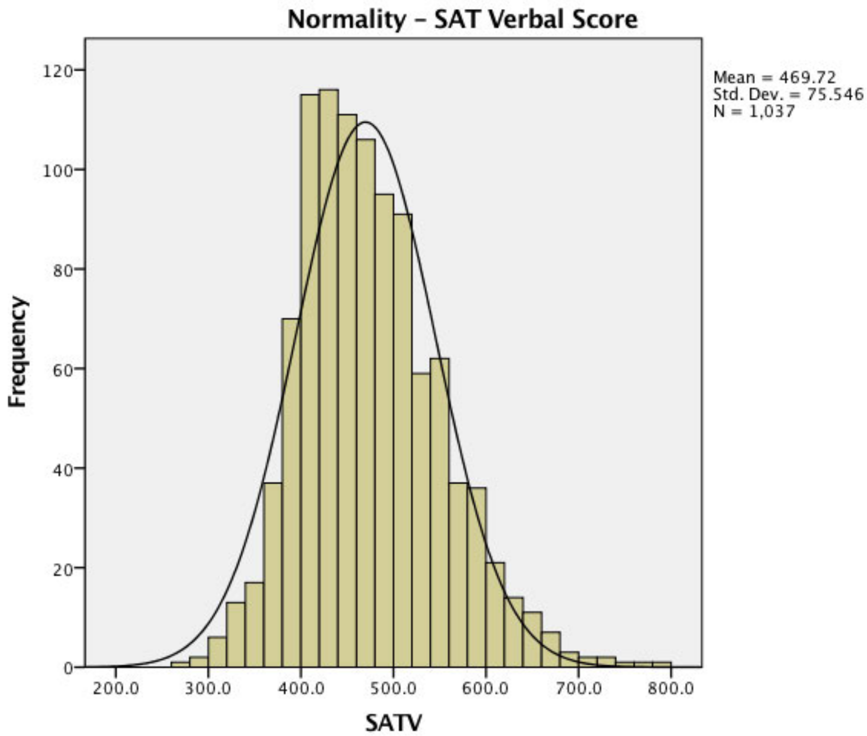

Figure 10. Normal Distribution of SAT Verbal Scores.

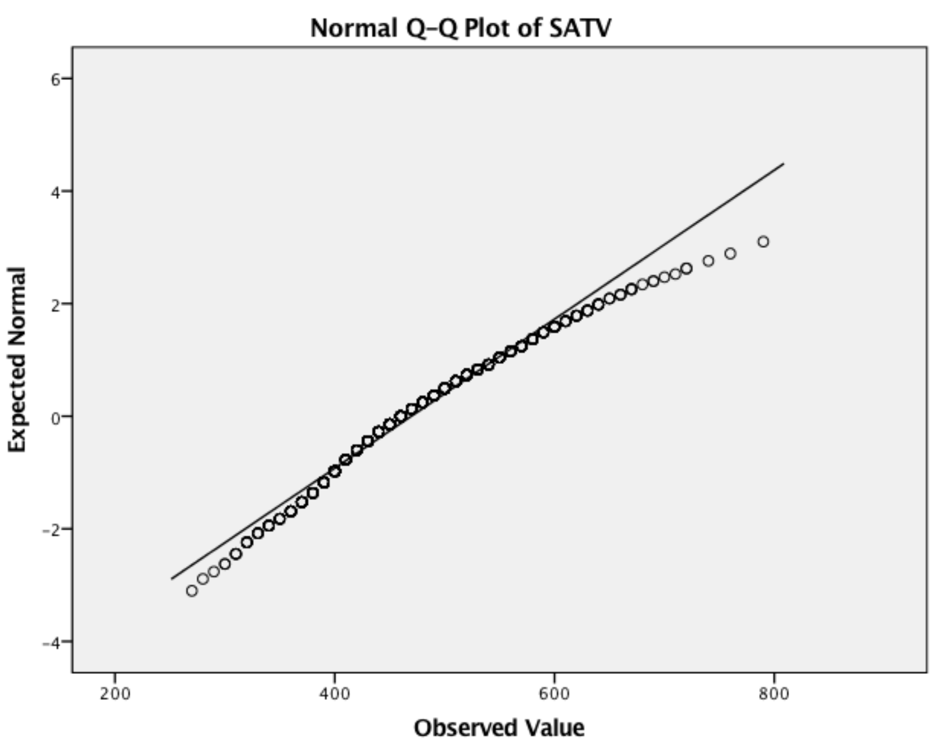

Figure 11. Q-Q Plot of SAT Verbal Scores.

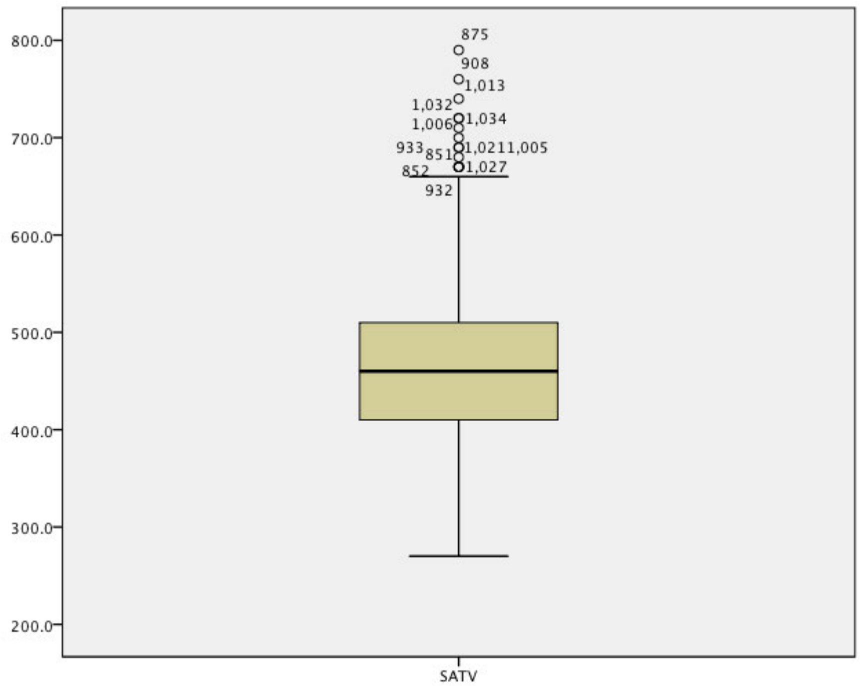

Figure 12. Boxplot of SAT Verbal Scores.

ACT mathematics scores. The descriptives of the independent variable ACT mathematics scores (ACTM), shown in Table 9, presents a calculated mean of 19.92, and a standard deviation of 3.81. The histogram (Figure 13) is close in resemblance to the normal curve, with slight positive skewness to the right. In terms of skewness, the QQ-plot (Figure 14) displays sagging in the first and last segments, and bubbling in the second segment, possibly attributed to outliers. The results of the Kolmogorov-Smirnov test presents a p-value less than 0.05, suggesting the data have violated the assumption of normality; however, this is quite common in larger samples. The boxplot (Figure 15) gives visual identification of a lack of symmetry in both halves of the box and whiskers, identifying three outliers in the data. To further examine skewness and kurtosis, their absolute values are examined. The absolute value of skewness is 0.68, and the absolute value of kurtosis is 0.01, which is not significant enough to discredit normality.

Table 9

Descriptive Statistics for ACT Mathematics Scores

		Statistic	Std. Error
Mean		19.92	0.12
95% Confidence	Lower Bound	19.68	
Interval for Mean	Upper Bound	20.15	
5% Trimmed Mean		19.73	
Median		19.00	
Variance		14.55	
Std. Deviation		3.81	
Minimum		13.00	
Maximum		34.00	
Range		21.00	
Interquartile Range		6.00	
Skewness		0.68	0.08
Kurtosis		-0.06	0.15

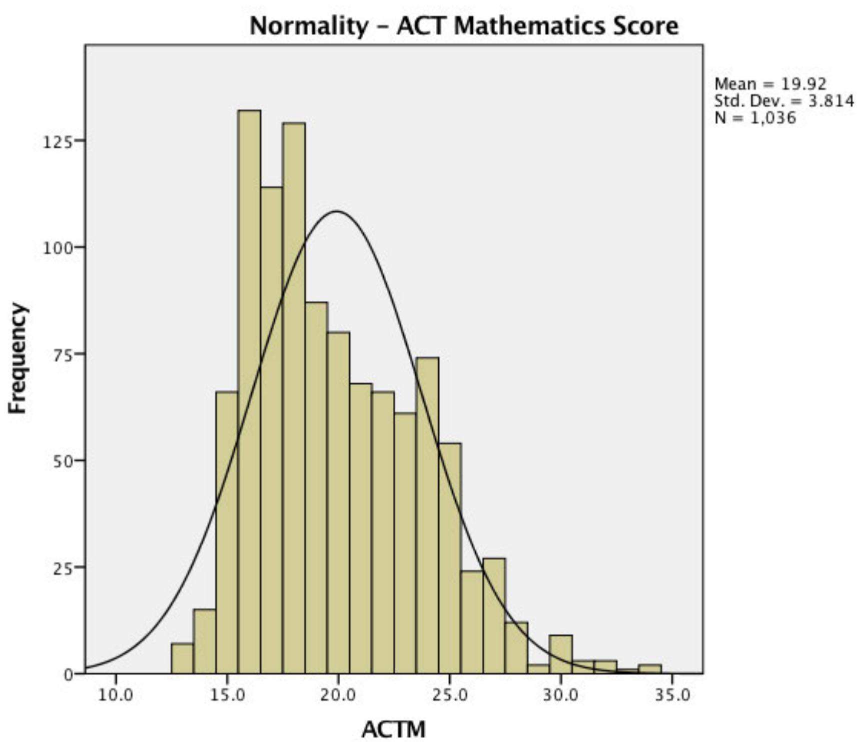

Figure 13. Normal Distribution of ACT Mathematics Scores.

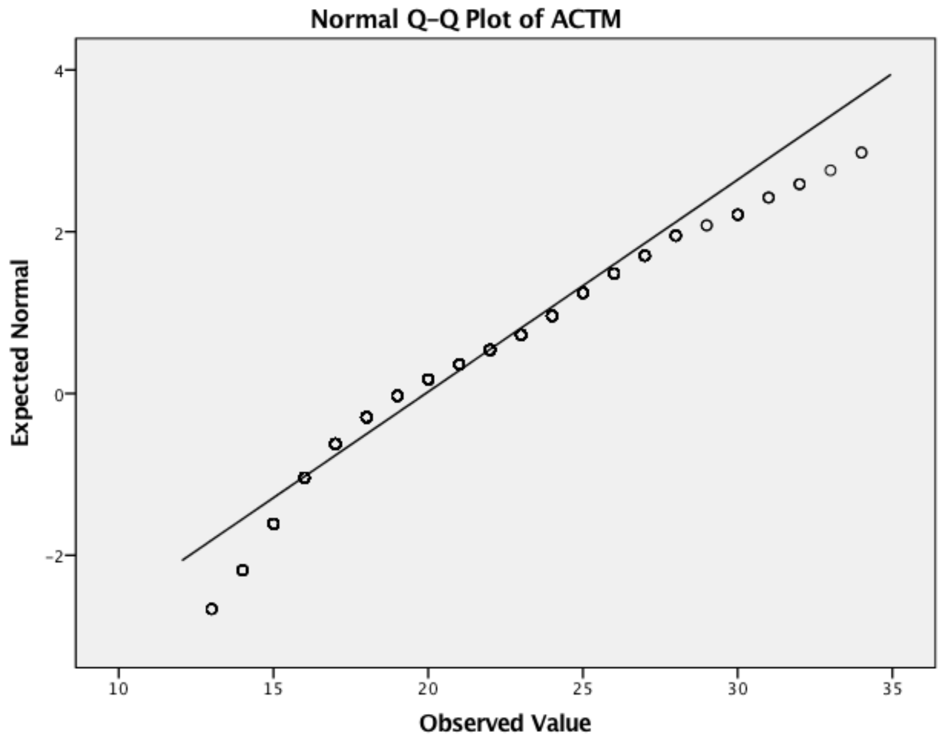

Figure 14. Q-Q Plot of ACT Mathematics Scores

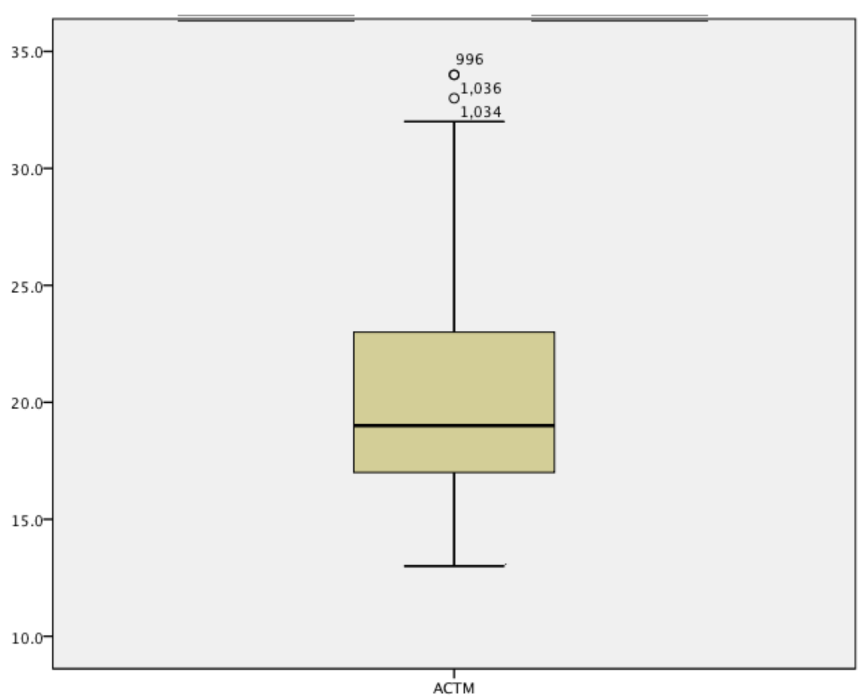

Figure 15. Boxplot of ACT Mathematics Scores.

ACT verbal scores. The descriptives of the independent variable ACT verbal scores (ACTV), shown in Table 10, presents a calculated mean of 19.08, and a standard deviation of 4.81. The histogram (Figure 16) is close in resemblance to the normal curve. In terms of skewness, the QQ-plot (Figure 17) displays fairly linear data, with slight sagging in the last segment, possibly attributed to outliers. The results of the Kolmogorov-Smirnov test presents a p-value less than 0.05, suggesting the data have violated the assumption of normality; however, this is quite common in larger samples. The boxplot (Figure 18) gives visual identification of close symmetry in both halves of the box and whiskers, identifying eight outliers in the data, but they are not extreme. To further examine skewness and kurtosis, their absolute values are examined. The absolute value of skewness is 0.14, and the absolute value of kurtosis is 0.01, which is not significant enough to discredit normality.

Table 10

Descriptive Statistics for ACT Verbal Scores

		Statistic	Std. Error
Mean		19.08	0.15
95% Confidence	Lower Bound	18.78	
Interval for Mean	Upper Bound	19.37	
5% Trimmed Mean		19.03	
Median		19.00	
Variance		23.13	
Std. Deviation		4.81	
Minimum		7.00	
Maximum		35.00	
Range		28.00	
Interquartile Range		6.00	
Skewness		0.14	0.08
Kurtosis		-0.01	0.15

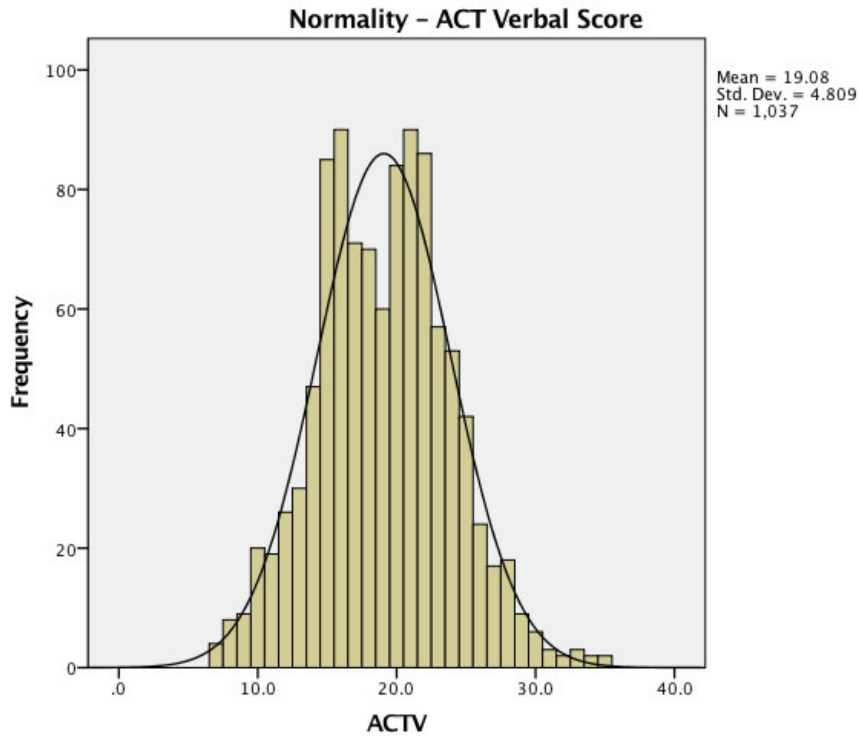

Figure 16. Normal Distribution of ACT Verbal Scores.

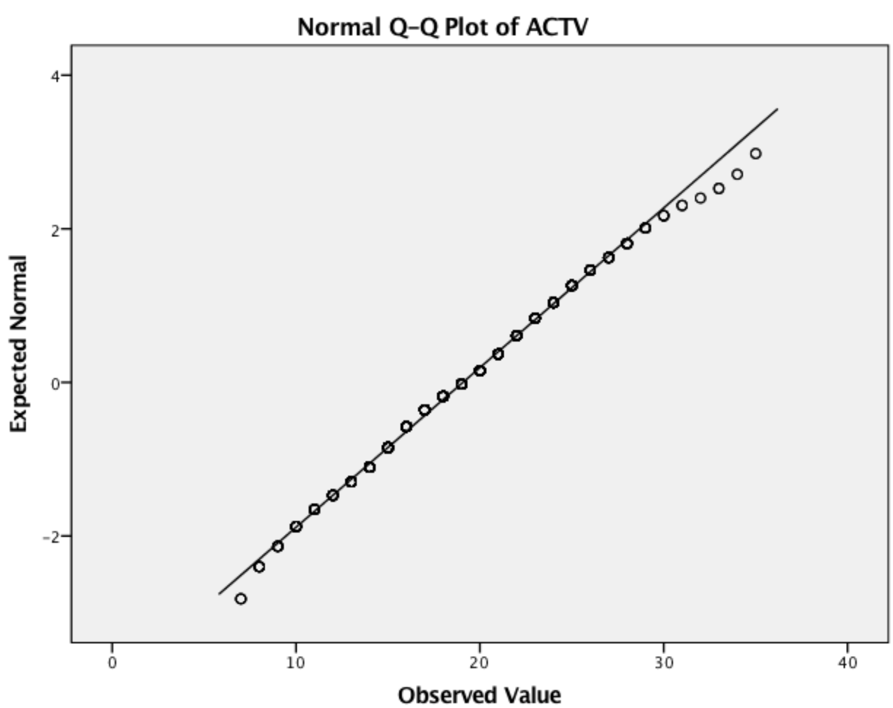

Figure 17. Q-Q Plot of ACT Verbal Scores.

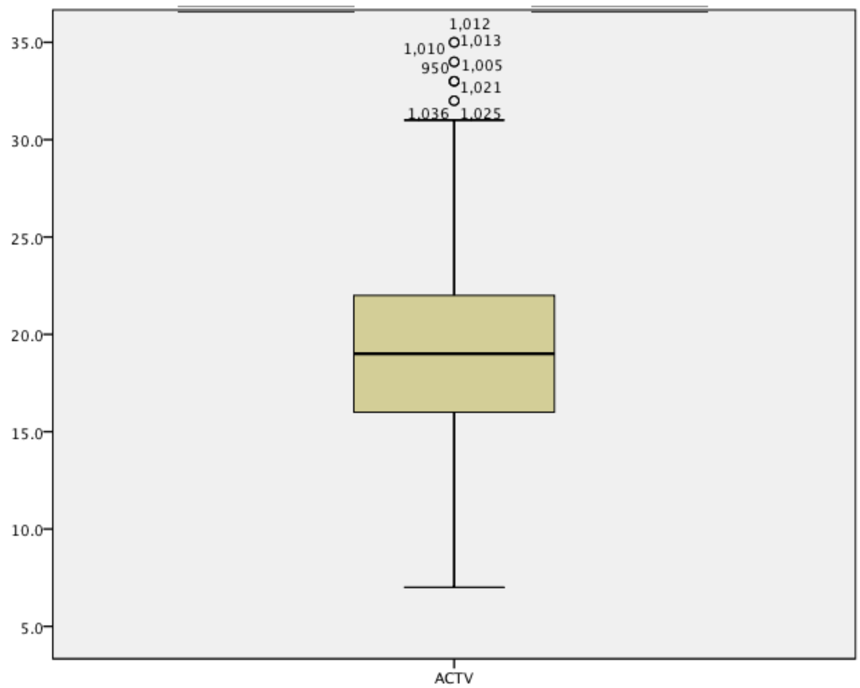

Figure 18. Boxplot – ACT Verbal Scores.

High school GPA. The descriptives of the independent variable high school GPA (HS GPA), shown in Table 11, present a calculated mean of 3.51, and a standard deviation of 0.34. The histogram (Figure 19) does exhibit slight negative skewness to the left. In terms of skewness, the QQ-plot (Figure 20) displays bubbling in the first half of the plot, possibly attributed to outliers. The results of the Kolmogorov-Smirnov test presents a p-value less than 0.05, suggesting the data have violated the assumption of normality; however, this is quite common in larger samples. The boxplot (Figure 21) gives visual identification of a lack of symmetry in both halves of the box and whiskers, identifying 15 outliers in the data. To further examine skewness and kurtosis, their absolute values are examined. The absolute value of skewness is 0.84, and the absolute value of kurtosis is 1.02, which is not significant enough to discredit normality.

Table 11

Descriptive Statistics for High School GPA

		Statistic	Std. Error
Mean		3.51	0.01
95% Confidence	Lower Bound	3.49	
Interval for Mean	Upper Bound	3.53	
5% Trimmed Mean		3.53	
Median		3.55	
Variance		0.12	
Std. Deviation		0.34	
Minimum		2.00	
Maximum		4.00	
Range		2.00	
Interquartile Range		0.45	
Skewness		-0.84	0.08
Kurtosis		1.02	0.15

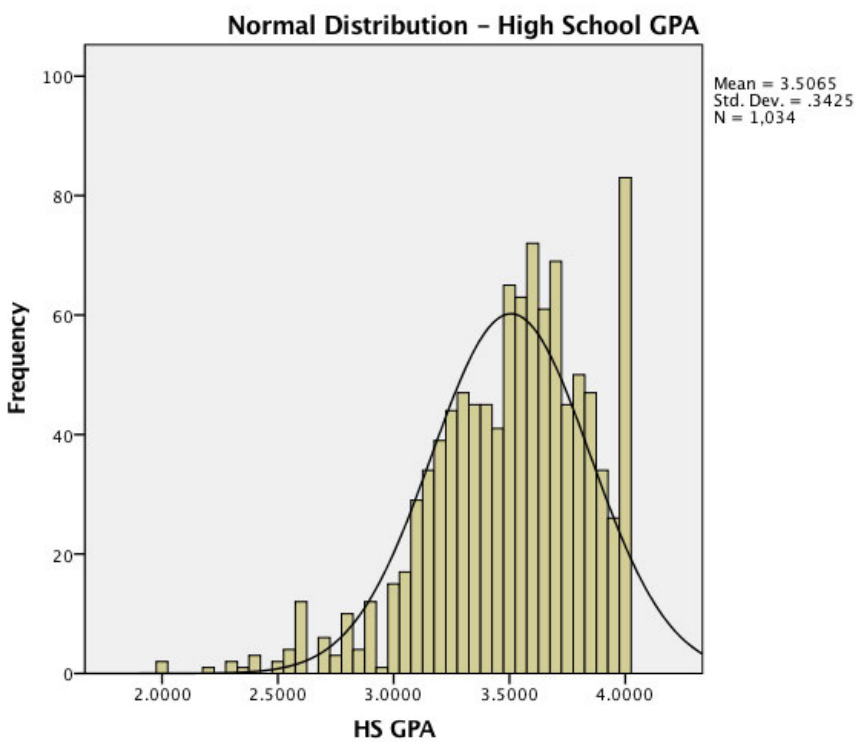

Figure 19. Normal Distribution of High School GPA.

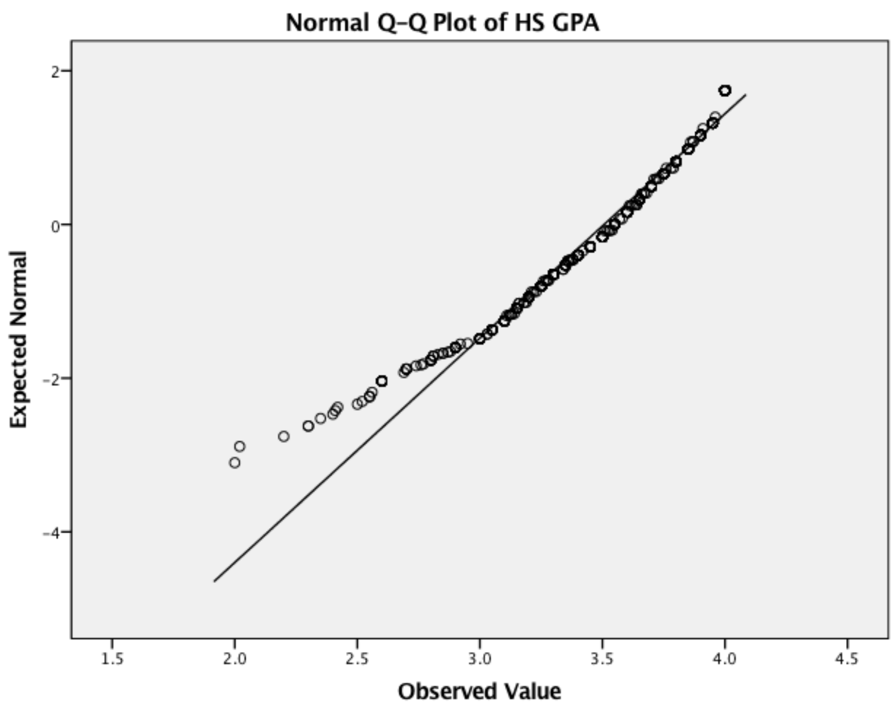

Figure 20. Q-Q Plot of High School GPA.

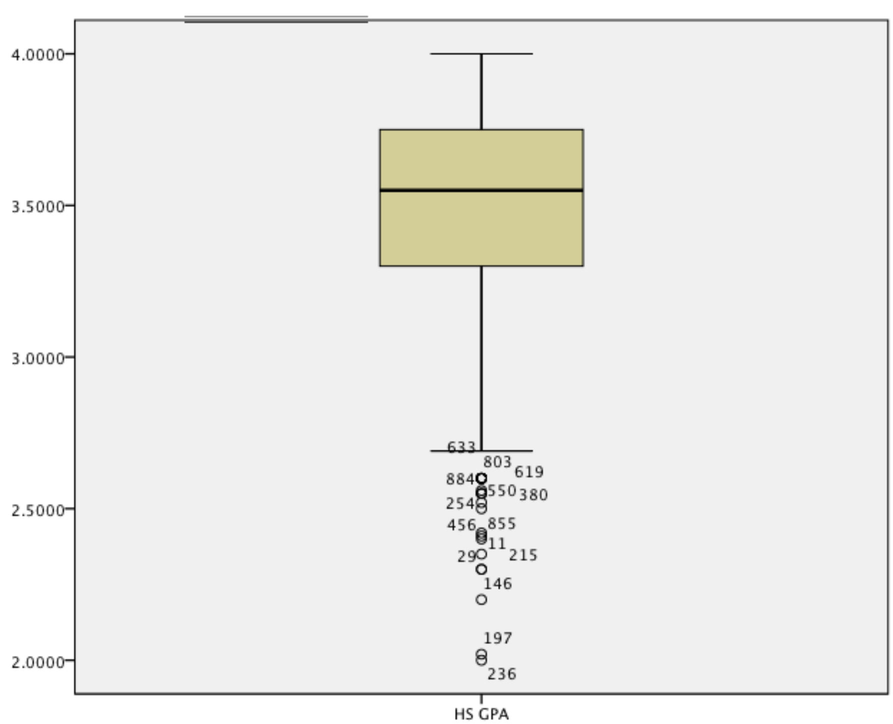

Figure 21. Boxplot of High School GPA.

High school rank. The descriptives of the independent variable high school rank (HS Rank), shown in Table 12, present a calculated mean of 115.61, and a standard deviation of 113.42. The histogram (Figure 22) exhibits significant positive skewness to the right. In terms of skewness, the QQ-plot (Figure 23) displays heavy bubbling in the first half of the plot, with sagging in last segment, possibly attributed to outliers. The results of the Kolmogorov-Smirnov test presents a p-value less than 0.05, suggesting the data have violated the assumption of normality; however, this is quite common in larger samples. The boxplot (Figure 24) gives visual identification of a lack of symmetry in both halves of the box and whiskers, identifying over 20 outliers in the data, two of which are extreme. To further examine skewness and kurtosis, their absolute values are examined. The absolute value of skewness is 1.69, and the absolute value of kurtosis is 3.93, which is not significant enough to discredit normality.

Table 12

Descriptive Statistics for High School Rank

		Statistic	Std. Error
Mean		115.61	4.01
95% Confidence	Lower Bound	107.75	
Interval for Mean	Upper Bound	123.47	
5% Trimmed Mean		103.93	
Median		76.00	
Variance		12864.83	
Std. Deviation		113.42	
Minimum		1.00	
Maximum		837.00	
Range		836.00	
Interquartile Range		132.30	
Skewness		1.69	0.09
Kurtosis		3.93	0.17

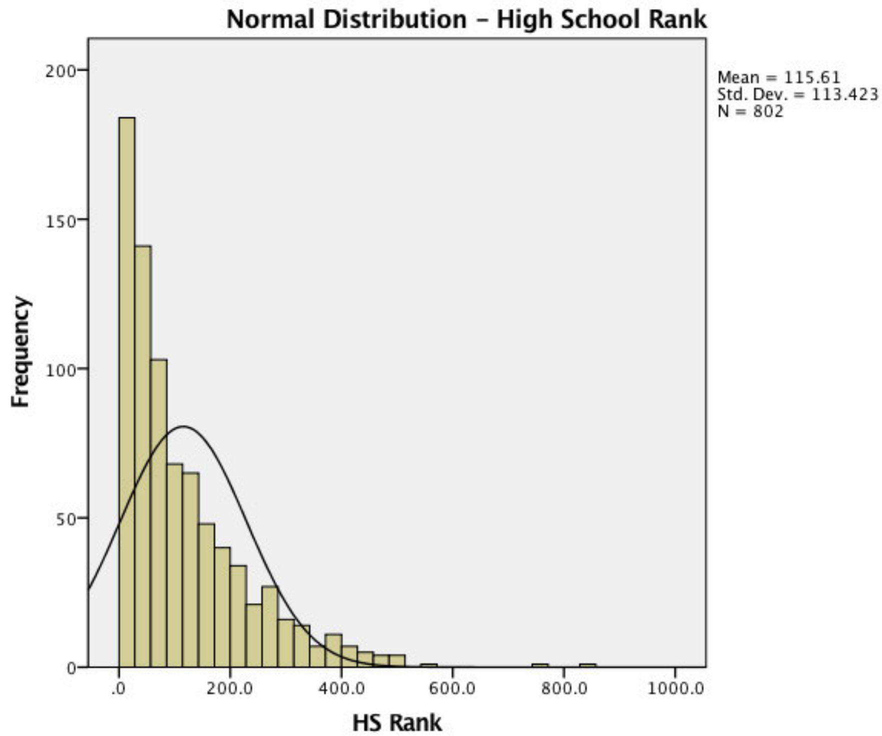

Figure 22. Normal Distribution of High School Rank.

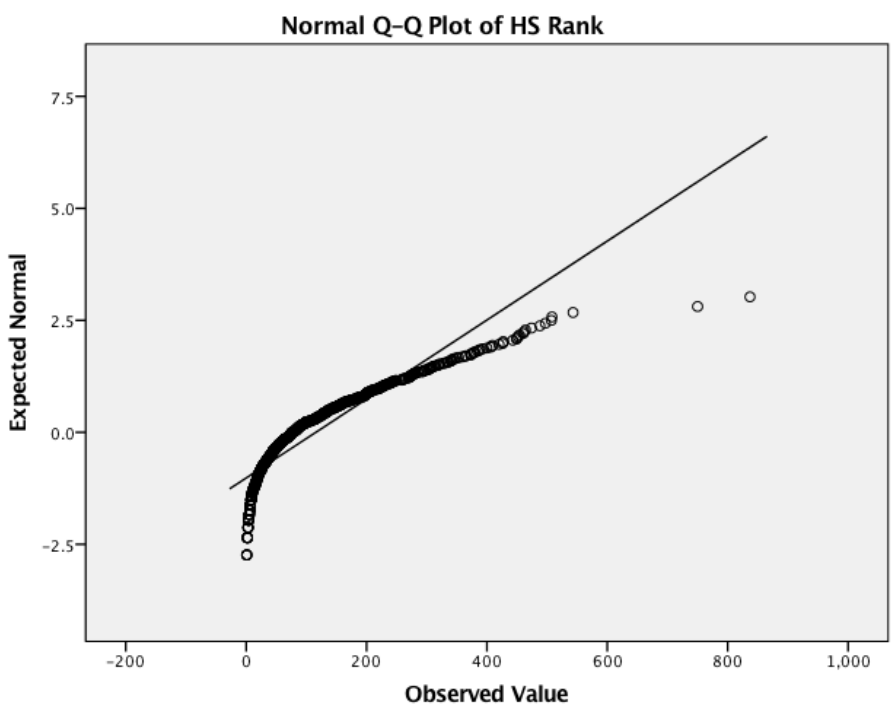

Figure 23. Q-Q Plot of High School Rank.

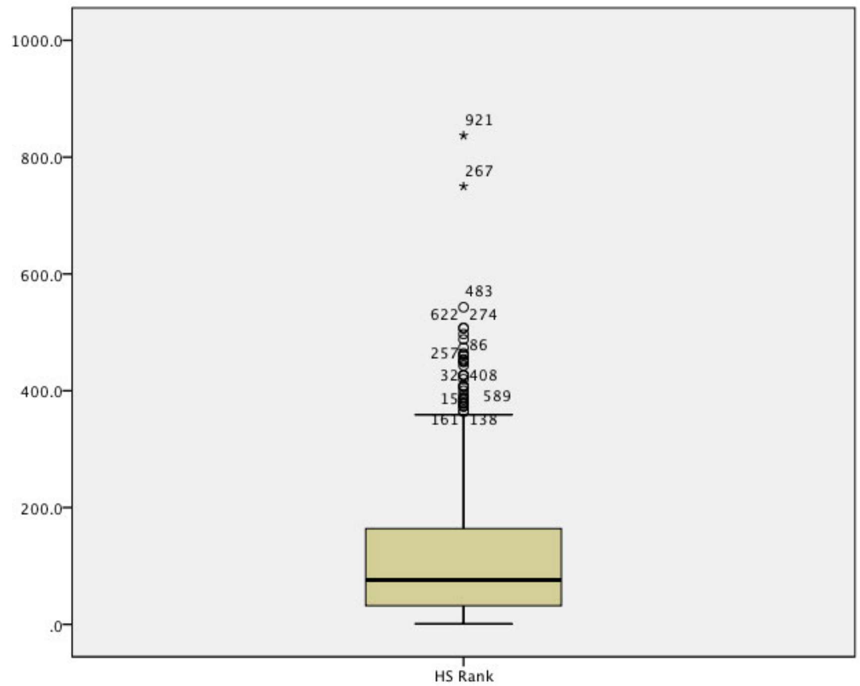

Figure 24. Boxplot of High School Rank.

Gender. The variable for gender was recoded from a string variable to a numeric variable for the purpose of analysis. Gender was coded into two groups: 0 – male, 1 – female. Descriptives of the independent variable gender, shown in Table 13, present a calculated mean of 0.69, and a standard deviation of 0.46. The histogram is close in resemblance to the normal curve. In terms of skewness, the QQ-plot displays fairly linear data. The results of the Kolmogorov-Smirnov test presents a p-value less than 0.05, suggesting the data have violated the assumption of normality; however, this is quite common in larger samples. The boxplot does not give visual identification of close symmetry in both halves of the box and whiskers, as we only have two genders to plot. To further examine skewness and kurtosis, their absolute values are examined. The absolute value of skewness is 0.81, and the absolute value of kurtosis is 1.35, which is not significant enough to discredit normality.

Table 13

Descriptive Statistics for Gender

		Statistic	Std. Error
Mean		0.69	0.01
95% Confidence	Lower Bound	0.66	
Interval for Mean	Upper Bound	0.72	
5% Trimmed Mean		0.71	
Median		1.00	
Variance		0.22	
Std. Deviation		0.46	
Minimum		0.00	
Maximum		1.00	
Range		1.00	
Interquartile Range		1.00	
Skewness		-0.81	0.08
Kurtosis		-1.35	0.15

Ethnicity. The variable for ethnicity was recoded from a string variable to a numeric variable for the purpose of analysis. Initially, ethnicity was coded into nine groups; however, four groups were too small for three-way ANOVA to determine interactions based on ethnicity. Ethnicity was recoded into six groups for successful analysis:

1) Asian
2) Black
3) Hispanic
4) Unknown
5) White
6) Other (to include American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Nonresident Alien, and Two or more races).

Descriptives of the independent variable ethnicity, shown in Table 14, present a calculated mean of 3.41, and a standard deviation of 1.03. The histogram is close in resemblance to the normal curve. In terms of skewness, the QQ-plot displays fairly linear data, with slight sagging in the

last segment, possibly attributed to outliers. The results of the Kolmogorov-Smirnov test presents a p-value less than 0.05, suggesting the data have violated the assumption of normality; however, this is quite common in larger samples. The boxplot gives visual identification of a lack of symmetry in both halves of the box, but the whiskers are symmetrical, identifying six outliers in the data. To further examine skewness and kurtosis, their absolute values are examined. The absolute value of skewness is 0.57, and the absolute value of kurtosis is 0.19, which is not significant enough to discredit normality.

Table 14

Descriptive Statistics for Ethnicity

		Statistic	Std. Error
Mean		3.41	0.03
95% Confidence	Lower Bound	3.35	
Interval for Mean	Upper Bound	3.48	
5% Trimmed Mean		3.40	
Median		3.00	
Variance		1.05	
Std. Deviation		1.03	
Minimum		1.00	
Maximum		6.00	
Range		5.00	
Interquartile Range		1.00	
Skewness		0.57	0.08
Kurtosis		0.19	0.15

Income. The variable for income was recoded from a string variable to a numeric variable for the purpose of analysis. Income was coded into 10 groups:

0) $0-$10k
1) $10k-$20k
2) $20k-$30k
3) $30k-$40k
4) $40k-$50k
5) $50k-$60k
6) $60k-$70k
7) $70k-$80k
8) $80k-$90k
9) $90k-$100k
10) $100k+.

Descriptives of the independent variable income, shown in Table 15, present a calculated mean of 5.64, and a standard deviation of 3.44. The histogram (Figure 25) does not resemble normal distribution, showing a disparity in income levels. In terms of skewness, the QQ-plot (Figure 26) displays pronounced bubbling in the first half, and pronounced sagging in the last half, possibly attributed to outliers. The results of the Kolmogorov-Smirnov test presents a p-value less than 0.05, suggesting the data have violated the assumption of normality; however, this is quite common in larger samples. The boxplot (Figure 27) gives visual identification of an absence of exact symmetry in both halves of the box and whiskers; however, there are no outliers. To further examine skewness and kurtosis, their absolute values are examined. The absolute value of

skewness is 0.06, and the absolute value of kurtosis is 1.45, which is not significant enough to discredit normality.

Table 15

Descriptive Statistics for Income

		Statistic	Std. Error
Mean		5.64	0.11
95% Confidence	Lower Bound	5.42	
Interval for Mean	Upper Bound	5.86	
5% Trimmed Mean		5.70	
Median		6.00	
Variance		11.82	
Std. Deviation		3.44	
Minimum		0.00	
Maximum		10.00	
Range		10.00	
Interquartile Range		7.00	
Skewness		-0.06	0.08
Kurtosis		-1.45	0.16

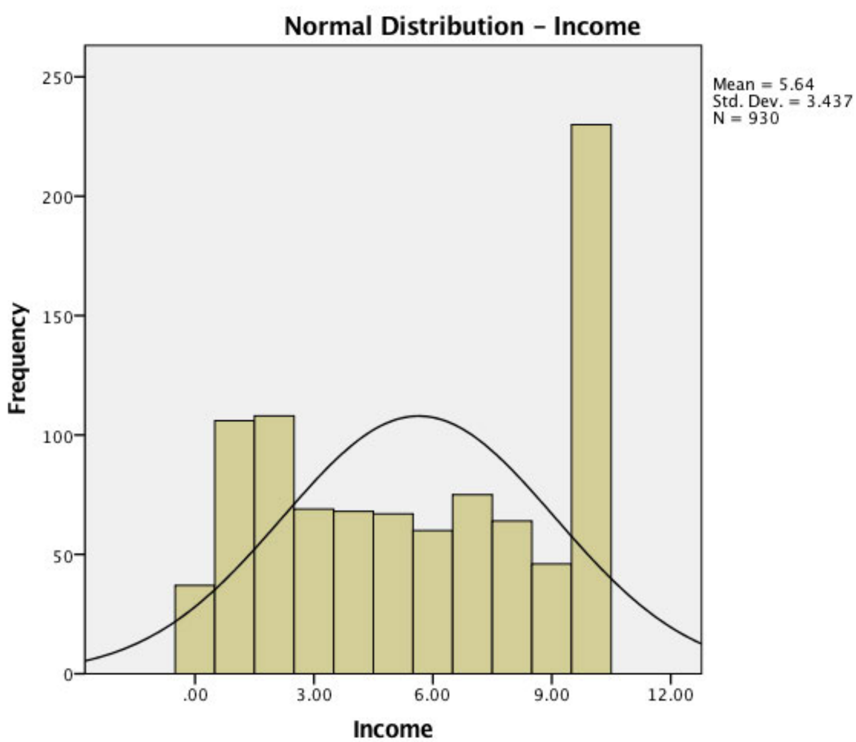

Figure 25. Normal Distribution of Income.

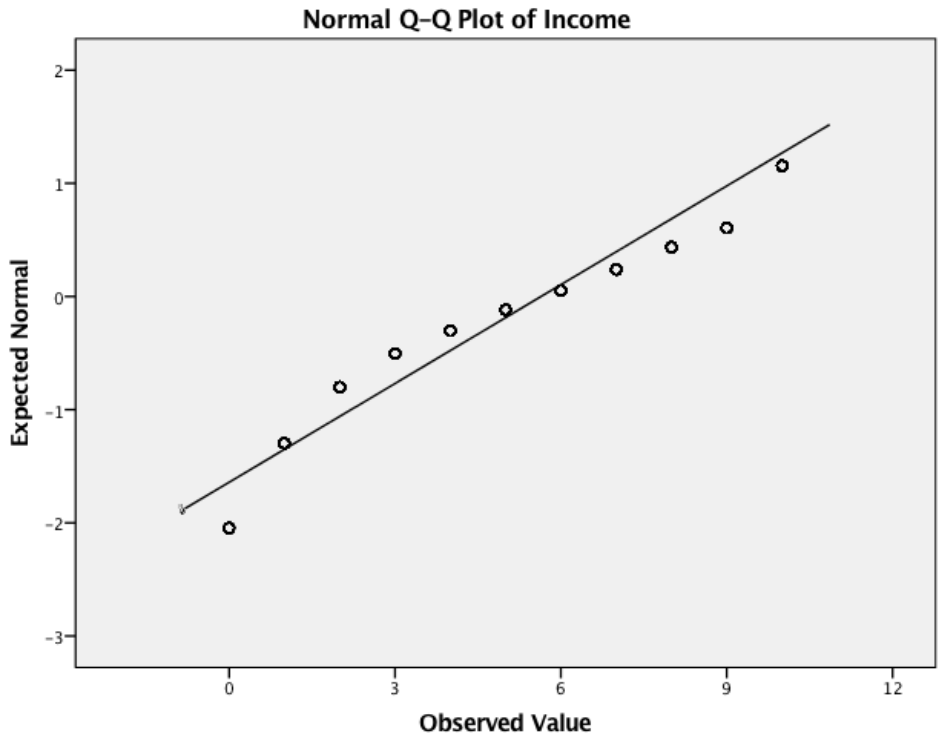

Figure 26. Q-Q Plot of Income.

Figure 27. Boxplot of Income.

Research Questions

Research question one. What is the extent to which mathematics and verbal standardized test scores predict the performance of students in their first-year mathematics courses?

The data were first evaluated with respect to practical limitations of discriminant analysis. The dependent variable was categorical and dichotomous (passing and not passing), the sample size of the smallest group exceeded the number of predictor variables, evaluation of assumptions of linearity (Figure 28) and normality were satisfactory for predictor variables, and the variable used to discriminate between groups was not completely redundant. A direct discriminant analysis was performed using several scale measurements available in the data set, which the researcher, based on previous studies and limitations, deemed most viable for placing students in mathematics courses. The best predictive model was built using a process of layering predictor variables in what ultimately totaled to nine predictive models until the best model for group assignment emerged.

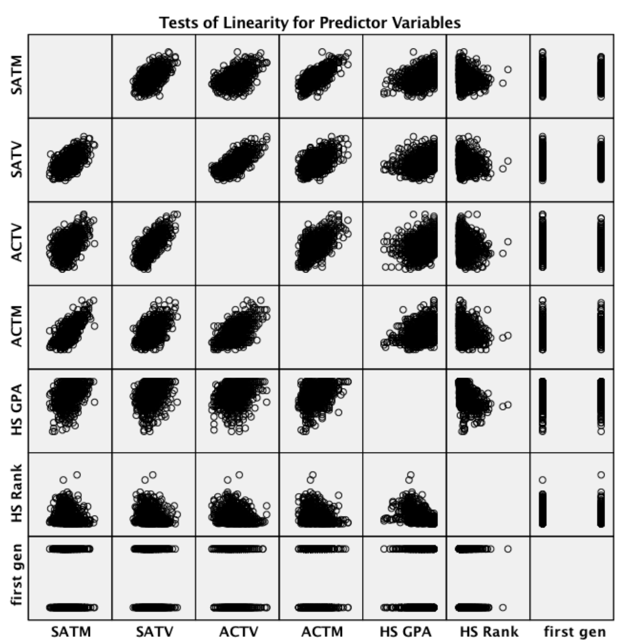

Figure 28. Tests of Linearity for Predictor Variables for Discriminant Analysis.

The first six models looked at each predictor variable alone – these included SAT mathematics scores, SAT verbal scores, ACT mathematics scores, ACT verbal scores, high school GPA, and high school rank. Predictive Model Seven included the SAT mathematics and verbal scores, along with high school GPA, and high school rank. Predictive Model Eight included the ACT mathematics and verbal scores, along with high school GPA, and high school rank. Predictive Model Nine included all six predictor variables. It is important to note that although first generation student status was thought to be significant as a predictor variable, the discriminant function in SPSS did not label it as a valid predictor. In fact, when placed in a separate test of Predictive Model Nine, it decreased the overall predictive ability and therefore was completely omitted from analysis.

Predictive model one. The first predictor model tested used only student SAT mathematics scores (SATM). All 1,037 cases were present in the analysis. Refer to Table 16 for more information regarding group statistics. Group membership was categorized as passing (1) and not passing (0) based on letter grades and grade point scale used by the university of interest. Passing was considered any letter grade greater than or equal to a “C”, with a grade point value greater than or equal to 2.00. Not passing was considered any letter grade less than or equal to a “D+”, with a grade point value less than or equal to 1.30. Tests for equality of group means suggest that group mean differences are significant for the predictor variable SATM ($p < .05$). Log determinants are almost equal for the two groups, confirming the assumption of equality of covariance matrices. Box’s M test of equality of covariance matrices present a significance greater than 0.05, suggesting that data do not differ significantly from multivariate normal, thus we may proceed with analysis.

The discriminant analysis function in SPSS automatically determined the optimal combination of variables in the first function calculated, providing the best level of discrimination possible between groups. Computationally, a canonical correlation analysis was then performed to determine any successive functions and canonical roots. Classification of groups was then made possible from the first canonical function. The summary of canonical discriminant functions calculated an eigenvalue of 0.03, suggesting that SAT mathematics scores contribute to 3% of the overall variance of the discriminant function. The calculated canonical correlation was 0.17, suggesting a low correlation (17%) between SAT mathematics scores and group membership. Wilk's lambda was 0.971, which suggests variance is explained by factors other than the difference between group means. The significance level calculated ($p < .05$), indicates that the group mean differences are significant and the model is a good fit for the data.

Since only one predictor variable was used in this predictive model, SPSS built a linear standardized canonical discriminant function consisting of a coefficient for the variable SATM and a constant. As that distance of SAT mathematics scores is being compared to the group mean for discrimination, the average SAT mathematics scores of each group (functions at group centroids) can be used to calculate a cut score for students. This number will be smaller than the actual SAT mathematics scores as the coefficient by which they are being multiplied is a fraction. Based on the information provided by functions at group centroid for this model, a cut score for placement can be calculated using the following formula:

$$\text{cut score} = \left(\frac{\text{function 1 value of not passing} + \text{function 1 value of passing}}{2}\right).$$

For SAT mathematics scores alone, the cut score is calculated by the following:

$$\text{cut score} = \left(\frac{-0.410+0.074}{2}\right) = -0.168.$$

This means that students who have a DF score above -0.168 are classified into the passing group, and are expected to pass, and students who have a DF score below -0.168 are classified into the not passing group, and are not expected to pass. The discriminant function calculated by SPSS correctly classified only 58.6% of the original grouped cases. Cross-validation results indicate that when used to generalize an independent data set, the discriminant function calculated by SPSS correctly classified 58.6% of grouped cases (Table 17). As noted in Chapter 3, cross-validation is the process of testing a model on more than one sample. This technique is often undertaken to assess the reliability and generalizability of the findings, which is particularly crucial in discriminant analysis because the solutions are often unreliable (Hair, Anderson, Tatham, & Black, 1995).

Predictive Models One through Six all included only one predictor variable for analysis. Predictive Model 7 included the predictor variables SATM, SATV, HS GPA, and HS Rank. Predictive Model 8 included the predictor variables ACTM, ACTV, HS GPA, and HS Rank. The last predictive model included all six predictor variables, and provided the highest percentage of correctly classified cases. All models followed the above procedure. Group statistics and classification results are displayed in Tables 18-33. A summary of the results for all predictive models is presented in Table 34.

Table 16

Group Statistics for SAT Mathematics Scores

	Mean	Standard Deviation	Valid N (listwise) Unweighted/Weighted
Not Passing	448.10	66.94	158
Passing	483.43	74.01	879
Total	478.05	74.04	1037

Table 17

Classification Results for SAT Mathematics Scores

			Predicted Group Membership		
		Math 1 Course	Not Passing	Passing	Total
Original	Count	Not Passing	105	53	158
		Passing	376	503	879
	%	Not Passing	66.5	33.5	100.0
		Passing	42.8	57.2	100.0
Cross-Validated	Count	Not Passing	105	53	158
		Passing	376	503	879
	%	Not Passing	66.5	33.5	100.0
		Passing	42.8	57.2	100.0

a. 58.6% of original grouped cases correctly classified.
b. 58.6% of cross-validated grouped cases correctly classified.

Table 18

Group Statistics for SAT Verbal Scores

	Mean	Standard Deviation	Valid N (listwise) Unweighted/Weighted
Not Passing	449.19	70.05	158
Passing	472.33	76.24	879
Total	469.72	75.55	1037

Table 19

Classification Results for SAT Verbal Scores

		Math 1 Course	Predicted Group Membership Not Passing	Passing	Total
Original	Count	Not Passing	97	61	158
		Passing	451	428	879
	%	Not Passing	61.4	38.6	100.0
		Passing	51.3	48.7	100.0
Cross-Validated	Count	Not Passing	97	61	158
		Passing	451	428	879
	%	Not Passing	61.4	38.6	100.0
		Passing	51.3	48.7	100.0

a. 50.6% of original grouped cases correctly classified.
b. 50.6% of cross-validated grouped cases correctly classified.

Table 20

Group Statistics for ACT Mathematics Scores

	Mean	Standard Deviation	Valid N (listwise) Unweighted/Weighted
Not Passing	18.16	3.04	157
Passing	20.23	3.86	879
Total	19.92	3.81	1036

Table 21

Classification Results for ACT Mathematics Scores

		Math 1 Course	Predicted Group Membership Not Passing	Passing	Total
Original	Count	Not Passing	119	38	157
		Passing	431	448	879
	%	Not Passing	75.8	24.2	100.0
		Passing	49.0	51.0	100.0
Cross-Validated	Count	Not Passing	119	38	157
		Passing	431	448	879
	%	Not Passing	75.8	24.2	100.0
		Passing	49.0	51.0	100.0

a. 54.7% of original grouped cases correctly classified.
b. 54.7% of cross-validated grouped cases correctly classified.

Table 22

Group Statistics for ACT Verbal Scores

	Mean	Standard Deviation	Valid N (listwise) Unweighted/Weighted
Not Passing	17.61	4.35	158
Passing	19.34	4.84	879
Total	19.08	4.81	1037

Table 23

Classification Results for ACT Verbal Scores

			Predicted Group Membership		
		Math 1 Course	Not Passing	Passing	Total
Original	Count	Not Passing	92	66	158
		Passing	387	492	879
	%	Not Passing	58.2	41.8	100.0
		Passing	44.0	56.0	100.0
Cross-Validated	Count	Not Passing	92	66	158
		Passing	387	492	879
	%	Not Passing	58.2	41.8	100.0
		Passing	44.0	56.0	100.0

a. 56.3% of original grouped cases correctly classified.
b. 56.3% of cross-validated grouped cases correctly classified.

Table 24

Group Statistics for High School GPA

	Mean	Standard Deviation	Valid N (listwise) Unweighted/Weighted
Not Passing	3.34	0.36	157
Passing	3.54	0.33	877
Total	3.51	0.34	1034

Table 25

Classification Results for High School GPA

			Predicted Group Membership		
		Math 1 Course	Not Passing	Passing	Total
Original	Count	Not Passing	95	62	157
		Passing	283	594	877
	%	Not Passing	60.5	39.5	100.0
		Passing	32.3	67.7	100.0
Cross-Validated	Count	Not Passing	95	62	157
		Passing	283	594	877
	%	Not Passing	60.5	39.5	100.0
		Passing	32.3	67.7	100.0

a. 66.6% of original grouped cases correctly classified.
b. 66.6% of cross-validated grouped cases correctly classified.

Table 26

Group Statistics for High School Rank

	Mean	Standard Deviation	Valid N (listwise) Unweighted/Weighted
Not Passing	162.06	132.04	127
Passing	106.87	107.45	675
Total	115.61	113.42	802

Table 27

Classification Results for High School Rank

			Predicted Group Membership		
		Math 1 Course	Not Passing	Passing	Total
Original	Count	Not Passing	60	67	127
		Passing	198	477	675
	%	Not Passing	47.2	52.8	100.0
		Passing	29.3	70.7	100.0
Cross-Validated	Count	Not Passing	60	67	127
		Passing	198	477	675
	%	Not Passing	47.2	52.8	100.0
		Passing	29.3	70.7	100.0

a. 67.0% of original grouped cases correctly classified.
b. 67.0% of cross-validated grouped cases correctly classified.

Table 28

Group Statistics for SAT Model with All Predictors

	Predictor Variable	Mean	Standard Deviation	Valid N (listwise) Unweighted/Weighted
Not Passing				
	HS Rank	163.22	113.92	126
	HS GPA	3.33	0.35	126
	SATM	448.73	64.67	126
	SATV	448.10	69.96	126
Passing				
	HS Rank	107.00	107.48	674
	HS GPA	3.57	0.30	674
	SATM	486.15	474.70	674
	SATV	470.68	76.03	674
Total				
	HS Rank	115.86	113.456	800
	HS GPA	3.53	0.32	800
	SATM	480.25	74.43	800
	SATV	467.13	75.52	800

Table 29

Classification Results for SAT Model with All Predictors

		Math 1 Course	Predicted Group Membership Not Passing	Passing	Total
Original	Count	Not Passing	80	46	126
		Passing	217	457	674
	%	Not Passing	63.5	36.5	100.0
		Passing	32.2	67.8	100.0
Cross-Validated	Count	Not Passing	80	46	126
		Passing	219	455	674
	%	Not Passing	63.5	36.5	100.0
		Passing	32.5	67.5	100.0

a. 67.1% of original grouped cases correctly classified.
b. 66.9% of cross-validated grouped cases correctly classified.

Table 30

Group Statistics for ACT Model with All Predictors

	Predictor Variable	Mean	Standard Deviation	Valid N (listwise) Unweighted/Weighted
Not Passing				
	HS Rank	163.22	113.92	126
	HS GPA	3.33	0.35	126
	ACTM	17.40	4.38	126
	ACTV	18.11	3.01	126
Passing				
	HS Rank	107.00	107.48	674
	HS GPA	3.567	0.30	674
	ACTM	19.07	4.82	674
	ACTV	20.31	9.20	674
Total				
	HS Rank	115.86	113.46	800
	HS GPA	3.53	0.32	800
	ACTM	18.81	4.79	800
	ACTV	19.96	3.87	800

Table 31

Classification Results for ACT Model with All Predictors

			Predicted Group Membership		
		Math 1 Course	Not Passing	Passing	Total
Original	Count	Not Passing	85	41	126
		Passing	220	454	674
	%	Not Passing	67.5	32.5	100.0
		Passing	32.6	67.4	100.0
Cross-Validated	Count	Not Passing	85	41	126
		Passing	225	449	674
	%	Not Passing	67.5	32.5	100.0
		Passing	33.4	66.6	100.0

a. 67.4% of original grouped cases correctly classified.
b. 66.8% of cross-validated grouped cases correctly classified.

Table 32

Group Statistics for Combined SAT and ACT Models with All Predictors

	Predictor Variable	Mean	Standard Deviation	Valid N (listwise) Unweighted/Weighted
Not Passing				
	HS Rank	163.22	113.92	126
	HS GPA	3.33	0.35	126
	SATM	448.73	64.67	126
	ACTV	17.40	4.38	126
	ACTM	18.11	3.01	126
	SATV	448.10	69.956	126
Passing				
	HS Rank	107.00	107.48	674
	HS GPA	3.57	0.30	674
	SATM	486.15	474.70	674
	ACTV	19.07	4.82	674
	ACTM	20.31	3.92	674
	SATV	470.68	76.03	674
Total				
	HS Rank	115.86	113.46	800
	HS GPA	3.53	0.32	800
	SATM	480.25	74.43	800
	ACTV	18.81	4.79	800
	ACTM	19.96	3.87	800
	SATV	467.13	75.52	800

Table 33

Classification Results for Combined SAT and ACT Models with All Predictors

			Predicted Group Membership		
		Math 1 Course	Not Passing	Passing	Total
Original	Count	Not Passing	85	41	126
		Passing	211	463	674
	%	Not Passing	67.5	32.5	100.0
		Passing	31.3	68.7	100.0
Cross-Validated	Count	Not Passing	83	43	126
		Passing	215	459	674
	%	Not Passing	65.9	34.1	100.0
		Passing	31.9	68.1	100.0

a. 68.5% of original grouped cases correctly classified.
b. 67.8% of cross-validated grouped cases correctly classified

Table 34

Summary of Discriminant Analysis Results

Discriminant Function	Eigenvalue	Canonical Correlation	Wilk's λ	Sig.	Cut Score	Classification Results % Correct	Cross- Validated
$0.014 \times SATM - 6.550$	0.03	0.17	0.97	**<0.01**	-0.168	58.6%	58.6%
$0.013 \times SATV - 6.235$	0.01	0.08	0.99	**0.01**	-0.079	50.6%	50.6%
$0.267 \times ACTM - 5.321$	0.04	0.20	0.96	**<0.01**	-0.193	54.7%	54.7%
$0.210 \times ACTV - 3.998$	0.02	0.13	0.98	**<0.01**	-0.126	56.3%	56.3%
$2.979 \times HS\ GPA - 10.447$	0.04	0.20	0.96	**<0.01**	-0.199	66.6%	66.6%
$0.009 \times HS\ Rank - 1.035$	0.03	0.18	0.97	**<0.01**	0.169	67.0%	67.0%
$2.203 \times HS\ GPA + 0.006 \times SATM - 0.002 \times HS\ Rank - 0.001 \times SATV - 9.614$	0.09	0.29	0.92	**<0.01**	-0.282	67.1%	66.9%
$-0.002 \times HS\ Rank + 2.056 \times HS\ GPA - 0.011 \times ACTV + 0.122 \times ACTM - 9.251$	0.10	0.30	0.91	**<0.01**	-0.29	67.4%	66.8%
$-0.002 \times HS\ Rank + 2.021 \times HS\ GPA + 0.002 \times SATM + 0.009 \times ACTV + 0.104 \times ACTM - 0.002 \times SATV - 8.956$	0.10	0.30	0.91	**<0.01**	-0.293	68.5%	67.8%

Research question two. What is the extent to which standardized test scores display differential validity among subgroups? Null hypothesis: There will be no difference between means of student performance on standardized test scores between students of different genders, ethnicities, and income levels.

A three-way ANOVA was used to determine the differential validity of overall student performance on the SAT and ACT on the basis of gender, ethnicity, and income. Normality was

assessed by examining histograms and the skewness and kurtosis for each of the independent variables, and homogeneity of variance assumptions. Two separate three-way ANOVAs were done to address each of these areas. The three main effects examined were gender, with two levels, ethnicity, with six levels (some categories were combined due to small member sizes), an income, with ten levels.

Three-way ANOVA for SAT. Tables 35 and 36 provide a summary of SAT scores by gender for each income level. Levene's test of equality of error variances suggests that group variances are significant for all independent variables ($p < .05$). A critical alpha of 0.05 was used for the following analyses.

Table 35

Three-way ANOVA Summary of SAT Total Mean Scores for Females by Income Level

	Income Level	Mean	Standard Deviation	Frequency
	$0 – $10k	927.89	140.26	37
	$10k – $20k	900.19	125.52	106
	$20k – $30k	934.30	120.21	107
	$30k – $40k	922.06	121.39	68
	$40k – $50k	950.88	127.90	68
	$50k – $60k	949.25	134.73	67
	$60k – $70k	968.33	122.50	60
	$70k – $80k	976.22	108.54	74
	$80k – $90k	940.48	139.79	63
	$90k – $100k	998.26	159.91	46
	$100k+	990.62	141.70	230
Total		954.68	134.67	926

Table 36

Three-way ANOVA Summary of SAT Total Mean Scores for Males by Income Level

	Income Level	Mean	Standard Deviation	Frequency
	\$0 – \$10k	951.00	124.58	10
	\$10k – \$20k	890.91	116.82	33
	\$20k – \$30k	938.71	123.47	31
	\$30k – \$40k	919.23	131.73	26
	\$40k – \$50k	961.43	180.68	14
	\$50k – \$60k	949.64	163.02	28
	\$60k – \$70k	991.77	152.37	17
	\$70k – \$80k	1004.00	104.64	25
	\$80k – \$90k	982.63	148.36	19
	\$90k – \$100k	930.00	122.42	15
	\$100k+	1003.80	144.23	79
Total		962.42	141.53	297

The tests of between-subjects effects, Table 37, contains the following information of main effects and interactions, starting from the three-way interaction to main effects. There was a statistically significant three-way interaction between gender, ethnicity and income level on the SAT [F (25, 824) = 1.840, p = .008, η^2 = 0.242]. This indicates that there is a statistically significant difference in the means of SAT scores between/among different combinations of gender, ethnicity, and income. There was no significant two-way interaction between ethnicity and income on the SAT [F (45, 824) = 1.261, p = .121, η^2 = 0.298]. This indicates that there is no statistically significant difference in the means of SAT scores between/among different combinations of ethnicity and income levels. There was a statistically significant two-way interaction between gender and income on the SAT [F (10, 824) = 2.173, p = .018, η^2 = 0.114]. This indicates that there is a statistically significant difference in the means of SAT scores between/among different combinations of gender and income levels. There was no significant two-way interaction between gender and ethnicity on the SAT [F (5, 824) = 1.060, p = .381, η^2 = 0.028]. This indicates that there is no statistically significant difference in the means of SAT

scores between/among different combinations of gender and ethnicity. The main effect of the within subjects variable of income level is significant [*F* (10, 824) = 2.319, $p < .011$, $\eta^2 = 0.122$]. This indicates that there is a statistically significant difference in the means of SAT scores among different income levels. The effect size based on eta-squared is large for this variable. The main effect of the within-subjects variable of ethnicity is significant [*F* (5, 824) = 7.455, $p < .001$, $\eta^2 = 0.196$]. This indicates that there is a statistically significant difference in the means of SAT scores among different ethnicities. The effect size based on eta-squared is large for this variable. The main effect of the within-subjects variable of gender is not significant [*F* (1, 824) = 1.674, $p = .949$, $\eta^2 < 0.0001$]. This indicates that there is not a statistically significant difference in the means of SAT scores between genders. The effect size based on eta-squared is extremely small for this variable. The independent variables (gender, ethnicity and income), explain 31.8% of the variance in SAT scores, a very large effect size.

Table 37

Tests of Between-Subjects Effects for Three-way ANOVA for SAT Total Scores

Source	df	Mean Square	F	Sig.	η^2	% of variance due to main effects
Gender	1	64.52	<0.01	0.95	0.00	
Ethnicity	5	115474.01	7.45	**<0.01**	0.20	
Income	10	35911.20	2.32	**0.01**	0.12	
Gender * Ethnicity	5	16416.27	1.06	0.38	0.03	
Gender * Income	10	33660.13	2.17	**0.01**	0.11	
Ethnicity * Income	45	19528.94	1.26	0.12	0.30	
Gender * Ethnicity * Income	25	28502.86	1.84	**0.01**	0.24	
Error	824	15488.69				
Total	926					31.8
Corrected Total	925					

$r^2 = 0.24$ (Adjusted $r^2 = 0.15$)

There is a need for follow-up post-hoc tests on income and ethnicity because these significant effects have more than two levels. Post-hoc tests will examine where the differences lie. Tukey HSD test is the most widely used post-hoc test because it is not as conservative as the Scheffé test and increases the likelihood of detecting mean differences (Laerd Statistics, n.d.). Multiple comparisons between ethnicities are shown in Table 38. Using a critical alpha of 0.05, it is determined that there is a statistically significant difference in the means of SAT scores between Asian students compared to the mean SAT scores of Black, Hispanic, and students of unknown ethnicity. There is also a statistically significant difference in the means of SAT scores between White students and the mean SAT scores of Black, Hispanic, and students of unknown ethnicity.

Using a critical alpha of 0.05, it is determined that there is not a statistically significant difference in the means of SAT scores between income levels the $60k+ range. There is a statistically significant difference in the means of SAT scores between income levels of $60k-$80k, and the mean SAT scores of income levels from $10k-$20k. There is a statistically significant difference in the means of SAT scores between income levels of $70k-$80k, and the mean SAT scores of income levels from $10k-$20k. There is a statistically significant difference in the means of SAT scores between income levels of $90k-$100k, and the mean SAT scores of income levels from $10k-$20k. The last set of statistically significant difference in the means of SAT scores exists between income levels in excess of $100k, and the mean SAT scores of income levels from $10k-$40k.

Table 38

Multiple Comparisons for Tukey HSD Post-Hoc Test on Ethnicity (SAT Total Scores)

Ethnicity 1 (x)	Ethnicity 2 (y)	Mean Difference	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Asian	Black	162.09	29.076	<0.01	79.03	245.14
	Hispanic	133.92	24.49	<0.01	63.97	203.87
	Unknown	150.66	28.50	<0.01	69.26	232.05
	White	56.68	25.82	0.24	-17.06	130.41
	Other	90.30	40.08	0.22	-24.18	204.78
Black	Asian	-162.09	29.08	<0.01	-245.14	-79.03
	Hispanic	-28.17	17.26	0.58	-77.46	21.12
	Unknown	-11.43	22.58	1.00	-75.94	53.08
	White	-105.41	19.09	<0.01	-159.94	-50.88
	Other	-71.79	36.12	0.35	-174.95	31.37
Hispanic	Asian	-133.92	24.49	<0.01	-203.87	-36.97
	Black	8.17	17.26	0.58	-21.12	77.46
	Unknown	16.74	16.26	0.91	-29.70	63.18
	White	-77.24	10.90	<0.01	-108.37	-46.11
	Other	-43.62	32.54	0.76	-136.56	49.32
Unknown	Asian	-150.66	28.50	<0.01	-232.05	-69.26
	Black	11.43	22.58	1.00	-53.08	75.94
	Hispanic	-16.74	16.26	0.91	-63.18	29.70
	White	-93.98	18.20	<0.01	-145.95	-42.01
	Other	-60.36	35.65	0.54	-162.19	41.47
White	Asian	-56.68	25.82	0.24	-130.41	17.06
	Black	105.41	19.10	<0.01	50.88	159.94
	Hispanic	77.24	10.90	<0.01	46.11	108.37
	Unknown	93.98	18.19	<0.01	42.01	145.95
	Other	33.62	33.55	0.92	-62.20	129.44
Other	Asian	-90.30	40.08	0.22	-204.78	24.18
	Black	71.79	36.12	0.35	-31.37	174.95
	Hispanic	43.62	32.54	0.76	-49.32	163.56
	Unknown	60.36	35.65	0.54	-41.47	162.19
	White	-33.62	33.55	0.92	-129.44	62.20

Three-way ANOVA for ACT. Tables 39 and 40 provide a summary of ACT scores by gender for each income level. Levene's tests of equality of error variances suggest that group variances are significant for all independent variables ($p < .05$). A critical alpha of 0.05 was used for the following analyses.

Table 39

Three-way ANOVA Summary of ACT Total Mean Scores for Females by Income Level

Income Level	Mean	Standard Deviation	Frequency
$0 – $10k	18.89	3.79	27
$10k – $20k	18.43	3.59	72
$20k – $30k	19.18	3.54	73
$30k – $40k	18.95	3.75	41
$40k – $50k	19.89	3.00	54
$50k – $60k	19.53	2.87	38
$60k – $70k	20.07	2.93	42
$70k – $80k	20.48	3.43	48
$80k – $90k	19.73	3.07	44
$90k – $100k	21.48	4.02	31
$100k+	20.91	3.77	146
Total	19.88	3.58	616

Table 40

Three-way ANOVA Summary of ACT Total Mean Scores for Males by Income Level

Income Level	Mean	Standard Deviation	Frequency
$0 – $10k	18.40	5.23	10
$10k – $20k	19.00	3.27	33
$20k – $30k	19.17	3.28	30
$30k – $40k	19.04	3.40	25
$40k – $50k	18.91	4.37	11
$50k – $60k	19.80	4.45	25
$60k – $70k	20.00	3.74	17
$70k – $80k	21.44	2.83	23
$80k – $90k	20.95	4.82	19
$90k – $100k	19.20	2.91	15
$100k+	21.28	3.89	76
Total	20.07	3.87	284

The tests of between-subjects effects, Table 41, contains the following information of effects and interactions, starting from the three-way interaction to main effects. There was a statistically significant three-way interaction between gender, ethnicity and income on the ACT [F (25, 798) = 1.864, p = .007, η^2 = 0.0473]. This indicates that there is a statistically significant

difference in the means of ACT scores between/among different combinations of gender, ethnicity, and income level. There was no significant two-way interaction between ethnicity and income on the ACT [F (45, 798) = 1.239, p = .139, η^2 = 0.0565]. This indicates that there is no statistically significant difference in the means of ACT scores between/among different combinations of ethnicity and income levels. There was a statistically significant two-way interaction between gender and income on the ACT [F (10, 798) = 2.385, p = .009, η^2 = 0.0242]. This indicates that there is a statistically significant difference in the means of ACT scores between/among different combinations of gender and income levels. There was no significant two-way interaction between gender and ethnicity on the ACT [F (5, 798) = 0.478, p = .793, η^2 = 0.0025]. This indicates that there is no statistically significant difference in the means of ACT scores between/among different combinations of gender and ethnicity. The main effect of the within subjects variable of income level is significant [F (10, 798) = 2.178, p = .017, η^2 = 0.0220]. This indicates that there is a statistically significant difference in the means of ACT scores between different income levels. The effect size based on eta-squared is small for this variable. The main effect of the within subjects variable of ethnicity is significant [F (5, 798) = 7.547, p < .001, η^2 = 0.0383]. This indicates that there is a statistically significant difference in the means of ACT scores between different ethnicities. The effect size based on eta-squared is small for this variable. The main effect of the within subjects variable of gender is not significant [F (1, 798) = 0.237, p = .627, η^2 = 0.0002]. This indicates that there is not a statistically significant difference in the means of ACT scores between genders. The effect size based on eta-squared is extremely small for this variable. The independent variables (gender, ethnicity and income level), explain 6.05% of the variance in ACT scores, a medium effect size.

Table 41

Tests of Between-Subjects Effects for Three-way ANOVA for ACT Total Scores

Source	df	Mean Square	F	Sig.	η^2	% of variance due to main effects
Gender	1	2.71	0.24	0.63	<0.01	
Ethnicity	5	86.46	7.55	**<0.01**	0.04	
Income	10	24.95	2.18	**0.02**	0.02	
Gender * Ethnicity	5	5.47	0.48	0.79	<0.01	
Gender * Income	10	27.33	2.38	**0.01**	0.02	
Ethnicity * Income	45	14.19	1.24	0.14	0.06	
Gender * Ethnicity * Income	25	21.36	1.86	**0.01**	0.05	
Error	798	11.46				
Total	900					6.1
Corrected Total	899					

r^2 = 0.25 (Adjusted r^2 = 0.15)

There is a need for follow-up post-hoc tests on income and ethnicity because these significant effects have more than two levels. Post-hoc tests will examine where the differences lie. Tukey HSD test is the most widely used post-hoc test because it is not as conservative as the Scheffé test and increases the likelihood of detecting mean differences (Laerd Statistics, n.d.). Multiple comparisons between ethnicities are shown in Table 42. Using a critical alpha of 0.05, it is determined that there is a statistically significant difference in the means of ACT scores between Asian students compared to the mean ACT scores of Black, Hispanic, and students of unknown ethnicity. There is also a statistically significant difference in the means of ACT scores between White students and the mean ACT scores of Black, Hispanic, and students of unknown ethnicity.

Table 42

Multiple Comparisons for Tukey HSD Post-Hoc Test on Ethnicity (ACT Total Scores)

Ethnicity 1 (x)	Ethnicity 2 (y)	Mean Difference (x - y)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Asian	Black	4.35	0.81	<0.01	2.04	6.65
	Hispanic	3.94	0.69	<0.01	2.01	5.88
	Unknown	3.88	0.79	<0.01	1.623	6.13
	White	1.31	0.72	0.45	-0.74	3.35
	Other	3.12	1.10	0.05	-0.03	6.24
Black	Asian	-4.35	0.81	<0.01	-6.65	-2.04
	Hispanic	-0.40	0.48	0.96	-1.78	0.98
	Unknown	-0.47	0.63	0.98	-2.26	1.33
	White	-3.04	0.53	<0.01	-4.56	-1.51
	Other	-1.24	0.99	0.81	-4.06	1.59
Hispanic	Asian	-3.94	0.68	<0.01	-5.88	-2.01
	Black	0.40	0.48	0.96	-0.98	1.78
	Unknown	-0.67	0.45	1.00	-1.35	1.22
	White	-2.64	0.30	<0.01	-3.50	-1.77
	Other	-0.84	0.89	0.94	-3.37	1.69
Unknown	Asian	-3.88	0.79	<0.01	-6.13	-1.63
	Black	0.47	0.63	0.98	-1.33	2.26
	Hispanic	0.07	0.45	1.00	-1.22	1.35
	White	-2.57	0.50	<0.01	-4.01	-1.13
	Other	-0.77	0.97	0.97	-3.55	2.01
White	Asian	-1.31	0.72	0.45	-3.35	0.74
	Black	3.34	0.53	<0.01	1.51	4.56
	Hispanic	2.64	0.30	<0.01	1.77	3.50
	Unknown	2.57	0.50	<0.01	1.13	4.01
	Other	1.80	0.91	0.36	-0.81	4.41
Other	Asian	-3.11	1.10	0.05	-6.24	0.03
	Black	1.24	0.99	0.81	-1.59	4.06
	Hispanic	0.84	0.89	0.94	-1.69	3.37
	Unknown	0.77	0.97	0.97	-2.01	3.55
	White	-1.80	0.91	0.36	-4.41	0.81

Using a critical alpha of 0.05, it is determined that there is not a statistically significant difference in the means of ACT scores between income levels until the $70k+ range. There is a statistically significant difference in the means of ACT scores between income levels of $70k-$80k, and the mean ACT scores of income levels from $10k-$20k. There is a statistically

significant difference in the means of ACT scores between income levels of $90k-$100k, and the mean ACT scores of income levels from $10k-$20k. The last set of statistically significant difference in the means of ACT scores exists between income levels in excess of $100k, and the mean ACT scores of income levels from $0-$40k.

Summary of Results

The purpose of this quantitative study was to examine the predictive ability of standardized test scores, and determine if differential validity existed among subgroups. Using discriminant analysis, it was established that the following percentages of students were correctly classified into passing and not passing groups:

- SAT mathematics scores—58.6%,
- SAT verbal scores—50.6%,
- ACT mathematics scores—54.7%, and
- ACT verbal scores—56.3%.

New predictive models were created using standardized test scores in combination with students' high school GPA, and high school rank to increase correct classification into passing and not passing groups to:

- 67.1% using SAT mathematics and verbal scores, high school GPA, and high school rank,
- 67.4% using ACT mathematics and verbal scores, high school GPA, and high school rank, and
- 68.5% using both SAT and ACT mathematics and verbal scores, high school GPA, and high school rank.

Using three-way ANOVA, it was determined that there was a significant three-way interaction between gender, ethnicity, and socioeconomic status (income), and a significant two-way interaction between gender and socioeconomic status (income) for both the SAT and ACT. An analysis of main effects determined that ethnicity and socioeconomic status (income) displayed statistically significant differences in the mean scores of students on the SAT and ACT. Main effects attributed to 31.8% of the variance of SAT total scores (significant), and 6.1% for ACT total scores.

Chapter 5: Discussion, Implications, and Recommendations

Discussion

The institution used for this study has a unique student population. Due to its status as an HSI, serving predominantly minority students, the process of analysis of this study might be generalizable to other institutions with similar student populations, and with comparable new-freshmen profiles. The results from this study had two major components. The first was to determine the extent to which mathematics and verbal standardized test scores predicted the performance of students in their first-year mathematics courses. The second was to determine the extent to which standardized test scores displayed differential validity among subgroups. Although standardized test scores alone do not greatly predict student performance in their first mathematics course, they do contribute to determining the overall performance of a student. Both the SAT and ACT display differential validity among subgroups, which is of concern as they are used for admission into the institution, and are the primary tool for placement of students into mathematics and English courses. The results from each of these parts will be compared to the literature previously published on the subjects of predictive ability, bias, placement, the ELL student population, and course and program effectiveness.

Predictive ability of standardized tests. The models examined to evaluate the predictive ability of the SAT and ACT validated the researcher's assumption that standardized test scores should not be used as the sole basis for placement. Each test score correctly classified less than 60% of the original grouped cases. Previous research suggests that there may be a possible relationship between a student's high school GPA and his or her success in college (Astin, 1985, 1991; Pascarella & Terenzini, 1991). The target population for this study confirmed this theory – providing substantial proof that high school GPA alone correctly classified 66.6% of the original

grouped cases, in comparison to 58.6% for SAT mathematics scores and 54.7% for ACT mathematics scores. "While conceding the importance of high school record as an admissions criterion, advocates of standardized admissions tests nevertheless state that, used as a supplement to high school record, tests provide additional information that can aid admissions officers and improve decision making" (Geiser & Santelices, 2007, p. 24). In this study, the combination of standardized test scores, high school rank, and high school GPA, provided an increase to the percent of correctly classified original grouped cases to:

- 67.1% for the SAT model,
- 67.4% for the ACT model, and
- 68.5% for a combined SAT and ACT model

suggesting their validity for use in placement.

Differential validity of standardized tests. Contrary to the results of Korpershoek et al. (2015), the SAT and ACT both yielded insignificant main effects of gender of students on test scores. In accordance with previous research done by Hagedorn et al. (1999) the SAT and ACT are bias on the basis of at least two of the three identified subgroups, ethnicity and socioeconomic status (income), showing significant main effects on student test scores. Hagedorn, Siadat, Fogel, Nora, & Pascarella (1999) stated, "students in remedial math placements were more likely to come from families with lower incomes and lower educational levels, and were more likely to receive less encouragement to enroll in college" (p. 278). One of these claims can be validated by analysis of the target population evaluated. It is important to note that regardless of income level, students are placed in developmental mathematics courses at smaller percentages than college-level mathematics courses; however, the income levels at $30,000 and below, consistently show larger numbers of enrollment in developmental

mathematics courses. School districts rely heavily on financing from property taxes. In accordance with previous research and demographic studies, unequal distribution of funding for K-12 education is highly presumed (Irons & Harris, 2007; Singh & Al Fadhli, 2011).

Correlation between mathematics and reading. Research indicates that student proficiency in reading can be a strong indicator of success in mathematics (Jiban & Deno, 2007). The overall predictive ability of SAT mathematics scores accurately classified 58.6% of students into passing and not passing groups, and SAT verbal scores accurately classified 50.6% of students into passing and not passing groups. The 8% difference in predictive ability is not significant enough to credit a correlation between reading ability and mathematics performance. However, it does signify that for the SAT, a student's mathematics score should continue to be used over verbal scores for placement as an indicator of success. The overall predictive ability of the ACT mathematics scores accurately classified 54.7% of students into passing and not passing groups, and ACT verbal scores accurately classified 56.3% of students into passing and not passing groups. The 1.6% difference in predictive ability gives sufficient proof that a correlation may exist between reading ability and mathematics performance. In terms of predictive ability of the ACT, mathematics and verbal scores could both be considered for placement.

Implications

Placement. The institution of interest, until Fall 2014, only used standardized test scores for placement of students, which has proven to be 58.6% accurate when using SAT mathematics scores, and 54.7% accurate when using ACT mathematics scores for placement of students. Previous studies suggest other student information that has proven to be helpful in the placement of students is high school GPA, consideration of mathematics courses taken in high school, as well as the length of time from last mathematics course taken (Breland et al., 2002; Geiser &

Santelices, 2007; Zwick, 2004). Geiser and Santelices (2007) claim that grades in college-preparatory subjects are the “best” indicator of student performance. Researchers have noted that standardized test scores add significantly to performance prediction (Breland et al., 2002; Burton & Ramist, 2001; Camara & Echternact, 2000; Geiser & Santelices, 2007; Zwick, 2004). The results of this study have proven that the predictability of student placement can be increased if standardized tests scores are used in conjunction with a student’s high school GPA and high school class rank to increase correct placement to 67.0% using the SAT, and 68.4% using the ACT. This validates previous research by proving the capability and ability of a student to successfully complete coursework should not be determined by one cut score (Geiser & Santelices, 2007; Geiser & Studley, 2002). It is suggested that the institution gather and implement the use of high school grades in placement procedures, as this may increase the correct placement into developmental and college-level mathematics courses.

In lieu of the general standardized test scores, content specific exams (i.e. AP exams and SAT subject tests) have been suggested as an alternative for indicating student preparation for, placement, and success in college (Dougherty, Mellor, & Jian, 2006; Geiser & Santelices, 2006). These exams are not taken by a great majority of students applying for admissions into college, thus, although they may be effective, they are not practical. However, the faculty of the courses being taught in colleges are the individuals who know and understand well the deficiencies that the students are coming in with to higher education, as well as what specific skillsets are required for success in all college-level courses. It is suggested that the mathematics department of the institution consider implementing the use of a content specific placement test for students as an alternative means for placement. This would enable assessment of the fundamental skills required to be successful in college-level courses. It could also give the institution more

information regarding any gaps in student education that may not be currently addressed. These recommendations are suggested for other HSIs and institutions with large volumes of remedial students.

Bias. Demographic studies have shown that schools in low socioeconomic and/or predominantly minority areas may not be delivering the same quality instruction as schools in affluent areas (Irons & Harris, 2007; Singh & Al Fadhli, 2011). With such an unequal distribution of financial resources in the education system, socioeconomic aspects of an educational system should be taken into consideration when considering the admission and placement of students. Although these situations are external factors that this institution cannot control, they pose a concern to both admission and placement processes. It has been estimated that at least 60% of a university's student body is required to take at least one developmental reading, writing, or mathematics course, sometimes two or more concomitantly (Attewell, Lavin, Domina, & Levey, 2006; Bailey, Jeong & Cho, 2010; Cullinane & Treisman, 2010). Of these three courses, developmental mathematics courses have the highest rate of attrition (Adelman, 2004). The results of differential validity of standardized tests points to the possible relationship between high school quality, ethnicity, and socioeconomic background – ethnicity being the greatest concern. After identification of the significant main effect of ethnicity on student scores, the researcher ran a separate set of discriminant analyses.

Using the SAT, ACT, and combined predictor models, the researcher analyzed the results for each individual ethnicity. The results for every analysis indicated that the Hispanic student population was at the greatest disadvantage when standardized test scores are used for placement purposes. The SAT model with predictor variables correctly classified 63.3% of Hispanic students. The ACT model with predictor variables provided correct classification of 64.1%, and

the combined SAT/ACT model with predictor variables correctly classified only 64.7% of Hispanic students. These same models correctly classified 100% of students in the American Indian or Alaska Native, Nonresident Alien, and Two or more races ethnicities, and between 73.2-91.7% for all other ethnicities – suggesting that placement procedures are extremely skewed with respect to the student population that the institution is nationally recognized for serving. Attewell et al. (2006) provided compelling evidence that minority, low income, and ELL students are overrepresented in remedial courses. After analysis of course enrollment for this target population, it was calculated that 75.8% of students enrolled in MATH 0318, and 62.6% of students enrolled in MATH 0319 are Hispanic students. This compared to 52.79% - the percentage of Hispanics in the student body. The income levels at $30,000 and below, consistently show larger numbers of enrollment in developmental mathematics courses. Samway and McKeon (1999) suggest that higher education institutions revisit their admissions and placement procedures – this research has validated that argument.

In his study, Olivas (1979) noted that a shortage in minority instructors, specifically of African-American and Hispanic ethnic background, posed a particular problem for students feeling accepted and welcomed in higher education institutions. Although this study was performed more than thirty years ago, a NCES study (2010), detailed the breakdown of higher education institutions' faculty as follows: 80% Caucasian, 7% African American, 6% Asian/Pacific Islander, 4% Hispanic, and 1% American Indian/Alaska Native. The institution's faculty breakdown in Fall 2013 was: 65.3% White, 19.7% Hispanic, 13% Other, and 2% Nonresident Alien. Perhaps this institution, as well as other HSIs and institutions serving minority populations, should consider increasing the number of minority faculties that are consistent with the demographic change in the United States.

Course changes and community college outreach. The mathematics department at the institution of interest is continually adapting to the changing student population. In Fall 2014, it implemented new placement procedures for students. It is currently in the process of a developmental course change to address the prevalence of extreme deficiencies in mathematics exhibited by students. It is suggested that the institution pilot the course with one or two sections prior to implementing a campus-wide course change.

The birth of the junior/community college offered one solution to address the problem with these extreme deficiencies. It has attempted to serve the developmental student population of this city by offering a multi-level developmental course sequence. Although a multi-level developmental course sequence may or may not be a valid solution, it is suggested that the institution, specifically the mathematics department, reach out to the local community colleges in a collaborative effort to analyze placement procedures and success and/or failure of all developmental course sequence(s) offered. In doing so, there may be a way to create a placement tool that addresses the needs of the city and its student population as a whole.

ELL student population. According to Das (2008), the ELL student population in his study was the worst performing group out of their peers. It is not presumed that the entire Hispanic population of this study is ELL. However, this study did prove that the Hispanic population has the greatest disadvantage in terms of test bias. All HSIs should consider shifting their admissions and placement procedures to address the demographic changes in the United States. It is suggested that in an effort to address the concerns and needs of the ELL population, this and all other institutions should begin tracking ELL student status and course patterns in an attempt to serve this growing population.

The silver lining. Roueche and Roueche (1993) define effective courses and academic programs as those having a successful completion rate of 60-70%. This guideline has been used in previous research to determine whether other programs and curricula are valid and reliable (e.g., Alexander, 2000; Amey & Long, 1998; Aycaster, 2001; Bishop, 1992; Peak, 1996). Using this guideline as a measure of effectiveness with this target population, 89.0% of students in MATH 0318 (Intro to Geometry, Probability and Statistics), 77.8% of students in MATH 0319 (Intro to Algebra), 74.9% of students in MATH 1304 (College Algebra), 88.9% of students in MATH 1306 (Geometry and the Imagination), and 88.7% of students in MATH 2303 (Introduction to Probability & Statistics) successfully completed their first college mathematics course. Although the procedures used for placement may not be highly effective and show bias, the curriculum, instruction, and services being offered to students are proving to be effective.

Recommendations for Future Research

Accuplacer. In the initial data received from the institution of interest, the largest ethnicity group providing testing information for the ACCUPLACER was the Nonresident Alien student population. It should be noted that since the great majority of these students do not take the SAT and ACT for admissions and placement purposes, only a small sample of that entire population was included in this study. It is suggested that as an increase in this growing population becomes more prevalent, a study should be conducted to evaluate the predictive ability and differential validity of the ACCUPLACER. Based on the results and implications of this study, it can be presumed that the ACCUPLACER may have a very small predictive ability and may exhibit differential validity.

Hispanic ELL vs. Hispanic non-ELL. As noted earlier, it is not presumed that all students in the Hispanic population are ELL. It was suggested that institutions begin tracking

ELL student status, as well as course-taking patterns. The implication of this change is the suggestion that a longitudinal study be performed on the Hispanic student population as a whole to compare the enrollment and retention rates, course taking patterns, and graduation rates between Hispanic ELL and Hispanic non-ELL students. This study was not able to identify an ELL subgroup in the Hispanic population. However, it would be very valuable to determine if trends for both groups are the same or significantly different.

Perceptions of bias. As bias is prevalent in this study, it is suggested that a qualitative study be performed to examine any perceptions of bias among the minority populations on campus, to include the perceptions of students and instructors.

Reading comprehension. Verbal scores for both the SAT and ACT showed little predictive ability. Other studies have successfully used reading comprehension specific tests, to include Maze in K-12 education, and ACCUPLACER in college for analysis purposes. It is suggested that a study be performed to examine the ability of a reading comprehension test to predict student success in their first English course. These results can then be compared with the placement procedures currently used at institutions.

Non-traditional student and adult education campuses. In terms of other minority groups, a study by Hoyt (1999) suggested that older, non-traditional students have higher dropout rates than traditional students. It is suggested that a study be performed to examine possible cause for higher dropout rates. Based on this study, potential subgroup bias may exist on the basis of standardized test scores. Another explanation may be perceptions of bias on campus, which should also be examined in this growing student population.

Re-examination of predictive ability and differential validity. If this or other institutions implement changes in their admissions and placement procedures, a follow-up study

should be conducted to re-examine both predictive ability and differential validity. It is suggested that these institutions wait until they have significant enrollment numbers to conduct such studies, which should include data that spans over an extended period of time – more than one academic year.

References

Adelman, C. (2004). *Principal indicators of student academic histories in postsecondary education, 1972-2000*. Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved from the ERIC database. (ED483154)

Agnese, L. (2009). *The changing university: Adapting to meet the goals of vision 2012*. Retrieved from: http://www.uiw.edu/president/stateofu09.html.

Alexander, G. (2000, February 28). *Developmental math program recommendation*. Retrieved from: http://www.sheboygan.uwc.edu/developmental-math/dev_data/recomm.htm.

Amey, M., & Long, P. (1998). Developmental course work and early placement: Success strategies for underprepared community college students. *Community College Journal of Research & Practice*, *22*(1), 3 – 10. doi:10.1080/1066892980220101

Astin, A. W. (1985). *Achieving educational excellence*. San Francisco, CA: Jossey-Bass.

Astin, A. W. (1991). *Assessment for excellence*. New York, NY: ACE MacMillan.

Attewell, P., Lavin, D., Domina, T., & Levey, T. (2006). New evidence on college remediation. *Journal of Higher Education*, *77*(5), 886-924.

Aycaster, P. (2001). Factors impacting success in community college developmental mathematics courses and subsequent courses. *Community College Journal of Research & Practice*, *25*(5/6), 403 – 416. doi:10.1080/106689201750192256

Bailey, T., Jeong, D. W., & Cho, S. (2010). Referral, enrollment, and completion in developmental education sequences in community college. *Economics of Education Review*, *29*(2), 255-270. doi:10.1016/j.econedurev.2009.09.002

Beal, C. R., Adams, N. M. & Cohen, P. R. (2010). Reading proficiency and mathematics problem solving by high school English language learners. *Urban Education*, *45*, 58-75. doi:10.1177/0042085909352143

Beswick, G., Rothblum, E. D., & Mann, L. (1988). Psychological antecedents of student procrastination. *Australian Psychologist*, *23*(2), 207-217. doi:10.1080/00050068808255605

Bishop, J. (1992). *Evaluating developmental prerequisites*. Unpublished manuscript, Carteret Community College, Morehead City, North Carolina.

Boero, P., Douek, N., & Ferrari, P. L. (2008). Developing mastery of natural language: Approaches to some theoretical aspects of mathematics. In L. D. English (Ed.), *Handbook of international research in mathematics education* (2nd ed., pp. 262–295). New York, NY: Routledge.

Bolt, S. E., & Thurlow, M. L. (2007). Item-level effects of the read-aloud accommodation for students with reading disabilities. *Assessment for Effective Intervention*, 33(1). 15–28. doi:10.1177/15345084070330010301

Borasi, R., Siegel, M., Fonzi, J., & Smith, C. F. (1998). Using transactional reading strategies to support sense-making. *Journal for Research in Mathematics Education*, *29*(3), 275-305. doi:10.2307/749791

Boylan, H. R. (Ed.) (1988). The historical roots of developmental education. *Research in Developmental Education*, *4*(3), 3-13. Retrieved from the ERIC database. (ED341434)

Breen, M. J., Lehman, J., & Carlson, M. (1984). Achievement correlates of the Woodcock-Johnson reading and mathematics subtests, KeyMath, and Woodcock Reading in an elementary aged learning disabled population. *Journal of learning disabilities*, *17*(5), 258-261. doi:10.1177/002221948401700501

Breland, H. M. (1998). *National trends in the use of test scores in college admissions*. Princeton, NJ: Educational Testing Service.

Breland, H., Maxey, J., Gernand, R., Cumming, T., & Trapani, C. (2002). *Trends in college admission 2000: A report of a national survey of undergraduate admission policies, practices, and procedures*. Retrieved from Educational Testing Service website: http://www.ets.org/research/policy_research_reports/publications/report/2002/cnrr

Bridgeman, B., Pollack, J., & Burton, N. (2004). Understanding what SAT Reasoning Test™ scores add to high school grades: A straightforward approach (ETS Research Report Series, 2004(2), i-20). Retrieved from Educational Testing Service website: http://www.ets.org/research/policy_research_reports/publications/report/2004/hsir

Brown, B. A. & Ryoo, K. (2008). Teaching science as a language: A “content-first” approach to science teaching. *Journal of Research in Science Teaching*, *45*(5), 529-553. doi:10.1002/tea.20255

Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.

Burton, N. W., & Ramist, L. (2001). *Predicting success in college: SAT studies of classes graduating since 1980* (Research Report No. 2). Retrieved from The College Board website: https://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2001-2-predicting-college-success-sat-studies.pdf

Camara, W. J., & Echternacht, G. (2000). *The SAT I and high school grades: Utility in predicting success in college* (Research Notes, RN-10). Retrieved from The College Board website: http://research.collegeboard.org/publications/content/2012/05/sat-i-and-high-school-grades-utility-predicting-success-college

Canfield, J. H. (1889). *Opportunities of the rural population for higher education*. Nashville, TN: National Council on Education.

Carter, T. & Dean, E. (Apr-Jun 2006). Mathematics intervention for grades 5-11: Teaching mathematics, reading, or both? *Reading Psychology*, *27*(2-3), 127-146. doi:10.1080/02702710600640248

Case, B (2003). *Accommodations on Stanford 10 for limited English proficient (LEP) students* (Pearson Assessment Report). Retrieved from Pearson website: http://images.pearsonassessments.com/images/tmrs/tmrs_rg/AccommodationsforLEP.pdf?WT.mc_id=TMRS_Accommodations_on_Stanford_10

Colby, S. L., & Ortman, J. M. (2015). *Projections of the size and composition of the US population: 2014 to 2060* (Current Population Reports, Population Estimates and Projections). Retrieved from U.S. Census Bureau website: https://www.census.gov/content/dam/Census/library/publications/2015/demo/p25-1143.pdf

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd Ed.). Thousand Oaks, CA: Sage.

Crouse, J., & Trusheim, D. (1988). *The case against the SAT*. Chicago, IL: University of Chicago Press.

Cullinane, J., & Treisman, P. U. (September, 2010). *Improving developmental mathematics education in community colleges: A prospectus and early progress report on the Statway initiative*. Paper prepared for the National Center for Postsecondary Research (NCPR) Developmental Education Conference, New York, NY. Retrieved from the ERIC database. (ED533871)

Das, K. P. (2008) Reading and mathematics connection: English language learner students' perspective. *Journal of Mathematical Sciences & Mathematics Education*, 3(2), 48-55. Retrieved from: http://www.tsusmell.org/downloads/Products/Articles/MELL_Das_JMSME_Article.pdf

Dougherty, C., Mellor, L., & Jian, S. (2006). *The relationship between advanced placement and college graduation*. (2005 AP Series, Report 1 – February 2006). Austin, TX: National Center for Education Accountability. Retrieved from the ERIC database. (ED519365)

Driscoll, A. (2007). Beyond access: How the first semester matters for community college student's aspirations and persistence (Policy Brief 07-2). Retrieved from the ERIC database. (ED498154).

Flick, L. B. & Lederman, N. G. (2002). The value of teaching reading in the context of science and mathematics. *School Science & Mathematics*, *102*(3), 105-106. doi:10.1111/j.1949-8594.2002.tb17903.x

Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlet, C., & Karns, K. M. (2000). Supplementing teacher judgments of mathematics test accommodations with objective data sources. *School Psychology Review*, *29*(1), 65. Retrieved from the ERIC database. (EJ665781)

Fuchs, L.S., Fuchs, D., & Prentice, K. (2004). Responsiveness to mathematical problem-solving instruction: Comparing students at risk of mathematics disability with and without risk of reading disability. *Journal of Learning Disabilities*, 37, 293-306. doi:10.1177/00222194040370040201

Geiser, S., & Santelices, M. V. (2006). The role of advanced placement and honors courses in college admissions. In Gandera, P. C., Orfield, G., & Horn, C. L. (Eds.), *Expanding opportunity in higher education: Leveraging promise* (pp. 75-114). Albany, NY: State University of New York Press.

Geiser, S., & Santelices, M. V. (2007). *Validity of high-school Grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes* (Research & Occasional Paper Series: CSHE. 6.07). Center for Studies in Higher Education. Retrieved from the ERIC database. (ED502858)

Geiser, S. & Studley, R. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. doi:10.1207/S15326977EA0801_01

Grimm, K. J. (2008). Longitudinal associations between reading and mathematics achievement. *Developmental Neuropsychology*, *33(3)*, 410-426. doi:10.1080/87565640801982486

Grimm, L. G., & Yarnold, P. R. (Eds.) (2000). *Reading and understanding more multivariate statistics.* Washington, DC: American Psychological Association.

Guthrie, J. W. (2001). *No Child Left Behind Act of (2001) – The original ESEA, the new act*. Retrieved from: http://education.stateuniversity.com/pages/2295/No-Child-Left-Behind-Act-2001.html

Hagedorn, L. S., Siadat, M. V., Fogel, S. F., Nora, A., & Pascarella, E. T. (1999). Success in college mathematics: Comparisons between remedial and nonremedial first-year college students. *Research in Higher Education*, *40*(3), 261 – 284. doi:10.1023/A:1018794916011

Hair Jr., J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis with readings*. Englewood Cliffs, NJ: Prentice Hall.

Hawkins, D. A., & Lautz, J. (2005). State of College Admission. *National Association for College Admission Counseling*. Retrieved from the ERIC database. (ED489862)

Helwig, R., Rozek-Tedesco, M. A., Tindal, G., Heath, B., & Almond, P. J. (1999). Reading as an access to mathematics problem solving on multiple-choice tests for sixth-grade students. *The Journal of Educational Research*, *93*(2), 113-125. doi:10.1080/00220679909597635

Hobbs, F., & Stoops, N. (2002). US Census Bureau, Census 2000 Special Reports, Series CENSR-4. *Demographic trends in the 20th century*, *24*, 27-28. Retrieved from: https://www.census.gov/prod/2002pubs/censr-4.pdf

Hoyt, J. (1999). Remedial education and student attrition. *Community College Review*, *27*(2), 51-72. Retrieved from the ERIC database. (EJ597652)

Irons, E. J., & Harris, S. (2007). *The challenges of No Child Left Behind: Understanding the issues of excellence, accountability, and choice*. Lanham, MD: Rowman & Littlefield Education.

Jerman, M. E. & Mirman, S. (Mar. 1974). Linguistic and computational variables in problem solving in elementary mathematics, *Educational Studies in Mathematics,* 5(3), pp. 317-362.

Jiban, C. L., & Deno, S. L. (2007). Using math and reading curriculum-based measurements to predict state mathematics test performance are simple one-minute measures technically adequate? *Assessment for Effective Intervention*, *32*(2), 78-89. doi:10.1177/15345084070320020501

Jordan, N.C., Hanich, L.B., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development*, 74, 834-850.

Jordan, N.C., Kaplan, D., & Hanich, L.B. (2002). Achievement growth in children with learning difficulties in mathematics: Findings of a two-year longitudinal study. *Journal of Educational Psychology*, 94, 586-597. Retrieved from the ERIC database. (EJ653378)

Kerrigan, M. R., & Slater, D. (2010). *Collaborating to create change: How El Paso community college improved the readiness of its incoming students through Achieving the Dream* (Culture of Evidence Series, Report No. 4). New York, NY: Columbia University, Teachers College, Community College Research Center. Retrieved from the ERIC database. (ED509711)

Kintsch, W. & Greeno, J. G. (Jan. 1985). Understanding and solving word arithmetic problems. *Psychological Review*, *92*(1), 109-129. Retrieved from the ERIC database. (EJ313052)

Klitgaard, R. (1985). *Choosing Elites*. New York, NY: Basic Books, Inc.

Korpershoek, H., Kuyper, H., & Van Der Werf, G. (2015). The relation between students' math and reading ability and their mathematics, physics, and chemistry examination grades in secondary education. *International Journal of Science and Mathematics Education*, *13*(5), 1013-1037. doi:10.1007/s10763-014-9534-0

Kyttälä, M. (2008). Visuospatial working memory in adolescents with poor performance in mathematics: Variation depending on reading skills. *Educational Psychology*, *28*(3), 290-307. Retrieved from the ERIC database. (EJ790954)

Laerd Statistics. (n.d.). One-way ANOVA. Retrieved from: https://statistics.laerd.com/statistical-guides/one-way-anova-statistical-guide-4.php

Lamb, J. H. (2010). Reading grade levels and mathematics assessment: An analysis of Texas mathematics assessment items and their reading difficulty. *The Mathematics Educator*, *20*(1), 22-34.

Lee, O. (2004). Teacher change in beliefs and practices in science and literacy instruction with English language learners. *Journal of Research in Science Teaching*, *41*(1), 65-93. Retrieved from the ERIC database. (EJ760025)

Lee, O., Deaktor, R. A., Barnett, J. E. H. , Cuevas, P. & Enders, C. (2005). An instructional intervention's impact on the science and literacy achievement of culturally and linguistically diverse elementary students. *Journal of Research in Science Teaching*, *42*(8), 857-887. doi:10.1002/tea.20071

Lemann, N. (2004). A history of admissions testing. In Zwick, R. (Ed.), *Rethinking the SAT: The future of standardized testing in university admissions* (pp. 5-14). New York, NY: Routledge.

Lohman, D. F. (2004). Aptitude for college: The importance of reasoning tests for minority admissions. In Zwick, R. (Ed.), *Rethinking the SAT: The future of standardized testing in university admissions* (pp. 41-55). New York, NY: Routledge.

McCombs, J. S., Kirby, S. N., Barney, H., Darilek, H., & Magee, S. J. (2004). *Achieving state and national literacy goals, a long uphill road: A report to Carnegie Corporation of New York*. Santa Monica, CA: RAND Corporation. Retrieved from: http://www.rand.org/pubs/technical_reports/TR180-1.html

Mattern, K. D., Patterson, B. F., & Kobrin, J. L. (2012). *The validity of SAT scores in predicting first-year mathematics and English grades*. (Research Report 2012-1). Retrieved from The College Board website: https://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2012-1-sat-predicting-1st-year-mathematics-english-grades.pdf

Menken, K. (2006). Teaching to the test: How No Child Left Behind impacts language policy, curriculum, and instruction for English language learners. *Bilingual Research Journal*, *30*(2), 521-546. Retrieved form the ERIC database. (EJ750294)

Nathan, M. J., Kintsch, W. & Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. *Cognition and Instruction*, *9*(4), 329-389. doi:10.1207/s1532690xci0904_2

National Center for Education Statistics. (April 2010). *Digest of education statistics (2009): Chapter 3 – postsecondary education*. Retrieved from: http://www.nces.ed.dgov/programs/digest/d09/ch_3.asp.

National Commission of Excellence in Education. (April 1983). *A nation at risk: The imperative for educational reform*. Retrieved from: http://www2.ed.gov/pubs/NatAtRisk/risk.html.

Newcombe, N. S., Ambady, N., Eccles, J., Gomez, L., Klahr, D., Linn, M., & Mix, K. (2009). Psychology's role on mathematics and science education. *American Psychologist*, *64*(6), 538-550. doi:10.1037/a0014813

Ní Ríordáin, M. & O'Donoghue, J. (May 2009). The relationship between performance on mathematical word problems and language proficiency for students learning through the medium of Irish. *Educational Studies in Mathematics*, *71*(1), 43-64. doi:10.1007/s10649-008-9158-9

Norušis, M. J. (2009). *PASW statistics 18 statistical procedures companion*. Prentice Hall.

Olivas, M. A. (1979). *The dilemma of access: Minorities in two-year colleges*. Washington, DC: Harvard University Press.

Onwuegbuzie, A. J. (2004). Academic procrastination and statistics anxiety. *Assessment & Evaluation in Higher Education*, *29*(1), 3-19. doi:10.1080/0260293042000160384

Pallant, J. (2007). *SPSS survival manual: A step-by-step guide to data analysis using SPSS version 15*. Maidenhead, Berkshire, England: McGraw-Hill Education.

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students: Findings and insights from twenty years of research*. San Francisco, CA: Jossey-Bass.

Peak, L. (1996). *Pursuing excellence: A study of U.S. eighth grade mathematics and science teaching, learning, curriculum, and achievement in international context*. Washington, D. C.: National Center for Education Statistics.

Peterson, J. J. (1983). *The Iowa testing programs: The first fifty years*. Iowa City, IA: University of Iowa Press.

Pitts, R. J. (1952). Relationships between functional competence in mathematics and reading grade levels, mental ability, and age. *Journal of Educational Psychology*, 43, 486–492. doi:10.1037/h0055218

Poiani, E. L. (1999). Does developmental mathematics work? In B. Gold, et al. (Eds.), *Assessment practices in undergraduate mathematics* (pp. 224 – 228). Washington, D. C.: Mathematical Association of America.

Popham, W. J. (2001). Teaching to the test? *Educational Leadership*, *58*(6), 16-20.

Posner, D. (2004). What's wrong with teaching to the test? *Phi Delta Kappan*, *85*(10), 749-751. doi:10.1177/003172170408501009

Powell, S. R., Fuchs, L. S., Fuchs, D., Cirino, P. T., & Fletcher, J. M. (2009). Do word-problem features differentially affect problem difficulty as a function of students' mathematics difficulty with and without reading difficulty? *Journal of Learning Disabilities*, 42, 99–110. doi:10.1177/0022219408326211

Reikerås, E. K. L. (2006). Performance in solving arithmetic problems: A comparison of children with different levels of achievement in mathematics and reading. *European Journal of Special Needs Education*, *21*(3), 233-250. doi:10.1080/08856250600810633

Richardson Jr., R. C., Martens, K. J., Fisk, E. C., Okun, M. A., & Thomas, K. J. (1981). Draft report for the study of literacy development in the community college. *Report to the National Institute of Education, contract*, (400-78), 0071. Retrieved from the ERIC database. (ED217925)A Report on Literacy Development in Community Colleges: Technical Report

Roueche, J. E., & Roueche, S. D. (1993). *Between a rock and a hard place: The at-risk student in the open-door college*. Washington, DC: Community College Press.

Roueche, J. E., & Snow, J. J. (1977). *Overcoming learning problems*. San Francisco, CA: Jossey-Bass.

Rutherford-Becker, K. J., & Vanderwood, M. L. (2009). Evaluation of the relationship between literacy and mathematics skills as assessed by curriculum-based measures. *The California School Psychologist*, *14*, 23-34. Retrieved from the ERIC database. (EJ878358)

Samway, K. D., & McKeon, D. (1999). *Myths and realities: Best practices for language minority students*. Portsmouth, NH: Reed Elsevier.

Saxon, D. P., Sullivan, M., Boylan, H., & Forrest, D. (2005). Developmental education facts, figures, and resources. *Research in Developmental Education*, *19*(4), 1-4.

Shaw, J. M., Bunch, G. C. & Geaney, E. R. (Oct. 2010). Analyzing language demands facing English learners on science performance assessments: The SALD framework. *Journal of Research in Science Teaching*, *47*(8), 909-928.

Singh, M., & Al-Fadhli, H. (2011). Does school leadership matter in the NCLB era? *Journal of Black Studies*, *42*(5), 751-767. doi:10.1177/0021934710372895

Smedley, B. D., Hector, F. M., & Harrell, S. P. (July/August 1993). Minority-status stresses and the college adjustment of ethnic minority freshmen. *Journal of Higher Education*, *64*(4), 434 – 452. doi:10.2307/2960051

Stern, E. (1993). What makes certain arithmetic word problems involving the comparison of sets so difficult for children? *Journal of Educational Psychology*, *85*(1), 7-23. doi:10.1037//0022-0663.85.1.7

Stewart, D. M. (April 1998). Why Hispanic students need to take the SAT. *Education Digest, 63*(8), 33.

Stoddard, T., Pinal, A., Latzke, M. & Canaday, D. (2002). Integrating inquiry science and language development for English language learners. *Journal of Research in Science Teaching*, *39*(8), 664-687.

Szafran, R. F. (2012). *Answering questions with statistics*. Thousand Oaks, CA: Sage Publications

Tabachnick, B. G., & Fidell, L. S. (2006). *Using multivariate statistics* (5th Ed.). Needham Heights, MA: Allyn & Bacon, Inc.

Texas Higher Education Coordinating Board & Texas Education Agency. (2010). *Texas high school graduates from FY 2009 enrolled in Texas public or independent high education fall 2009*. Austin, TX: Texas Higher Education Coordinating Board. Retrieved from THECB website: http://www.thecb.state.tx.us/reports/pdf/2012.pdf

Thiel, T., Peterman, S., & Brown, M. (2008). Addressing the crisis in college mathematics: Designing courses for student success. *Change: The Magazine of Higher Learning*, *40*(4), 44-49. Retrieved from the ERIC database. (EJ799925)

Thompson, E. N. (1967). Readability and accessory remarks: Factors in problem solving in arithmetic. (Doctoral dissertation, Stanford University, 1967). Dissertation Abstracts International, 28, 2464A

Thurber, R.S., Shinn, M.R., & Smolkowski, K. (2002). What is measured in mathematics tests? Construct validity of curriculum-based mathematics measures. *School Psychology Review*, 31, 498-513. Retrieved from the ERIC database. (EJ667623)

Turkan, S. & Liu, O. L. (2012). Differential performance by English language learners on an inquiry-based science assessment. *International Journal of Science Education*, *34(*15), 2343-2369. doi:10.1080/09500693.2012.705046

University of the Incarnate Word. (n.d.). *About the University of the Incarnate Word*. Retrieved from:
http://www.uiw.edu/home.about/

Vilenius-Tuohimaa, P. M., Aunola, K. & Nurmi, J. (2008). The association between mathematical word problems and reading comprehension. *Educational Psychology*, *28*(4), 409-426. Retrieved from the ERIC database. (EJ799203)

Walker, C. M., Zhang, B., & Surber, J. (2008). Using a multidimensional differential item functioning framework to determine if reading ability affects student performance in mathematics. *Applied Measurement in Education*, 21, 162-181. doi:10.1080/08957340801926201

Webber, C. (1989). The mandarin mentality: civil service and university admissions testing in Europe and Asia. In Gifford, B. R. (Ed.), *Test policy and the politics of opportunity allocation: The workplace and the law* (pp. 33-59). Norwell, MA: Springer.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: problems and remedies. In Hoyle, R. H. (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 56–75). Newbery Park, CA: Sage.

Whilden, B. E. (2010). *The elementary and secondary education act: A primer on reauthorization in 2010*. Retrieved from: http://www.congressweb.com/aascu/docfiles/ESEA%20PRIMER%20FINAL.pdf

Wright, D. A., & Cahalan, M. W. (1985). Remedial/Developmental Studies in Institutions of Higher Education Policies and Practices, 1984. Paper presented at the Annual Conference of the American Educational Research Association (Chicago, IL, April 1, 1985) Retrieved from the ERIC database. (ED263828)

Zwick, R. (2004). *Rethinking the SAT: The future of standardized testing in university admissions*. New York, NY: Routledge.

Zwick, R. (2007). *College admission testing*. National Association for College Admission Counseling. Report Feb. 2007 pp. 1-44. Retrieved from NACAC website: http://www.nacacnet.org/mediacenter/standardizedtesting/documents/standardizedtesting whitepaper.pdf

Zwick, R. (2012). The Role of Admissions Test Scores, Socioeconomic Status, and High School Grades in Predicting College Achievement. *Pensamiento Educativo,* 49(2), 23-30.

Zwick, R. & Himelfarb, I. (2011). The effect of high school socioeconomic status on the predictive validity of SAT scores and high school grade-point average. *Journal of Educational Measurement*, 78, 101-121. Retrieved from the ERIC database. (EJ929738)

APPENDICES

Appendix A

UIW Application for IRB Approval
Part I: Application Form

This application is to be used for initial application for IRB review only. Sufficient time must be allowed for review. Incomplete applications will be returned without review. For a list of application components, see the IRB Manual.

Submit this completed form as part of the application to the Office of Research Development electronically for IRB review. **Do not submit applications directly to the IRB representative**, as this form will be electronically routed to them for review after it has been checked for completion and logged into the IRB database. Signatures will be applied electronically once the application is approved.

Principal Investigator

A Principal Investigator (PI) must be designated for any human subjects research. The PI is responsible for ensuring university and federal regulatory compliance for all research activities and research personnel associated with this protocol. For the responsibilities of the PI, refer to the UIW IRB Manual.

Name:	Phone #:	E-mail:	Mailing Address:
Jayme Agozie			

College/School or Department:	CITI Training Date:	PIDM (UIW ID):
Dreeben School of Education	10/30/2015	

Is the PI a student?

☐ NO

☒ If, YES, a faculty supervisor must be designated for this research protocol. Include a signed copy of the **Faculty Supervisor Agreement** with this application.

Faculty Supervisor

Name:	Phone #:	E-mail:	CPO:
Dr. Glenn E. James			

College/School or Department:	CITI Training Date:	PIDM (UIW ID):
Provost	09/04/2012	

Other Project Personnel

List all other project personnel, including co-investigators, research associates, and student researchers who will be recruiting, consenting, collecting data, or working with data collected from human subjects. Use "enter"/"return" key to list personnel on separate lines.

Name:	Role in Research:	CITI Training Date:	Email:	PIDM (if student):
Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter text.

Research Information

Title of Study:
Broken System, Broken Dreams? The Extent of Differential Validity and Prediction of First-Year College Mathematics Performance Among Subgroups

Research Category: ☒Exempt ☐Expedited ☐Full Board

This research will be conducted:

☐ On the UIW campus or UIW facilities

☒ Off campus (list all locations where research will be conducted):
Data will be requested from the University of the Incarnate Word's Office of Institutional Research and shall be void of all identifiable student information. Analysis will be done at the principal investigator's home office.

Number of Subjects:	Number of Controls:	Total Duration of Study Activities:
Click here to enter text.	Click here to enter text.	Click here to enter text.

This research will involve the following (check all that apply):

☐ Inmates of penal institutions
☐ Institutionalized intellectually handicapped
☐ Institutionalized mentally disabled
☐ Committed patients
☐ Intellectually handicapped outpatient
☐ Mentally disabled outpatient

- ☐ Pregnant women
- ☐ Fetus in utero
- ☐ Viable fetus
- ☐ Nonviable fetus
- ☐ Dead fetus
- ☐ In Vitro fertilization
- ☐ Minors (under 18)

Funding Disclosures
Funding source(s): ☐Internal ☐External ☐Pending ☒None
List all funding sources (pending and awarded): Click here to enter text.
The funding provides for (select all that apply): ☐ Investigator release time or compensation ☐ Research materials ☐ Graduate assistants, student workers, or other project employees ☐ Travel ☐ Other: Click here to enter text.
Financial Conflict of Interest
Does any member of the project team hold financial interest in the funding organization or any similar organization (stocks, board membership, etc)?
☒ NO ☐ If YES, describe below: Click here to enter text.

This Section for Office of Research Development Use Only
Signatures will be applied electronically upon approval

Investigator Signature(s) & Assurances		
I certify that the information above is accurate and complete. I will request prior IRB approval for any changes to the approved protocol and/or informed consent forms, and will not implement those changes until I receive IRB approval. I will report any adverse effects to the IRB immediately. I agree to comply fully with the ethical principles and regulations regarding the protection of human subjects in research.		
Principal Investigator:		
Name: Jayme Agozie	Signature: DocuSigned by	Date: 11/23/2015
Faculty Supervisor (if Principal Investigator is a student):		
Name: Glenn James	Signature: DocuSigned by	Date: 11/23/2015

Approval Signature(s)		
IRB Representative/Reviewer:		
Name: Ana Wandless-Hagendorf	Signature: DocuSigned by	Date: 11/23/2015
IRB Chair (or Chair's Designee):		
Name:	Signature:	Date:

UIW Application for IRB Approval
Faculty Supervisor Agreement

Please read this information and complete the requested fields. **Print, sign, and scan** to submit electronically or **submit in hard copy** to CPO 1216. **A signed copy of the Student Researcher Agreement is required for any research protocol with a student PI.** Incomplete applications will be returned without review.

Application Information			
Title of Study: Broken System, Broken Dreams? The Extent of Differential Validity and Prediction of First-Year College Mathematics Performance Among Subgroups			
Student PI Information			
Name: Jayme Agozie	Phone #:	E-mail:	PIDM (UIW ID):
College/School or Department: Dreeben School of Education		Program of Study: PhD	
Faculty Supervisor Information			
Name: Dr. Glenn E. James	Phone #:	E-mail:	CPO:
College/School or Department: Provost			

I certify that the student named above is knowledgeable of the regulations and policies governing research with human subjects and has sufficient training and experience to conduct this study as described in the proposed protocol.

I furthermore certify the following:

- I have reviewed this application;
- I will maintain knowledge of the direction and completion of the project;
- I will assure the student investigator remains in compliance with UIW and federal human subjects protection policies;
- I assure the student investigator will promptly file for revision, amendment, annual continuing review, or completion of the supervised protocol and will provide assistance to them as needed;
- I assure both the student investigator and I will promptly report any significant or untoward adverse effects to the UIW IRB;
- If this protocol is to be conducted as part of a course, I will ensure the student investigator is informed of the requirement to file appropriate documents at the end of the course; and
- If at any time I am unable to proceed as Faculty Supervisor (e.g., end of the course during which research was planned, sabbatical leave, or exit from the University), I will assist the student in designating an alternate Faculty Supervisor for the remainder of the study.

Student Investigator and Faculty Supervisor Signatures		
Student Principal Investigator:		
Name: Jayme Agozie	Signature:	Date: 11/16/2015
Faculty Supervisor:		
Name: Dr. Glenn E. James	Signature:	Date: 11/16/2015

UIW Application for IRB Approval
Part II: Research Protocol

Provide the requested information and develop your research protocol in accordance with requirements specified in the UIW IRB Manual. Submitted protocols must be in the following format: single-spaced, 11-12 pt. sans-serif (e.g., Arial, Calibri, Helvetica) font. For explanations on each section, follow the Help link.

Submit this completed form as part of the application to the Office of Research Development electronically for IRB review. **Do not submit applications directly to the IRB representative**, as this form will be electronically routed to them for review after it has been checked for completion and logged into the IRB database.

Section 1: Purpose Help

The purpose of this study is to examine the extent to which verbal and mathematics standardized test scores display differential validity and predict first-year college mathematics course performance between various subgroups. The principal investigator also plans to explore other potential predictive variables that are available in Bannerweb and FAFSA data tables accessed by the Office of Institutional Research.

Section 2: Background and Significance Help

This study is based heavily on previous research in the areas of student success, predictive validity of standardized test scores, and their bias. It is well known that institutions rely heavily on cognitive assessments for admissions and the placement of students into courses (Breland et al., 2002). Using only the scores of cognitive assessments can lead to misplacement of students, and have an impact on funding, resources, retention, and matriculation rates (Geiser & Santelices, 2007; Geiser & Studley, 2002). The use of standardized test scores has raised concerns of bias, namely race and ethnicity, gender, and socioeconomic status (Breland, 1998; Breland et al., 2002; Crouse & Trusheim, 1988; Geiser & Santelices, 2007; Lohman, 2004; Zwick, 2007; Zwick, 2012; Zwick & Himelfarb, 2011). An increase in scoring errors on the SAT has also led many to question the validity of the use of cognitive assessments for admissions and course placement (Zwick, 2007).

This study has significance at both the national and local level. On a national level, previous studies have not: 1) considered the predictive ability of verbal scores in mathematics course performance in higher education; 2) examined the demographic shift in the U.S. and the possible impact it may have on the predictive validity of standardized test scores for ELL students; and 3) examined course-taking patterns for Hispanic students. On a local level, there has been a noticeable increase in the number of students: 1) enrolled in developmental courses; 2) enrolled as ELL and or international students, and 3) required to repeat mathematics courses. The hope of this study is to find a correlation that may address these concerns. By performing analyses on these correlations, there may be justification in adopting new placement and admissions procedures to address potential subgroup bias.

Section 3: Location, Facility and Equipment to Be Used Help

The principal investigator plans to collect data solely from the University of the Incarnate Word's Office of Institutional Research. The prinicpal investigator plans to use an existing version of SPSS on her

personal computer, which is located off-campus in a secure office. No other location, facility or equipment is to be used.

Section 4: Subjects and Informed Consent Help

"Physical" subjects will not be used in this study. Only student data that has been previously collected by the Office of Institutional Research is to be compiled and used for the purpose of this research. For the purposes of this study, the data set received from the Office of Institutional Research shall be void of any and all identifiable student information.

Section 5: Subject Compensation Help

There will be no subject compensation.

Section 6: Duration Help

The principal investigator anticipates complete analysis of data and write-up of findings to be concluded within one academic year of approval.

Section 7: Research Design (Description of the Experiment, Data Collection and Analysis) Help

The principal investigator plans to collect data solely from the University of the Incarnate Word's Office of Institutional Research. It is understood that the requested data fields may be collected from either Bannerweb or FAFSA data tables accessed by the Office of Institutional Research. For the purposes of this study, the data set received from the Office of Institutional Research shall be void of any and all identifiable student information.

This study will follow a correlational research design. The data fields to be requested from the Office of Institutional Research include demographic data (age, gender, race and ethnicity), standardized test scores (TAKS, SAT, ACT, Accuplacer, and TOEFL), semester of admission, enrollment status as an international student, enrollment status as a first generation college student, Mathematics course grades (0318, 0319, 1304, 1306, 1311, and 2303), English course grades (1311L, 1311, and 1312), and other UIW course grades. Correlations between the variables in the data are to be researched to determine the nature and strength of their relationships to standardized test scores. Due to the dichotomous, categorical nature of the dependent variable for this research study, success in a first-year mathematics course, the principal investigator will also be using binary logistic regression for further analysis.

Section 8: Risk Analysis Help

No risk.

Section 9: Confidentiality Help

All data and its analysis will be kept in encrypted files within a folder on the principal investigator's personal computer. The computer is in a secure location and requires a passcode for access. A backup copy of the data and its analysis will be kept in the possession of the principal investigator in the form of an encrypted flash drive. All data and analysis will be provided to the dissertation committee upon request and only in review of progress towards completion of the research study.

Section 10: Literature Cited Help

Breland, H. M. (1998). National trends in the use of test scores in college admissions. Princeton, NJ: Educational Testing Service.

Breland, H., Maxey, J., Gernand, R., Cumming, T., & Trapani, C. (2002). Trends in college admission 2000: A report of a national survey of undergraduate admission policies, practices, and procedures. *Retrieved May, 13*, 2012.

Crouse, J., & Trusheim, D. (1988). *The case against the SAT*. University of Chicago Press.

Geiser, S., & Santelices, M. V. (2007). Validity of High-School Grades in Predicting Student Success beyond the Freshman Year: High-School Record vs. Standardized Tests as Indicators of Four-Year College Outcomes. Research & Occasional Paper Series: CSHE. 6.07. *Center for Studies in Higher Education*.

Geiser, S. & Studley, R. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment*, 8(1), 1-26.

Lohman, D. F. (2004). Aptitude for college: The importance of reasoning tests for minority admissions. *Rethinking the SAT: The future of standardized testing in university admissions*, 41-55.

Zwick, R. (2007). *College admission testing*. National Association for College Admission Counseling.

Zwick, R. (2012). The Role of Admissions Test Scores, Socioeconomic Status, and High School Grades in Predicting College Achievement. *Pensamiento Educativo*, 49(2), 23-30.

Zwick, R. & Himelfarb, I. (2011). The effect of high school socioeconomic status on the predictive validity of SAT scores and high school grade-point average. *Journal of Educational Measurement*, 78, 101-121.

Appendix B

UNIVERSITY OF THE INCARNATE WORD

11/23/2015

Jayme Agozie

Dear Jayme:

Your request to conduct the study *Broken System, Broken Dreams? The Extent of Differential Validity and Prediction of First-Year College Mathematics Performance Among Subgroups* was approved by exempt review on 11/23/2015. Your IRB approval number is 15-11-004.

Please keep in mind these additional IRB requirements:

- This approval is for one year from the date of the IRB approval.
- Request for continuing review must be completed for projects extending past one year. Use the **IRB Continuation/Completion form**.
- Changes in protocol procedures must be approved by the IRB prior to implementation except when necessary to eliminate apparent immediate hazards to the subjects. Use the **Protocol Revision and Amendment form**.
- Any unanticipated problems involving risks to subjects or others must be reported immediately.

Approved protocols are filed by their number. Please refer to this number when communicating about this protocol.

Approval may be suspended or terminated if there is evidence of a) noncompliance with federal regulations or university policy or b) any aberration from the current, approved protocol.

Congratulations and best wishes for successful completion of your research. If you need any assistance, please contact the UIW IRB representative for your college/school or the Office of Research Development.

Sincerely,

Ana Wandless-Hagendorf

Ana Wandless-Hagendorf, PhD, CPRA
Research Officer
University of the Incarnate Word IRB