12-2004

A study of education reform in Texas: State mandated testing results, effects, and the University of Texas-Pan American

Monica M. Trevino
University of Texas-Pan American

A STUDY OF EDUCATION REFORM IN TEXAS: STATE MANDATED TESTING RESULTS, EFFECTS, AND THE UNIVERSITY OF TEXAS - PAN AMERICAN

A Thesis

by

MONICA M. TREVIÑO

Submitted to the Graduate School of The University of Texas Pan American in partial fulfillment of the requirements for the degree of

MASTER OF PUBLIC ADMINISTRATION

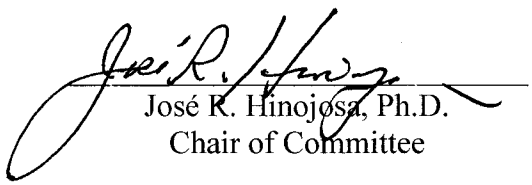
December 2004

Major subject: Public Administration

A STUDY OF EDUCATION REFORM IN TEXAS: STATE MANDATED TESTING RESULTS, EFFECTS, AND THE UNIVERSITY OF TEXAS – PAN AMERICAN

A Thesis
by
MONICA M. TREVIÑO

Approved as to style and content by:



José R. Hinojosa, Ph.D.
Chair of Committee

John R. Milford, MPA
Committee Member

Espiridion "Al" Borrego, Ph.D.
Committee Member

December 2004

ABSTRACT

Treviño, Monica M., A Study of Education Reform in Texas: State Mandated Testing Results, Effects, and The University of Texas Pan American. Master of Arts (MA), December, 2004, 97pp., 14 tables, 0 illustrations, references, 77 titles.

The study will use scores on TASP, and ACT as predictors of first year college success. First year college success will be measured by first-year year college grade point average and number of class hours dropped. We expect to find a high correlation between standardized test scores and college success.

In order to continue yearly state test administration efforts, data needs to be found supporting that these efforts are sufficient. If this evidence is not present, then state mandated achievement testing needs to be reconsidered for use as a requirement for receipt of a high school diploma, school accountability, and as a measurement for future student success. The research will be accomplished through use of research questions and research statements.

DEDICATION

...for everyone who supported me throughout this process...

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

> "...Two prevailing goals in American education have been 1) providing equal access to public education and 2) efficient delivery (Haladyna, Haas, & Allison, 1998, p.2)."

One key component in meeting those goals has been through evaluation. Evaluation is the vehicle to observing individual student learning and comprehension, and gauging how that learning compares to that of his peers (Dietel, Herman, & Knuth, 1991). Further, evaluation is critical for maintaining standards and helping institutions meet objectives for equity in curriculum, instruction, and across institutions of learning for all students (Cizek, & Seldon, 1992; McKenna, 1977; Ravitch, 1995). Educators may employ various means to make this estimation thereby allowing for methods such as in class quizzes or observations and often requiring more sophisticated methods such as standardized tests (Dietel, et al, 1991).

The implication that a student's current knowledge is based on the amount of his previous learning suggests that educators can compare student achievement over time and make decisions based on the outcomes (Dietel et al, 1991). Judging those outcomes equally among diverse groups of students in terms of quality and quantity of learning gained by each participant is a critical aspect of standardized education assessments and is important to the education of all students (Dietel et al, 1991). A dearth of research

supports and negates this justification for various yearly test administration efforts. This research will concentrate on how well standardized tests predict first year college success at The University of Texas - Pan American, a South Texas institution with a predominately Hispanic student population. First year college success will be measured using students' first year college grade point average and first year number of class hours dropped. The expectation is that standardized test scores will predict college success. If this evidence is not present, then the accuracy of state mandated achievement testing for school accountability, student receipt of a high school diploma, or as a measurement for future student success, is not universal and needs to be reevaluated.

Standardized Testing

The "standardized" label describes exams whose scoring methods and exam content is similar for all test takers. The exam is developed to measure specific competencies and is administered to all test takers in a similar, pre determined fashion. Standardized tests measure student knowledge and academic preparedness, and predict student success (Dietel et al, 1991). They should not adversely affect members of a particular race, gender, or socio economic background (Butterfield, 1995; Gipps; 1994, McKenna, 1977).

The use of tests to evaluate student achievement and learning and to influence education policies and public opinion (Butterfield, 1995; Gipps, 1994; McKenna, 1977) makes producing the most fair and accurate instrument a priority. To assure that despite race, gender, or socio-economic status, every student has the same level of expectations and opportunity (Texas Higher Education Coordinating Board, 2000) education agencies rely heavily on field testing, sampling a large number of students before an assessment is adopted and implemented (Texas Education Agency, 2001). This sampling is especially

important in evaluating the instruments in education, as standardized tests are usually developed by a testing company. The type of test developed is usually either a norm-referenced or a criterion-referenced test (Butterfield, 1995; Gipps, 1994; McKenna, 1977). Norm-referenced exams are multiple choice exams designed to compare test takers' performance on a bell curve (Butterfield, 1995; Gipps, 1994; McKenna, 1977). Criterion-referenced tests are designed to judge each test taker individually, with scores calculated on the basis of mastery on test objectives (Butterfield, 1995; Gipps; 1994, McKenna, 1977). Standards-based/standards referenced exams are a more recent variation of the criterion reference test. They are often administered statewide to gauge student comprehension of state curriculum objectives as predetermined by education experts (Fair Test, 2002).

History of Standardized Tests in the United States of America (U.S.)

The movement toward standardized testing in the U.S. has a long history (Perrone, 1991; Haladyna, et al, 1998; Madaus & O'Dwyer, 1999). In the late 1800s achievement tests were developed to evaluate individual test takers (Haladyna, et al, 1998). This would change in the coming decades as identifying students' academic potential and ensuring equal standards of learning increased in focus and ability testing gained popularity (Haladyna, et al, 1998). In the 1900s intelligence testing gained popularity as a way to evaluate students. Criticism of the use of the tests grew (Haladyna, et al, 1998), but so did its use.

According to Haladyna, there was a disconnect between the standardized achievement test and school based learning objectives. This led to the development of criterion-referenced exams, which were able to gauge the level of classroom learning

obtained by students (Haladyna, et al, 1998). Criterion reference tests are developed around a core set of learning objectives that educators determine. They judge each test taker individually and calculate scores by the mastery of test objectives (Butterfield, 1995; Gipps; 1994; Haladyna, et al, 1998; McKenna, 1977). The major criticism of this type of standard assessment was and continues to be the ability of educators to merely teach the tested objectives, which can curtail creativity and learning in classrooms (Haladyna, et al, 1998).

With the various tests developed and claims of proponents and opponents voiced, the early 1980s report by The National Committee on Excellence in Education validated the use of achievement testing by calling for increased use of the measurement tool (National Commission on Excellence in Education, 1983). According to the report, *A Nation at Risk*, United States students were not performing at the level of their international counterparts, reportedly achieving less in areas such as science achievement testing and college graduation (National Commission on Excellence in Education, 1983). The need for large scale changes in the American education system was evident and with the commitment of various education partners, the committee was sure that American leaders and students would deliver (National Commission on Excellence in Education, 1983). After all, education is one of the most significant areas addressed by state and local governments.

As such, when government and its constituents became increasingly concerned that American students were falling short of international standards, focus turned to the states to spearhead the changes necessary to ensure education quality. The states responded, creating or amending their accountability systems (Bauer, 1995). They moved

to increase their state education standards, increase graduation requirements, improve salary amounts offered to teachers, and/or implement standardized assessments for students and teachers (Bauer, 1995; Finn & Rebarber, 1992). This shift to stronger accountability has not occurred without controversy. As state education systems continued work on their strategies to address concerns made public in *A Nation at Risk*, criticism of standardized multiple-choice testing emerged (Madaus & O'Dwyer, 1999).

Once again, the use of the tests did not decrease. This is likely attributable to the strong public support of education and a concern for the state of education in the United States. Polls indicate that voters support issues related to education and oppose issues or candidates perceived as unsupportive of education (Pitsch, 1996). The American public strongly supports education (Pitsch, 1996) and although the state is legally responsible for this education, voter support catapults the issue into the political arena, having a profound effect on the state and national education reform policies that influence the state of American education (Dorn, 1998; Pitsch, 1996). One such effect has been the adoption of high stakes testing.

High Stakes Testing

Recent trends in American education include using standardized test result data to make decisions about the future of a student, school, and/or education policy or process. Placing this level of importance on the outcomes of testing is called "high stakes testing" (Fair Test, 2003). A good example of the use of high stakes testing is states' utilization of student performance on standardized exams as a requirement for graduation (Fair Test, 2003).

In Texas, the adoption of standardized testing to measure student achievement has been a major development over the past several decades (Sergiovanni, 1990; Toch, 1992; Texas Education Agency [TEA], 2000) and, in recent years, their strategy has extended to incorporate high stakes expectations. These tests are used because they gauge student knowledge, predict student preparedness for the university, and hold schools accountable for student learning. In Texas, improvements attributed to high stakes testing are welcomed changes in the state's education system. However, the changes in scores may or may not reflect increased learning or improved equity in the system, and may or may be an indicant of future student performance. A common factor in the discussions is the interpretation of student learning and student impact through use of statistical measures dependent on student performance (Dorn, 1998).

Texas Testing Trends

The effects of the publication of *A Nation at Risk* combined with the Texas State Department of Education's history of controversy before and during the 1980s movement prompted changes, including the incorporation of standardized testing. Texas was still under scrutiny for its education financing system (Haney, 2000; Kramer, 2002). Perhaps the state's expectations of schools were unrealistic given the lack of financial support offered to the districts to comply with state requirements (Kramer, 2002). The differences in funding dollars per pupil ranged significantly from poor to affluent districts; taxable property value ranged from approximately 20 thousand dollars per student, to about 14 million dollars per student (Kramer, 2002). Consequently, the state endured extensive litigation resulting in the public school finance system being ruled as violating the Texas Constitution (*Edgewood v Kirby*, 1988). Serious restructuring of the system occurred

based on the belief that all students, regardless of the wealth of their districts, deserve a quality education.

The resulting strategies have included higher distribution of wealth among districts and enriching the Texas school curriculum with items such as "The Essential Elements," the curriculum guide for Texas schools before the Texas Essential Knowledge and Skills (Haney, 2000; TEA, 2001); the "no pass no play" rule, the requirement of minimum grade attainment to qualify for participation in school extra curricular activity (Zirkel & Gluckman, 1993); and the increased reliance on standardized tests as a measure of student achievement (Haney, 2000; TEA, 2001). Of these modifications, perhaps the most controversial continues to be the use of standardized testing to gauge student knowledge, predict student preparedness for post secondary education, and to hold schools accountable for student learning.

These efforts seem to be working: Scholars have noted that students within this evolving system have made significant progress on various academic measures, including performance on standardized exams (Fuller & Johnson, 2001). The Texas Education Agency reports various improvements: That student high school completion rates have increased (TEA, 1999); that the gap between minority and Caucasian students has narrowed (TEA, 1999); that dropout rates have decreased (TEA, 2000); and that the number of graduating students taking college admission tests has increased and so have their test scores (TEA, 2001). Additionally, in addressing the State's higher education system, a number of authors describe higher education structure in Texas as "ever changing" and dynamic (Richardson et al., 1999).

Interest in standardized testing will continue and concerns about the units used to measure student achievement and whether they are the most fair and appropriate evaluation techniques will remain (Haladyna, et al, 1998). The public is interested in the amount of knowledge retained by American students and in how comparable their achievement is relative to their U.S. and international counterparts. This makes education an important political issue (Pitsch, 1996). Politicians spotlighting their commitment to education may hold positive regard for standardized tests, particularly when compared to other methods of education reform (Linn, 2000). It is imperative that they have as much information as possible to guide them in making these decisions about standardized test use in education reform.

In addressing the foundation of the study, it is important to include information about some of the standardized exams that are relevant to the study. The 1998 student cohort examined in the current study had to pass the Texas Assessment of Academic Skills (TAAS) or approved equivalent high school exit exam, and the Texas Academic Skills Program exam (TASP). Each student was also required to take the American College Testing Program Assessment (ACT), or Scholastic Aptitude Test (SAT), prior to consideration for acceptance into a post secondary institution in Texas.

Texas Assessment of Academic Skills

The TAAS, started in 1991, is a criterion-reference exam used to measure Texas statewide curriculum in elementary, junior high, and high school. Students are tested in several subjects, including English and mathematics (Just for Kids, 2001).

Texas Academic Skills Program

The Texas Academic Skills Program (TASP) is more than a standardized exam administered to gauge student knowledge. It is an entire program that signals Texas institutions of the level of basic skills an incoming college student possesses so the university can determine and provide supplemental instruction in areas where the student could use development in order to do well at their institution. The reason this exam and program is important is because without these skills, a university student is believed to be at a disadvantage with respect to completing their coursework. The TASP exam is the precursor to a large scale advisement process and supplemental programs required by the state so that students at their institutions will be successful at the university level. The university is then responsible for providing that advisement and educational support. Admission to a university is not affected based on TASP exam scores. However, there are limits to how many college courses a student may enroll in prior to passing the exam (TEA, 2001).

American College Testing Program Assessment

The American College Testing Program Assessment (ACT) is a standardized, multiple choice test meant to predict first year college grades. The ACT is composed of four tests which are scored individually. Additionally, a composite score is formed to give an overall indication of a student's performance. The ACT was developed as an alternative to the SAT (ACT, 2001).

Scholastic Aptitude Test

The Scholastic Aptitude Test (SAT) is a standardized exam used by colleges as part of their admission criteria. It measures verbal and quantitative aptitudes (College

Board, 2001). Research has shown that high school graduates who obtain certain scores on this exam have a better chance of obtaining a B or better in their first year coursework.

> "The Stanford Achievement Test is recognized as the first of the large-scale publishers' tests (Haladyna, in press). First published in 1923, it continues to be an acknowledged leader in its field....The ACT Assessment and the Scholastic Aptitude Test (SAT) became the nation's leading college admissions tests. (Haladyna, et al, 1998, p.2)"

What is certain is that the issue of standardized testing has a prominent position within discussions on educating the youth. The rhetoric over the state of education often includes summarizations of complex issues such as student achievement and persistence, quality of education, equal learning opportunities for diverse populations and education policy implications (Dorn, 1998; Pitsch, 1996). Are our students pursuing and obtaining higher education degrees? Are all students getting a quality education? Do minority students get the same level of teaching as non-minority students? Standardized testing serves as a benchmark for discussion and change.

STATEMENT OF THE PROBLEM

The problem for the study is as follows:

There is a dearth of information defining the meaning of high performance on standardized exams, thus limiting the extent to which standardized exams can be used and leaving room for growing controversy over current uses. For example, despite evidence that Texas education reform has been effective, the use of tests to measure knowledge and preparedness, to serve as an exit requirement, to determine course placement and/or to predict academic performance is controversial. Some critics question the effectiveness of Texas' use of standardized tests.

Additionally, the majority of research that has been conducted has been at universities without Hispanic student majority populations. With the Hispanic population

growing as rapidly as it is the importance of understanding the impact of standardized tests on this population increases. If standardized testing gauges student knowledge and preparedness, then it should be a good predictor of student success at The University of Texas-Pan American. The problem for the study is to test how predictive of student first year performance standardized exams are in a predominately Hispanic sample.

PURPOSE OF THE STUDY

Uncertainty as to the direct reasons for student score improvements and conflicting views over the exact impact of standardized tests on all students place policy initiatives reliant on standardized test results under great scrutiny. That being said, if state test administration policies are to remain in effect, it seems imperative to find data supporting that these efforts are important to the education of students. Texas students enrolling at Texas universities should pass the TASP exam and should perform well in their first year of college.

The purpose of the study will be to analyze the predictiveness of standardized tests on the 1998 University of Texas- Pan American College of Science and Engineering population. Ethnicity of the sample is taken into account although the vast majority of the sample is Hispanic. The predictiveness of standardized test scores in a predominately minority population will be interesting to observe, particularly because studies are commonly conducted in universities where the student majority is Caucasian-non Hispanic. The study will use regression analysis to test the relationship between the independent and dependent variables.

HYPOTHESIS AND RESEARCH QUESTIONS

Standardized tests measure the amount of preparation students have in a given subject (Corbett & Wilson, 1991). In Texas, high school students with the intention of pursuing a university degree must show proficiency in reading, writing, and mathematics in order to pass the exit TAAS exam (or an approved substitute exam). To pursue higher education in Texas, students must meet the requirements of the school in which they will be enrolling, which will be inclusive of certain scores on a college entrance exam, and take the TASP exam. Scores on these exams will determine placement in certain courses, based on the belief that these exams are representative of a student's accumulation of knowledge and are predictive of how a student will perform their first year at a university. The study intends to show that test scores can predict a certain level of success with regard to first-year G.P.A. and number of course hours dropped. The research questions of the study are based on the assumption that standardized tests are good predictors of first year student success at a university as well as on the following research statements. To make this determination, prediction models based on the following four research statements will be evaluated. There will likely be varied achievement levels based on gender and ethnicity, as studies show that these factors affect student score outcomes (Weiss, 1998), but success levels of students passing these exams should be high.

The research statements being tested are:

1) Students with higher TASP scores are more likely to have a higher GPA their first year of college.

2) Students with higher ACT scores are more likely to have a higher GPA their first year of college.
3) Students with higher TASP scores drop fewer classes during long (fall and spring) semesters.
4) Students with higher ACT scores drop fewer classes during long (fall and spring) semesters.

Statistical Data Analysis

The data collected is for the 1998-1999 school year from the Office of Admissions and Records at The University of Texas - Pan American and through student records on the Student Information System (SIS). The data set includes only entering freshmen in the College of Science and Engineering. Each student is identified numerically in the order in which they are listed for the College of Science and Engineering. To test the predictive ability of tests on G.P.A. and number of class hours dropped, regression analyses will be performed.

Multiple regression analysis will be used to test the whether the standardized test scores predict student success outcomes like GPA and class hours dropped. A predictor model is developed for each subtest/outcome combination, resulting in a total of 12 models. The ACT composite score was treated as a subtest during statistical analysis; despite that it is the arithmetic mean of its constituent subtests in an effort to remain consistent with its use in academic settings.

OVERVIEW OF THE SUBJECTS

Research indicates that forty percent of a four-year institution's entering class will not earn a degree. Of that group more than half of the attrition is seen within the first year

(Tinto, 1993). Reasons for student withdrawal vary however, academic difficulty accounts for the main factor in over thirty percent of all students who do not graduate (Tinto, 1993). Many students are under-prepared for university coursework.

In 1997, of all Texas high school graduates who graduated with the advanced "honors" degree, about forty-eight percent were college ready. Only forty-two percent of students graduating with an advanced degree were college ready, and for students graduating with a regular degree, only twenty-six were college ready as determined by the Texas Academic Skills Program (TASP) test (Lopez, 2000). With academic difficulty contributing to attrition at higher education institutions, knowing whether an entering student is prepared for their first year is of great importance to an institution.

Use of standardized test scores to determine college readiness is understandable. The University of Texas - Pan American (UTPA), a four-year public institution located in deep South Texas, uses score information from the TASP and ACT as part of its student placement process. This branch of the UT-System is near the Mexican border, in Hidalgo County and will be the base for this research. The state demographics are unique in that there is a large minority population. Education strategies have had to be inclusive of this diversity but controversy has been imminent. Much of the literature on standardized testing in Texas focuses on the impact (positive or negative) that these assessments have on the youth in the state.

The University of Texas - Pan American's student population is also predominantly Hispanic (86% in fall 1999 as per the Lujan et al, 1999). In addition for sixty-three percent of the student body no parent has obtained a college degree. The majority of students qualify for and receive needs based financial aid (sixty-five percent

as per Lujan et al, 1999). A significant fraction of the student body consists of non-traditional students, as indicated by the average undergraduate student age of 23 (Lujan et al, 1999). Students involved in the study are 1998 entering freshmen.

Limitations

The limitations of this study are primarily in the sample used for the study. The study is restricted to the population of College of Science and Engineering students who entered The University of Texas- Pan American in Fall 1998. Socioeconomic status is not accounted for and there is very little ethnic diversity within the sample. As the study is a cross sectional study of first year students in a particular college of study, it does not account for all students, graduation, or job placement.

THEORETICAL FRAMEWORK

This research is dependent on assumptions inherent in prediction research. First, the assumption made is that the exams are valid and reliable instruments. Second, test material is curriculum based therefore scores on standardized exams reflect ability and/or knowledge needed to be successful during first year coursework. Third is the assumption that grades are fair, accurate, and reflective of knowledge and/or needed skills (Noble, 1991).

Some of the prediction formulas found in the literature are useful for the current study. In 1987, Laing, Engen, & Maxey use the following models to analyze the relationship between highschool courses taken and ACT scores:

1	Related ACT Test	Number of Related Courses
2	Related ACT Test	High School Rank
3	Related ACT Test	High School Grade Point Average

In Noble and Sawyer (1987) academic prediction models are used to analyze the relationship between ACT scores plus self reported high school grades to freshmen college grades. Their research uses multiple regression techniques to analyze:

1 ACT Subtest scores

2 High school grades

3 ACT Subtest scores and High School Grades

SIGNIFICANCE OF THE STUDY

With legislators promoting standardization and accountability to ensure that no student is left behind, knowing whether standardized testing is a long term solution, or popular one, is pertinent to enact effective policy. Minority population increases reflect the need for equity and reaffirm the importance of studying the uses and effects of standardized testing. As standardized exam scores are used to predict first-year student success in higher education (Noble & Sawyer, 1987; Noble, 1991), the relationship between student performance on these exams and their performance at the university is important. The study will focus on a population of Texas students, predominately minority, to assess the predictive ability of standardized exam scores on student success at The University of Texas - Pan American. The study is based on literature pertaining to the effects of state standardized testing, issues of concern with standardized testing, and their relationship to student success. This study builds on previous research on education reform in the state of Texas and adds to the literature currently available about that state.

There is a great deal of literature that suggests that students' performance on standardized exams is reflective of the quality of education they have received, thus being a good benchmark of their knowledge and potential performance in an academic setting

(TEA, 2000). In addition, the literature suggests that although gaps in test performance levels may be found on the basis of gender, ethnicity, and socio economic status, raising standards and setting higher expectations of all students will help to narrow the gaps (Ravitch, 1995). Chapter two will review literature pertinent to the current study. The chapter will include information on the adoption of standardized tests to measure student learning and school accountability, increase access and equity in education, to predict student preparedness for the university, and to develop systems to remediate students who do not meet set standards, as part of education strategies designed to improve learning for all students.

Chapter three will detail the research and methodology employed by the study. An explanation about the data gathering and analysis will be provided. A discussion on the research objectives and research questions will be given. The hypotheses being tested stem from the assumption that standardized tests are good predictors of student first-year success at a university. To determine this, the study must answer whether there is a significant relationship between standardized test scores and students' first-year performance at a university.

Chapter four will present the results of the study. The chapter compliments the methodology section of the research by detailing the results of the multiple regression models. Special attention will be given to models yielding statistically significant results.

Chapter five will discuss the implications of the research findings. If evidence suggests a statistically significant relationship between higher test scores and student first year performance, then the use of standardized tests is validated. If this evidence is not found, then the use standardized tests for school accountability, student receipt of a high

school diploma, or as a measurement for future student success, needs to be reevaluated. Suggestions for future research will be given. The policy implications of the study will also be discussed.

CHAPTER TWO

LITERATURE REVIEW

> "Good assessment information provides accurate estimates of student performance and enables teachers or other decision makers to make appropriate decisions" (Dietel et al, 1991).

From measuring student knowledge and academic preparedness (Ravitch, 1995), to determining the likelihood of student persistence, predicting first-year college grades, determining course placement, and assuring equity in education, assessments help maintain standards and meet objectives (Neill, M. D. & Medina, N. J., 1992; McKenna,1977). Education stakeholders have employed these assessment techniques to evaluate and compare students (Dietel et al, 1991). The ideology that the use of standardized assessments is accurate and fair for all students should be studied to address the criticism that standardized tests hinder the student achievement they purport to evaluate (Valenzuela, 1999; Haney, 2000).

While it is the increase in national attention on education reform that reinforced the need for change at the state level (Finn, 1992), student access to education and accountability of schools have become the foci in states that are incorporating the use of tests (Dorn, 1998). In Texas the policy decision has received compliments and criticism from supporters and skeptics. Supporters believe Texas education improvements are due to implementation of standards for the statewide curriculum (TEA, 2001; Scheurich, et al, 2000; Trueba, 2001; National Center for Developmental Education, 1996; Flores &

Clark, 2003). Further, they regard the state's implementation of standardized high school exit exams and that the increase in scores for minorities as a byproduct of the Texas system's aggressive approach to increasing standards for all students in its schools (TEA, 2001; Scheurich et al., 2000; Trueba, 2001; National Center for Developmental Education, 1996; Flores & Clark, 2003).

Political and Social Significance

Sherman Dorn's research on the political entrenchment of standardized tests in the evaluation and accountability of schools outlines the events leading to the American education system's dependence on statistics for measurement of success (Dorn, 1998). Regardless of circumstances, teachers and students are judged on their performance on standardized tests but the results are positive (Ravitch, 1995) in spite of the controversies that have emerged. Dorn (1998) posits that these tests have become the dominant way to evaluate schools and thus the debate is no longer whether the system will last, but by whom will it be controlled.

Costs

There have been a considerable number of studies conducted to determine the true costs of standardized tests (Phelps, 2000). Most of the studies contend that standardized testing is relatively inexpensive, most costs totaling under 25 dollars per student and under 1 percent of a school's budget for students (Phelps, 2000). Performance-based testing and other tests requiring professional grading are more expensive (Phelps, 2000). When considering cost as part of test considerations it seems that for a relatively reasonable price, administrators can obtain reliable information about their student population (Phelps, 2000).

The arguments presented by Phelps make good theoretical sense. Yes, testing costs should be limited to general material and scoring costs with a few marginal expenses, however, one cannot discard arguments made by critics such as Walt Haney who charges that teachers are teaching test taking strategies and concentrating solely on the kinds of material that will be tested (Haney, 2000). Additionally, there are claims that educators are not happy with the current testing system and that it costs billions of dollars (Dietel et al, 1991). In this respect, the cost of testing is complicated by costs associated with student and professional creativity, motivation and improvement.

Phelps addresses some of these concerns by citing studies that show a link between standardized testing programs and student success (Phelps, 2000). He accounts for costs associated with development, shipment, distribution and administration of the exams (Phelps, 2000). The conclusion is that standardized testing is a cost effective endeavor. For the purposes of the current study, one could argue that the cost effectiveness of the tests in higher education is dependent upon how well they predict first-year student success.

The Effects of Traditional Tests

Evaluation is a critical aspect of education, particularly because of the potential impact education can have on the future of a person or community. Historically information has been gathered for assessment by various education stakeholders, for whom comparing scores among students and measuring student learning is critical (Fair Test, 2003; Dorn, 1998; Dietel et al, 1991). Proponents of standardized testing will argue that the use of the exams provides opportunities to increase equity, improve classroom effectiveness, and reduce deficit thinking. They state that if a student knows he is

expected to perform at a certain level, he will strive to achieve that goal. Critics cite evidence indicating that testing does not lead to better educational policies and practices. What is certain is that the effect of these tests on education stakeholders such as teachers, students, parents, administrators and politicians (Dietel et al, 1991) and on the shaping of our accountability systems is extensive (Dorn, 1998).

Outside assumptions about standardized testing include both positive and negative rhetoric. According to James Scheurich and colleagues, three historic opportunities have emerged as a result of accountability movements (Scheurich et al, 2000): First, the systems have provided a window of opportunity to focus on underprivileged youth which was not previously experienced. Second, the systems have brought to light the deep commitment to education. Third, they also seem to have helped minority groups (Scheurich et al, 2000). He proposes that accountability systems have painted a true picture of our successes and failures in reaching low income students and students of color. He stresses that prior to implementation of accountability systems there was no concrete way to determine where the system was lacking. Now student success and failure patterns are easily determined and available for all stakeholders. Efforts can be made to remedy deficiencies.

Standardized tests measure the amount of preparation students have in a given subject (Corbett & Wilson, 1991). It is argued that those who support high-stakes testing do so under the assumption that school performance has been negatively impacted because insufficient attention has been paid to it (Sheldon & Biddle, 1998). While Scheurich does not specifically state that education was negatively impacted by a lack of attention, his praise of the systems' influence on focusing all stakeholders on education

lends credibility to Sheldon and Biddle. Under that premise, supporters understand that in order to change student behaviors and expectations, standards have to be assessed in some manner (Ravitch, 1995, Sheldon & Biddle, 1998; Bauer, 2000). While the argument can be made that there was never a problem with student learning, that discussion is best made in an entirely different research paper. For the purpose of the current study the assumption is that the 1980s national report and other trends supported by literature and policymakers did indeed warrant address. Continuing under that premise, the study presented by Scheurich seems particularly relevant. However, the argument does not answer the relevance of these tests to first-year success in higher education, an issue that will be addressed by the study.

Proponents of aforementioned arguments can be grouped with those who believe Texas education improvements are due to implementation of standards for the statewide curriculum and the state's implementation of standardized high school exit exams. Both groups believe in the benefits of standardized testing.

> "To solve such problems, according to this view, we need to set high standards for students, assess students' performance with standardized tests, and reward or punish students, their teachers, and their schools, depending on whether those standards are met." (Sheldon & Biddle, 1998; Bauer, 2000)

Their data seems particularly supportive of the effect of testing and accountability systems on minority and low income students. It supports the Texas accountability system and others like North Carolina as having raised student achievement (TEA, 2001; Scheurich et al, 2000; Trueba, 2001; National Center for Developmental Education, 1996; Flores & Clark, 2003). There are those who disagree. Haney argues that the data supporting Texas education advancements is an illusion (Haney, 2002). To him and

others, testing has not been helpful to student learning, it has been a limiting and discriminatory tool affecting students and teachers alike.

> "The higher the stakes, the greater the pressure that is placed on teachers and administrators to devote more and more time to prepare students to do well on the tests. As a consequence, narrowly focused tests that emphasize recall have led to a similar narrowing of the curriculum and emphasis on rote memorization of facts with little opportunity to practice higher-order thinking skills." (Dietel et al, 1991)

Some examples of the consequences of these tests include the idea that teachers teach to the test, limiting their instruction time to test related material and structure. Critics of testing see this as a severely limiting aspect of current education systems. Instructors are not encouraged to seek innovative modes of instruction as traditional, seemingly effective tools are used. To change or attempt new instructional practices that may take time to develop and implement properly would not be supported by the timetables of test administration. This technique would be ineffective in many professional environments. Unfortunately, the constraints of an accountability system that is dependent on standardized testing also restrict teachers who are likely to structure their in class assessments around the format of timed standardized tests. Instructors are forced to produce results and thus instructional creativity is not encouraged or supported (Dorn, 1998). This provides little flexibility for a teachers grading of student learning and stifles thoughtful and creative responses that students would provide in short answer or essay assessment as opposed to structured multiple choice responses (Dietel et al, 1991).

Also a topic of discussion is the Texas Academic Skills Program (TASP), a requirement for college entrance in Texas. Despite support for the instrument itself, the implementation of the TASP is found to be lacking. There is poor evaluation of

remediation programs, poor communication between secondary and higher education programs, and remedial education instruction quality is lacking. Some effects cited by the consultants are disturbing. There are students who remain unable to earn an associates degree due to TASP restrictions. Most students who fail the TASP pass the exam on their second attempt (the exception being math) however students who fail the TASP more than twice are very likely to drop out of college. Additionally, minority students are more likely to fail in subsequent attempts (National Center for Developmental Education, 1996). The implications of the findings indicate that despite intending to help students persist and achieve success in higher education, the Texas system may be severely limiting. What is particularly disturbing is that this result is likely not related to the test itself, rather the implementation of the test remediation policy. The TASP exam scores gathered for the current study may help provide information as to whether the students in the group persisted despite high or low TASP scores.

While assessments are used to improve teaching and learning, and promote remediation where necessary, one argument against testing states that it does not measure or promote higher thinking (Dorn, 1998; Dietel et al, 1991; Amrein & Berliner, 2002; Klein et al, 2000; McNeil & Valenzuela, 2000; Haney, 2000; Koretz & Barron, 1998). If this is accurate, then the idea that testing promotes learning of all students would be jeopardized, as certain students would remain unchallenged. Gifted or advanced students would not benefit from testing programs.

To counter, there is literature that supports that passing a standardized exam is a byproduct of a teacher's efforts, independent of special test taking related instruction. As a result, there should be no undue stresses on educators to change their instructional

approaches (Finn, 1992). Gifted students should remain challenged and less advanced students should receive the necessary support to learn, automatically resulting in improved test scores.

Despite the controversy, standards are believed to assure every student the same level of expectations (Ravitch, 1995; TEA, 1995; Lopez, 2000) and provide every student with quality education (Ravitch, 1995) while highlighting the quality of education in our academic systems. Much of the current research supports the use of standardized exams in gauging a student's current academic achievement, how that compares to the academic level of his peers, and predicting an entering student's first year performance at a university (Corbett & Wilson, 1991; Gottfredson & Crouse, 1986; Pui Wa-Lei & Schultz, 2001; Noble & Sawyer, 1987). Also, studies using test results show that student performance on standardized tests are similar whether the test is high stakes or low stakes (Greene, Winters & Forster, 2003). This lends credibility to the idea that high stakes promotes achievement, as students were able to transfer their learning on to both exams (Greene et al, 2003).

To further solidify this claim, Greene and colleagues tested the idea that high stakes testing results are reliable by comparing student score results on both a high stakes and low stakes test. Their results indicate that high stakes data is reliable.

The issue of accuracy and fairness

The accuracy and fairness of assessing students is one of the most debated aspects of testing (Linn, 2000). Supporters praise the test development process and results claiming that expert opinion is an integral part of the development and evaluation process (TEA, 2001); that the evaluation of students provides information to educators and

administrators that can be used to shape curriculum and instruction to improve learning for students, particularly in areas that a test may determine the student is lacking proficiency in. They see them as the best gauge we have of/for student learning. Critics are unconvinced, often citing examples of test bias and adverse impact (Haney, 2000; Valenzuela, 2000), and questioning testing design (Haney et al, 1999; Koretz & Barron, 1998; Dorn, 1998), classroom teaching and learning effectiveness (Amrein & Berliner, 2002; McNeil and Valenzuela, 2000; Haney, 2000; Dorn, 1998), and test integrity (Cizek, 2001; Dorn, 1998).

A test is considered biased if a student's scores depend on something other than what is being tested (something outside knowledge of a situation). Generally a test is thought to be biased when the members of a group (usually of a historically disadvantaged group) do not perform as well on it as others (Steele, 1997). However, even a test that is not biased may adversely impact a group. Adverse impact in education is the reflection of a negative affect on a particular group of persons as a result of a particular requirement (Haney, 2000; Steele, 1997).

> "There is a large and growing body of research on high stakes testing much of which illustrates its deleterious effects (Fair Test, 2003)."

One of those effects is adverse impact. Standardized tests are often seen as having adverse affects on groups such as females, minorities, and the poor (Haney, 2000). Walt Haney argues that Texas students are affected adversely through the state's testing system, particularly at the high school level. It is at that level that students are required to take the Texas Assessment of Academic Skills (TAAS) exam in order to complete graduation requirements. Minority students tend to score lower on these exams. In 1998

TAAS results reflected that over 30,000 minority students failed the tenth grade exit exam (Haney, 2000). Haney purports that there is a definite relationship between testing and adverse impact. If groups don't persist, they are going to be impacted negatively later in the job market as academic success is directly related to job performance.

Much of the literature on standardized testing in Texas focuses on the impact (positive or negative) that these assessments have on the youth in the state. The state demographics are unique in that there is a large minority population. Education strategies have had to be inclusive of this diversity but controversy has been imminent. In a three year qualitative case study, Angela Valenzuela criticizes high-stakes testing as an alienating factor in Texas education. According to the study, low achieving students of Mexican origin are crippled from receiving a high school diploma and from obtaining a college education because the TAAS, the state's high school exit exam, is not reflective of all students' true capabilities (Valenzuela, 2000). For example, limited English proficiency students who tend to have higher grades than their peers, often have trouble passing the English TAAS section (Valenzuela, 2000). The author asserts that education is not only about content mastery, but about the education of a person as a whole (Valenzuela, 2000).

In the early 1990s, a consulting team charged with evaluating the TASP focused on: The TASP exam itself, how the test is implemented, what effects result from the TASP, and whether there is a need for the TASP exam (National Center for Developmental Education, 1996). As discussed in the literature, these tests are only useful when they measure what they were developed to assess, (Gipps, 1994). Upon review of available data and investigation of various aspects of the TASP in Texas, the

findings were supportive of the TASP. Although some feedback lends credibility to critics, the study supports that the exam is a reliable instrument that meets expectations surrounding its design (National Center for Developmental Education, 1996).

Every year, universities across the state are seeing a significant number of students in remediation (THECB, 2001). Remediation is useful in assisting students to reach higher academic learning when it is needed. In Texas, students are required to take the TASP exam to see whether they are prepared for university coursework or if they need to participate in a remediation program. It is believed that the TASP and other such standardized exams are good gauges of student preparedness however discussions in the literature review add that standardized tests are limited in identifying student remediation needs (Patten, 2000). The report compiled by consultants in 1996 supports the need to gauge the academic level a student is at and states that the TASP is a good way to do so, however, it seems that the implementation is in need of being made more effective. The lack of clear evaluation of the remediation program was cited as a concern. Without an effective remediation program there may be students who are, as noted in some of the literature, adversely impacted despite the noble intent of the test (National Center for Developmental Education, 1996).

The cause of differences in scores among various groups is also the subject of debate. Some contributors, according to the literature, are: limits to educational access, socioeconomic disadvantages, segregating social practices, restrictive cultural orientations and lack of role models and proportional disadvantages faced by these groups (Steele, 1997). If students from a minority group believe they are not as good as

their non minority peers in certain subjects, they may not do as well. This is known as stereotype threat.

Stereotype threat has been looked at academically, particularly as drop out rates have increased, leaving minorities greatly impacted. What has been found is that minorities at some point stop being identified with the academic domain. Somewhere historically disadvantaged groups (generally ethnic minorities and females) start finding areas of identifying themselves. Every time they start to do well in those areas considered to be outside their realm of expertise, they feel threatened. (Steele, 1997) There is an attack on their self esteem. They start being encouraged not to be part of the academic domain. Stereotype threat is not the only issue involved but they are all a part of preserving their self-esteem and identifying with something that promotes our self esteem (Spencer, Steele, & Quinn, 1999; Steele, 1997).

While these conditions affect more than an individual's performance on standardized tests, they are still an important aspect of such education techniques and the only relevant aspect to this study. If members of minority groups are affected by stereotype threat, African Americans (Steele & Aronson, 1995), Latin Americans (Gonzales, Blanton, & Williams, 2002), members of low socioeconomic groups (Croziet & Claire, 1998), and women (Oswald & Harvey, 2000; Walsh, Hickey, & Duffy, 1999; Spencer et al, 1999), then standardized testing may not be related to first year college success. In the case of students sampled from The University of Texas - Pan American, it will be interesting to see differences among ethnic groups in terms of scores being that the majority of students attending are from a minority group.

Whatever the causes of these differences may be, what is certain is that they exist. In addition, standards are purported to assure all students, regardless of school, ethnicity, or socio economic status, equally challenging scholarship, prepared instructors, and equal learning expectations (Cizek & Seldon, 1992) but the number of minorities pursuing and completing four year programs is minimal (McNeill, 2000; Weiss, 1998). Methods geared toward achieving equity in education while improving educational quality are major topics in the literature.

Nationally studies show that factors like gender, ethnicity, and socio economic backgrounds affect academic success (Weiss, 1998). This is particularly true in math and science areas (Weiss, 1998). There are those who argue that minority students are thriving under the new system and that the increase in scores for minorities in the state of Texas may be a byproduct of the Texas system's aggressive approach to increasing standards for all students in its schools (TEA, 2001) yet there is also evidence supporting the argument that testing negatively affects certain groups of students, stifles professional creativity, and restricts higher learning skills for students (AEA, 2003). Supporters provide evidence of improved instruction and increased learning in schools yet critics raise concern that state education systems may be preparing our minority students for a test they may never pass (McNeil, 2000).

Prediction Research

Measuring the quality of learning a student has gained is invaluable to education stakeholders, particularly when debating mechanisms to enhance education at all levels. With every idea brought forth at the state level, the education system is placed under intense scrutiny and expected not only to account for the student learning that takes place

in the classroom, but also to shift gears quickly and efficiently to implement whatever changes are being integrated into their curriculum. Given the nature of the system, the integration of standardized tests into the Texas education system seems understandable. These exams are purported to measure student learning, hold schools accountable for that learning, and increase access and equity in education. Among the standardized tests the American College Testing Program Assessment (ACT) and Scholastic Aptitude Test (SAT) even claim that these tests are predictive of first-year academic performance at the university.

Questions about whether high school students are prepared for college are of concern for education stakeholders. Over a million students take the ACT each year. Students who take more coursework in high school score higher on related exams (Laing et al, 1987) because the tests are curriculum based. The more a student is exposed to coursework in a particular area, the more learning he should gain and thus score higher (Laing et al, 1987).

> "The utility of ACT test scores and high school coursework, as measured by grades or courses taken by students, for predicting college course grades rests on several assumptions: 1) ACT test scores and high school coursework and grades either directly measure or are closely related to the academic skills and knowledge required for success in particular courses. 2) College course grades are of sufficient reliability and validity so that they measure real and relevant educational outcomes, rather than random, irrelevant factors (Noble, 1991)."

Regression models are often used to analyze variables related to predicting various results from ACT data. In the Laing study, 3 variables are used as predictors. The first was how many courses a student took in a designated area prior to taking the ACT; the second was the student's rank in his graduating class; and the third was the student's

cumulative high school grade point average. The findings supported the hypothesis that the number of courses students take in high school do predict ACT scores. When accounting for income and racial/ethnic groups, the results were similar. The courses significantly contributed to increased ACT scores. The same is true for gender, although it was noted that more males took related coursework. American Indian/Alaska Native were the least likely to have taken the related coursework in high school, while Asian American/Pacific Islander were most likely to have done so. For the purposes of the current study, it can be assumed that the higher the ACT score, the more related coursework students have had in high school (Laing et al., 1987).

College Admissions decisions are improved by using the Scholastic Aptitude Test (SAT) to supplement the high school record (SAT). Aptitude and achievement tests are valid for predicting later educational and economic success (Gottfredson & Crouse, 1986). Studies have shown that a student's performance on standardized examinations do predict higher grade point averages for first year college students. In a study by Pui-Wa Lei et al (2001), researchers attempted to measure student academic achievement by placing a weight value on certain courses to account for grade inflation and fairly evaluate the predictive ability of the ACT. The results, that ACT scores are significant predictors of achievement, seem to substantiate their use. Perhaps, as related studies indicate (see Noble & Sawyer, 1987), this is due to the relationship between curriculum and standardized tests also found positive results when using the ACT as a predictor of student freshmen course grades (Noble & Sawyer, 1987). Noble and Sawyer found that specific college course grades can be predicted using ACT assessment test scores and high school courses completed in related subjects as predictors (Noble & Sawyer, 1987).

The ability to accurately predict student first year performance is pertinent to good admission policy (Noble, 1991). One of the most useful studies conducted on the topic of predicting college includes the use of ACT scores and self reported high school grades to see whether using the ACT and student high school grades accurately predict college grades (Noble, 1991). The author uses multiple regression techniques to test the predictive ability of the ACT. The results support the use of ACT scores in predicting student first year performance and support the hypotheses of the current study (Noble, 1991).

There are arguments countering the usefulness of standardized tests in college admissions. Crouse indicates that the SAT is redundant in college admissions (Gottfredson & Crouse, 1986). Atkinson prefers the use of achievement testing because it gauges what a student knows based off what the student has had the opportunity to know however his research does support the predictive validity of the SAT (Atkinson, 2001). Atkinson's study indicates the SAT accounted for 21.1 percent of the variance in combination with high school grades at the University of California. The writing portion was the best single predictor for freshmen grades (Atkinson, 2001).

The majority of students who attend The University of Texas Pan American are graduates from a Texas high school. In order for students to be eligible to attend The University of Texas Pan American they must meet the state's requirements for graduation or GED, take the TASP exam, and take the ACT or SAT. Given their residency status, it is reasonable to assume that this population has been impacted by the culture of the state's education system; therefore they will be used in the current study. The current study will employ the following prediction formulas:

	Independent Variable (Subtest Score)	Dependent Variable (Student Success)
1	TASP Reading	GPA
2		Hours Dropped
3	TASP Math	GPA
4		Hours Dropped
5	TASP Writing	GPA
6		Hours Dropped
7	ACT Verbal	GPA
8		Hours Dropped
9	ACT Math	GPA
10		Hours Dropped
11	ACT Composite	GPA
12		Hours Dropped

Chapter three will provide further explication of the model and tools utilized for the study.

Standardized exams such as the TAAS, TASP, ACT, and SAT help various education officials promote and measure learning. Their use in education is not going to go away. Political, social and financial investments into accountability systems that are highly dependent upon standardized exams seem to be increasing. The continuing importance of high performance on these exams provides an important reason to fully understand their utility.

CHAPTER III

METHODOLGY

This chapter describes the subject, data, and variables that will be used to conduct the research study. As the purpose of the study is to test analyze whether standardized tests predict first year college success the study will use regression analysis to test the independent and dependent variables. Several analyses will be run to gauge the predictive value of the three TASP tests and ACT verbal, math and composite scores. For the purposes of the study, first year G.P.A., and hours dropped, will be the operationalized variables. G.P.A. is chosen as an indicator of college success because it is a measurable performance for determining college success. Hours dropped was selected as a variable due to the impact that dropping courses can have on a student's satisfactory academic progress. Students who drop more hours in a term are likely to be affected in areas such as financial aid and progress to sophomore status within one year. The impact that this can have on a student, and the university they attend, is large. The study measures success by G.P.A. and dropped classes.

Aptitude and achievement tests are considered valid predictors of student success (Gottfredson & Crouse, 1986). As mentioned in previous chapters, prediction research is often conducted to provide this information to education stakeholders, particularly those interested in knowing a student's academic achievements and academic potential. Models

intended to measure student academic potential often use standardized test scores as values in their model. Other information commonly found in prediction models include: high school class performance, class rank, number of courses taken in specific academic areas, socio economic background, gender, ethnicity, first year college performance, and more (see Noble & Sawyer, 1987 Pui-Wa Lei et al; Noble, 1991). Multiple regression techniques are used to analyze these variables and standardized test scores. The results generally support the use of standardized test scores in predicting student first year performance and support the research questions listed for the current study (Noble, 1991).

Participants

The subjects analyzed will be 370 students that make up the 1998 College of Science and Engineering (COSE) entering freshmen at The University of Texas – Pan American (UTPA). The UTPA student population is predominantly Hispanic (86% in fall 1999) and the COSE cohort is reflective of this demographic. In addition, sixty-three percent of the student body is first generation college students, and sixty-five percent receive financial aid based on need. A significant fraction of the student body consists of non-traditional students, as indicated by the average undergraduate student age of 23 (Lujan, 1999). Students involved in the study are 1998 entering freshmen.

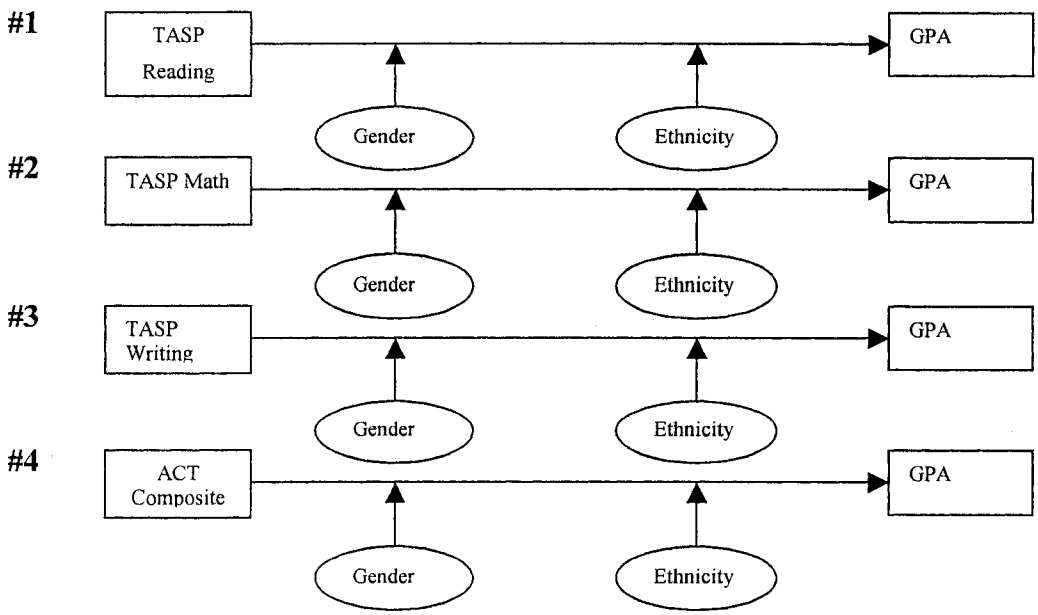
Data

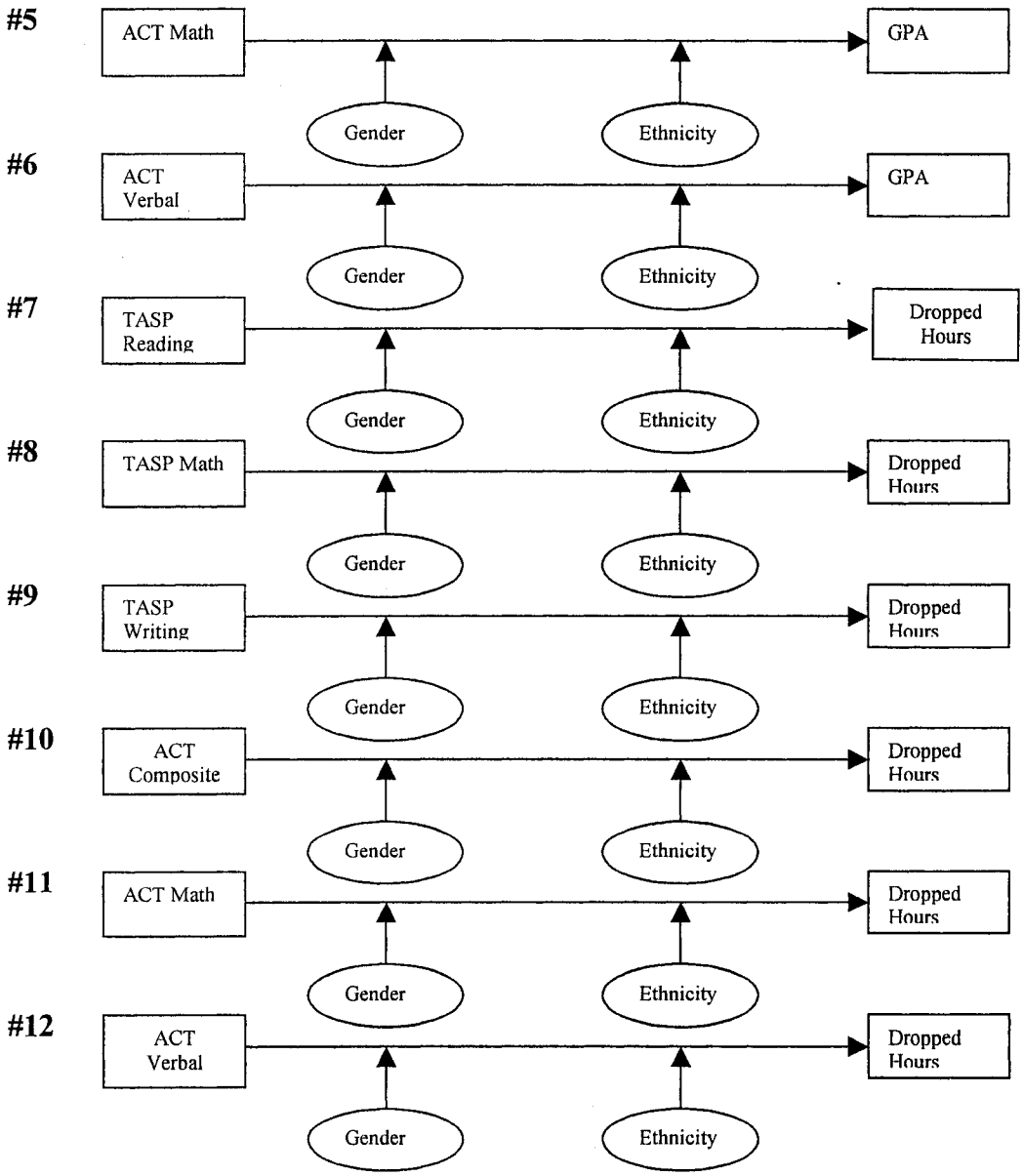
Participant data is drawn from an existing database of student information, collected from the Office of Admissions and Records at The University of Texas - Pan American and through student records on the Student Information System (SIS). Participant data is limited to students enrolling into the College of Science and Engineering for the Fall semester of 1998. Each student is assigned an arbitrary

numerical case number, eliminating any identifying information, such as social security numbers, or university identification numbers. The data collected is then examined on the SIS in order to remain consistent with the University of Texas - Pan American definition of an entering freshman. Any student who is not a 1998-1999 entering freshman will be excluded from the data set. Data will be eliminated from this data set for 11 students, as these are not correctly classified as entering freshmen. After these deletions, the final dataset consists of information from 370 out of 381 students originally identified as entering freshmen.

The Models

The following prediction models were used to determine the relationship between the scores a student received on various sections of the TASP and ACT. The models are based on hypotheses related to student academic performance during their first year of college and rests on the assumption that these tests are valid instruments that do predict student performance. Each model is weighted to account for gender and ethnicity.





Variables

There are two dependent variables in this study and six independent variables. The dependent variables are: first year college grade point average and number of class hours dropped. The independent variables are: TASP writing, TASP reading and TASP math subtests, as well as the ACT verbal, ACT math and ACT composite scores.

The first of two dependent variables obtained for this study is first year grade point average (GPA) and will serve as an indicant of student success. All values for this variable are positive numbers ranging from 0 to 4.0 and each student is considered more successful to the extent that their GPA score increases. On occasion, a student will repeat coursework in an effort to improve GPA. For these cases, GPA used in subsequent analyses is recomputed to account for this repetition. In no case did a student obtain a lower score for a course when taken a second time, so if recalculation of GPA results in any change, it will be an improvement.

Dropped hours serve as a second indicant of student success. Each student is considered less successful, to the extent that they drop an increasing number of credit hours. The final form of this variable is a sum of hours dropped in the fall, spring and summer semesters for each student's first academic year. All values for this variable are positive whole numbers with a minimum value of 0 representing a complete absence of dropped credit hours. This variable has no upper boundary as students can drop as many credit hours as they register for in a given semester, though this is generally constrained by financial considerations, as well as university restrictions on the maximum number of credit hours a student may register for.

A number of test scores obtained through student records represent the independent variables for this study; the TASP writing, reading and math subtests, as well as the ACT verbal, math and composite scores. With regard to the ACT, the most recently reported test score will be used in all analyses without regard to previous test scores. In relation to the TASP, as students can repeat sections independently of others, the highest score was taken from sections independently repeated. Unlike ACT results,

TASP results do not include a composite score as they are not provided by the test designers. Gender and ethnicity information will also be used as collected to account for any variance in student success attributable to these variables. It is also important to note that it was determined that the vast majority of students from the population being studied were confirmed graduates of Texas high schools. This information was also obtained by looking up each case individually in the SIS system mentioned previously.

Research Questions

Is there a relationship between standardized test scores and first year college success? If the literature is accurate in the support of standardized tests measuring the amount of preparation students have in a given subject (Corbett and Wilson, 1993), and that the more preparation a student has in an area (Laing, 1998), the higher his ACT will be, then these scores should translate into college success results.

In Texas, high school students with the intention of pursuing a university degree must show proficiency in reading, writing, and mathematics in order to pass the exit TAAS exam (or an approved substitute exam) at the high school level (TEA, 2002). Having passed the TAAS, a student graduating from a Texas school should ideally pass the TASP exam, the ACT, and experience a successful first year transition to the university. The research questions being tested stem from the assumption that standardized tests are good predictors of student first year success at a university. The best technique by which to accomplish this is using "an interval level statistical technique" known as multiple regression, "that uses several independent variables to predict or explain one dependent variable based on minimizing squared error (Meier & Brudney, 1997)." The research statements are:

1) Students with higher TASP scores are more likely to have a higher GPA their first year of college.
2) Students with higher ACT scores are more likely to have a higher GPA their first year of college.
3) Students with higher TASP scores drop fewer classes during long (fall and spring) semesters.
4) Students with higher ACT scores drop fewer classes during long (fall and spring) semesters.

Statistical Analysis

There is one regression equation for each model:

a. GPA = a + TASPr(b1) + Gend(b2) + Ethni(b3) +TASPr*Gend(b4)+ TASPr*Ethni(b5) +Gend*Ethni(b6) + TASPr*Gend*Ethni(b7)

b. GPA = a + TASPm(b1) + Gend(b2) + Ethni(b3) +TASPm*Gend(b4)+ TASPm*Ethni(b5) +Gend*Ethni(b6) + TASPm*Gend*Ethni(b7)

c. GPA = a + TASPw(b1) + Gend(b2) + Ethni(b3) +TASPw*Gend(b4)+ TASPw*Ethni(b5) +Gend*Ethni(b6) + TASPw*Gend*Ethni(b7)

d. GPA = a + ACTc(b1) + Gend(b2) + Ethni(b3) + ACTc*Gend(b4)+ ACTc*Ethni(b5) +Gend*Ethni(b6) + ACTc *Gend*Ethni(b7)

e. GPA = a + ACTm(b1) + Gend(b2) + Ethni(b3) + ACTm*Gend(b4)+ ACTm*Ethni(b5) +Gend*Ethni(b6) + ACTm *Gend*Ethni(b7)

f. GPA = a + ACTv(b1) + Gend(b2) + Ethni(b3) + ACTv*Gend(b4)+ ACTv*Ethni(b5) +Gend*Ethni(b6) + ACTv *Gend*Ethni(b7)

g. Dhrs = a + TASPr(b1) + Gend(b2) + Ethni(b3) +TASPr*Gend(b4)+

g. Dhrs = a + TASPr(b1) + Gend(b2) + Ethni(b3) +TASPr*Gend(b4)+ TASPr*Ethni(b5) +Gend*Ethni(b6) + TASPr*Gend*Ethni(b7)

h. Dhrs = a + TASPm(b1) + Gend(b2) + Ethni(b3) +TASPm*Gend(b4)+ TASPm*Ethni(b5) +Gend*Ethni(b6) + TASPm*Gend*Ethni(b7)

i. Dhrs = a + TASPw(b1) + Gend(b2) + Ethni(b3) +TASPw*Gend(b4)+ TASPw*Ethni(b5) +Gend*Ethni(b6) + TASPw*Gend*Ethni(b7)

j. Dhrs = a + ACTc(b1) + Gend(b2) + Ethni(b3) + ACTc*Gend(b4)+ ACTc*Ethni(b5) +Gend*Ethni(b6) + ACTc *Gend*Ethni(b7)

k. Dhrs = a + ACTm(b1) + Gend(b2) + Ethni(b3) + ACTm*Gend(b4)+ ACTm*Ethni(b5) +Gend*Ethni(b6) + ACTm *Gend*Ethni(b7)

l. Dhrs = a + ACTv(b1) + Gend(b2) + Ethni(b3) + ACTv*Gend(b4)+ ACTv*Ethni(b5) +Gend*Ethni(b6) + ACTv *Gend*Ethni(b7)

This is the conceptual framework being used to conduct a thorough, manageable analysis. Multiple regression analysis is used to determine whether the standardized test scores predict student success outcomes like G.P.A. To resolve a clearer picture of the predictiveness of the tests, a regression equation was developed for each subtest/outcome combination, resulting in a total of 12 equations. The ACT composite score was treated as a subtest during statistical analysis; despite that it is the arithmetic mean of its constituent subtests in an effort to remain consistent with its use in academic settings.

It has previously been noted that standardized test scores can vary reliably across gender and ethnic classifications therefore these variables were included in these analyses. Rather than treat these demographic characteristics as covariates, these analyses tested whether gender and ethnicity moderated the relationship between the

predictors and student outcomes. To test for moderation, interaction variables were computed as products of each of its constituent factors (Pedhazur, 1997), so each regression equation included a predictor by gender, predictor by ethnicity and predictor by gender by ethnicity interaction. All predictors scores were mean centered prior to computation of the interaction terms (Aiken & West, 1991).

Once all the interaction terms were computed, they were entered into a multiple regression equation in a hierarchical manner, so that both demographic variables were included in the first step, the predictor was entered in the second step, all two-way interactions were entered in the third step and the three-way interaction was entered in the last step.

CHAPTER 4

RESULTS

Overall descriptive statistics for all variables of interest in this study are summarized in Table 1. The minimum values of 0 for G.P.A. and dropped hours represent true zero values, not missing data; At least one student included in this data set received a 0.00 GPA, and at least one student in this sample refrained from dropping any classes in the semester considered for analysis. Sample size changes found across variables in the study are due to data missing from the original dataset, or otherwise not reported.

Table 1

Overall Descriptive Statistics for Independent and Dependent Variables

	N	Min	Max	Mean	SD
GPA	368	0.00	3.97	1.88	1.02
Dropped Hours	302	0	23	3.55	4.03
TASP Reading	284	136	298	237.48	30.10
TASP Math	286	121	297	236.02	31.32
TASP Writing	274	120	300	229.16	31.02
ACT Verbal	369	5	29	17.64	4.50
ACT Math	369	12	31	19.56	3.89
ACT Composite	369	8	29	18.94	3.79

Table 2 summarizes the descriptive characteristics of each cell in the subsequent analyses. Out of a total of 370 students in this data set, the vast majority are classified as Latino/Hispanic: Only 5.7% of the total sample is classified as Caucasian, African-American, or any other ethnic group. The division of this sample across gender lines is less disparate: 63% of the students in this sample reported their gender as male.

Table 2

Gender by Ethnicity: Cell Means for Dependent and Independent Variables

Variable	Gender	N	Mean	SD
GPA	Male			
	Latino	220	1.77	1.031
	Non-Latino	15	1.83	1.084

Table 2 (continued).

Variable	Gender	N	Mean	SD
	Female			
	Latino	127	2.06	.973
	Non-Latino	6	2.76	.768
Dropped Hours	Male			
	Latino	181	3.90	4.221
	Non-Latino	12	4.92	4.010
	Female			
	Latino	102	2.76	3.556
	Non-Latino	6	3.67	4.803
TASP reading	Male			
	Latino	171	235.30	30.951
	Non-Latino	7	250.57	29.421
TASP reading	Female			
	Latino	102	240.08	28.457
	Non-Latino	4	241.25	35.734
TASP Math	Male			
	Latino	173	239.29	31.216
	Non-Latino	7	241.43	36.244
	Female			
	Latino	102	230.08	30.392
	Non-Latino	4	236.50	41.122

Table 2 (continued).

Variable	Gender	N	Mean	SD
TASP Writing	Male			
	Latino	165	225.58	3.429
	Non-Latino	7	235.71	36.904
	Female			
	Latino	98	234.29	28.101
	Non-Latino	4	240.00	.000[a]
ACT Verbal	Male			
	Latino	220	17.16	4.510
	Non-Latino	15	21.73	4.832
	Female			
	Latino	127	17.80	4.117
	Non-Latino	6	21.50	5.431
ACT Math	Male			
	Latino	220	19.98	4.006
	Non-Latino	15	22.60	3.795
	Female			
	Latino	127	18.38	3.224
	Non-Latino	6	21.67	5.715
ACT Composite	Male			
	Latino	220	18.92	3.87
	Non-Latino	15	22.53	3.980

Table 2 (continued).

Variable	Gender	N	Mean	SD
	Female			
	Latino	127	18.41	3.313
	Non-Latino	6	22.17	3.462

[a] Lack of variance in this cell is due to low cell size: data was not available for 2 participants, while the other two reported an identical score.

Brief examination of predictors of student success by gender and ethnicity reveals that non-Latino males drop the highest number of credit hours and demonstrate the lowest levels of persistence, while Latino males obtain the lowest GPA of the four demographic groups. Conversely, non-Latino females obtain the highest GPA and demonstrate the highest levels of persistence, while Latino females drop the fewest credit hours.

Similar examination of the sub-test scores reveals that non-Latino males outperform the other 3 groups on the TASP reading and mathematics subtests and Non-Latino females obtain the highest TASP writing scores. Latino males obtain the lowest scores on both the TASP reading and writing subtests and Latino females score the lowest on TASP math. ACT scores yield a similar pattern of results, so that non-Latino males perform the best in both the verbal and mathematics, Latino males yield the lowest ACT verbal scores and Latino females obtain the lowest ACT math scores.

Predictors of GPA

Results of the multiple regression analysis testing the relationship between TASP reading and GPA are summarized in Table 3. These results suggest including the

interactions between predictors does not yield a significant increase in the proportion of GPA variance explained. Step one features a regression model including only the two demographic variables. In this model, only gender is statistically predictive of student GPA. Including the TASP reading sub-test results in a significant increase in proportion of variance explained, as illustrated by Step 2 in the hierarchical regression model. Inclusion of any interactions between first order variables does not yield an improvement in prediction. Though the relationship between the predictor and GPA is likely partially moderated by the demographic characteristics of the student, the test score yields the largest and most highly statistically significant standardized regression coefficient (β). As the inclusion of any interactions in the regression equation does not enhance the predictive ability of the equation, these will be eliminated from the predictive model, yielding an equation where gender, ethnicity and TASP reading account for 7.8% of the total observed variance in student GPA.

Table 3

Results of Hierarchical Regression Analysis for TASP Reading (TASP R) and GPA

Variable	B	SE B	β	ΔR^2
Step 1				.019
Gender	.281	.121	.137*	
Ethnicity	.129	.304	.025	
				.059***
Step 2				
Gender	.247	.118	.120*	
Ethnicity	.047	.296	.009	

Table 3 (continued).

Variable	B	SE B	β	Δ R^2
TASP R	.008	.002	.244***	
Step 3				.005
Gender	.229	.121	.11	
Ethnicity	-.214	.396	-.041	
TASP R	.009	.002	.270***	
TASP R * Ethnicity	.004	.010	.024	
TASP R * Gender	-.003	.004	-.049	
Gender * Ethnicity	.588	.624	.070	
Step 4				.011
Gender	.228	.121	.111	
Ethnicity	-.418	.409	-.081	
TASP R	.008	.002	.255***	
TASP R * Ethnicity	.020	.014	.122	
TASP R * Gender	-.001	.004	.104	
Gender * Ethnicity	.875	.641	.104	
TASP R * Gender * Ethnicity	-.004	.021	-.143	

*$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

Results of a hierarchical regression analysis using TASP math as a predictor are summarized in Table 4. This analysis suggests that a regression equation including higher-order interactions accounts for a significant amount of the variance in the dependent variable above that explained by a model excluding them. Step 4 in this model

accounts for 13.3% of the observed student GPA variance. The full model features three significant predictors: Gender, TASP math and the TASP math*gender*ethnicity interaction, though again the sub-test yields the largest and most statistically significant regression coefficient.

Table 4

Results of Hierarchical Regression Analysis for TASP Math (TASP M) and GPA

Variable	B	SE B	β	Δ R^2
Step 1				.022*
Gender	.297	.120	.14**	
Ethnicity	.118	.301	.02	
Step 2				.097***
Gender	.387	.115	.19***	
Ethnicity	.082	.287	.016	
TASP M	.010	.002	.315***	
Step 3				.002
Gender	.379	.118	.18**	
Ethnicity	-.043	.363	-.008	
TASP M	.009	.002	.291***	
TASP M * Ethnicity	.000	.008	-.006	
TASP M * Gender	.002	.004	.041	
Gender * Ethnicity	.347	.600	.041	
				.012*

Table 4 (continued).

Variable	B	SE B	β	Δ R^2
Step 4				
Gender	.383	.118	.18**	
Ethnicity	-.113	.363	-.022	
TASP M	-.009	.002	.272***	
TASP M * Ethnicity	.012	.011	.084	
TASP M * Gender	.004	.004	.072	
Gender * Ethnicity	.420	.598	.050	
TASP M * Gender * Ethnicity	.034	.017	-.144*	

*$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

Results of the multiple regression analysis testing the predictive validity of TASP writing are summarized in Table 5. Inclusion of a three way interaction was not possible in this analysis due to an unforeseen problem with available data: One cell of this analysis featured a complete lack of variance due to very small sample size, missing data as well as identical data points for what data was available. The most complete regression model available for this data includes all three two-way interactions, but inclusion of these interactions does not account for a significant increase in proportion of GPA variance explained. The most parsimonious model is step two of the hierarchical regression model. This model includes only gender, ethnicity and the TASP writing score, though only the latter yields a significant regression coefficient. All together, this model accounts for 3.5% of the observed variance in student GPA, suggesting that the remaining 96.5% of variance is explained by factors not included in this model.

Table 5

Results of Hierarchical Regression Analysis for TASP Writing (TASP W) and GPA

Variable	B	SE B	β	Δ R^2
Step 1				.016
Gender	.259	.124	.12*	
Ethnicity	.101	.304	.02	

Table 5 (continued).

Variable	B	SE B	β	Δ R^2
Step 2				.019*
Gender	.221	.124	.10	
Ethnicity	.062	.302	.012	
TASP W	.005	.002	.141*	
Step 3				.014
Gender	.211	.126	.103	
Ethnicity	-.234	.385	-.046	
TASP W	.004	.002	.135	
TASP W * Ethnicity	.020	.011	.114	
TASP W * Gender	-.001	.004	-.023	
Gender * Ethnicity	.386	.629	.047	

Note: small sample size precluded development of a three way interaction so only 3 hierarchical steps were included in this model.

*$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

Results of the multiple regression analysis testing the predictive validity of ACT math are summarized in Table 6. Much like results for the TASP subtests, inclusion of interactions into the regression model does not usually enhance the model's ability to account for variations in student GPA. Regression models including interactions between the predictor variable fail to yield significant increases in R^2, so these are excluded from the final model. Step two of this regression equation features gender, ethnicity and the sub-test score, though only gender and the test score are reliable predictors of GPA. This model accounts for 8.6% of the total observed variance in student GPA, suggesting that 91.4% of this variance is explained by factors not included in this model.

Table 6

Results of Hierarchical Regression Analysis for ACT Math (ACT M) and GPA

Variable	B	SE B	β	Δ R^2
Step 1				.025**
Gender	.316	.110	.14*	
Ethnicity	.243	.227	.05	
Step 2				.071***
Gender	.430	.108	.20***	
Ethnicity	.040	.108	.009	
ACT M	.072	.014	.276***	
Step 3				.005
Gender	.413	.113	.19	
Ethnicity	-.121	.306	-.028	
ACT M	.067	.016	.254***	

Table 6 (continued).

Variable	B	SE B	β	Δ R^2
ACT M * Ethnicity	.001	.053	.001	
ACT M * Gender	.020	.030	.040	
Gender * Ethnicity	.536	.497	.067	
Step 4				.000
Gender	.413	.114	.19	
Ethnicity	-.126	.335	-.029	
ACT M	.066	.016	.253***	
ACT M * Ethnicity	.003	.071	.003	
ACT M * Gender	.020	.032	.041	
Gender * Ethnicity	.546	.553	.06	
ACT M * Gender * Ethnicity	-.004	.108	-.003	

*$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

A similar pattern of results are observed in the analysis of ACT verbal and GPA; these results are summarized in Table 7. Specifically, inclusion of interactions does not yield a significant increase in proportion of variance in GPA explained. Step two in this model includes the two demographic variables and the sub-test score, though ethnicity is not a significant predictor of student GPA. This model accounts for 9.6% of the observed variance in GPA.

Table 7

Results of Hierarchical Regression Analysis for ACT Verbal (ACT V) and GPA

Variable	B	SE B	β	Δ R^2
Step 1				.025**
Gender	.316	.110	.14**	
Ethnicity	.243	.227	.05	
Step 2				.061***
Gender	.281	.107	.13**	
Ethnicity	-.006	.226	-.001	
ACT V	.058	.012	.255***	
Step 3				.008
Gender	.247	.109	.11	
Ethnicity	-.395	.322	-.090	
ACT V	.060	.015	.263***	
ACT V * Ethnicity	.044	.047	.063	
ACT V * Gender	-.014	.025	-.035	
Gender * Ethnicity	.759	.496	.094	
Step 4				.002
Gender	.247	.110	.11*	
Ethnicity	-.499	.343	-.114	
ACT V	.058	.015	.255***	
ACT V * Ethnicity	.071	.056	.103	
ACT V * Gender	.008	.026	-.021	

Table 7 (continued).

Variable	B	SE B	β	Δ R^2
Gender * Ethnicity	1.082	.619	.13	
ACT V * Gender * Ethnicity	-.088	.101	-.069	

*$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

Results of the multiple regression analysis testing the predictive validity of ACT composite are summarized in Table 8. As with the analyses involving ACT subtests, models including interactions between variables do not account for significant increases in GPA explained above that offered by a model excluding them. Step 2 is the most complete regression model yielding meaningful increases in prediction and it features gender, ethnicity and ACT composite, though again, ethnicity is not a significant predictor of the dependent variable. This model accounts for 9.9% of the observed variance in student GPA.

Table 8

Results of Hierarchical Regression Analysis for ACT composite (ACT C) and GPA

Variable	B	SE B	β	Δ R^2
Step 1				.025
Gender	.316	.110	.149*	
Ethnicity	.243	.227	.055	
Step 2				.074**
Gender	.354	.106	.167**	
Ethnicity	-.032	.225	-.007	
ACT C	.075	.014	.280**	

Table 8 (continued).

Variable	B	SE B	β	Δ R^2
Step 3				.007
Gender	.320	.109	.151*	
Ethnicity	-.416	.324	-.095	
ACT C	.070	.017	.262**	
Gender * Ethnicity	.650	.496	.081	
ACT C * Ethnicity	.061	.055	.075	
ACT C* Gender	.002	.030	.003	
Step 4				.435
Gender	.323	.109	.152*	
Ethnicity	-.521	.350	-.118	
ACT C	.068	.017	.255**	
Gender * Ethnicity	.927	.610	.115	
ACT C * Ethnicity	.092	.068	.113	
ACT C * Gender	.008	.031	.017	
ACT C *Gender * Ethnicity	-.090	.115	-.062	

Predictors of Hours Dropped

Results of the hierarchical regression calculation for the TASP reading subtest are presented in Table 9. No significant results were obtained for any model: Neither the demographic variables by themselves, or in combination with the test scores accounted for a significant proportion of the variance in hours dropped. Looking through the

resulting regression coefficients, it is clear that no single coefficient reaches traditional conventions of statistical significance.

Table 9

Results of Hierarchical Regression Analysis for TASP Reading (TASP R)) and Hours Dropped

Variable	B	SE B	β	Δ R^2
Step 1				.017
Gender	-.537	.543	-.066	
Ethnicity	2.177	1.266	.115	
Step 2				.000
Gender	-.532	.545	-.065	
Ethnicity	2.194	1.273	.116	
TASP R	-.002	.009	-.012	
Step 3				.001
Gender	-.539	.566	-.066	
Ethnicity	1.706	1.869	.090	
TASP R	-.002	.011	-.012	
Gender * Ethnicity	.633	2.704	.021	
TASP R * Ethnicity	.018	.047	.031	
TASP R * Gender	-.002	.019	-.009	
Step 4				.001
Gender	-.533	.567	-.066	
Ethnicity	2.073	2.060	.109	

Table 9 (continued).

Variable	B	SE B	β	Δ R^2
TASP R	-.001	.011	-.008	
Gender * Ethnicity	.190	2.901	.006	
TASP R * Ethnicity	-.002	.066	-.003	
TASP R * Gender	-.004	.020	-.017	
TASP R *Gender * Ethnicity	.040	.093	.042	

Results of the hierarchical regression calculation for the TASP math subtest are presented in table 10. Here again, no significant results were obtained for any model. Looking through the resulting regression coefficients, it is clear that for this test also, no single coefficient reaches traditional conventions of statistical significance.

Table 10

Results of Hierarchical Regression Analysis for TASP Math (TASP M) and Hours Dropped

Variable	B	SE B	β	Δ R^2
Step 1				.017
Gender	-.524	.536	-.065	
Ethnicity	2.174	1.257	.115	
Step 2				.002
Gender	-.586	.544	-.072	
Ethnicity	2.216	1.260	.117	
TASP M	-.006	.009	.046	

Table 10 (continued).

Variable	B	SE B	β	Δ R^2
Step 3				.010
Gender	-.593	.558	-.073	
Ethnicity	1.623	1.771	.086	
TASP M	.003	.011	.021	
Gender * Ethnicity	.846	2.668	.029	
TASP M * Ethnicity	.024	.044	.040	
TASP M * Gender	-.026	.018	-.120	
Step 4				.000
Gender	-.593	.559	-.073	
Ethnicity	1.642	1.971	.087	
TASP M	-.003	.012	.021	
Gender * Ethnicity	.827	2.811	.028	
TASP M * Ethnicity	.022	.070	.038	
TASP M * Gender	-.026	.019	-.120	
TASP M *Gender * Ethnicity	.002	.090	.002	

Results of the hierarchical regression calculation for the TASP writing subtest are presented in Table 11. As a result of the lack of variance discussed previously, no three way interaction test was possible. No significant results were obtained for any model or coefficient.

Table 11

Results of Hierarchical Regression Analysis for TASP Writing (TASP W) and Hours Dropped

Variable	B	SE B	β	Δ R^2
Step 1				.018
Gender	-.555	.550	-.068	
Ethnicity	2.162	1.267	.115	
Step 2				.001
Gender	-.519	.557	.064	
Ethnicity	2.214	1.275	.118	
TASP W	-.004	.009	-.032	
Step 3				.012
Gender	-.495	.577	-.061	
Ethnicity	3.316	1.880	.177	
TASP W	.003	.011	.021	
Gender * Ethnicity	.072	2.617	.002	
TASP W * Ethnicity	-.086	.059	-.199	
TASP W * Gender	-.015	.019	-.065	

Results of the hierarchical regression calculation for the ACT verbal subtest are presented in Table 12. For this regression model, only the step including the demographic variables by themselves accounts for a significant proportion of the variance in hours dropped. A closer look at the individual coefficients suggests that

student gender is the most important factor in predicting the number of hours dropped in the first year. No other variables were significant.

Table 12

Results of Hierarchical Regression Analysis for ACT Verbal (ACT V) and Hours Dropped

Variable	B	SE B	β	Δ R^2
Step 1				.022*
Gender	-1.157	.481	-.138*	
Ethnicity	.977	.973	.058	
Step 2				.002
Gender	-1.141	.482	-.136	
Ethnicity	1.146	1.000	.067	
ACT V	-4.055E-02	.054	-.044	
Step 3				.007
Gender	-1.100	.499	-.131	
Ethnicity	1.831	1.605	-.108	
ACT V	1.553E-02	.069	.017	
Gender * Ethnicity	.215	2.120	.007	
ACT V * Ethnicity	-.197	.237	-.070	
ACT V * Gender	-.126	.114	-.081	
Step 4				.494
Gender	-1.094	.499	-.130	
Ethnicity	2.500	1.880	.147	
ACT V	2.2554E-02	.070	.024	

Table 12 (continued).

Variable	B	SE B	β	Δ R^2
Gender * Ethnicity	-1.068	2.831	-.037	
ACT V * Ethnicity	-.351	.327	-.125	
ACT V * Gender	-.146	.118	-.094	
ACT V * Gender * Ethnicity	.326	.475	.071	

* $p < .05$

Results of the hierarchical regression calculation for the ACT math subtest are presented in Table 13. For this regression model, only the step including the demographic variables by themselves accounts for a significant proportion of the variance in hours dropped. A closer look at the individual coefficients suggests that student gender is the most important factor in predicting the number of hours dropped in the first year. No other variables were significant.

Table 13

Results of Hierarchical Regression Analysis for ACT Math (ACT M) and Hours Dropped

Variable	B	SE B	β	Δ R^2
Step 1				.022*
Gender	-1.157	.481	-.138*	
Ethnicity	.977	.973	.058	
Step 2				.000
Gender	-1.117	.496	-.133*	
Ethnicity	.928	.986	.055	
ACT M	.021	.062	.020	

Table 13 (continued).

Variable	B	SE B	β	Δ R^2
Step 3				.024
Gender	-1.143	.513	-.136*	
Ethnicity	2.432	1.466	.143	
ACT M	.108	.075	.104	
Gender * Ethnicity	-.046	2.113	-.002	
ACT M * Ethnicity	-.571	.299	-.145	
ACT M * Gender	-.193	.139	-.103	
Step 4				.008
Gender	-1.163	.512	-.138*	
Ethnicity	5.101	2.259	.300*	
ACT M	.120	.075	.115	
Gender * Ethnicity	-3.103	2.886	-.108	
ACT M * Ethnicity	-1.511	.676	-.384*	
ACT M * Gender	-.234	.141	-.125	
ACT M *Gender * Ethnicity	1.168	.754	.230	

* $p < .05$

Results of the hierarchical regression calculation for the ACT composite are presented in Table 14. Only the first step, containing the demographic variables in isolation, accounted for a significant proportion of the variance in hours dropped. Looking through the resulting regression coefficients, it appears that this is due in large part to the influence of the gender variable.

Table 14

Results of Hierarchical Regression Analysis for ACT Composite (ACT C) and Hours Dropped

Variable	B	SE B	β	Δ R^2
Step 1				.022*
Gender	-1.157	.481	-.138*	
Ethnicity	.977	.973	.058	
Step 2				.000
Gender	-1.171	.484	-.139*	
Ethnicity	1.053	1.001	.062	
ACT C	-.022	.065	-.020	
Step 3				.010
Gender	-1.164	.499	-.139*	
Ethnicity	2.417	1.663	.142	
ACT C	.042	.079	.039	
Gender * Ethnicity	.101	2.138	.003	
ACT C * Ethnicity	-.399	.296	-.118	
ACT C * Gender	-.132	.139	-.069	
Step 4				.001
Gender	-1.168	.499	-.139*	
Ethnicity	3.142	2.139	.185	
ACT C	.048	.080	.044	
Gender * Ethnicity	-1.012	2.972	-.035	

Table 14 (continued).

Variable	B	SE B	β	ΔR^2
ACT C * Ethnicity	-.591	.462	-.174	
ACT C * Gender	-.150	.143	-1.049	
ACT C *Gender * Ethnicity	.325	.603	.062	

* $p < .05$

As noted throughout the chapter, the analyses of each model indicate that the use of the ACT and TASP exam as predictors of student success are primarily valid for student first year G.P.A. The models for credit hours dropped generally indicated that no subtest is an adequate predictor for that outcome variable. As for student persistence, there was some validity in using the student persistence regressions, however, it was when including the gender and ethnicity interactions.

CHAPTER 5

CONCLUSIONS

Texas has included the adoption of standardized tests to measure student learning, hold schools accountable for that learning, increase access and equity in education, predict student preparedness for the university, and develop systems to remediate students who do not meet set standards. This is part of its commitment to improving learning for all students. There is a great deal of literature that suggests that students' performance on standardized exams is reflective of the quality of education they have received, thus being a good benchmark of their knowledge and potential performance in an academic setting (TEA, 2000). In addition, the literature suggests that although gaps can be found among test performance levels based on gender, ethnicity, and socioeconomic status, raising standards and setting higher expectations of all students will help to narrow this gap (Ravitch, 1995). The problem is that to argue that standardized testing is pertinent to education improvement, one needs substantial evidence, including information about exactly what high performance on these exams means. Lack of this evidence, as well as evidence not indicating positive relations between student exam performance and student success, limits the extent to which standardized exams can be used and leaves room for controversy over current uses.

With that, the findings discovered in this support the use of standardized exams as an indicator of student first-year G.P.A. However, the research conducted both by this author and many others researching testing reinforces the care that must be taken to assure proper use of standardized tests. In the current study, the TASP Reading GPA model proved to be a significant predictor of student G.P.A. performance. After accounting for gender and ethnicity, the model indicated that TASP Reading accounts for 7.8 % of the G.P.A. variance. The TASP Math accounted for 13.3 percent of the variance while TASP writing accounted for 5%. The indication is that the Texas TASP can prove useful in predicting student first year grades. However, it is clear through the percent of the variance not accounted for that a case by case analysis would suggest that the exam is not always predictive of student performance. Some students will perform well on the exams but do poorly at the university; others will score poorly and do well.

Implications for Testing Policy Advances

Policy makers must be careful not to develop policies that may underscore the value of standardized assessment tools. The education of students should not be limited by policy. One area of concern is the propensity of instructors to focus their instruction on objectives that standardized exams measure. Policy that is dependent on test performance may lead to training of students instead of educating them. With the need for critical thinking skills, problem solving, and initiative, this result could be detrimental to the progression of our youth from education arenas to work places.

Standardized testing for accountability purposes has become socially acceptable. The tests are used for a variety of reasons however policy makers should be made

aware of limitations inherent in over reliance on tests to fix education problems. The current study indicates a large percentage of the variance is not accounted for, even though the study weighed in gender and ethnicity during the testing of each model. In this situation, there needs to be greater attention paid to discovering what does account for the rest of the variance. Perhaps there are social issues, economic issues, or other issues that policy makers should consider that will have a greater impact on student learning and success. This understanding should provide great information for future policy matters.

Another area of concern is the transition of the conceptualization of the policy to the area of implementation. Policy makers must be aware of potential flaws in the execution of testing policy and allow flexibility for safety nets. Reporting for the purpose of oversight is important whether an administrator is concerned with the implementation of an exit exam, college entrance exam, or other test mandate because a flawed system can negatively impact student persistence. Provisions within a policy should include that findings of poor implementation will result in suspending negative consequences for exam takers affected by the poor implementation. The intent of testing programs is to gauge student knowledge and increase their chances at being successful therefore it is important to avoid negatively impacting students that policy is trying to help.

The current study is on The University of Texas – Pan American College of Science and Engineering entering freshman student population in Fall 1998. Socioeconomic status was not accounted for and there was very little ethnic diversity within the sample. Further, the cross sectional design of this study of first-year

students in a particular college of study, does not account for all students, graduation or job placement.

WORKS CITED

ACT Assessment. (2001, August) Retrieved August, 22, 2001, from http://www.act.org/aboutact/history.html

ACT Assessment. (2001, August) Retrieved August, 22, 2001, from http://www.act.org/aboutact/history.html

Aiken, Leona S. & West, Stephen G. (1991). Multiple Regression: Testing and Interpreting Interactions. Thousand Oaks, California. Sage Publications.

American Evaluation Association (AEA). (2003). American Evaluation Association Position Statement on High Stakes Testing In PreK-12 Education. (Retrieved June 11, 2003 from: http://www.eval.org/hst3.htm).

American Psychological Association. (2001). Publication manual of the American Psychological Association (5th ed.). Washington, DC: Author.

Amrein, Audrey L. & Berliner, David C. (2002). High Stakes Testing, uncertainty and student learning. Education Policy Analysis Archives, 10 (18). (Retrieved from: http://www.greatlakescenter.org/pub/H-S%20Analysis%20final.pdf on June 7, 2003.

Atkinson, Richard C. (2001). Achievement versus Aptitude in College Admissions. *Issues in Science and Technology, 18 n2, 31-36.*

Bauer, Scott C. (2000). Should Achievement Tests be Used to Judge School Quality? Education Policy Analysis Archives, 8 (46). (Retrieved from: http://epaa.asu.edu/epaa/v8n46.html on June 11, 2003).

Butterfield, Sue. (1995). Educational Objectives & National Assessment. Philadelphia, Open University Press.

Cizek, G. J. & Seldon, R. (1992). Standardized tests should not be eliminated from schools. In D. L. Bender & L. Bruno (Ed.), *Education in America: Opposing Viewpoints* (pp. 50- 56). San Diego, California: Greenhaven Press.

Corbett, D. H. & Wilson, B. L. (1991). *Testing, reform, and rebellion.* Norwood, New Jersey: Ablex Publishing.

College Board. (2001). How the SAT is made. *New York: The College Board.* (Retrieved from www.collegeboard.org/html/pdf/howsatmade.pdf on June 11, 2003).

Croziet, J. C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin, 24*, 588-594.

Dietel, R.J.; Herman, J.L. & Knuth R.A. (1991) What Does Research Say About Assessment? Oak Brook, IL: North Central Regional Educational Library (NCREL) Available at: http://www.ncrel.org/sdrs/areas/stw_esys/4assess.htm

Dorn, Sherman (1998). The Political Legacy of School Accountability Systems. *Education Policy Archives, 6, 1,* 1-34.

Edgewood v Kirby, 761 S.W.2d 859, Lexis 3292. (Tex. App. December 14, 1988).

Fair Test Organization. The National Center for Fair & Open Testing. April 12, 2002 (2003) Retrieved October, 16, 2003 from http://www.fairtest.org/facts/Dangerous%20Consequences.html

Finn, C. & Rebarber, T. (1992). *Education reform in the 90's*. New York: Macmillan Publishing.

Flores, B. B., & Clark, E. R. (2003, March 3). Texas voices speak out about high-stakes testing: Preservice teachers, teachers, and students. *Current Issues in Education, 6(3).* Retrieved from: http://cie.ed.asu.edu/volume6/number3/ on June 7, 2003.

Fuller, E. J. & Johnson, J. F. (2001). Can state accountability system drive improvements in school performance for children of color and children from low-income homes? Education and Urban Society, 33(3), 260-283.

Gipps, C. V. (1994). *Beyond testing; Towards a theory of educational assessment.* Washington DC: Falmer Press.

Gonzales, Patricia M., Blanton, Hart, & Williams, Kevin J. (2002) The Effects of Stereotype Threat and Double-Minority Status on the Test Performance of Latino Women. *Personality and Social Psychology Bulletin* 28(5) 659-670.

Gottfredson, Linda S., & Crouse, James. (1986) Validity versus Utility of Mental Tests: Example of the SAT. *Journal of Vocational Behavior 29, 363-378.*

Greene, Jay P, Winters Marcus, & Forster Greg. (February, 2003) Testing High Stakes Tests: Can We Believe the Results of Accountability Tests?. *Manhattan Institute for Policy Research (33) 1-19.* 52 Vanderbilt Ave. New York, NY 10017. (Also available at: http://www.manhattan-institute.org/html/cr_33.htm).

Haladyna, Thomas, Haas, Nancy, & Allison, Jeanette. (1998). Continuing tensions in standardized testing. Childhood Education, 75(5), 262 – 273.

Haney, W. (2000). The Texas miracle in education. *Education Policy Analysis Archives, 8, no. 41*. Retrieved February 8, 2002, from http://olam.ed.asu.edu/epaa/v8n41/part2.html

Hughes, J. F. (Ed). (1975). Education and the State. Washington DC. American Council on Education.

Johnson, J. & Tresisman, U. & Fuller, E. (2000) Testing in Texas. *School Administrator, 57*. 20-26.

Just for Kids (JFK). (2001) (Retrieved from: http://www.just4kids.org/ on December 5, 2001). Author.

Klein, Stephen P, Hamilton, Laura S, McCaffrey, Daniel F., & Stecher, Brian M. (2000) "What do test scores in Texas tell us?" *Education Policy Analysis Archives, 8(49.)* (Retrieved from: http://epaa.asu.edu/epaa/v8n49/ on March, 2, 2001).

Koretz, D., & Barron, S. I. (1998). The validity of gains on the Kentucky Instructional Results Information System (KIRIS). Santa Monica, CA: RAND, MR-1014-EDU. (Retrieved from: http://www.rand.org/publications/MR/MR1014/ on June 11, 2003).

Kramer, Liz. (2002) Achieving equitable education through the courts: a comparative analysis of three states. *Journal of Law and Education 31(1)1-51.*

Laing, Joan, Engen Harold B., & Maxey, James. (January, 1987) Relationships between ACT test scores and high school courses. *ACT Research Report Series 87(3.)*

Linn, Robert, L. (March, 2000). "Assessments and Accountability," *Educational Researcher* 29(2) 4-17 (Retrieved from: http://www.aera.net/pubs/er/arts/29-02/linn01.htm on June 11, 2003).

Lopez, O. (2000). The relationship of the Texas high school curriculum to college readiness. Retrieved January 22, 2001 from http://www.just4kids.org/us/us_home.asp

Lujan, S. S., Vasques, H. & Rios, X. (1999). Study of first time students to UTPA; Report of 1998 entering students. Edinburg: *The University of Texas Pan American, Office of Institutional Effectiveness.*

Lujan, S. S., Vasques, H. & Rios, X. (2000). Study of first time students; Report of 1999 entering students. Edinburg: *The University of Texas Pan American, Office of Institutional Effectiveness.*

Madaus, George F. & O'Dwyer, Laura M. (1999). A short history of performance assessment: Lessons learned. *Phi Delta Kappan 73: 226-31.*

McKenna, Bernard. (March, 1977). "Guidelines as Cautions for Considering Criterion Referenced Testing". *Standardized Testing Issues: Teachers' Perspectives. Reference and Resource Series. (35) Corp Author(s): National Education Association, Washington, DC.*

McNeil, Linda. (November, 2000). Sameness, bureaucracy, and the myth of educational equity: The TAAS system of testing in Texas Public Schools. *Hispanic Journal of Behavioral Sciences 22(4). 508-523.*

McNeil, Linda & Valenzuela, Angela. (2000)"Harmful Effects of the TAAS System of Testing in Texas: Beneath the Accountability Rhetoric," (As found in: Kornhaber, M.; Orfield, Gary.; & Kurlaendar, Michal [eds.] Raising Standards or Raising Barriers? Inequality and High-Stakes Testing in Public Education) *New York: Century Foundation.* (Also available at: http://www.ruf.rice.edu/~ctreduc/TAASArticle.pdf)

Meir, Kenneth J. & Brundy, Jeffrey L. (1997). Applied Statistics for Public Administration (4th ed.). Orlando Florida. Harcourt Brace College Publishers.

National Center for Developmental Education; Director Boylan, Hunter R. (1996, September). An Evaluation of the Texas Academic Skills Program (TASP). (Available from The Texas Higher Education Coordinating Board at: www.thecb.state.tx.us/reports/HTML/0282/boynof.htm).

National Commission on Excellence in Education. (1983). A Nation at Risk: The Imperative for Educational Reform. Washington, DC: Author. (Available at: http://www.ed.gov/pubs/NatAtRisk/risk.html)

Neill, M. D. & Medina, N. J. (1992). Eliminating standardized tests would improve schools. In D. L. Bender and L. Bruno (Ed.), *Education in America: Opposing Viewpoints* (pp. 42- 49). San Diego, California: Greenhaven Press.

Noble, J.P. (1991) Predicting college grades from ACT assessment scores and high school course work and grade information. *ACT Research Report Series. 91-3.*

Noble, J. P. &, Sawyer, R. (1987). Predicting grades in specific college freshman courses from ACT test scores and self reported high school grades. *ACT Research Report Series. 87-20.*

Oswald, D. L. and Harvey, R. (2000/2001). Hostile environments, stereotype threat and math performance among undergraduate women. *Current Psychology: Developmental – Learning – Personality - Social*, 19, 338-356.

Patten, Peggy. (2000). Standardized Testing in Schools. *Parent News Archives*. (Retrieved on November 2003 from: http:npin.org/pnews/2000/pnew100/feat100.html

Pedhazur, Elazar. (1997) Multiple Regression In Behavioral Research: Explanation and Prediction (3rd Ed.). Orlando, Florida. Harcourt College Publishers.

Perrone, Vito. (1991) On Standardized Testing. ERIC Digest. Retrieved on September 2002 from: http:www.ericeece.org/pubs/digests/1991/perron91.html

Phelps, Richard. (2000). Estimating the Cost of Standardized Student Testing in the United States. *Journal of Education Finance 25(3)Winter.*

Pitsch, Mark. (1996) Polls Confirm Key Role of Education in Political Arena. *Education Week, 15,39.* (Retrieved from: http://www.edweek.com/ew/vol-15/39polls.h15 on January 8, 2004).

Pui Wa-Lei, B. D. & Schultz, M.E. (2001). Alternatives to the grade point average as a measure of academic achievement in college. *ACT Research Report Series, 2001-4.*

Ravitch, D. (1995). *National standards in American education.* Washington D.C.: The Brookings Institution.

Richardson, R. C. Jr., Bracco, K. R., Callan, P. M. & Finney, J. E. (1999). *Designing state higher education systems for a new century.* Phoenix, Arizona: Oryx Press.

Scheurich, James Joseph; Skrla, Linda Johnson, Joseph F. (2000) Thinking carefully about equity and accountability. Phi Delta Kappan 82, 4, 293-299.

Sergiovanni, T. J. & Moore, J. E. (1990). *Target 2000: A compact for excellence in Texas schools.* San Antonio, TX: The Watercress Press.

Sheldon, K. M. & Biddle, B. J. (1998). Standards, accountability, and school reform: Perils and pitfalls. Teachers College Record, 100(1), 164-180.

Spencer, S. J., Steele. C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4-28.

Steele, C. M., & Aronson. J. (1995). Stereotype threat and the intellectual performance of African Americans. *Journal of Personality and Social Psychology, 69,* 797-811.

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

Texas Education Agency. (1995). *Academics 2000, Education Improvement Plan. Texas Education Agency Publications Distribution. Available at: 1701 North Congress Avenue. Austin Texas, 78701. (*Retrieved March 10, 2001 from http://www.tea.state.tx.us/) Author.

Texas Education Agency. (1999). 1996-97 Report on High School Completion Rates. *Office of Policy Planning and Research Division of Research and Evaluations. Texas Education Agency Publications Distribution. Available at: 1701 North Congress Avenue. Austin Texas, 78701.* (Retrieved from: www.tea.state.tx.us/research/abstracts1.html on August 2, 2001). *Author.*

Texas Education Agency. (2000). *Expanding the Scope of the Texas Public School Accountability System.* (Retrieved from: http://www.tea.state.tx.us/research/pdfs/prr9.pdf#xml=http://www.tea.state.tx.uswww.tea.state.tx.us/cgi/texis/webinator/search/xml.txt?query=+Expanding+the+Scope+of+the+Texas+Public+School+Accountability+System&db=db&id=1008a2a26c241edf on March 22, 2001) Author.

Texas Education Agency. (2000). *Interim report on Texas public schools. Texas Education Agency Publications Distribution. Available at: 1701 North Congress Avenue. Austin Texas, 78701. Author.*

Texas Education Agency. (April, 2001). 2001 Accountability Manual. *Texas Education Agency Publications Distribution. Available at: 1701 North Congress Avenue. Austin Texas, 78701.* (Also available at: http://www.tea.state.tx.us/perfreport/). Author.

Texas Higher Education Coordinating Board (2000-2001) 1996 TASP Annual Report. THECB. Available at: Texas Higher Education Coordinating Board. PO Box 12788 Austin, TX. 78711. (Retrieved from http://www.thecb.state.tx.us/reports/html/0208/ar96.htm on February 19, 2002) Author.

Tinto, V. (1993) *Leaving college: Rethinking the causes and cures of student attrition.* (2nd ed.). Chicago: University of Chicago Press.

Trueba, Henry. (2001). Conclusion: Polar Positions on the Texas Assessment of Academic Skills (TAAS): Pragmatism and the Politics of Neglect. *Education and Urban Society 33(3) 333-344*

Toch, T. (23, March 1996). Reformer Ross. US Online News. Retrieved December 20, 2001, from http://www.usnews.com/usnews/news/ross.htm

Valencia, R. R. (Ed.). (1991). *Chicano School failure and success.* New York: The Falmer Press.

Valenzuela, Angela. (1999). *Subtractive schooling: U.S. Mexican youth and the politics of caring.* Albany, New York: State University of New York Press.

Valenzuela, Angela. (2000). The significance of the TAAS Test for Mexican Immigrant and Mexican American adolescents: A case study. *Hispanic Journal of Behavioral Sciences22(4) 524-539.*

Walsh, M., Hickey, C., & Duffy, J. (1999). Influence of item content and stereotype situation on gender differences in mathematical problem solving. *Sex Roles, 41,* 219-240.

Weis, Lois. (1998) (ed). Class, Race, and Gender in American Education. SUNY Press, Albany, New York.

Zirkel, Perry A., & Gluckman, Ivan B. (January, 1993). "No pass, no play" rules. *Principal 72. 62-63.*

APPENDIX A

BIBLIOGRAPHY

BIBLIOGRAPHY

ACT Assessment. (2001, August) Retrieved August, 22, 2001, from http://www.act.org/aboutact/history.html

ACT Assessment. (2001, August) Retrieved August, 22, 2001, from http://www.act.org/aboutact/history.html

American Evaluation Association (AEA). (2003). American Evaluation Association Position Statement on High Stakes Testing In PreK-12 Education. (Retrieved June 11, 2003 from: http://www.eval.org/hst3.htm).

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Amrein, Audrey L. & Berliner, David C. (2002). High Stakes Testing, uncertainty and student learning. Education Policy Analysis Archives, 10 (18). (Retrieved from: http://www.greatlakescenter.org/pub/H-S%20Analysis%20final.pdf on June 7, 2003.

Anderson, C. J. (1998) *Fact book on higher education: 1997 edition.* Phoenix, Arizona: ORYX Press.

Aronson, J.; Lustina, M. J.; Good, C.; Keough, K.; Steele, C. M.; & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*, 29-46.

Atkinson, Richard C. (2001). Achievement versus Aptitude in College Admissions. Issues in Science and Technology, 18 n2, 31-36.

Bauer, Scott C. (2000). Should Achievement Tests be Used to Judge School Quality? Education Policy Analysis Archives, 8 (46). (Retrieved from: http://epaa.asu.edu/epaa/v8n46.html on June 11, 2003).

Blascovich, J., Spencer, S. J., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science, 12,* 225-229

Butterfield, Sue. (1995). Educational Objectives & National Assessment. Philadelphia, Open University Press.

Cizek, G. J. & Seldon, R. (1992). Standardized tests should not be eliminated from schools. In D. L. Bender & L. Bruno (Ed.), *Education in America: Opposing Viewpoints* (pp. 50- 56). San Diego, California: Greenhaven Press.

Corbett, D. H. & Wilson, B. L. (1991). *Testing, reform, and rebellion.* Norwood, New Jersey: Ablex Publishing.

College Board. (2001). How the SAT is made. *New York: The College Board.* (Retrieved from www.collegeboard.org/html/pdf/howsatmade.pdf on June 11, 2003).

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self protective properties of stigma. *Psychological Review, 96,* 608-630

Croziet, J. C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin, 24,* 588-594.

Danso, H. A. & Esses, V. M. (2001). Black experimenters and the intellectual performance of white participants: The tables are turned. *Journal of Experimental Social Psychology, 37,* 158-165.

Dietel, R.J.; Herman, J.L. & Knuth R.A. (1991) What Does Research Say About Assessment? Oak Brook, IL: North Central Regional Educational Library (NCREL) Available at: http://www.ncrel.org/sdrs/areas/stw_esys/4assess.htm

Dorn, Sherman (1998). The Political Legacy of School Accountability Systems. *Education Policy Archives, 6, 1,* 1-34.

Edwards, G. C. (1999). Who influences whom? The president, congress, and the Media: *American Political Science Review, 93*, 327-343.

Eckstein, M. & Noah, H. (1993). *Secondary school examinations*. New Haven and London. Yale University Press.

Edgewood v Kirby, 761 S.W.2d 859, Lexis 3292. (Tex. App. December 14, 1988).

Fair Test Organization. The National Center for Fair & Open Testing. April 12, 2002 (2003) Retrieved October, 16, 2003 from http://www.fairtest.org/facts/Dangerous%20Consequences.html

Finn, C. & Rebarber, T. (1992). *Education reform in the 90's*. New York: Macmillan Publishing.

Flores, B. B., & Clark, E. R. (2003, March 3). Texas voices speak out about high-stakes testing: Preservice teachers, teachers, and students. *Current Issues in Education, 6(3).* Retrieved from: http://cie.ed.asu.edu/volume6/number3/ on June 7, 2003.

Fuller, E. J. & Johnson, J. F. (2001). Can state accountability system drive improvements in school performance for children of color and children from low-income homes? Education and Urban Society, 33(3), 260-283.

Gipps, C. V. (1994). *Beyond testing; Towards a theory of educational assessment.* Washington DC: Falmer Press.

Gonzales, Patricia M., Blanton, Hart, & Williams, Kevin J. (2002) The Effects of Stereotype Threat and Double-Minority Status on the Test Performance of Latino Women. *Personality and Social Psychology Bulletin* 28(5) 659-670.

Gottfredson, Linda S., & Crouse, James. (1986) Validity versus Utility of Mental Tests: Example of the SAT. *Journal of Vocational Behavior 29, 363-378.*

Greene, Jay P, Winters Marcus, & Forster Greg. (February, 2003) Testing High Stakes Tests: Can We Believe the Results of Accountability Tests?. *Manhattan Institute for Policy Research (33) 1-19.* 52 Vanderbilt Ave. New York, NY 10017. (Also available at: http://www.manhattan-institute.org/html/cr_33.htm).

Haladyna, Thomas & Haas, Nancy Allison, J. (1991). Continuing tensions in standardized testing. Childhood Education, 75(5), 262 – 273.

Haney, W. (2000). The Texas miracle in education. *Education Policy Analysis Archives, 8, no. 41.* Retrieved February 8, 2002, from http://olam.ed.asu.edu/epaa/v8n41/part2.html

Hughes, J. F. (Ed). (1975). Education and the State. Washington DC. American Council on Education.

Johnson, J. & Tresisman, U. & Fuller, E. (2000) Testing in Texas. *School Administrator, 57.* 20-26.

Just for Kids (JFK). (2001) (Retrieved from: http://www.just4kids.org/ on December 5, 2001). Author.

Kingdon, J. (1984). *Agendas, alternatives, and public policies*. Boston, MA: Little Brown.

Klein, Stephen P, Hamilton, Laura S, McCaffrey, Daniel F., & Stecher, Brian M. (2000) "What do test scores in Texas tell us?" *Education Policy Analysis Archives, 8(49.)* (Retrieved from: http://epaa.asu.edu/epaa/v8n49/ on March, 2, 2001).

Koretz, D., & Barron, S. I. (1998). The validity of gains on the Kentucky Instructional Results Information System (KIRIS). Santa Monica, CA: RAND, MR-1014-EDU. (Retrieved from: http://www.rand.org/publications/MR/MR1014/ on June 11, 2003).

Kramer, Liz. (2002) Achieving equitable education through the courts: a comparative analysis of three states. *Journal of Law and Education 31(1)1-51.*

Laing, Joan; Engen Harold B.; & Maxey, James. (January, 1987) Relationships between ACT test scores and high school courses. *ACT Research Report Series 87(3.)*

Liebman, J. S., & Sabel, C. F. (2001). The emerging model of public school governance and legal reform: Beyond redistribution and privatization. Retrieved on January 10, 2002, from http://www.law.columbia.edu/sabel/papers/school.html

Linn, Robert, L. (March, 2000). "Assessments and Accountability," *Educational Researcher* 29(2) 4-17 (Retrieved from: http://www.aera.net/pubs/er/arts/29-02/linn01.htm on June 11, 2003).

Lopez, O. (2000). The relationship of the Texas high school curriculum to college readiness. Retrieved January 22, 2001 from http://www.just4kids.org/us/us_home.asp

Lujan, S. S., Vasques, H. & Rios, X. (1999). Study of first time students to UTPA; Report of 1998 entering students. Edinburg: *The University of Texas Pan American, Office of Institutional Effectiveness.*

Lujan, S. S., Vasques, H. & Rios, X. (2000). Study of first time students; Report of 1999 entering students. Edinburg: *The University of Texas Pan American, Office of Institutional Effectiveness.*

Madaus, George F., O Dwyer, Laura M. (1999). A short history of performance assessment: Lessons learned. *Phi Delta Kappan 73: 226-31.*

Major, B., Spencer. S., Schmader, T., Wolfe, C., & Crocker, J. (1998). Coping with negative stereotypes about intellectual performance: The role of psychological disengagement. *Personality and Social Psychology Bulletin, 24*, 34-50.

McKenna, Bernard. (March, 1977). "Guidelines as Cautions for Considering Criterion Referenced Testing". *Standardized Testing Issues: Teachers' Perspectives. Reference and Resource Series. (35) Corp Author(s): National Education Association, Washington, DC.*

McNeil, Linda. (November, 2000). Sameness, bureaucracy, and the myth of educational equity: The TAAS system of testing in Texas Public Schools. *Hispanic Journal of Behavioral Sciences 22(4). 508-523.*

McNeil, Linda & Valenzuela, Angela. (2000)"Harmful Effects of the TAAS System of Testing in Texas: Beneath the Accountability Rhetoric," (As found in: Kornhaber,

M.; Orfield, Gary.; & Kurlaendar, Michal [eds.] Raising Standards or Raising Barriers? Inequality and High-Stakes Testing in Public Education) *New York: Century Foundation.* (Also available at: http://www.ruf.rice.edu/~ctreduc/TAASArticle.pdf)

Meir, Kenneth J. & Brundy, Jeffrey L. (1997). Applied Statistics for Public Administration (4th ed.). Orlando Florida. Harcourt Brace College Publishers.

National Center for Developmental Education; Director Boylan, Hunter R. (1996, September). An Evaluation of the Texas Academic Skills Program (TASP). (Available from The Texas Higher Education Coordinating Board at: www.thecb.state.tx.us/reports/HTML/0282/boynof.htm).

National Commission on Excellence in Education. (1983). A Nation at Risk: The Imperative for Educational Reform. Washington, DC: Author. (Available at: http://www.ed.gov/pubs/NatAtRisk/risk.html)

Neill, M. D. & Medina, N. J. (1992). Eliminating standardized tests would improve schools. In D. L. Bender and L. Bruno (Ed.), *Education in America: Opposing Viewpoints* (pp. 42- 49). San Diego, California: Greenhaven Press.

Noble, J.P. (1991) Predicting college grades from ACT assessment scores and high school course work and grade information. *ACT Research Report Series. 91-3.*

Noble, J. P. &, Sawyer, R. (1987). Predicting grades in specific college freshman courses from ACT test scores and self reported high school grades. *ACT Research Report Series. 87-20.*

Osborne, J. W. (2001). Testing stereotype threat: Does anxiety explain race and sex differences in achievement? *Contemporary Educational Psychology, 26,* 291-310.

Oswald, D. L. and Harvey, R. (2000/2001). Hostile environments, stereotype threat and math performance among undergraduate women. *Current Psychology: Developmental – Learning – Personality - Social*, 19, 338-356.

Parsons, M. D. (1997). *Power and politics: Federal higher education policymaking in the 1990s.* New York: State University Press.

Patten, Peggy. (2000). Standardized Testing in Schools. *Parent News Archives.* (Retrieved on November 2003 from: http:npin.org/pnews/2000/pnew100/feat100.html

Perrone, Vito. (1991) On Standardized Testing. ERIC Digest. Retrieved on September 2002 from: http:www.ericeece.org/pubs/digests/1991/perron91.html

Phelps, Richard. (2000). Estimating the Cost of Standardized Student Testing in the United States. *Journal of Education Finance 25(3)Winter.*

Phelps, Richard. (2000). Why Testing Experts Hate Testing. *Thomas B. Fordham Foundation* (Retrieved from: http://www.edexcellence.net/library/phelps.htm on August 14, 2003).

Pitsch, Mark. (1996) Polls Confirm Key Role of Education in Political Arena. *Education Week, 15,39.* (Retrieved from: http://www.edweek.com/ew/vol-15/39polls.h15 on January 8, 2004).

Pui Wa-Lei, B. D. & Schultz, M.E. (2001). Alternatives to the grade point average as a measure of academic achievement in college. *ACT Research Report Series, 2001-4.*

Pulakos, E. D. and Schmitt, N. (1996). An evaluation of two strategies of reducing adverse impact and their effects on criterion-related validity. *Human Performance, 9,* 241-258.

Ravitch, D. (1995). *National standards in American education.* Washington D.C.: The Brookings Institution.

Richardson, R. C. Jr., Bracco, K. R., Callan, P. M. & Finney, J. E. (1999). *Designing state higher education systems for a new century.* Phoenix, Arizona: Oryx Press.

Salinas, Moises F. (1998). Stereotype Threat: The Role of Effort Withdrawal and Apprehension on the Intellectual Underperformance of Mexican-Americans. *Dissertation Abstracts International, A Humanities and Social Sciences. (59) 1-162.* (Available at: http://wwwlib.umi.com/dissertations/preview_all/9838106)

Sawyer, R. (1987). Determining minimum sample sizes for estimating prediction equations for college freshman grade point average. *ACT Research Report Series87-4.*

Scheurich, James Joseph; Skrla, Linda Johnson, Joseph F. (2000) Thinking carefully about equity and accountability. Phi Delta Kappan 82, 4, 293-299.

Sergiovanni, T. J. & Moore, J. E. (1990). *Target 2000: A compact for excellence in Texas schools.* San Antonio, TX: The Watercress Press.

Sheldon, K. M. & Biddle, B. J. (1998). Standards, accountability, and school reform: Perils and pitfalls. Teachers College Record, 100(1), 164-180.

Simmons, O. G. (1974). *Anglo-Americans and Mexican Americans in south Texas.* New York: Arno Press.

Spencer, S. J., Steele. C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4-28.

Steele, C. M., & Aronson. J. (1995). Stereotype threat and the intellectual performance of African Americans. *Journal of Personality and Social Psychology, 69,* 797-811.

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

Steele, C. M. (1998). Stereotyping and its threats are real. *American Psychologist, 53,* 680-681.

Texas Education Agency. (1995). *Academics 2000, Education Improvement Plan. Texas Education Agency Publications Distribution. Available at: 1701 North Congress Avenue. Austin Texas, 78701. (*Retrieved March 10, 2001 from http://www.tea.state.tx.us/) Author.

Texas Education Agency. (1997). *Academics 2000, Progress Report. Texas Education Agency Publications Distribution. Available at: 1701 North Congress Avenue. Austin Texas, 78701. Author.*

Texas Education Agency. (1999). *Academics 2000, Progress Report. Texas Education Agency Publications Distribution. Available at: 1701 North Congress Avenue. Austin Texas, 78701. Author.*

Texas Education Agency. (1999). 1996-97 Report on High School Completion Rates. *Office of Policy Planning and Research Division of Research and Evaluations. Texas Education Agency Publications Distribution. Available at: 1701 North*

Congress Avenue. Austin Texas, 78701. (Retrieved from: www.tea.state.tx.us/research/abstracts1.html on August 2, 2001). *Author.*

Texas Education Agency. (2000). *Expanding the Scope of the Texas Public School Accountability System.* (Retrieved from: http://www.tea.state.tx.us/research/pdfs/prr9.pdf#xml=http://www.tea.state.tx.usww.tea.state.tx.us/cgi/texis/webinator/search/xml.txt?query=+Expanding+the+Scope+of+the+Texas+Public+School+Accountability+System&db=db&id=1008a2a26c241edf on March 22, 2001) Author.

Texas Education Agency. (2000). *Interim report on Texas public schools. Texas Education Agency Publications Distribution. Available at: 1701 North Congress Avenue. Austin Texas, 78701. Author.*

Texas Education Agency. (April, 2001). 2001 Accountability Manual. *Texas Education Agency Publications Distribution. Available at: 1701 North Congress Avenue. Austin Texas, 78701.* (Also available at: http://www.tea.state.tx.us/perfreport/). Author.

Texas Higher Education Coordinating Board (2000-2001) 1996 TASP Annual Report. THECB. Available at: Texas Higher Education Coordinating Board. PO Box 12788 Austin, TX. 78711. (Retrieved from http://www.thecb.state.tx.us/reports/html/0208/ar96.htm on February 19, 2002) Author.

Thompson, S. (2001). The authentic standards movement and its evil twin. *Phi Delta Kappan, 82-5.* 358-363. Retrieved April 14, 2002 from Academic Search Elite database.

Thurston, L. M, & Roe, W. H. (1957). *State school administration.* New York: Harper and Brothers Publishers.

Tinto, V. (1993) *Leaving college: Rethinking the causes and cures of student attrition.* (2nd ed.). Chicago: University of Chicago Press.

Trueba, Henry. (2001). Conclusion: Polar Positions on the Texas Assessment of Academic Skills (TAAS): Pragmatism and the Politics of Neglect. *Education and Urban Society 33(3) 333-344*

Toch, T. (23, March 1996). Reformer Ross. US Online News. Retrieved December 20, 2001, from http://www.usnews.com/usnews/news/ross.htm

Valencia, R. R. (Ed.). (1991). *Chicano School failure and success.* New York: The Falmer Press.

Valenzuela, Angela. (1999). *Subtractive schooling: U.S. Mexican youth and the politics of caring.* Albany, New York: State University of New York Press.

Valenzuela, Angela. (2000). The significance of the TAAS Test for Mexican Immigrant and Mexican American adolescents: A case study. *Hispanic Journal of Behavioral Sciences22(4) 524-539.*

Walsh, M., Hickey, C., & Duffy, J. (1999). Influence of item content and stereotype situation on gender differences in mathematical problem solving. *Sex Roles, 41,* 219-240.

Weis, Lois. (1998) (ed). Class, Race, and Gender in American Education. SUNY Press, Albany, New York.

Zirkel, Perry A., & Gluckman, Ivan B. (January, 1993). "No pass, no play" rules. *Principal 72. 62-63.*

APPENDIX B

VITA

VITA

Monica Marie Treviño

Educational Background

The University of Texas - Pan American
Address: 1201 W University DR., Edinburg, TX 78541
Major: English
Degree: BA English, December 1998

Professional Positions

Northern Illinois University (8/2003 – Current)
Assistant Director, University Resources for Latinos (URL)

Aurora University (8/2002-8/2003)
Admission/Advisement Specialist, College of Education/Office of Admission

The University of Texas - Pan American (6/1999-8/2002)
Student Development Specialist, College of Science and Engineering/University Retention Advisement Program (URAP)

Languages

Bilingual: Spanish speaking, reading and writing with basic translation ability

Professional Affiliations

Illinois Association for College Admission Counseling (IACAC) (2001-2002)
National Academic Advising Association (NACADA) (1999-2001, 2002-2003)
National Association of Student Personal Administrators (NASPA) (1999-2001, 2002-2003, 2003-2004)