5-2009

A Quantitative Study of School Characteristics that Impact Student Achievement on State Assessments and those Assessments' Associations to ACT Scores in Tennessee.

Phillip L. Swanson
East Tennessee State University

A Quantitative Study of School Characteristics That Impact Student Achievement on State Assessments and Those Assessments' Associations to ACT Scores in Tennessee

---

A dissertation

presented to

the faculty of the Department of Educational Leadership and Policy Analysis

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Doctor of Education

---

by

Phillip L. Swanson

May 2009

Dr. Catherine H. Glascock, Chair

Dr. Cecil Blankenship

Dr. Eric Glover

Dr. James Lampley

Keywords: ACT, Value-Added, Lottery, Grade-Point Averages, Meritocracy, Tennessee

ABSTRACT

A Quantitative Study of School Characteristics That Impact Student Achievement on State Assessments and Those Assessments' Associations to ACT Scores in Tennessee

by

Phillip L. Swanson

The purpose of this study was to determine whether a significant relationship exists between particular home, student, and school characteristics and ACT scores and the relationship these characteristics subsequently have with the TVAAS grades assigned to each high school's ACT scores. These home, student, and school variables were socioeconomic status, percentage of minority, graduation rate, per-pupil expenditure, Gateway English II scores, and Gateway Algebra I scores. By looking at these variables and the influence they hold on the education of students, I sought to ascertain the fairness that is present when schools and districts are given grades through the TVAAS assessment.

The population in this study was students in the Tennessee high schools that had given the Gateway English II test, Gateway Algebra I test, and had TVAAS ACT composite grades. I also examined the influence that variables such as socioeconomic status, percentage of minority students, graduation rate, and per-pupil expenditure have on achievement. Before doing the first phase of this project, I set about to see if assumptions of normality were met. I then analyzed data to establish that certain home, student, and school variables affect achievement. After doing that, I was able to show a strong relationship between these 6 home, student, and school variables and achievement.

After establishing predictor variables, I examined the predictor variables and their relationship with the TVAAS ACT composite scores of Above, No Detectable Difference, and Below. These designations from the Tennessee Department of Education are "grades" for the schools and districts. The analyses indicated that, indeed, some of these home, student, and school variables such as socioeconomic status and percentage of minority students still have a relationship with the grades, despite the claim that TVAAS measures teacher effectiveness almost exclusively.

This study concluded with recommendations that further modifications need to be done with the TVAAS grades on ACT composite scores. The conclusions in this dissertation merit consideration from Dr. William Sanders as well as the assessment division of the Tennessee Department of Education.

DEDICATION

This study is dedicated:

To my parents, Leo and Myrle Swanson, who were not afforded the opportunity to be educated beyond the eighth grade yet were truly life long learners. They always made sure that I was able to participate in any activity I desired and insisted on excellence in those activities. Their dedication to their children and community were commendable to all who knew them.

To my daughter, Savannah Swanson, who toils every day to be the best student she can be. I hope my work has been the example you will need as you grow toward adulthood.

To my brothers, sister, and the many nieces and nephews who work hard to be the type of citizens all should aspire to.

Finally, I would like to dedicate this work to all students who may come from backgrounds that are more challenging than some face. Education is the key to becoming the person you can be, to achieving true equality--whether in the classroom or in life. Although some people have to climb harder than others do to find success, the mountain is still there awaiting you to conquer it.

ACKNOWLEDGMENTS

There are many to acknowledge in the fulfillment of this life-long dream. I would like to acknowledge my committee: Dr. Catherine Glascock, chair, Dr. Cecil Blankenship, Dr. Eric Glover, and Dr. James Lampley. Each has contributed in his or her own way to my reaching this final educational destination.

Dr. Glascock, who took over this project, after the prospectus, while also coming to lead the department, reassured me that I could accomplish this. Thank you for your help.

Dr. Lampley was always there for advice on the statistical part of this project. Thank you for your responsiveness.

Drs. Blankenship and Glover made astute observations that enabled me to improve my product to this point. Thank you, gentlemen, for that assistance.

My gratitude also goes out to Dr. Jasmine Renner, previous chair. She helped guide me through the prospectus. Thanks for that.

I would also like to thank Dr. Susan Twaddle and Debby Bryan for their professional expertise that allowed me to have some level of confidence that I could get this right! Please know I appreciate you.

Finally, I would like to thank the Walters State Cohort. From the first day that I traveled that long road to Morristown, I felt surrounded by friends who could help if help was needed. In this doctoral program of higher education, I was fortunate to be surrounded by several people who were from the K-12 educational arena. Thanks go to April Sell, Jill Reuschel, Eydie Pettigrew, and LaKisha Waters. I want to especially acknowledge Lakisha as she played a vital role in keeping me straight in this difficult process. Thanks to all who have helped.

CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The state of Tennessee began its venture into accountability in 1992. Originally, the state met the accountability requirements of the Tennessee General Assembly through a process called the Tennessee Value Added Assessment System (TVAAS). With the passage of the *No Child Left Behind Act* (2001), the state had to meet additional requirements for accountability. The TVAAS legislation, for example, required testing of norm referenced data, whereas *No Child Left Behind* requirements called for criterion referenced data. After 1 year of giving "two tests in one," the state was able to merge the criterion data with the norm referenced data and meet requirements of both the state legislature and the federal statute of *No Child Left Behind*. Tennessee now produces a State Report Card containing test scores and demographics for its public schools (Tennessee Department of Education, 2007b). These scores, analyzed by state and individual districts, are given an original analysis from the state to be scrutinized at the local level. If the information is not deemed to be correct, the local school district can appeal the findings of the state. These appeals can be based on natural disasters such as flooding that might have occurred during the testing period or coding errors such as a student with disabilities not bubbling in the section indicating he or she was a student with disabilities. After this "cleansing" process, the local district concurs with the state's analysis and the scores are then deemed to be final. Subsequently, the Report Card is released to the media and the public by the state through the state's website (Tennessee Department of Education, 2007b).

Secondary education plays a vital role in our country and is being subjected to increasing scrutiny. Over the past decade, state department of education officials have led the national movement to raise standards and graduation requirements, improve assessments, and strengthen accountability. Far too many young people leave our schools today without the skills and knowledge needed to compete in college, careers, and life. One goal is to help every state close the expectations gap so that all students graduate ready for success (Achieve, 2008). By all

indications, high schools in America can improve and there are tremendous efforts to ensure that improvement will occur. One of the biggest initiatives has been the American Diploma Project that is being implemented in states throughout the country (Achieve). Of the states, 33 have now joined with Achieve in the American Diploma Project Network to tackle the challenge of preparing graduates for the postsecondary world (Achieve). To inform their work, Achieve is conducting a study of the American College Testing (ACT), the Scholastic Achievement Test (SAT), and the most commonly used college placement tests to better understand what they measure and how they compare to high school tests and expectations. As a result of the American Diploma Project, there should be an increased emphasis on college entrance exams. ACT and SAT will receive more importance than ever before with the emphasis of a prekindergarten through grade 16 alignment (Achieve). As postsecondary institutions receive students from the nation's secondary schools, many business and postsecondary leaders have questioned the education these students attained.

By looking at the history of college placement in America, it is apparent that we are coming full circle as we renew our emphasis in secondary schools on college placement tests. Tennessee should systematically use the ACT for all students to enhance their college readiness skills. Our educational system from prekindergarten through grade 16 should be more aligned with the coming emphasis on test scores for college placement (Achieve, 2008).

Since the advent of the ACT in 1959, students across America have taken the test and lived with the results of the ACT scores (ACT, 2008). The stakes are rising for our nation's students. Presently, the American Diploma Project is being implemented in 33 states and more will soon adopt the program (Achieve). Tennessee has joined the movement with the adoption of the Tennessee Diploma, the state's version of the American Diploma. A new curriculum and new standards will be in place with the ninth grade class of 2009-10. The ramifications of the Tennessee Diploma will be far reaching for the state's students.

The curriculum is a vital key to producing students who can complete college (Adelman, 1999). According to ACT (2008a), studies reveal that a core ready curriculum will lead to

greater achievement and. consequently, to greater college success for students across the nation. Tennessee's students are no exception to this research-based conclusion.

Another aspect of the educational system in Tennessee is the lottery scholarship, or more specifically, the HOPE scholarship. The HOPE Scholarship has been in place in Tennessee since 2004. Students must meet certain requirements to qualify for the lottery scholarships including having a grade point average of 3.0 or attaining a score of 21 on the ACT.

Tennessee's original accountability measuring stick was the Tennessee Value-Added Assessment System (TVAAS). This system was originally the state's only accountability measure. The Tennessee General Assembly had put this in place long before the federal legislative accountability system of *No Child Left Behind*. The TVAAS system attempted to measure students' growth from one year to another based on what the innovators of the TVAAS system called an average year's schooling experience (Sanders & Horn, 1994). Sanders (1998) and his colleagues have made a national name for themselves and their accountability system now called EVAAS, with the growth model being used in nine states for average yearly progress. By looking at the history of testing and the role the ACT and TVAAS now play in Tennessee, educators might come to understand more about the best way to educate secondary students in Tennessee.

This study focused on the school characteristics that have enabled schools to have the greatest impact on ACT scores, thus qualifying the greatest number of students for the lottery scholarships. Characteristics such as socioeconomic status, percentage of minority students, graduation rate, and per-pupil expenditure were explored. The schools' free or reduced school lunch prices were one element of the study. Student ethnicity was another characteristic studied. I examined a percentage of minority students in this study to see what relationship existed between test scores and the percentage of student minorities throughout the state. Another characteristic measured was the graduation rate of the high schools. Finally, I examined the relationship between system per-pupil expenditures and test scores. These various characteristics do affect student achievement and, consequently, schools' and systems' reputations.

"These times, they are a'changin," Dylan (1964) once sang, and indeed, with the coming of the Tennessee Diploma, Tennessee's version of the American Diploma, we need to determine those instructional practices that have been effective and discard those that have not been effective.

Statement of the Problem

Tennessee has adopted the American Diploma with additional emphasis on the ACT test with an accountability package that grades schools through the Tennessee Value Added Assessment System (TVAAS). There continues to be little empirical evidence as to the equity of the evaluation of high schools in Tennessee on the basis of the TVAAS data and the *No Child Left Behind* requirements. Schools continue to be graded on the data produced per state statute whereas the Tennessee Diploma is effectively a national diploma based on a national curriculum. Because a new Tennessee diploma is coming that will emphasize college readiness, there exists a concern that graduation rates will be affected, career-technical education will suffer, and students will be negatively affected.

The importance of the ACT test is increasing throughout the country, while educators in Tennessee continue to be subject to the TVAAS analyses that provide a grade for teachers, schools, and school districts. These psychometric measurements are being used to determine success or failure of school systems, teachers, and students. The purpose of this study was to determine whether a significant relationship exists between particular home, student, and school characteristics and ACT scores and the relationship these characteristics subsequently have with the TVAAS grades assigned to each high school's ACT scores. These home, student, and school variables were socioeconomic status, percentage of minority, graduation rate, per-pupil expenditure, Gateway English II scores, and Gateway Algebra I scores.

Significance of the Study

This study focused on the relationship between student demographics and their relationship with the testing data. By examining these relationships, the state department, systems, and schools should understand better how to evaluate students, schools, and systems in a more equitable manner. Schools are being reconstituted all across the state impacting schools' and systems' reputations and school personnel's lives. Consequently, it is imperative that our accountability system be as fair as possible to all involved. The upside to the Tennessee Diploma Project should be that more students will learn more things and will have a greater opportunity for college completion. There is also a concern as we begin to test all students on the ACT that the public will perceive public education is a failure and may call for counter-productive reforms such as vouchers. The Tennessee Diploma Project might benefit from a study that sheds light on those variables that affect learning in the school environment. The data used in this study can be found online and compared to the State Department of Education's website containing the 2007 Report Card (State of Tennessee, 2008).

Limitations and Delimitations

The data reflect only those schools that are public high schools with grades 9-12. Some school systems, such as the Athens City School system, have ninth grade students and are subject to the Gateway tests, but those schools were not included in this study. There are 410 high schools listed in the *Public School Review* in the state. Many schools were omitted from the study for reasons such as being vocational centers, adult high schools, or other "atypical" high schools. Other schools were eliminated due to incomplete reporting of Gateway or ACT scores. This data set contains 265 secondary schools and constitutes a significant percentage of the Tennessee schools that served secondary students in 2007 as reported by the State Department of Education.

Another limitation that existed in this study was ascertaining the percentage of students who took the test in a given school. I inquired at different schools and found it difficult to

determine the percentage of students who took the ACT in those schools. The ACT was not required in 2007-08 of all students making it difficult to conduct a completely accurate evaluation of all schools. Despite this limitation, however, analysis should reveal significant associations among the demographic and test score data.

Finally, another limitation in this study is the percentage of minority students' data. This study is defining this as being all students who are not white. The Asian/Pacific Islander category, therefore, is counted in the percentage of minority data. Asian Americans tend to score higher than the white subgroup, African American subgroup, and the Hispanic subgroup.

Research Questions and Null Hypotheses

1. Which home, student, and school variables have the strongest relationship with 2007 ACT scores for Tennessee high schools?
2. Is there a relationship between the home, student, and school variables (socioeconomic status, percentage of minority, graduation rate, per-pupil expenditure, Gateway English II scores, and Gateway Algebra I scores) with the three graded categories of TVAAS (Below, Above, No Detectable Difference)?

From research question number one, the following null hypotheses were tested:

$Ho1_1$: There is no relationship between socioeconomic status and a high school's composite ACT score.

$Ho1_2$: There is no relationship between percentage of minorities and a high school's composite ACT scores.

$Ho1_3$: There is no relationship between graduation rates and a high school's composite ACT scores.

$Ho1_4$: There is no relationship between per-pupil expenditures and high school's composite ACT scores.

$Ho1_5$: There is no relationship between Gateway English II scores and a high school's composite ACT scores.

$Ho1_6$: There is no relationship between Gateway Algebra I scores and a high school's ACT composite scores.

From research question number two, the following null hypotheses were tested.

$Ho2_1$: There is no relationship between socioeconomic status and the three graded categories of TVAAS (Below, Above, No Detectable Difference).

$Ho2_2$: There is no relationship between percentage of minorities and the three graded categories of TVAAS (Below, Above, No Detectable Difference).

$Ho2_3$: There is no relationship between graduation rate and the three graded categories of TVAAS (Below, Above, No Detectable Difference).

$Ho2_4$: There is no relationship between per-pupil expenditures and the three graded categories of TVAAS (Below, Above, No Detectable Difference).

$Ho2_5$: There is no relationship between Gateway English II scores and the three graded categories of TVAAS (Below, Above, No Detectable Difference).

$Ho2_6$: There is no relationship between Gateway Algebra I scores and the three graded categories of TVAAS (Below, Above, No Detectable Difference).

Definitions of Terms

1. *Above (status):* As used in the study, this is a term to describe the percentage of students in this school who made significantly more progress in this subject than students in the average school in the state (Tennessee Report Card, 2007, p. 1).
2. *Below (status):* As used in the study, this is a term to describe the percentage of students in this school who made significantly less progress in this subject than students in the average school in the state (Tennessee Report Card, p. 1).
3. *Free- or Reduced-Price Meals*: This is a federal program that provides free- or reduced-priced meals to children based upon their family's income. Effective July 1, 2007, children in a family of four making less than $26,845 are eligible for free

meals; those making less than $38,203 qualify for reduced prices (Food & Nutrition Service, 2008, p. 1).

4. *Gateway Exams:* Students who enter their freshmen year in 2001-02 must pass three Gateway tests--mathematics, science, and language arts--before graduation to earn a high school diploma (Tennessee Department of Education, 2005, p. 1).
5. *NDD (status)*: As used in the study, this is a term to describe the progress of students made in a school that was not detectably different (NDD) from the progress of students in the average school in the state (Tennessee Report Card, p. 1).
6. *Tennessee Value Added Assessment System (TVAAS)*: TVAAS is a tool that gives feedback to school leaders and teachers on student progress. It allows districts to follow student achievement over time and provides schools with a longitudinal view of student performance. TVAAS provides valuable information for teachers to make informed instructional decisions (Tennessee Report Card, p. 1).
7. *Value-Added Assessment*: Value-added measures student progress within a grade and subject that demonstrates the influence the school has on the students' performance. This reporting provides diagnostic information for improving educational opportunities for students at all achievement levels (Tennessee Report Card, p. 1).

Overview of the Study

Chapter 1 presented an introduction, statement of the problem, significance of the study, limitations and delimitations, research questions and hypotheses, and definitions of terms used in the study. Chapter 2 contains a review of the related literature. Chapter 3 addresses the research methodology including data collection and data analysis. Chapter 4 presents the data analysis and Chapter 5 contains the summary of findings, conclusions, and recommendations for practice and further research on the subject.

CHAPTER 2

LITERATURE REVIEW

Introduction

The purpose of this study was to determine whether a significant relationship exists between particular home, student, and school characteristics and ACT scores and the relationship these characteristics subsequently have with the TVAAS grades assigned to each high school's ACT scores. These home, student, and school variables were socioeconomic status, percentage of minority, graduation rate, per-pupil expenditure, Gateway English II scores, and Gateway Algebra I scores. A review of the relevant literature indicated a need to know more about the TVAAS scores and the grades that the state department of education assigns each secondary school. Sanders and Horn (1994) indicated those who have developed the TVAAS assessment system purport to be able to distinguish teacher effectiveness without factors such as socioeconomics and ethnicity significantly impacting schools' scores.

This literature review traces the history of testing from past psychometric practices to the present day practices of accountability. It also explores the philosophy of intelligence testing and achievement testing and exposes a philosophy called a meritocracy. The current state of accountability measures in Tennessee is reviewed and this study traces those measures to their conception. The Tennessee Education Lottery Scholarship Programs are also explored. Finally, the review of literature examines the role that specific school characteristics play in a school's performance and its assessment from the state department of education. By examining the past practices and understanding more about the mistakes from those practices, officials might be empowered to make better decisions about future practices in the area of public education in our secondary schools.

History of Testing

In ancient Greece, Socrates tested his students through conversations. Answers were not scored as right or wrong, they just led to more dialogue. Many intellectual elites in the 5th and 4th centuries B.C. cared more about finding the path to higher knowledge than they did in producing a correct response (Matthews, 2006). To them, accuracy was for shopkeepers. So how did one go from that concept to an educational model shaped--and perhaps even ruled--by standardized, normed, charted, graphed, regressed, calibrated, and validated testing? Critics have said standardized testing has robbed schools of the creative clash of intellects that make Plato's dialogues still absorbing. Bracey (2004), research columnist for the Phi Delta Kappan education journal, said, "There is a growing technology of testing that permits us now to do in nanoseconds things that we shouldn't be doing at all" (p. 716). Standardized exams have many sources. In imperial China in the A.D. 7th century, government job applicants had to write essays about Confucian philosophy and compose poetry. In Europe, the invention of the printing press and modern paper manufacturing fueled the growth of written exams according to Mathews (2006). By 1845 in the United States, public education advocate Horace Mann was calling for standardized essay testing. Spelling tests, geography tests, and math tests blossomed in schools although they were rarely standardized. I have explored the pathways such as intelligence tests and achievement tests that have dominated the testing movement from the past to now.

Ever wonder how many brush strokes it takes to create a painting? Have you thought of how to measure boredom, attraction to the opposite sex, the efficacy of prayer, or the intelligence of earthworms? According to Ludlow (2008), Sir Francis Galton, founder of psychometrics, wondered about these things and set out to develop procedures and instruments by which such questions could be answered and replicated. He counted everything that appeared to have any form of regularity. He counted brush strokes of the painter who painted his portrait, points of similarities of twins, the attractiveness or “turn-offs” of women, facial characteristics, and other things. In fact, he was probably one of the first to measure the phenomenon of snoring. He

seemed to have always carried a notebook and some type of ingenious device capable of pricking a piece of paper by which he recorded, unobtrusively, various aspects of events occurring around him. He even performed arithmetic by taste and smell. His inquisitive nature must have run in the family, as Sir Francis Galton was a cousin to Charles Darwin ("Francis Galton: The First Modern Attempt," 2003).

The field of psychometrics has been studied extensively. It began with the classical German psychophysics of the 1800s, Ernst Heinrich Weber, Wilhelm Wundt, and Gustav Fechner; and moved into the 1900s ability-testing movement with James McKeen Cattell, Alfred Binet, and Charles Spearman; and then into the psychological scaling methods associated with Louis Leon Thurstone. Although the field of "intelligence testing" has evolved a great deal, the practice of giving intelligence assessments persists today. Modern test theory texts are introduced where standard presentations include something like "The field of psychometrics has a history of growth and development extending over some 75 years since the early work of Binet in France and Spearman in England" (Thorndike, 1982, p. 1) and "psychometric methods" is simply defined as "procedures for psychological measurement" (Guilford, 1954, p 1).

Galton's (1879) interests in mental operations led him to propose a "new instance of psychometry" (p. 149). In his article, "Psychometric Experiments," he defined "psychometry" as the "art of imposing measurement and number upon operations of the mind" (p. 149). He then argued, "Until the phenomena of any branch of knowledge have been subjected to measurement and numbers, it cannot assume the status of dignity of a science"(p 150). His work illustrated what he called the psychometric side of anthropology.

According to Ludlow (2008), in 1879, Galton did work with word associations, a common practice in later years with psychologists and psychiatrists. He was also interested in mental tests. According to Pearson (1924), Galton "expressed the following words and thoughts and they seemed to illustrate the need to exercise caution before putting too much emphasis on a test:

> There are many faculties that may be said to be potentially constant in adults though they are not developed, owing to want of exercise. After adequate practice, a limit of efficiency would in each case be attained and this would be the *personal constant* (italics added); but it is obviously impossible to guess what that constant would be from the results of a single trial. No test professes to do more than show the efficiency of the faculty at the time it was applied, and many tests do even less than this. (pp. 371-372)

However, according to Sweet (2004), times were to change for Galton and his colleagues. Galton's protégé was Cattell, who assisted him on experiments in South Kensington, Great Britain. Cattell and Galton were "measuring" the amount of time that it took for subjects to complain about a "rubber tipped compass point" being pressed against their forehead. After this test was done, the findings were used to chronicle the intelligence of the subjects. This ended in 1901 when Clark Wissler, one of Cattell's graduate students, squished Galton's theory by showing that there was no correlation whatsoever between high scores in Cattell's tests and high academic achievement; a hypersensitive forehead was no guarantor of straight As.

Wissler is best known today as an influential American anthropologist, but his early training was in psychology. Fancher (1985) noted that Wissler's 1901 doctoral dissertation created an academic uproar by purportedly debunking some of the most influential intelligence theories of the time. The controversy was made scandalous by the fact that Wissler's findings discredited the research of his mentor, Cattell.

It is rare for a graduate student to single-handedly crush the morale of the professional establishment, but that is exactly what Clark Wissler did. At the time of his doctoral dissertation, Wissler was one of Cattell's graduate students at Columbia University. He obtained the psychophysical test scores from several hundred Columbia University and Barnard College students who had been Cattell's research subjects. He then used the newly perfected Pearson correlation coefficient to examine the relationship between each student's score on each of the tests and his or her undergraduate academic grades. Karl Pearson was a long-time friend and associate of Galton (Lohninger, 1999).

The surprising result was that there was virtually no correlation between scores on Cattell's tests and academic achievement (Human Intelligence, 2008). Perhaps equally surprising

was the fact that scores on Cattell's tests did not even appear to correlate with each other. Because the tests did not agree among themselves and they did not correlate with independent measures of mental ability (undergraduate academic grades), it did not seem possible that they could be valid measures of intelligence (Fancher, 1985; Sternberg, 1990). At the time of Wissler's dissertation, eugenics was gaining momentum and psychophysical measurement was the primary research paradigm for intelligence testing; however, this paradigm was about to change.

The eugenics movement was growing, and the majority of the psychological community was thoroughly invested in the findings of Galton and Cattell. Eugenics was the theory based on the work of Gregor Mendel's work with peas and genetics (PBS, 1998). Those who espoused the theory of eugenics believed that human breeding could affect the quality of the species. In other words, if the races and people with good qualities were encouraged to have children, the quality of the species would be improved. Conversely, those people who were seen to be lacking in positive qualities would be discouraged from having children. It was presented as a mathematical science that could be used to predict the traits and behaviors of humans, and in a perfect world, to control human breeding so that people with the best genes would reproduce and thus improve the species. It was an optimistic school of thought with a profound faith in the powers of science, presented as mathematics (PBS). Obviously, there were those in the testing community who attempted to use test results to support their beliefs in the benefits of eugenics. After Wissler's results became known, the psychology community gradually lost interest in psychophysical testing. This represented a significant shift. Although Cattell remained a psychologist, he also became disenchanted with psychophysical testing and spent the remainder of his career in relative obscurity (Fancher, 1985). Galton continued to be interested in eugenics and hereditary theory until he died in 1911 (Fancher). Wissler's findings were obviously devastating to the two men whom had been Wissler's mentors.

Critics have pointed to flaws in Wissler's research. According to Fancher (1985), Charles Spearman professed doubts about the findings and suggested that a "correction formula"

would allow the relationships to be assessed more accurately (pp. 88-89). Although Clark Wissler left the field of psychology a century ago, psychology has not left him behind. His doctoral dissertation permanently changed the dominant research paradigm for intelligence testing (Fancher). Testing of intelligence had taken a dramatic turn. Is it possible there is more to learn today about testing of students?

After the Wissler controversy, men like Alfred Benet became increasingly important players in the arena of testing. Sweet (2004) maintained that Benet, a French child psychologist, was instrumental in developing a means of testing students and establishing a numerical figure to represent the child's "intelligence." Binet established norms with 50 normal children, as picked by their teachers, doing 30 incrementally difficult tasks. One such task was passing a lighted match in front of the student's face and noting if his or her eyes followed the match. He also had the students draw from memory and construct sentences with three words to be used (Sweet). After the norms were established for 6-year olds, Binet compared older children on similar tasks and assigned a grade of mental age and chronological age. A 7-year-old child who could only do 6-year old's tasks would be seen as being a year behind. The term "intelligence quotient" became in vogue in 1812 when William Stern converted the mental age into a ratio between mental age and chronological age. If a child performed like a child of 7 years and was 8 years old, he or she would be assigned a quotient of 7/8 or .875. It did not take long for Lewis Terman of Stanford University to adopt Binet's idea, multiplied Stern's score by 100 to eliminate the decimal, and called the results the intelligence quotient (Sweet).

Fancher noted that these were the first tests that really worked to any substantial degree because they did make some effective diagnoses (1985). Fancher was a professor of psychology at York University, Ontario. Fancher observed that Binet was quite skeptical about the quantification of the results and the irony was that the IQ, the unitary number of intelligence, became so strongly identified with Binet's name (1985). Binet died in 1911 at the age of 54 with his journey half completed. His death prevented him from seeing some of the uses to which his ideas would be put: turning public vanity into cash and advancing the cause of racist eugenics.

The two most prominent names in the next generation of intelligence testers, Lewis Terman and David Wechsler, both got rich by selling IQ tests (Sweet, 2004).

According to Minton (1988), the development of the Stanford-Binet test was furthered by Terman. In 1916, Terman authored *The Measurement of Intelligence: An Explanation of and a Complete Guide for the Use of the Stanford Revision and Extension of the Binet-Simon Intelligence Scale*. Although there were other translations of the Binet-Simon available at this time, Terman's normative studies and his methodical approach are credited with the success of the Stanford-Binet (Minton). Working with Maud Merrill, first his student and later a fellow professor and research collaborator at Stanford University, Terman created two parallel forms of the Stanford-Binet. One form was called L for Lewis and M for Maud (Becker, 2003). In the 1950s, Merrill took the lead in revising the Stanford-Binet, selecting the best items from Forms L and M to include in a new version of the test. The two forms from 1937 were combined to create the Form L-M. This form was published in 1960 (Terman & Merrill, 1960) and was later renormed in 1973 (Terman & Merrill, 1973). This form added alternate items at all levels, but otherwise, the format remained similar to the 1937 forms (Becker). The Stanford-Binet Intelligence Scale: Fourth Edition (Thorndike, Hagen, & Sattler, 1986) moved from the age-scale format introduced by Binet to a point-scale format. The Fourth Edition also formalized the practice of multistage testing in which performance on the vocabulary scale determines the starting point for subsequent tests. Whereas some examiners used the vocabulary test for routing on earlier editions of the test, this was not official practice (Becker).

In 2003, the Fifth Edition was published (Roid, 2003). This edition was an attempt to carry on the tradition of the prior editions while taking advantage of current research in measurement and cognitive abilities. Like the Fourth Edition, the Fifth Edition includes multiple factors. These factors are modified from those on the Fourth Edition but represent abilities assessed by all former versions of the test. The Standford-Binet test is still the most frequent intelligence test used in elementary schools today.

Along with the extensive research literature on the Stanford-Binet, reviews of the test have been available since before the first *Mental Measurement Yearbook* was published, (Pratt, 1917) and now in its 17th edition (Buros, 2008). The tests were tweaked with each addition. Although it was better than the 1916 edition, Forms L and M were criticized for the quality of the scoring rules, the paucity of nonverbal content at the upper levels of the test, and the nonuniform standard deviation of IQ that led to different interpretations of IQ at different ages. The Fourth Edition was an attempt to address many concerns that had been raised with prior versions of the test while maintaining the same types of tasks and items (Becker, 2003).

The Stanford Binet is not without its critics. Gould (1981) maintained that the later tests were not administered like the earlier tests. He noted that students were originally given the tests in one-on-one settings. He also questioned the mass measurement of mankind and its implications racially. He challenged the notion of any one test being able to measure real property in the head (Gould).

The testing movement was not limited to people like Benet and Terman. According to Ravitch (2006), Nicholas Murray Butler, president of Columbia University, and Charles Eliot, president of Harvard University, persuaded their colleagues in American education of the importance of developing an organization to establish uniform curriculum standards and a uniform examination system. Their planning led to the creation of the College Entrance Examination Board in 1900. The College Entrance Examination Board, thus, played a huge role in the evolution of testing. The new organization created the best, most consistent, and most influential standards that American education had ever known. The work of "the Board," as it was known, had a powerful, uplifting influence on secondary education according to Ravitch (2006). Even though roughly only 1 of every 20 17-year-olds in 1900 finished high school, and even fewer expected to go to college, everyone who attended high school in that era studied the curriculum that was later called the college track. This resembles the current movement toward the American Diploma in which the one track (college path) is being pursued. Not everyone was thrilled with this new plan in the early 1900s. Ravitch (2006) noted that the president of

Princeton University had worried it would lead to a state examination system. Charles Eliot, president of Harvard University, assured him that was not even a remote possibility. That is rather ironic, because the College Board's SAT and the ACT are completed by a majority of our nation's high school students as a required state examination. There were concerns among the elite that the "have's" might not receive their entitled place among the finest universities. The president of Lafayette College complained that it might prevent the college from admitting the sons of wealthy benefactors and faculty members. According to Ravitch (2006), Columbia president Butler assured him that Lafayette, if it chose, could admit "only such students as cannot pass these examinations" (p. 1).

Now called the College Board, it "created the best, most consistent and most influential standards that American education has ever known," New York University educational historian Ravitch (2006) wrote in the *Chronicle of Higher Education* (p. 1). The board's early exams were written and graded by teachers and professors and had no multiple-choice questions. These essay exams, Ravitch (2006) wrote, led "everyone who went to high school, whether they were the children of doctors or farmers or factory workers . . . to study mathematics, science, English literature, composition, history and a foreign language, usually Latin" (p. 1).

The professors and teachers worked together in harmony for some time and there was a sense of collegiality amongst the secondary teachers and the college professors. They embraced the role of educational standard-bearers for the country, but this was to change. There began to be a debate among the powerful policy makers about the effectiveness and necessity to measure what a student had already learned. Would it not be better to measure what a student is capable of learning instead? As a result of this tension, the testing movement began the process of evolving again and another giant in the testing arena would emerge.

This Goliath was Yerkes (1920). Yerkes was raised in rural Pennsylvania in the late 1800s and desired to become a physician. He went to college and left the tough rural life to receive the education necessary to become a doctor. He hit a crossroad, however, when he was given an opportunity to attend Harvard, not to become a physician, but rather to study biology.

He chose to go to Harvard and began studying animal behavior and became fascinated with the study of comparative psychology. Yerkes received his PhD from the psychology department at Harvard in 1902. A contemporary of behaviorist, John Watson, Yerkes wrote his first book, *The Dancing Mouse*, in 1907, and went on to become the president of the of American Psychological Association (APA).

As president of the APA, Yerkes (1920) urged that organization to become active in the war effort. He became involved immensely in the war efforts during World War I. As chairman of the Committee on the Psychological Examination of Recruits, he developed the Army's Alpha and Beta Intelligence Tests, the first nonverbal group tests that were given to over one million United States soldiers during the war. These tests eschewed the use of one-on-one assessments and moved to mass testing with paper and pencil. Yerkes persuaded the Army to let him test all recruits. Although Binet was often critical of America's preoccupation with mass assessment (Ravitch, 2000), there were advantages to this practice. For many reasons, the launch of the Army Alpha and Beta testing program was a pivotal moment in the history of psychology. First, it provided psychometricians with the first group intelligence tests. Second, the publicity it generated popularized intelligence testing in the public and private sectors. Third, the program provided vast amounts of data to serve as fuel for future controversies over apparent racial differences in intelligence test scores and the supposed decline of America's national intelligence (Fancher, 1985). One of Yerkes' assistants, Carl Brigham, was a young psychologist who taught at Princeton.

After working on the Alpha and Beta intelligence tests, Brigham began the process of developing a test to measure scholastic aptitude and began to make the test more difficult. Psychologists like Brigham, Terman, Thorndike, and Yerkes claimed that the new tests could quickly make accurate predictions about students' innate ability. The test was first given in 1926 to a few thousand college applicants. The test that Brigham had developed was the SAT and unlike the College Boards' test, the questions were multiple-choice, not essay questions (Ravitch, 2006).

The College Board was under fire. Brigham himself was now on the College Board and never missed an opportunity to raise concerns about the test given by the Board. Faced with claims that its examinations were obsolete and not scientific like the new tests, the College Board engaged a group of psychologists to design a modern test. The committee of the Board, which included Brigham and Yerkes, continued to examine whether the College Board should adopt the SAT as their test for college entrance (Ravitch, 2006). In 1930, when Brigham joined the staff of the College Board, he continued to conduct research on the SAT. One of the key figures in the development, marketing, and popularization of group intelligence tests, he maintained not only that they measured fixed, innate intelligence but also that inherited intelligence varied by race and ethnicity (Ravitch, 2006).

According to Frontline PBS (1999b), in1933, James Bryant Conant, on becoming president of Harvard, decided to start a new scholarship program for academically gifted boys who did not come from the Eastern boarding schools that were the regular suppliers of Harvard's students. He gave Henry Chauncey, an assistant dean at Harvard, the task of finding a test to evaluate candidates for these scholarships. Chauncey met Brigham and came back to Conant with the recommendation that he use the SAT. Conant liked the test because he thought it measured pure intelligence, regardless of the quality of the taker's high school education (Frontline PBS, 1999b) and would allow students who had not had the privileges of private schooling or boarding schools an equal opportunity to qualify for attendance at Harvard.

The debate continued to rage for a few more years until a national crisis led to an end to the issue. On December 7, 1941, the course of history was changed in more than military matters. On that day, the presidents of Harvard, Princeton, and Yale were attending a routine meeting to discuss College Board affairs. When they learned about the attack on Pearl Harbor, they realized that many young men would be called into active duty. Consequently, there would not be time to administer the traditional written examinations to those who might want to take them; this was exactly the argument that the SAT proponents had made. So they agreed to drop the College Board's written examinations and to offer, instead, the SAT and multiple-choice

achievement tests. After 41 years of continuous service to American education, the reign of examinations written, read, and graded by teachers was over. One can not help but notice the irony of Binet's norming of students was at the behest of what teachers called normal. The events of December 7, 1941, influenced education in a great way and most have no idea that was the day that teachers lost some control over what students learn. The day of the multiple-choice objective test, technically valid and reliable, psychometrically sound, and machine-scored, had arrived (Ravitch, 2006).

Although some educators hoped that the change would be strictly a wartime measure, it was not. According to Ravitch (2006), those who admired the Boards valued them because of their clear educational purpose, their emphasis on writing and lucid expression, and the support they gave to a strong curriculum. The admirers did not say that multiple-choice questions could adequately replace essays in which students were expected to demonstrate knowledge of subject. Some educators admitted that multiple choice questions only measure what students do not know on a given day. They espoused that the essay as a more accurate way of exposing what the students do know. The College Board did not return to govern the setting of standards as had been the practice. As a matter of fact, the Board insisted that they were not to influence what is taught in schools. They acquiesced to the SAT and even started the Educational Testing Service (ETS) to oversee the now impartial assessment of student abilities (Ravitch, 2006).

This appears to be changing now. Lewin (2002) noted the SAT did return to an essay test on March 12, 2005, and established partnerships between high school teachers and college professors to grade the essays. This seemed to signal a trend toward the SAT being an achievement test, as opposed to an intelligence test. Consider what has taken place in California and the University of California's threats at using an alternative form of assessment to the SAT. The College Board had been rethinking the SAT I, the newest version of the SAT, for some years as more colleges, including Bates, Bowdoin, and Mount Holyoke, have dropped it from their requirements. However, according to Lewin (2002), the board was galvanized when Richard C. Atkinson, president of the University of California, proposed replacing the SAT I

with a new test that would more closely reflect the state's high school curriculum.

This seemed to hasten the College Board's reaction to pressure to change. As Lewin (2002) noted, Gaston Caperton, the president of the College Board, said the revised test would likely require students to provide a handwritten short essay and multiple-choice writing questions along with more advanced math problems based less on aptitude and reasoning and more on problem-solving learned in second year algebra or perhaps trigonometry. Currently, the exam's math problems cover only arithmetic, first-year algebra, and geometry. Caperton indicated that the writing was a new thing but something that had been recommended to be added as far back as 1993 (as cited in Lewin, 2002). As cited in Lewin (2002), Caperton added:

> Analogies have analytical thinking that is very important, but some people feel that reading comprehension can measure the same kind of intellectual skill, and maybe in a fairer way. Reading is more consistent with what people are learning in school, and more connected to the curriculum. (p. 1)

According to Lewin (2002), Caperton noted that what was learned in the classroom should be critically important to how one did on the test. This sounded much like the opinion of Ravitch (2006) about the standard-bearers of yore. Caperton went on to say, "That should help focus people on improving the classroom, making it more and more clear that the issue is not the test, it's an unequal educational system" (as cited in Lewin, 2002, p. 1). One can not help but surmise that people like Yerkes, Terman, and Binet would think negatively about testing to expected standards of learning, rather than one's aptitudes or intelligence.

Lemann, whose 1999 book, *The Big Test*, traced the rise of the SAT, said he was heartened by the proposed shift not because it would do much to change the system of admissions to the nation's most selective colleges, but because it might help improve education for millions of high school students. According to Frontline PBS (1999c), Lemann observed:

> A switch in the test will not change the composition of the freshman class at Harvard very much, but given the nature of our society, where everyone wants to be someone, the main college-admission test ends up being an organizing principle for much of American high school life. So if the test would now be billed as an achievement test, and you could tell people that the way to do better on it is to learn what's being taught in the classroom, it's a lever to improve the teaching at schools with systemically low scores, and a much healthier signal than a curriculum-free aptitude test. (p. 1)

This refrain is exactly what the advocates of the Tennessee Diploma were singing in 2008: Let's make college attainable to everyone.

While Caperton (as cited in Lewin, 2002) played down the extent of the likely changes, his description of the goals of the process reflected a profound change, turning what was once deemed an aptitude or intelligence test (until 1990, the SAT stood for Scholastic Aptitude Test) into an achievement test designed to measure what is actually learned in the classroom.

So the SAT, historically an intelligence test, is now morphing into an achievement test. What does that mean for other major testing companies as we embark toward the American Diploma? The American College Testing program was established in 1959 with the first test being given in the fall of 1959. The concept for the American College Testing program emerged in the 1950s, and the organization itself was founded in 1959. At that time, U.S. political and demographic developments were inspiring major changes in attitudes about, and approaches to, higher education. Prior to 1959, there was just one major national college-entrance testing program, and it focused on identifying the most academically able students for admission to the nation's selective universities. The remainder of college students were admitted either on the basis of scores earned on entrance exams offered by individual states or colleges or on the basis of family ties (ACT, 2008a).

In the late 1950s, large numbers of students were approaching college age and wanted to attend college. Financial aid to students was increasing, and most colleges wanted to expand their enrollments. It was in this environment that ACT's founders established The American College Testing Program and it became known as ACT in 1996 (ACT, 2008b). The ACT test is not an intelligence test. It is an achievement test that measures what students have learned in the areas of reading, math, English, and science. The ACT also gives an optional writing test for its takers. The organization is a not-for-profit one whose mission is "helping people achieve education and workplace success" (ACT, 2008a, n. p.). The influence of ACT is apparent in many areas of education including the partnership that ACT shares with the National Institute for Automotive Excellence (ASE), the organization that certifies automotive specialists throughout

America (ACT, 2008b).

From these analyses of the evolution of testing in the United States, it is apparent that the ACT and the SAT are here to stay. They are playing an increasingly important part in the high school curriculums and are even used as accountability measures in some states such as Maine (Gendren, 2007). We know where the current status of the testing companies stands. Testing is dominating our educational landscape under the concept of accountability. There must be some philosophical underpinnings that have led to our current status concerning testing in public schools in the United States.

American Meritocracy

The testing movement, as envisioned by Harvard president Conant (1943), would lead to a society in which only those students who "merited" entrance into the prestigious universities would be allowed. This would be a change from the entrance practices at places such as Columbia University and Harvard, where students were usually Caucasians and from families of great wealth. Conant visualized a new group of students composed of those who might not have the advantages of privilege and wealth. These students would take the College Board examinations now called the SAT, and on the basis of merit receive the coveted entrance into the Ivy League schools and other prestigious institutions of higher education. This meritocracy would emerge replacing the aristocracy of wealth that had existed for years in those institutions, and consequently in the "have's" of society. His vision would prove to be true but not in all the ways he envisioned.

Although today's high school seniors might find it hard to believe, Harvard, Yale, and other leading universities were not exactly bastions of the best and brightest before World War II. They educated primarily the progeny of the upper class--White, Protestant, male students from New York and New England's private schools, who were often more interested in debutante cotillions and sporting events than in the life of the mind. Many students brought servants with them to Cambridge and New Haven. Conant (1943), the president of Harvard University and one

of the most influential men of his day, wanted to replace this aristocracy of birth and wealth with what Thomas Jefferson called a "natural aristocracy" of the intellectually gifted from every walk of life who would be educated to high standards and then be given the responsibility of governing society. In *Wanted: American Radicals*, Conant had this to say about what America needed to emerge from World War II:

> No one needs to be told that the American radical will be a fanatic believer in equality. Yet it will be a peculiar North American brand of doctrine. For example, he will be quite willing in times of peace to let net salaries and earnings sail way above the $25,000 mark. He believes in equality of opportunity, not equality of rewards; but, on the other hand, he will be lusty in wielding the axe against the root of inherited privilege. To prevent the growth of a caste system, which he in abhors he will be resolute in his demand to confiscate (by constitutional methods) all property once a generation. He will demand really effective inheritance and gift taxes and the breaking up of trust funds and estates. And this point cannot be lightly pushed aside, for it is the kernel of his radical philosophy. (p. 3)

Opportunity for those who were not raised in wealth was the point that Conant (1943) emphasized. The creation of what Conant called "Jefferson's ideal," a new intellectual elite selected strictly on the basis of talent, or merit, and dedicated to public service, would, he believed, make America a more democratic country, a meritocracy if you will. These goals led him to Brigham to develop the SAT. This test would become for many students a narrow path to the best opportunities--and richest rewards--in American society.

This history of Conant's (1943) worthy goals is important and timely. A college education is fast becoming necessary to earn the middle-class salaries that workers won with less than a high school diploma in the days of America's industrial economy. The American Diploma that will weigh SAT and ACT scores so heavily in college entrance is strikingly similar to the era under which Conant established the original SAT. The rise of teenage Internet entrepreneurs notwithstanding, selective colleges and universities represent the way to the top of American society for the majority of those students who are accepted. According to Lemann (1999), these select institutions educate a disproportionate number of the nation's corporate lawyers, investment bankers, leading doctors, and influential academics, and they rely heavily on SAT scores in the admissions process. Although they do admit some students with low scores, these

are emphatically the exception. In telling the story of the people and events that shaped the postwar American meritocracy Lemann, a staff writer at *The New Yorker* and a contributing editor of *The Washington Monthly*, has given us valuable new points of reference with which to consider the role of the SAT in college admissions, affirmative action, and school reform.

Conant (1943) selected the SAT, which he said he believed to be a mental, or intelligence test, over achievement tests, created by the developer of the New York Regents exams to measure a student's grasp of course content. As Lemann (1999) pointed out, achievement tests favored unexceptional rich boys (girls were not part of Conant's meritocratic equation) whose parents could buy them top-flight high school instruction.

There was no national debate over Conant's (1943) drive to create an education-based meritocracy or to make education the official repository of opportunity in America as it is today. Conant achieved his coup with the help of a handful of close colleagues. Ironically, they were all members of what Lemann (1999) neatly termed the "Episcopacy," the social class whose defining institutions were the Protestant Episcopal Church, country clubs, New England boarding schools, Ivy League colleges, and, in their working lives, investment banks, major foundations, the foreign service, and university faculties--the very same crowd whose duller members Conant was trying exclude. Key among them was Henry Chauncey, a square-jawed Harvard assistant dean and descendent of Puritan clergy who would later serve as the founding president of the Educational Testing Service. Another was Devereaux Josephs, a classmate of Chauncey's at both the Groton School and Harvard who, as the President of the Carnegie Foundation, funded the creation of ETS for Conant. Together, they substantially redefined the nature of and route to success in America (Lemann). Lemann wrote, "It was like a slow-motion, invisible constitutional convention whose result would determine the American social structure" (as cited in Toch, 1999, p. 1).

After Harvard deployed the multiple-choice SAT successfully in pursuit of talent worth subsidizing with scholarships, Conant (1943) convinced other Ivy League schools to use it. When the essay exams that the Ivies used to test regular applicants were suspended during World

War II and replaced with the SAT, the test's influence expanded. When Conant's advocacy of a new national testing agency culminated with the opening of ETS in Princeton in 1948, his vision of a national test-based meritocracy was assured of becoming a reality.

Lemann (1999) recounted events in riveting detail. Lemann informed with stories of fascinating characters like Reynold Johnson, a young high school science teacher in Ironwood, Michigan, whose 1931 experiments led to the electrical devices that quickly score multiple-choice tests--this was a key catalyst to the rise of a national testing industry. Another character was Stanley Kaplan, the Brooklyn-born son of a plumber and a secretary who by happenstance launched today's vast SAT test-prep industry.

According to Lemann (1999), Kaplan had resorted to helping neighborhood students with their schoolwork to support himself after failing to get a place in medical school, even after graduating from City University at the top of his class at age 17. One day in 1946, a student asked him to help her with a test he had never heard of, and the rest is history.

Lemann's (1999) reporting has also yielded a big scoop. When Congress passed the Civil Rights Act of 1965, it ordered the federal Office of Education to study the educational status of Black students. The study's lead investigator was Coleman (1990), a University of Chicago sociologist, who concluded in a now-famous 1966 report that student performance was much more heavily influenced by families than by schools. Lemann revealed that ETS administered the tests on which Coleman based his conclusions and that ETS analysts largely rejected Coleman's interpretation of the results. School quality, they concluded, had a much larger influence on student achievement than Coleman acknowledged. They maintained that spending money to fix Black schools was a smart investment. But they did not argue their perspective publicly and Coleman's conclusion--that spending money on schools was not a smart way to raise Black student achievement--dominated the national education debate for the next two decades.

Subsequent research proved the ETS researchers correct; school quality influences student achievement more than Coleman (1990) acknowledged. But it was not until the

publication of *A Nation at Risk* and other reform reports in the mid-1980s that the nation began to agree that it was worth making a major effort to improve public schools.

To Lemann (1999), Conant's (1943) meritocracy has been a decidedly mixed blessing. It had certainly produced opportunities for millions of gifted students who would not have had opportunities by virtue of birth. Lemann noted that among the very first group of 10 Harvard National Scholars graduating in 1938 was James Tobin, the son of the sports-information director at the University of Illinois and a senior at Champaign High School, who would later win the Nobel Prize in Economics. In more recent years, Asian students have benefited tremendously from the SAT.

But Conant's (1943) vision of a governing elite selected through a new, education-based system and devoted to public service in a largely classless society was hopelessly naive. Not surprisingly, the new educated aristocracy has embraced the trappings of its newfound social superiority. Today's educated elite are disproportionately lawyers, bankers, and doctors, not the dedicated, European-style civil servants that Conant had hoped for. As Lemann (1999) said, the American meritocracy has become largely "a means of handing out economic rewards to a fortunate few" (as cited in Toch, p. 3).

Much more troubling is the perverse influence the SAT has had on the nation's elementary and secondary education system. Adapted by Carl Brigham, a Princeton psychology professor, from crude intelligence tests used to sort U.S. Army recruits in World War I, the SAT was a multiple-choice exam emphasizing word recognition (as is the test's verbal section today; the math section measures students' ability to reason mathematically and requires knowledge of basic arithmetic, geometry, and algebra) Lemann (1999). Lemann, however, revealed that as early as 1934, Brigham repudiated the basic premise that the tests measured solely native intelligence by recording, "The test scores very definitely are a composite including schooling, family background, familiarity with English, and everything else, relevant and irrelevant" (as cited in Toch, 1999, p. 3) that Brigham wrote in an unpublished manuscript Lemann dug out of the ETS archives. ETS and the College Board, the organization of schools and colleges that

sponsors the exam, acknowledged as much in 1994 when they finally changed the exam's name from Scholastic Aptitude Test to Scholastic Assessment Test (Toch).

Internal opposition to the SAT did not subside as the test's influence spread rapidly in the decades after Brigham's change of mind. In the 1960s, a researcher at the College Board who would later become ETS's senior expert on the technical aspects of testing, argued in a report titled "Criticisms of Testing: Background Papers" that colleges should use the SAT and other ETS tests for placement rather than selection. After the report had been printed, the entire press run was shredded--on whose orders, the author, Win Manning, never learned (Toch, 1999, p. 3).

By 1990 Manning was at ETS arguing that ETS should take steps to reduce affluent students' advantage on the SAT, according to Lemann (1999). Knowing that students from disadvantaged families tended to score lower on the test, he proposed comparing their actual scores to the scores they would be expected to achieve given their family backgrounds. He based this on the premise that students who greatly outperformed their class background on the test could be expected to do so in college as well. Manning argued that his idea would align the SAT more closely with Conant's original aim (Lemann).

According to Lemann (1999), college officials loved the idea. They saw Manning's new index as a way of diversifying their campuses without running afoul of the Supreme Court's Bakke decision in 1978: *Regents of the University of California v. Bakke* on racial quotas. Lemann recorded, however, that ETS's second ranking official and now the organization's president, Nancy Cole, responded by cutting off Manning's funding:

> Imagine the hell that would break loose if the idea were instituted and every lawyer's and doctor's kid in America got an envelope in the mail containing a score that had been adjusted downward to account for the parents' high socioeconomic status? (p. 3)

Just such a controversy did break out recently, in the wake of press reports that Manning's idea was rekindled within ETS. Almost immediately, the College Board's president attacked the so-called "Strivers" initiative with vague language about the importance of preserving the "art" of using SAT scores in admissions (Toch).

How meritocratic, then, is a test that measures neither innate ability nor course-specific

knowledge? The rise of a lucrative test-preparation industry built on families' willingness to pay thousands of dollars for courses that boost SAT scores has suggested the answer. Lemann (1999) pointed out, "The very privileged denizens of Park Avenue that Conant thought he was stripping of advantage [are] now trying like mad to manipulate testing and admissions on behalf of their children, and [are] having quite a good deal of success" (p. 4).

The lingering but false notion that the SAT measures native ability also has undercut teachers' and students' belief in the importance of hard work in schools. Indeed, much of what is measured on the test's verbal section is easily learned outside of school. Asian education systems, in contrast, are built on the belief that achievement comes from hard work rather than innate capacity. So, working closely with parents, Asian schools push all students, and not surprisingly, average performance is higher there than in American schools (Toch, 1999).

What then is the best way to achieve Conant's aim of lifting students from disadvantaged backgrounds into the meritocracy's jet stream so that the nation can reduce its reliance on affirmative-action measures such as ETS's Strivers' scheme? A logical step would be to replace the SAT with high school end-of-course exams based on rigorous state or national curricula. This is exactly what the American Diploma is proposing to do. Lemann (1999) concurred, stating, "Test-prep should consist of mastering the high school curriculum not learning tricks to outwit multiple-choice aptitude tests" (p. 4).

Tennessee Accountability System and TVAAS

A major aspect of accountability in Tennessee has been the system of student growth called TVAAS. This system is a measure of teacher effectiveness and has been studied, implemented, and copied all across the country. What began as a measure of accountability in the state of Tennessee has now spread to at least nine states for measuring school and system AYP. What is value-added and how did it get its start in this state? Ironically enough, the movement started with a lawsuit evolving from poor rural school systems that maintained they had inequitable and inadequate funding for the students in their school systems.

In the early 1990s, Tennessee Governor Ned McWherter wanted the legislature to reform the financing of public education in the state. His first idea, which floundered, was the creation of a state income tax. He finally settled for a half-cent increase in the state sales tax to bring many county systems into line with the rest of the state (*Educational Improvement Act*, 1992). Tennessee had no state income tax and was dependent on sales and use taxes and property taxes to fund public education.

Tennessee's system of funding with sales tax was found to be inadequate and inequitable by the state Supreme Court in *Tennessee Small School Systems (TSSS) v. McWherter*, 1993). The state is not wealthy; it has rural counties with child poverty rates among the highest in the nation. For example, Hancock County's child poverty rate was 49.9% in 1990 (U.S. Census Bureau, 2003). Hancock County was used as an example in a small school system's lawsuit against the state; a subsequent ruling that the state's method of funding education was unconstitutional paved the way for the Better Education Program (BEP) favored by McWherter. This lawsuit, dubbed "Robin Hood," was the first of several along the same principle of inequitable funding for smaller school systems. The Robin Hood lawsuit pitted small counties and school systems against the state.

In 1991, a trial court ruled in favor of the TSSS and declared that Tennessee school funding was in violation of the state constitution. The Tennessee General Assembly was assigned responsibility for the reform of school funding before June 30, 1992. An appeal was filed by the state in 1991, and in 1992, the Appeals Court reversed the trial court. The TSSS requested that the Tennessee Supreme Court review the case. The loss of this lawsuit by the state helped Governor McWherter drive reform in the state educational system, and as part of the reform enactment of a component of school accountability called the Tennessee Value-Added Assessment System (Dorn, 2001).

Along with the financing reform came a broad array of other efforts to improve education in an omnibus bill passed by the Tennessee legislature in 1992. One such measure was the creation of a statistical system for measuring student gains on achievement tests from 1 year to

the next, which was called the Tennessee Value Added Assessment System (TVAAS). Several legislators wrote the TVAAS into the bill based on advice from University of Tennessee statistician William Sanders, who had been testing a small version using data from a few Tennessee cities in the 1980s (Dorn, 2001).

TVAAS began reporting scores for systems and schools. The creation of a mechanistic system for producing effect sizes for individual schools--and eventually individual teachers--struck some as the epitome of distrust of teachers (Dorn, 2001). The hysteria among newspapers to publish the scores and then derive school rankings also struck many as the wrong way to make schools accountable, by some technocratic mechanism. According to Dorn, the critics wanted teachers to ask the hard questions, every day, of how to help students. Giving them abstract scores would neither help them nor encourage them to ask the hard questions. It creates, instead, a very high stakes environment that makes many teachers defensive.

Prompted by questions about TVAAS by teachers and administrators, the state comptroller investigated some of the results and suggested an external, independent review. Bock and Wolfe (1996) concluded that the basic statistical system was sound, but that estimates of school effects could vary widely, and that some of the tests used for TVAAS had too few items for reliability. These researches suggested that the use of teacher scores should wait until the state could verify that teacher scores confirmed principals' and other administrators' judgments of excellent and poor teachers. The state delayed the use of teacher scores for evaluation until the state had more research, and according to Bock and Wolfe, in a discussion at the 1996 American Educational Research Association meeting, the state was to add more questions to the social studies and science tests. Fisher (1996) also criticized the politics of TVAAS; however, no evidence was found in literature that his criticism of TVAAS made any difference in Tennessee.

In early December, 1996, some legislators said they wanted to dismantle TVAAS, but some business leaders were making clear that their support of educational reform depended on the maintenance of TVAAS. Dorn (2001) recorded that in the spring of 1997, Education

Commissioner Walters recommended eliminating all but 10th-grade subject tests in high school and making the 2nd grade tests voluntary, as well as scaling back the probationary measures for school systems. The legislature eventually passed amendments that delayed the subject tests in high school for at least a year, made the 2nd grade tests voluntary (with the promise to replace them with diagnostic tests in reading and other basic skills), and postponed formal probation for school systems until a year after being put on notice that their performance was inadequate. This was prior to the passage of *No Child Left Behind*. Since the inception of TVAAS in Tennessee, several states have implemented the system or a similar system to Tennessee's Sanders' (1998) Model.

Although there were several different value-added models in use, only the Sanders (1998) model originally mandated for use statewide: in Tennessee, since 1992, and most recently in Pennsylvania and Ohio, as well as in over 300 other school districts in 21 states. Under the value-added approach, test scores are projected for students and then compared to the scores they actually achieve at the end of the school year. Classroom scores that exceed projected values indicate effective instruction. Conversely, scores that are mostly below projections suggest that the instruction was ineffective. At the same time, this approach considers student factors such as the pattern of prior test scores, both those of the individual student as well as those of other students in the same class. If a student's present performance is below projected scores, while students with comparable previous academic history in the same classes have done well, this is evidence of the student effect--external variables such as the home environment--that is outside the control of teachers and schools (Sanders).

Because students' projected scores are based only on their prior academic records rather than on race or socioeconomic background, value-added does not introduce bias: in other words, low-income children are not expected to do poorly and high-income students are not expected to do well. Because value-added traces the same students over time, thus accounting for family and neighborhood characteristics that so strongly bias absolute test scores, educators are not being penalized for circumstances beyond their control according to advocates of the Sanders model

(Hershberg, 2005).

When value-added scores are collected for each classroom and averaged over 3 years, teachers have rich diagnostic information to improve their instruction and administrators have an empirical basis for evaluating teachers' effectiveness. When these classroom scores are aggregated over entire buildings and districts, principals and superintendents can be held accountable for students' learning results.

In 1992, Tennessee became the first state to adopt a value-added model statewide, the TVAAS. This system remains the best known, most detailed, and most statistically sophisticated example of implementing value-added assessment. The primary developer of TVAAS was Sanders (1998), formerly a professor of agriculture at the University of Tennessee. In the early 1980s, Tennessee was examining the possibility of awarding merit pay for teachers. In response to statements that it was impossible to evaluate teachers fairly based on student achievement, Sanders and a colleague theorized that a statistical model developed in agriculture (mixed model) could be used to discover how much a teacher's class had learned. They gained permission to examine 3 years of test data from the Knox County schools and found that by examining student growth rather than absolute test scores, and correlating data by classrooms, they could estimate teacher effectiveness in ways that were consistent from year to year and fit with the subjective impressions of school administrators (Sanders & Horn, 1994). Despite these findings, the study failed to attract much attention at the time. However, in 1992, and after the state had lost the small schools lawsuit, the Tennessee legislature undertook another round of education reform, one that would require raising taxes. Business interests were demanding that accountability for districts, schools, and teachers become part of the package (Ceperly & Reel, 1997). This time legislators were attracted to Sanders' proposal as a way to verify results. After inviting Sanders to speak, legislators amended the state's Educational Improvement Act to incorporate the Sanders' Model. Schools and systems are expected to have a mean gain in student learning that would meet or exceed the national mean gain. As of 1995, data were analyzed at the teacher level and used in teacher evaluations (Cepperly & Reel). Through 1997, Tennessee used data

from the CTBS/4 test by CTB/McGraw Hill, testing second through eighth graders. Since 1998, Tennessee has tested third through eighth graders using the *Terra Nova* test by CTB/McGraw Hill. *Terra Nova* is a nationally available test that uses both multiple choice and constructed response questions and provides both norm and criterion referenced results. Students are tested in reading, math, language, social studies, and science. System and school scores, expressed as an average of the last 3 years' gains, are made public. Scores are expressed as a percentage: a score of 100% reflects normal gains. The evaluation system for secondary schools has been developed, and currently includes three end-of-course tests for math, an English I end-of-course test, physical science end-of-course test, and a U. S. history end-of-course test. Three Gateway tests are given in Algebra I, English II, and Biology I for *No Child Left Behind* AYP purposes (Tennessee Department of Education, 2007a). A writing assessment is also given in English III. The state provides, as well, the opportunity to take the ACT free of charge. Actual student scores are compared with predicted scores based on their *Terra Nova* scores in earlier grades.

Another component of TVAAS and value added in general has been the growth model. The *No Child Left Behind Act* requires existing teachers to demonstrate competency in all core academic subject areas via a highly objective uniform state standard of evaluation (HOUSSE). One option for complying with this requirement is by using the *Teacher Effect Data*, a statistical means of estimating the teacher's impact (effect), or lack of impact on student achievement or learning, which is a component of TVAAS. The analysis of teacher effect data uses 3-year average gain comparisons: teacher vs. norm, teacher vs. state, and teacher vs. system as an estimated measure of the teacher's effect on student learning according to the accountability workbook adopted by the United States Department of Education (Tennessee Department of Education, 2001).

In November 2005, Secretary of Education Margaret Spellings announced a new pilot program that would allow selected states to use growth models to determine if their schools and districts were meeting *No Child Left Behind* performance targets. Tennessee was one of two original states whose proposed growth model was accepted by the U.S. Department of

Education. On May 17, 2006, Spellings formally announced approval of two high-quality growth model pilots that followed the principles of *No Child Left Behind*. The first two states that approved and have subsequently implemented these growth models were Tennessee and North Carolina. These were first implemented as pilots in 2005-06 and are still being used for AYP purposes today according to Spellings (United States Department of Education, 2006). The United States Department of Education (2006) recorded Spellings as saying:

> A growth model is a way for states that are already raising achievement and following the bright-line principles of the law to strengthen accountability," Secretary Spellings said. "North Carolina and Tennessee were recognized by our impressive group of peer reviewers to have written strong growth models that adhere to the core principles of No Child Left Behind." Numerous other states were denied their proposals to implement their own particular "growth model" plans.
>
> There are many different routes for states to take, but they all must begin with a commitment to annual assessment and disaggregation of data. Additionally, they all must lead to closing the achievement gap and every student reaching grade level by 2014. We are open to new ideas, but when it comes to accountability, we are not taking our eye off the ball. (n. p.)

Six other states were not approved for using the growth model in 2006 but have been subsequently approved according the United States Department of Education (2007).

According to the United States Department of Education (2007), Secretary Spellings said, "I believe that extending the growth model pilot for the 2007-2008 school-year will promote two important goals. It will allow states another effective way of measuring adequate yearly progress (AYP) by measuring individual student growth over time, and it will continue to expand the flexibility available to states under *No Child Left Behind*" (p.1). The growth model pilot was established by Secretary Spellings in November 2005 and was included in the President's *No Child Left Behind* reauthorization blueprint in 2007. Nine states currently have approved growth model proposals: North Carolina, Tennessee, Delaware, Arkansas, Florida, Iowa, Ohio, Alaska, and Arizona (United States Department of Education, 2007).

Growth models track individual student achievement from one year to the next, giving schools credit for student improvement over time. The pilot program enables the Department to rigorously evaluate growth models and ensure their alignment with *No Child Left Behind* and to

share these results with other states (United States Department of Education, 2006). In Tennessee, this enables schools to avoid being targeted under requirements of *No Child Left Behind*, although not in great numbers. Olson (2006) reported in *Education Week* that growth models, via the USDOE pilot program in 2006, do not help much. Olson recorded Smith, the director of accountability for the state education department, as saying, "In Tennessee, only eight schools' achievement of AYP was attributable to the growth model. I was not surprised, it's a stringent application of the projection model, but it's always worth doing and using even if it helps one school" (p.1). It appears this trend of growth will continue with Secretary Spellings encouraging other states to send in growth model proposals as late as February 2008 (United States Department of Education, 2007).

Value-added models have staunch advocates and critics as well. The advocates see a system of value-added as a method of giving school systems a chance of showing progress when they may have been very far behind the targets set for them under *No Child Left Behind*. Value-added refers to any one of several models that are used to interpret test scores in a way that evaluates the growth or progress in a student's academic achievement over time, usually over several academic years (Rubin, Stuart, & Zanutto, 2004). The concept of an assessment that measures a student's achievement growth over several years, commonly known as longitudinal assessment, has long existed in education (Goldschmidt, Choi, & Martinez, 2004). However, value-added assessment represents an approach to evaluating student achievement growth that is distinct from traditional growth models in several respects.

A fundamental concept of value-added assessment is the assertion that schools are responsible for providing each student with the equivalent of 1 year of growth, regardless of the level of education with which the student began the academic year (Callender, 2004; Carey, 2004; Hershberg, Simon, & Lea-Kruger, 2004). If the value-added measure reflects the true effect of teachers and schools on a student, it should be possible to determine whether their contribution to the students' growth was sufficient.

Each value-added model formulates what constitutes a year of growth for a student. In

the case of TVAAS, a year of growth is estimated using each student's individual achievement history (Bianchi, 2003; Hershberg et al., 2004). At the end of the year, if the student has achieved what was computed to be 1 year of growth, the student is considered to have received an effective education. If the student shows more than 1 year of growth, the student has received a highly effective education. Students that show less than 1 year of growth are considered to have received a less effective education (Hershberg et al.).

The quality of the student's education is considered to have long term implications. Relying on the findings from independent studies of TVAAS, supporters of value-added assessment claim that teacher effectiveness is far more important in a student's current and future achievement than is any other non-educational factor (Bianchi, 2003; McCaffrey, Lockwood, Koretz, & Hamilton, 2003). This assertion would seem to dispute many findings of the Coleman Report that suggested students' peers and home life play an integral part of their educational achievement (Bracey, 2004) as well as findings that ethnicity and socioeconomics may play a role in the grades on TVAAS (Webb, 2005). Supporters of value-added also asserted that, at later grade levels, students who receive less effective educations will continue to demonstrate low levels of achievement while students who receive effective educations will attain higher levels of achievement.

Proponents of value-added assessment proposed that these new types of data will make it possible to evaluate and compare the quality of schools that have widely different student populations (Meyer, 1996). This assertion is a marked departure from conclusions drawn from studies, particularly the well-known "Coleman Report" that claimed to show that a student's socioeconomic status and demographics, rather than teachers and schools, had the most impact on his or her achievement (Bracey, 2004). Since the publication of those studies, the evaluation of schools has typically relied on measures of educational inputs, such as funding and teacher certifications, rather than test results (Meyer).

Some educational researchers have pointed to value-added assessment as supporting the view that teachers play the dominant role in student achievement growth (Vaughan 2002).

However, in the new era of accountability led by the *No Child Left Behind Act*, educational researchers have supported the view that instruction has a real impact on students that can be measured with standardized assessments (Archer, 1999; Marzano, Pickering, & Pollock, 2001). Such conclusions support anecdotal evidence from parents and other stakeholders who observe a wide range of teacher effectiveness (McCaffrey et al., 2003).

Commentators foresee potential for using a value-added assessment system to guide large-scale and small-scale educational reform. The principle expectation for the value-added model is that the results will conclusively determine the impact that educators and education policies are having on their students (Drury & Doran, 2003; McCaffrey et al., 2003). With this evidence, effective teachers and education policies can be identified, and reform based on these findings can be instituted in education systems (Carey, 2004; Crane, 2002). For example, results from a value-added system could be used to transfer effective teachers to schools where they are needed (through financial incentives or other means), study the instructional practices of effective teachers, and offer professional development to less effective teachers (Carey; Drury and Doran; Hershberg et al., 2004). Other, more controversial, suggestions include holding educators accountable for student growth, such that highly effective teachers receive financial incentives and professional advancement while consistently ineffective teachers are sanctioned (Summers, 2002).

Accountability systems based on value-added assessment have already been instituted in some states. For example, school districts in Chattanooga, Tennessee, use TVAAS to identify and attract highly effective teachers using salary bonuses, housing benefits, and funding for graduate education (Carey, 2004). Indeed, Hamilton County's differentiated pay plan has been used as a model for Tennessee's State Board of Education to pass legislation mandating all school systems to implement a differentiated pay plan for hard to fill areas or hard to fill schools, meaning subjects or schools that have had a hard time hiring in certain areas. As most education systems rely on seniority to determine teacher transfers and salary schedules, implementing a reward or evaluation system based on value-added assessment data might represent a significant

change in policy (Carey, 2004; Drury & Doran, 2004).

Looking past the traditional pay system has paid off for some already. Now in its 7th year of offering bonuses to experienced teachers to transfer to struggling schools, the Hamilton County school district including urban Chattanooga has seen students' scores soar in their neediest schools (Delisio, 2004). Delisio recorded the superintendent of schools, Jesse Register, as saying in 2004, "We have seen statistically significant changes; the urban schools are catching up" (p. 1).

According to Delisio (2004), Gerry Dowler, who coordinates the national, state, and local teachers' unions in Tennessee, said Hamilton County is experiencing startling results in student achievement: Reading and math scores jumped an average of 10% to 12% in a year in nine priority schools since the influx of new teachers.

Under the incentive program, high-performing teachers can receive an additional $5,000 a year for teaching in low-performing schools, and principals receive $10,000. Teachers qualify for the program based on the value-added data system Tennessee uses to evaluate teachers. According to Delisio (2004), the system involves reviewing students' achievement at the beginning and end of the year by using a variety of factors to determine a teacher's effect on student growth. Administrators review the data to identify teachers who make the greatest gains with students.

One may wonder if the system of differentiated pay works. Delisio (2004) wrote, “Before the incentives, between 70% to 80% of the staff turned over in priority schools every year. We knew we would never reduce the achievement gap if we did not have stable staff" (p. 1). In the past, teachers in city schools got tenure and moved to the suburbs, leaving the inner city schools with hard-to-fill vacancies. Hamilton County merged with the Chattanooga district in 1997; whereas Hamilton County schools are suburban with a mostly white population, the Chattanooga schools are urban with a high minority enrollment (Delisio).

Hamilton County’s success was noted throughout the state. As chief negotiator of McMinn County, I was able to hear Hamilton County’s chief negotiator and its teachers’ union

chief negotiator speak in conjunction 3 years ago in support of the differentiated pay plan concept. The Tennessee Educator's Association (TEA) has not come to support this effort at this time. In their May, 2006 newsletter, the Tennessee Education Association (2006) endorsed the use of a differentiated pay plan with certain caveats. Namely, the use of a value-added component was to be excluded as a requirement for the teachers' union's support. The union stated in its newsletter, "Any additional compensation beyond the single salary schedule must not be based upon individual evaluations, student standardized test scores or value-added gain scores" (Tennessee Education Association, 2006, p.1). This represented a continuation of concerns mentioned a year earlier (Tennessee Education Association, 2005). The value-added concept was growing in acceptance, however. Clearly, the teachers' association said that the value-added component was not acceptable by 2006 under certain conditions as a means of differentiated pay. The Tennessee State Board of Education (2007), however, apparently was impressed enough with their success in Hamilton County that they are now requiring every system in the state to implement a differentiated pay plan.

Clearly, growth models such as Tennessee's Value-Added Assessment System have come of age. Although there have been many concerns about value-added, the concept has a great deal of promise. Sanders (1998) had a great deal of concern about models that are being implemented under the umbrella of value-added that lack the creditability that his Tennessee model possesses. As the concept widens across America, more research might occur and the product might continue to be strengthened.

Tennessee Education Lottery Scholarship Program

Tennessee has a long history of struggles to equitably support its kindergarten- through 12th-grade instructional program. Governor McWherter had advocated both a state income tax as well as a state lottery program. The citizens of Tennessee actually had a chance to vote for a para-mutual betting statute that would have allowed bingo, betting on horse racing, and the lottery. The proposal was defeated and remained in the background until the budget crisis of

early 2000 under Governor Don Sundquist. The likelihood of the lottery became a reality when it was passed 57% to 42% in November 2002 (Tennessee Student Assistance, 2004). The primary purpose of this legislation was for the creation of lottery-based scholarships for students to go to college or postsecondary technical schools.

The Tennessee Education Lottery Scholarship Program (2007) was intended to provide financial awards to offset costs associated with pursuing postsecondary education. There are several different types of TELS including the Tennessee HOPE Scholarship Award, Tennessee ASPIRE supplemental award, General Assembly Merit Scholarship supplemental award, Tennessee HOPE Access Grant award, and the Wilder-Naifeh Technical Skills Grant (Tennessee Education Lottery Scholarship Program, 2007).

Recipients of any TELS award as provided by these rules, except for the Tennessee Dual Enrollment Grant and the Wilder-Naifeh Technical Skills Grant could enroll as a full-time or part-time student at any eligible postsecondary institution. The amount of the award for part-time students shall be based on the hours attempted. Students enrolled in 6, 7, or 8 hours receive half of the award that full-time students receive. Students enrolled in 9, 10, or 11 hours receive three quarters of the award that a full-time student receives (Tennessee Education Lottery Scholarship Program, 2007).

The receipt of a Tennessee HOPE Scholarship, Tennessee HOPE Access Grant, Tennessee ASPIRE Award, Tennessee HOPE Foster Child Grant, General Assembly Merit Scholarship (GAMS), or Tennessee Dual Enrollment grant is contingent upon admission and enrollment at an eligible postsecondary institution (Tennessee Education Lottery Scholarship Program, 2007).

The Free Application for Federal Student Aid (FAFSA) or the Renewal FAFSA as authorized by the U. S. Department of Education to indicate eligibility for federal and state financial aid programs shall be the application for all first-year TELS awards. The FAFSA is the means by which eligible students reapply for TELS awards after their initial year of eligibility. The FAFSA must be submitted by mail or electronically as directed in the FAFSA instructions

(Tennessee Education Lottery Scholarship Program, 2007).

Eligibility for the Hope Scholarship depends on several things. To be eligible for a Tennessee HOPE Scholarship as an entering freshman, a student, who graduated from an eligible high school after December 1, 2003, upon having completed curriculum requirements of the high school for graduation, shall meet the requirements of T.C.A. § 49-4-907 (Tennessee Education Lottery Scholarship Program, 2007).

The ASPIRE award has eligibility requirements as well. Except as provided in T.C.A. § 49-4-931, any student eligible for the Tennessee HOPE Scholarship with an adjusted gross income attributable to the student that does not exceed the amount as described in the law will receive the ASPIRE award in addition to the base award (Tennessee Education Lottery Scholarship Program, 2007). The adjusted gross income attributable to the student shall be reviewed each academic year to determine continuing eligibility for the ASPIRE award. A student otherwise eligible for the Tennessee HOPE Scholarship and meeting the requirements of this rule will receive the ASPIRE award regardless of the student's eligibility for this grant in any prior year. A student eligible for both the ASPIRE award and the General Assembly Merit Scholarship will be awarded the ASPIRE award but will not simultaneously receive both awards (Tennessee Education Lottery Scholarship Program, 2007).

Qualifying for the TELS is done through ACT scores or high school grade-point averages. Entering freshmen must have a minimum of a 21 ACT (980 SAT) or an overall weighted minimum 3.0 grade point average. If a student ceases to be eligible for HOPE, except for GAMS and HOPE Access Grant, the student may regain HOPE, one time only. The state continues to track students' demographics that become eligible for the HOPE scholarships. The demographic breakdown of TELS recipients by gender, race, ethnicity, and postsecondary sector has remained steady over time with family income being the only exception. As the program continues, the percentage of students in the higher income bracket has grown. Although there might be some actual growth in students in the highest income bracket, it is also likely that inflation is pushing more students into that bracket (Tennessee Education Lottery Scholarship

Program, 2007).

Racial and gender differences regarding the level of academic preparation for a TELS award persisted among fall 2006 TELS first-time freshmen at public institutions; 55% of fall 2006 TELS first-time freshmen at public institutions met the ACT and GPA requirements; 26% met only the GPA standard, and 19% met only the ACT requirement. In addition, 59% of Caucasian awardees met both the GPA and ACT requirements compared to 35% of African American participants. African American awardees were most likely to meet the GPA requirement only (48%), and males were more likely than females were to qualify solely on the basis of ACT (26% to 13%) (Tennessee Education Lottery Scholarship Program, 2007).

It is important to note that maintaining eligibility in college or postsecondary institutions has requirements as well. To retain a TELS award authorized by the law, a student at an eligible postsecondary institution must continue to meet all applicable requirements for the scholarship and must reapply by completing the FAFSA or Renewal FAFSA as required by the statute for the applicable award for each academic year. Eligibility must also be reviewed at the end of the semester in which the student has attempted 24, 48, 72, or 96 semester hours. At the end of the semester in which the student has attempted 24 semester hours, the student must have achieved a cumulative grade point average of at least 2.75 to continue to receive the TELS award (Tennessee Education Lottery Scholarship Program, 2007). At the end of the semester in which the student has attempted 48, 72, or 96 semester hours, the student must achieve a cumulative grade point average of at least 3.0 to continue to receive the TELS award (Tennessee Education Lottery Scholarship Program, 2007). However, many students have lost their eligibility for the TELS awards.

In terms of Hope Scholarship retention for the 2nd year, some trends have existed in terms of qualification for the scholarship. Higher-income students retained the lottery scholarship at a higher rate than did their peers. Even though the programs have the same initial academic eligibility criteria, 57% of fall 2006 first-time freshmen HOPE recipients from families earning over $96,000 retained their awards into their 2nd year as compared to 42% of ASPIRE

recipients from families earning $12,000 and below. Better-prepared students retained the lottery scholarship at a higher rate than did their peers. For fall 2006 TELS first-time freshmen who qualified by meeting both the ACT and high school GPA criteria for initial eligibility, the fall 2007 scholarship retention rate was 62%. For those qualifying solely on the basis of ACT, the scholarship retention rate was 43% and for those qualifying solely on the basis of GPA, the scholarship retention rate was 40%. Among fall 2006 first-time freshmen who qualified on the basis of both high school GPA and ACT scores, scholarship retention rates for African American and Caucasian students were similar (Tennessee Education Lottery Scholarship Program, 2007).

A third of TELS recipients persist to their 4th year on the lottery scholarship. As information became available a better picture of retention rates has been drawn. Of fall 2004 TELS first-time freshmen, 50% retained their award into their 2nd year, 36% retained their award into their 3rd year, and 32% retained their award into their 4th year (Tennessee Education Lottery Scholarship Program, 2007).

Students losing scholarships is not good news, but it is interesting to note to what happens to those who have qualified for a scholarship and to those who lose their HOPE Scholarship while in college. Scholarship recipients are retained in college at a higher rate than are their peers. Of the fall 2004 TELS first-time freshmen, 65% were retained in college through their 4th year as compared to 52% of all students (Tennessee Education Lottery Scholarship Program, 2007).

The TELS program has likely induced students to attend colleges instate and has coincided with an increase in the average ACT score of incoming first-time freshmen at the University of Tennessee, Knoxville. Since the scholarship, the annual rate of growth in enrollment among Tennessee resident freshmen has accelerated at independent institutions and the UT system while decelerating at TBR universities, community colleges, and out-of-state institutions. Among recent Tennessee high school graduates who enrolled in college, the percentage choosing Tennessee institutions has increased from 82% before the lottery scholarship to 85% currently. The ACT profile of the entering freshman class has improved at

UT Knoxville, although not at other public institution types. We now have this picture of the TELS recipients and the impact that the scholarship program has had on postsecondary institutions. The scholarship qualifiers tend to come from more wealthy families and the wealthier a student's background, the more like he or she is to retain the scholarship. There are also revelations about race and gender and the qualification and retention rates of these subgroups (Tennessee Education Lottery Scholarship Program, 2007).

The 2006-07 academic year marked the 3rd year of the Tennessee Education Lottery Scholarship (TELS) program. More than 67,000 students received lottery-funded scholarships with total award allocations in excess of $191,000,000 (Tennessee Education Lottery Scholarship Program, 2008). The Dual Enrollment Grant program, which was added in 2005, has continued to grow rapidly with 8,300 high school students participating. The number of students using the Wilder-Naifeh Technical Skills Grant program dropped by 300 students from 10,023 to 9,721 from 2005-06 to 2006-07; however, the total funding for the program increased from $7.9 million to $8.1 million (Tennessee Education Lottery Scholarship Program, 2008).

It is estimated that the program will expend $233 million to serve some 78,000 students in 2007-08. Projected expenditures for 2008-09 were $238 million according to the Tennessee Education Lottery Scholarship Program (2008) annual report. According to the annual report, the program reached maturity with five classes of students in 2007-08.

The TELS program has grown steadily since its inception in 2004-05 and reached maturity in 2007-08. Monetarily, the program grew from expending $93 million in its initial year to $191 million in 2006-07 (Tennessee Education Lottery Scholarship Program, 2008). Enhanced by an additional year of freshmen students each year as well as the addition of a Dual Enrollment Grant for high school students, the number of students served grew from 40,000 in the program's inaugural year to 67,000 in 2006-07 (Tennessee Education Lottery Scholarship Program, 2008).

The demographic breakdown of TELS recipients by gender, race, ethnicity, and postsecondary sector has remained steady over time, with family income being the only

exception. As the program continues, the percentage of students in the higher income bracket has grown. Although there might be some actual growth in numbers of students in the highest income bracket, it is also likely that inflation is pushing more students into that bracket (Tennessee Education Lottery Scholarship Program, 2008).

The ACT profile of the entering freshman class has improved at UT Knoxville with a current ACT average of 25.2 as compared in 2004 ACT average of 23.9 (Tennessee Education Lottery Scholarship, p. 27, 2008), though not at other public institution types such as Memphis University.

The Tennessee Education Lottery Scholarship Program (2008) has had a tremendous impact on secondary education in Tennessee. Teachers reported they felt the pressure in their assessment practices. They have shared with me the burden of giving students a "C." Teachers understand that this grade would not meet the lottery scholarship requirement. Parents also understand that low grades reduce the likelihood that a scholarship might not be there when a student goes to college. With the adoption of the Tennessee Diploma Project, the emphasis will began to be placed on ACT scores and not simply GPAs.

CHAPTER 3

RESEARCH METHODOLOGY

Introduction

The purpose of this study was to determine whether a significant relationship exists between particular home, student, and school characteristics and ACT scores and the relationship these characteristics subsequently have with the TVAAS grades assigned to each high school's ACT scores. These home, student, and school variables were socioeconomic status, percentage of minority, graduation rate, per-pupil expenditure, Gateway English II scores, and Gateway Algebra I scores. This chapter describes the methodology and procedures used in this quantitative study to determine if a relationship exists between home, student, school characteristics, and ACT scores and the subsequent effect these predictor variables have in relation to TVAAS scores assigned to high schools in Tennessee. This chapter is organized into the following sections: population, data collection, research methodology, data analysis, research questions and null hypotheses, and summary.

Population

The population involved in this study was public high school students in Tennessee. Elementary schools, and middle schools were not part of the study because their students are not tested in the Gateway Algebra I, English II, and Biology I courses. The population included all public high schools in the state. Private schools are not subject to the accountability requirements of the Tennessee General Assembly. There were 265 high schools represented in these data. The information did not include an individual student's data, but rather contained averages of individual schools or school data. Therefore, there should not be any issues concerning individual student confidentiality. The data were comprised of school-level information except for per-pupil expenditure, which was district data. For example, the free- and

reduced-price school meals population constituted the low socioeconomic group for this study. This did not compromise any individual student but rather was a snapshot of a school's demographic makeup.

Data Collection

I contacted Connie Smith, the Executive Director of Education for Tennessee. I called Smith to discuss with her the possibility of gaining access to the high school data that are used to grade our schools and districts. After discussing my dissertation topic, she agreed to send me the high school information found on the state's report card in its entirety. This information was from the school year of 2006-07. I subsequently received the information from the Office of Accountability. The files were: Average Daily Membership, TVAAS, Graduation Rate, Dropout Cohort, NCLB status, Net Enrollment, LEA finance, Nutrition, and Teacher Credentials.

After receiving the files from the Department of Education, I began the process of merging the information into the Statistical Package of the Social Sciences (SPSS). The original files contained all the information from all public schools in Tennessee. The original file contained 2,036 schools including prekindergarten- through eighth-grade schools as well as high schools. I eliminated the elementary schools, middle schools, and other atypical schools such as vocational centers and adult high schools. I analyzed only those schools that were, in fact, public high schools that would constitute a typical high school in the state of Tennessee.

Research Methodology

Because there were no individual students involved in this study, certain portions of the Institutional Review Board process were not required. I used the Statistical Package for the Social Sciences (SPSS) for my statistical measurements. After purging inappropriate data from the State Department of Education data set, I began the process of doing the analyses. A descriptive analysis was done to obtain the minimum and maximum scores and the mean, range, and standard deviation for the 2007 ACT composite scores for 265 Tennessee High Schools.

Variables were checked for associations ACT composite scores, and each of the four ACT subtests: English, reading, math, and science. Multivariable linear regressions were computed with high schools' composite ACT scores as the dependent variable.

Data Analysis

For research question 1, I ran multiple regressions on home, student, and school characteristics of high schools that might have a relationship to ACT scores. The characteristics that were analyzed were socioeconomic status, percentage of minority students, graduation rate, per-pupil expenditure, Gateway Algebra I scores, and Gateway English II scores. These characteristics were analyzed to see which might be the best predictors of ACT scores.

After determining the best predictors of ACT scores from these analyses, I ran one way ANOVAs to identify the variables that tended to be the best predictors for answering research question 2. This question sought to determine if a correlation existed between the identified characteristics of socioeconomic status, percentage of minorities, graduation rate, per-pupil expenditures, Gateway Algebra I scores, and Gateway English II scores (based on previous analyses) and the TVAAS graded categories (Below, Above, No Detectable Difference). An ANOVA was done to determine if a correlation existed between the characteristics that were predictors and the TVAAS graded categories. These analyses would test the equity of grades given by the Tennessee Department of Education to schools throughout the state. Are these grades based on school characteristics or individual and school-wide teacher effectiveness, as TVAAS claims to be able to do?

Research Questions and Null Hypotheses

1. Which home, student, and school variables have the strongest relationship with 2007 ACT scores for Tennessee high schools?
2. Is there a relationship between the home, student, and school variables (socioeconomic status, percentage of minority, graduation rate, per-pupil expenditure,

Gateway English II scores, and Gateway Algebra I scores) with the three graded categories of TVAAS (Below, Above, No Detectable Difference)?

From research question number one, the following null hypotheses were tested:

$Ho1_1$: There is no relationship between socioeconomic status and a high school's composite ACT score.

$Ho1_2$: There is no relationship between percentage of minorities and a high school's composite ACT scores.

$Ho1_3$: There is no relationship between graduation rates and a high school's composite ACT scores.

$Ho1_4$: There is no relationship between per-pupil expenditures and high school's composite ACT scores.

$Ho1_5$: There is no relationship between Gateway English II scores and a high school's composite ACT scores.

$Ho1_6$: There is no relationship between Gateway Algebra I scores and a high school's ACT composite scores.

From research question number two, the following null hypotheses were tested.

$Ho2_1$: There is no relationship between socioeconomic status and the three graded categories of TVAAS (Below, Above, No Detectable Difference).

$Ho2_2$: There is no relationship between percentage of minorities and the three graded categories of TVAAS (Below, Above, No Detectable Difference).

$Ho2_3$: There is no relationship between graduation rate and the three graded categories of TVAAS (Below, Above, No Detectable Difference).

$Ho2_4$: There is no relationship between per-pupil expenditures and the three graded categories of TVAAS (Below, Above, No Detectable Difference).

$Ho2_5$: There is no relationship between Gateway English II scores and the three graded categories of TVAAS (Below, Above, No Detectable Difference).

$Ho2_6$: There is no relationship between Gateway Algebra I scores and the three graded categories of TVAAS (Below, Above, No Detectable Difference).

Summary

The purpose of this study was to determine whether a significant relationship exists between particular home, student, and school characteristics and ACT scores and the relationship these characteristics subsequently have with the TVAAS grades assigned to each high school's ACT scores. These home, student, and school variables were socioeconomic status, percentage of minority, graduation rate, per-pupil expenditure, Gateway English II scores, and Gateway Algebra I scores. This chapter contained a description of the population used in this study, the way the data were collected and analyzed, the research questions and null hypotheses, and a summary. By analyzing the state data, I obtained results that would reject or fail to reject the null hypotheses. The state's data were readily available through the cooperation of the Tennessee Department of Education. The following analyses might illuminate some issues for policy makers to consider as they shape the accountability system that will be used upon implementation of the new Tennessee Diploma in the 2009-10 school-year.

CHAPTER 4

ANALYSIS OF DATA

The analysis of data concentrated on home, student, and school variables most closely associated with ACT composite scores. These scores yielded predictive values to ascertain how strong a relationship these variables had with the ACT composite scores of their schools. The data were school level data. After determining the association between the home, student, and school variables and ACT composite scores, I then analyzed the correlation between the home, student, and school variables and the grades assigned to schools on the TVAAS ACT composite scores to see if a relationship also existed within the TVASS grades. The TVAAS grades assigned by the state department of education to schools and districts are supposed to factor out variables other than teacher effectiveness.

Research Question #1

Which home, student, and school variables have the strongest relationship with 2007 ACT scores for Tennessee high schools?

The following six predictor variables were used in a regression model with observed ACT composite scores as the dependent variable: (a) socioeconomic status was measured as the percentage of students who participated in the free- or reduced-price meals program, (b) percentage of minority students, (c) graduation rate, (d) per-pupil expenditure, (e) observed Gateway English II scores, and (f) observed Gateway Algebra I scores. All of the variables except for per-pupil expenditure were school level variables. Per-pupil expenditure was a school district level variable. A multiple regression model with the six predictors entered as a set was used to evaluate the following six hypotheses:

$Ho1_1$: There is no relationship between socioeconomic status and a high school's composite ACT score.

$Ho1_2$: There is no relationship between percentage of minorities and a high school's composite ACT scores.

$Ho1_3$: There is no relationship between graduation rates and a high school's composite ACT scores.

$Ho1_4$: There is no relationship between per-pupil expenditures and high school's composite ACT scores.

$Ho1_5$: There is no relationship between Gateway English II scores and a high school's composite ACT scores.

$Ho1_6$: There is no relationship between Gateway Algebra I scores and a high school's ACT composite scores.

Prior to examining the findings of the regression, preliminary analyses evaluated the appropriateness of the model by examining the residuals for violations of the assumptions of regression. First, the assumption of normality appeared to be met based on visual inspection of the histogram of standardized residuals as shown in Figure 1.

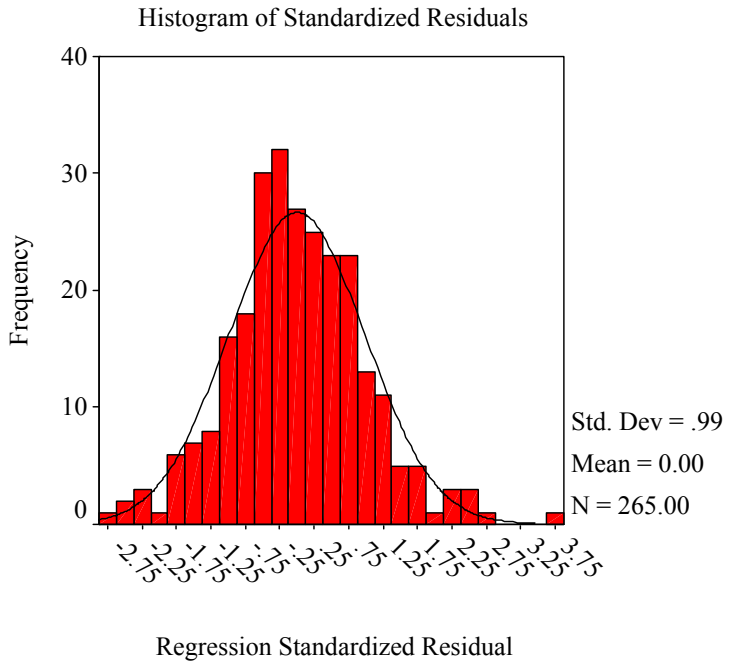


Figure 1. Histogram of the Standardized Residuals for the Regression Model

In addition, the one-sample Kolmogorov-Smirnov Test was used to test the distribution of the standardized residuals against a normal distribution. The null hypothesis that the distribution of the residuals from the regression model does not deviate from a normal distribution was retained (p = .73). Therefore, the assumption of normality was met. Second, there appeared to be no reason to question the assumption of linearity based on a visual examination of the normal probability plot (as shown in Figure 2) as indicated by the red dots falling very close to or on the green line.

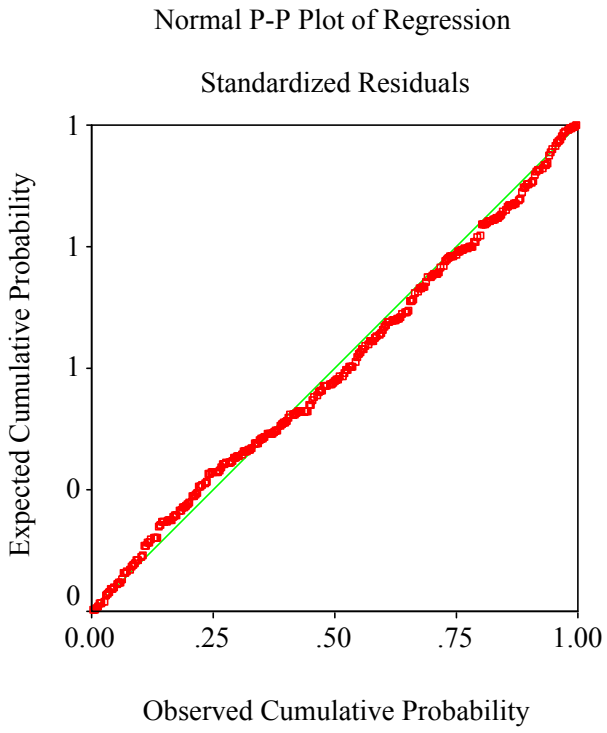


Figure 2. Normal Probability Plot of Regression Standardized Residuals

Likewise, the scatterplot of the standardized residuals and standardized predicted values as shown in Figure 3 revealed no discernible pattern of nonlinearity or unequal variances. Based

on the analyses of the residuals, I concluded that the assumptions of the regression model were met.

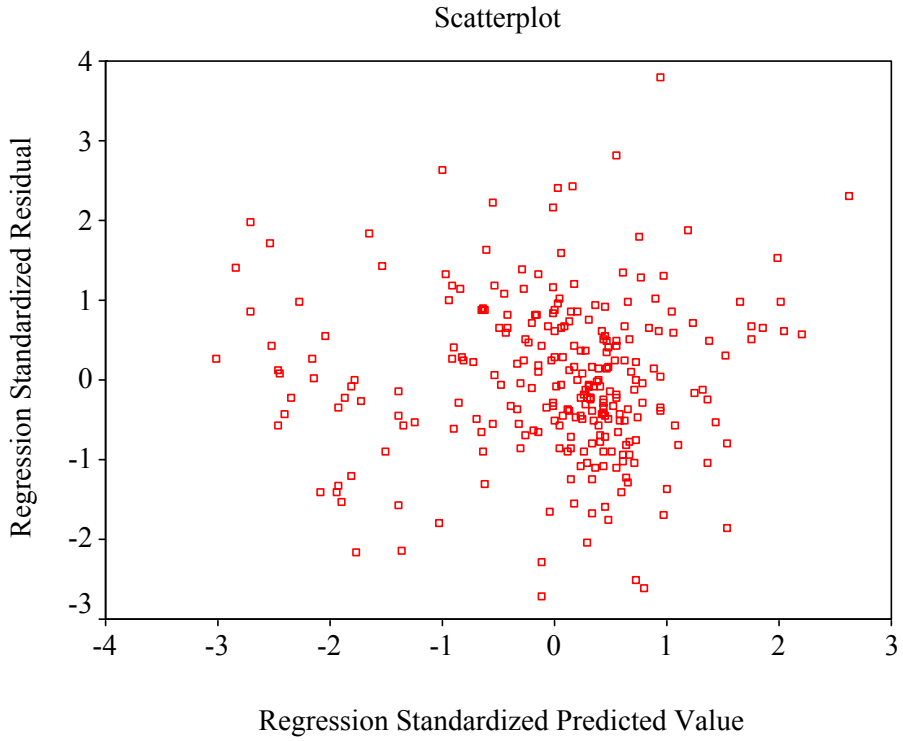


Figure 3. Scatterplot of the Standardized Residuals Regressed on the Standardized Predicted Values

Below one can see the r^2 change when a given predictor is entered in the model last representing the unique contribution to the variance in ACT composite scores accounted for by the predictor. However, small r^2 changes are, by definition, a direct consequence of the correlations among the predictors. The correlations among the predictor variables, all of which were significant, are shown in Table 1.

Table 1

Pearson Correlations Among the Predictor Variables

	SES	Percent Minority	Grad. Rate	PPE	English II
Percent Minority	.53*				
Graduation Rate	-.55*	-.62*			
Per-Pupil Expenditure	.42*	.66*	-.49*		
English II	-.67*	-.61*	.73*	-.42*	
Algebra I	-.53*	-.59*	.66*	-.51*	.72*

* Significant at the .01 level

The findings of the regression analysis showed that the six predictor variables as a set were significantly related to observed ACT composite scores, F (6, 258) = 157.18, $p < .01$. The R^2 for the model was .79 meaning that 79% of the variance in observed ACT composite scores was accounted for by the six predictors.

Table 2 presents indices to evaluate the relative strength of individual predictors. Table 2 also shows the coefficients for the regression model.

Table 2

Coefficients for the Regression Model Using ACT Composite Scores as the Dependent Variable

Predictors	*b*	*SE*	β	*t*	*p*	*r*	*Partial r*
(Constant)	-29.432	3.663		-8.04	<.01*		
Socioeconomic Status	-.015	.004	-.173	-4.30	<.01*		-.259*
Percent Minority	-.016	.003	-.247	-5.44	<.01*		-.321*
Graduation Rate	-.007	.008	-.004	-.09	.93		-.005
Per-Pupil Expenditure	.002	.000	.115	2.90	<.01*		.177*
English II	.009	.008	.620	11.70	<.01*		.589*
Algebra I	.003	.005	.031	.69	.49		.043

* Significant at the .01 level

As shown in Table 2, each of the six zero-order correlations between the predictors and ACT Composite scores was significant. However, after controlling for the other variables in the regression model, only four predictors remained significant: (a) observed Gateway English II, (b) percentage of minority students, (c) socioeconomic status measured as the percentage of students participating in the free- or reduced-price meals program, and (d) per-pupil expenditure.

Gateway English II scores significantly predicted ACT Composite scores, $\beta = .62$, $t\ (258) = 11.70$, $p < .01$. Therefore, the null hypothesis was rejected. The relationship between Gateway English II and ACT scores was positive. After controlling for the other variables in the model, the partial correlation between Gateway English II and ACT Composite scores was .59. The r^2 change for Gateway English II when entered into the model last was .11 indicating that Gateway English II scores contributed an additional 11% of the variance in ACT Composite scores over and above the variance accounted for by the five other predictors.

Percentage of minority students also significantly predicted ACT Composite scores after controlling for the other predictors, $\beta = -.25$, $t\ (258) = -5.44$, $p < .01$. Therefore, the null hypothesis was rejected. The relationship between percentage of minority students and ACT scores was negative. After controlling for the other variables in the model, the partial correlation between percentage of minority students and ACT Composite scores was -.321. The r^2 change for percentage of minority students when entered into the model last was .03 indicating that percentage of minority students accounted for an additional 3% of the variance in ACT composite scores over and above the variance accounted for by the other predictors.

Socioeconomic status, as measured by the percentage of students participating in the free- or reduced-price meals program, was also a significant predictor of ACT Composite scores, $\beta = -.17$, $t\ (258) = -4.30$, $p < .01$. Therefore, the null hypothesis was rejected. The relationship was negative. After controlling for the other variables in the model, the partial correlation between socioeconomic status and ACT Composite scores was -.26. The r^2 change for socioeconomic status when entered into the model last was .02 meaning socioeconomic status contributed an additional 2% of the variance accounted for in ACT composite scores over and above the variance accounted for by the other predictors.

The regression showed that graduation rate was not a significant predictor of ACT Composite scores, $\beta = -.004$, $t\ (258) = -.09$, $p = .93$. Also, Algebra I was not a significant predictor of ACT composite, $\beta = .03$, $t\ (258) = .69$, $p = .49$. Therefore, the null hypotheses for both graduation rates and Gateway Algebra I scores were retained.

Research Question #2

Is there a relationship between the home, student, and school variables (socioeconomic status, percentage of minority, graduation rate, per-pupil expenditure, Gateway English II, and Gateway Algebra I) with the three graded categories of TVAAS (Below, Above, No Detectable Difference)?

$Ho2_1$: There is no relationship between socioeconomic status and the three graded categories of TVAAS (Below, Above, No Detectable Difference).

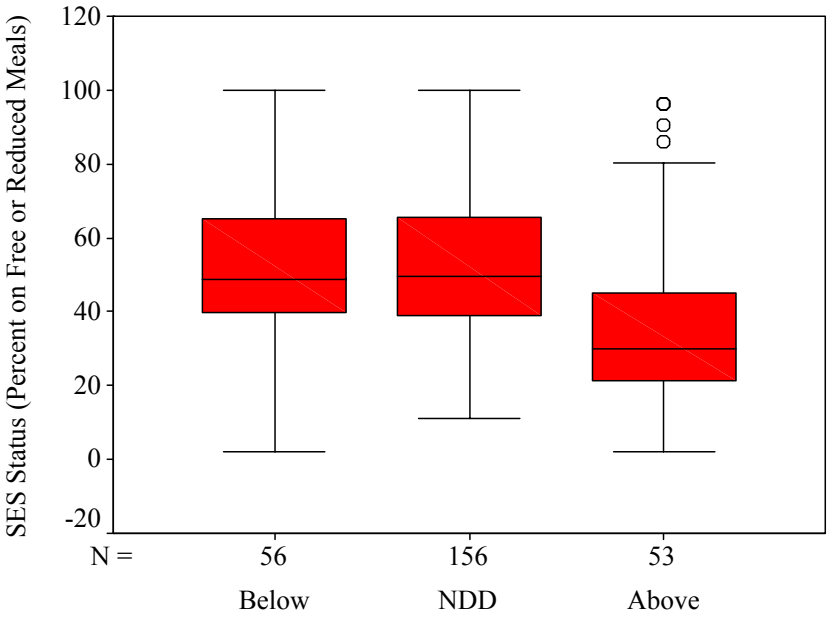
A one-way analysis of variance was conducted to evaluate the differences between the three graded categories of TVAAS and the schools' socioeconomic status, the test variable, measured as the percentage of students who qualified for the free- or reduced-priced meals program. The grouping variable was TVAAS, which had three levels (Below, Above, No Detectable Difference). The ANOVA was significant, F (2, 262) = 14.40, $p < .01$. Therefore, the null hypothesis was rejected. The measure of the strength of the relationship between TVAAS classification and socioeconomic status as measured by η^2 was medium (.10). That is, 10% of the variance in socioeconomic status was associated with TVAAS classifications.

Because the overall F test was significant, post hoc multiple comparisons were conducted to determine the pairwise differences in the socioeconomic status means of the three TVAAS classifications. A Tukey procedure was used because equal variances could be assumed, F (2, 262) = .41, $p = .67$. The Tukey procedure showed that the mean for socioeconomic status for schools that scored above on the TVAAS test was significantly different from both schools that scored below ($p < .01$) and schools that had no detectable difference ($p < .01$). In each case, the mean SES for schools that scored "above" on the TVAAS was lower than the mean for schools that scored below or those that had no detectable difference. There was no difference in the SES means between schools that scored below and those with no detectable difference on the TVAAS ACT Composite grades ($p = .86$). Table 3 shows the means and standard deviations for socioeconomic status by TVAAS ACT Composite classification. Figure 4 shows the boxplot for socioeconomic status by TVAAS ACT Composite classification.

Table 3

Means and Standard Deviations for Socioeconomic Status (Percentage of Students on Free- or Reduced-Price Meals Program) by TVAAS ACT Composite Classification

TVAAS Act Composite Classification	*n*	*M*	*SD*
Below	56	51.69	20.15
No Detectable Difference	156	53.48	21.44
Above	53	35.34	23.23
Total	265	49.48	22.61



o = an observation between 1.5 times to 3.0 times the interquartile range

Figure 4. Boxplot for Socioeconomic Status (Percentage of Students on the Free- or Reduced-Price Meals Program) by TVAAS Act Composite Classification

$Ho2_2$: There is no relationship between percentage of minority students and the three graded categories of TVAAS (Below, Above, No Detectable Difference).

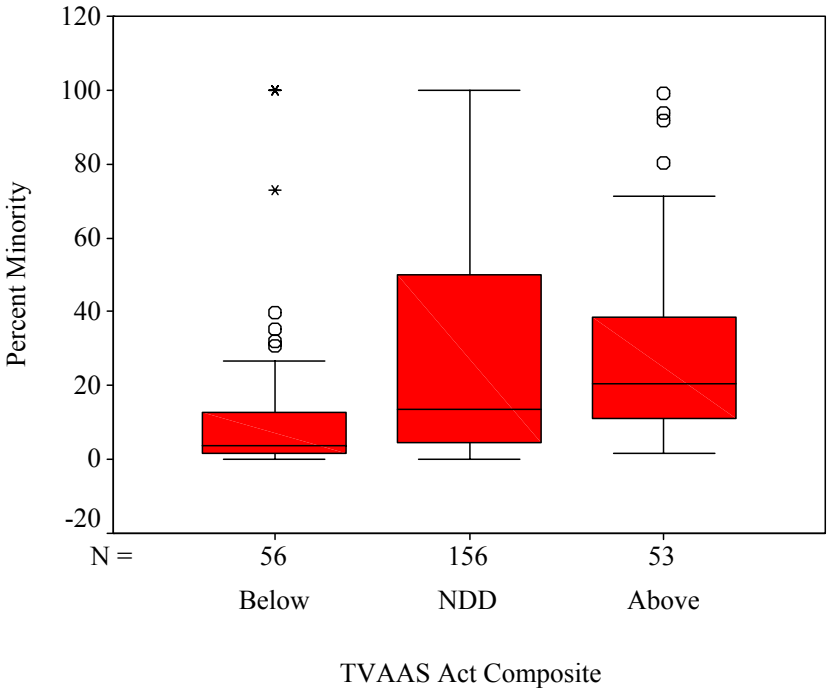
A one-way analysis of variance was conducted to evaluate the differences between the three graded categories of TVAAS and schools' percent of minority students, the test variable, measured as the percentage of students who are not White. The grouping variable was TVAAS which had three levels (below, no detectable difference, and above). The ANOVA was significant, F (2, 262) = 7.16, $p < .01$. Therefore, the null hypothesis was rejected. The measure of the strength of the relationship between TVAAS classification and percent of minority students as measured by η^2 was small (.05). In other words, only 5% of the variance in schools' percent minority was associated with TVAAS classifications.

Because the overall F test was significant, post hoc multiple comparisons were conducted to determine the pairwise differences in the means of the three TVAAS classifications. Because the Levene's test for equal variances was significant, F (2, 262) = 14.56, $p = <.01$, the Dunnett post hoc test was used. The Dunnett does not assume equal variances. The Dunnett procedure showed that the mean for percent minority for schools that scored below on the TVAAS test was significantly different from schools that had no detectable difference ($p < .01$) and from schools that scored above on the TVAAS. ($p < .01$). As shown in Table 4, the mean percentage of minority students in schools that scored below on the TVAAS ACT Composite test was much lower than the means for schools that scored no detectable difference and above on the TVAAS grades. The schools with no detectable difference and schools that scored above on the TVAAS were not statistically different ($p = .93$) regarding percent of minority students in the schools. The means and standard deviations for the three TVAAS groups are reported in Table 4. The boxplot for the distribution of percent minority by the three TVAAS classifications is shown in Figure 5.

Table 4

Means and Standard Deviations for Percentage of Minority by TVAAS ACT Composite Classification

TVAAS Act Composite Classification	*n*	*M*	*SD*
Below	56	13.47	24.41
No Detectable Difference	156	31.59	34.53
Above	53	29.10	25.48
Total	265	27.26	31.69



o = an observation between 1.5 times to 3.0 times the interquartile range

***** = an observation more than 3.0 times the interquartile range

Figure 5. Boxplot for Percentage of Minority Students by TVAAS ACT Composite Classification

$Ho2_3$: There is no relationship between graduation rate and the three graded categories of TVAAS (Below, Above, No Detectable Difference).

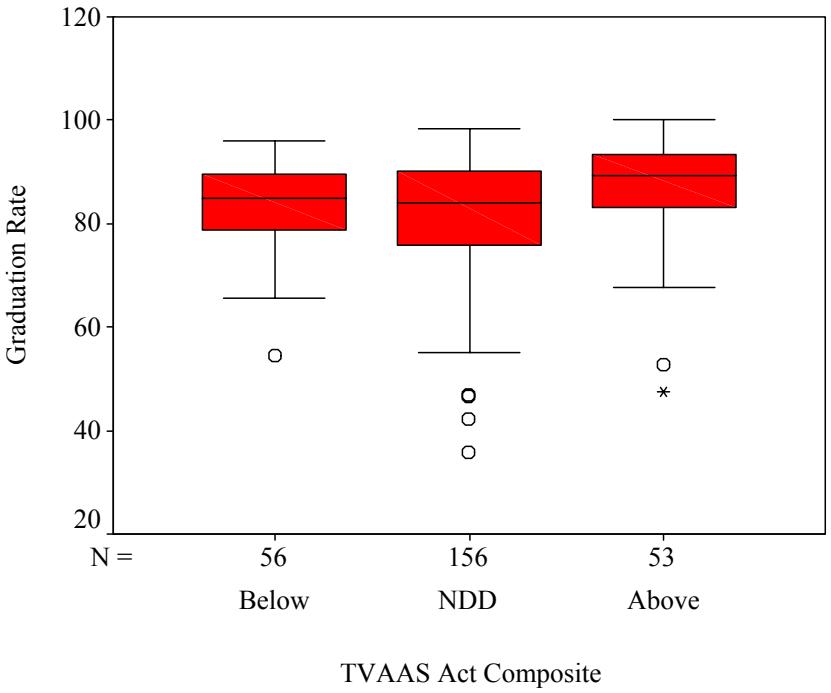
A one-way analysis of variance was conducted to evaluate the differences between the three graded categories of TVAAS and schools' graduation rate, the test variable, measured as the percentage of students that graduated on time. The grouping variable was TVAAS, which had three levels (Below, Above, No Detectable Difference). The ANOVA was significant, F (2, 262) = 3.56, p = .03. Therefore, the null hypothesis was rejected. The measure of the strength of the relationship between TVAAS classification and graduation rate as measured by η^2 was small (.03). That is, only 3% of the variance in graduation rates is accounted for by TVAAS classifications.

Because the overall F test was significant, post hoc multiple comparisons were conducted to determine the pairwise differences in the means of the three TVAAS classifications. Because the Levene's test for equal variances was significant, F (2, 262) = 3.48, p = .03, the Dunnett post hoc test was used. The Dunnett does not assume equal variances. The Dunnett procedure showed that the mean for graduation rate for schools that scored below on the TVAAS ACT grade was not significantly different from schools that showed no detectable difference (p = .51) and from schools that scored above (p =.35). As shown in Table 5, the mean graduation rate in schools that scored below on the TVAAS ACT Composite test was slightly higher than schools that scored no detectable difference and slightly lower than the mean for schools that scored above on the TVAAS grades. However, there was a significant difference in the mean graduation rates of schools that scored no detectable difference and schools that scored above on the TVAAS (p = .03). As shown in the table the mean graduation rate in schools that scored no detectable difference was almost five percentage points lower than those schools that scored above on the TVAAS grades. The means and standard deviations for the three TVAAS groups are shown in Table 5. The boxplot for the distribution of graduation rates by the three TVAAS classifications is shown in Figure 6.

Table 5

Means and Standard Deviations for Graduation Rates by TVAAS ACT Composite Classification

TVAAS Act Composite Classification	*n*	*M*	*SD*
Below	56	83.55	8.15
No Detectable Difference	156	81.75	11.78
Above	53	86.32	10.68
Total	265	83.04	11.00



o = an observation between 1.5 times to 3.0 times the interquartile range

***** = an observation more than 3.0 times the interquartile range

Figure 6. Boxplot for Graduation Rate by TVAAS ACT Composite Classification

$Ho2_4$: There is no relationship between per-pupil expenditures and the three graded categories of TVAAS (Below, Above, No Detectable Difference).

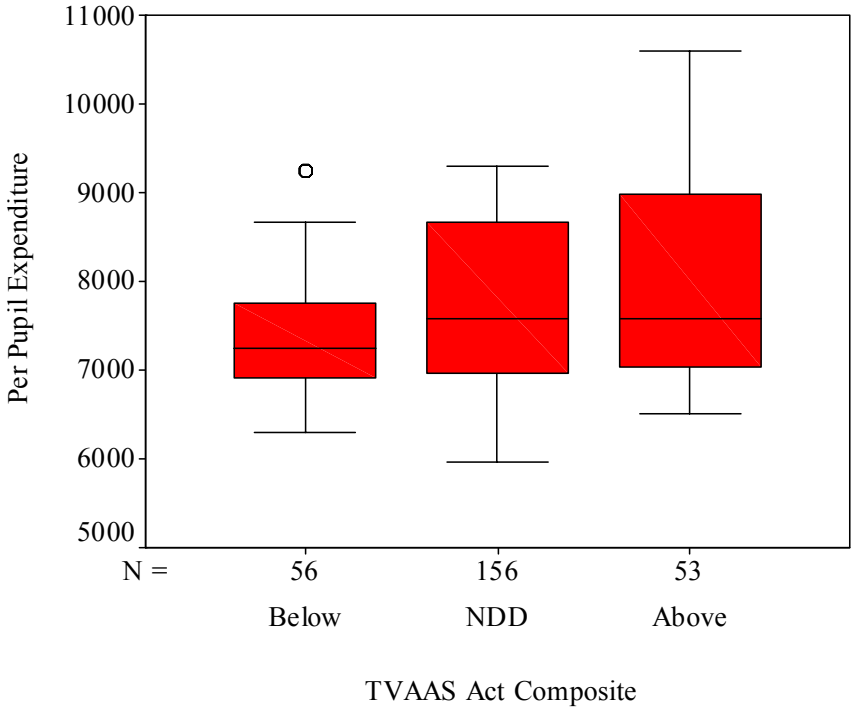
A one-way analysis of variance was conducted to evaluate the differences between the three graded categories of TVAAS and schools' per-pupil expenditure, the test variable, measured as the districts' per-pupil expenditure. The grouping variable was TVAAS which had three levels (below, no detectable difference and above). The ANOVA was significant, F (2, 262) = 5.18, p = .01. Therefore, the null hypothesis was rejected. The measure of the strength of the relationship between the graded TVAAS classifications and per-pupil expenditure as measured by η^2 was small (.04). In other words, 4% of the variance in per-pupil expenditure is shared with TVAAS classifications.

Because the overall F test was significant, post hoc multiple comparisons were conducted to determine the pairwise differences in the means of the three TVAAS classifications. Because the Levene's test for equal variances was significant, F (2, 262) = 8.45, $p < .01$, the Dunnett post hoc test was used. The Dunnett does not assume equal variances. The Dunnett procedure showed that the mean for per-pupil expenditure for schools that scored below on the TVAAS ACT grade was significantly different from schools that scored no detectable difference ($p < .01$) and from schools that scored above on the TVAAS test (p = .01). The schools that scored no detectable difference and above were not significantly different (p = .83) regarding per-pupil expenditures in the schools. As shown in Table 6, the mean per-pupil expenditures in schools that scored below on the TVAAS ACT Composite test was just over $400 less than the mean for schools that scored no detectable difference and $523 less than schools that scored above on the TVAAS grades. The means and standard deviations for the three TVAAS groups are reported in Table 6. The boxplot for the distribution of per-pupil expenditure by the three TVAAS classifications is shown in Figure 7.

Table 6

Means and Standard Deviations for Per-Pupil Expenditure by TVAAS ACT Composite Classification

TVAAS Act Composite Classification	*n*	*M*	*SD*
Below	56	7,374.81	693.07
No Detectable Difference	156	7,776.01	962.71
Above	53	7,897.88	1,011.67
Total	265	7,715.60	937.69



o = an observation between 1.5 times to 3.0 times the interquartile range

Figure 7. Boxplot for Per-Pupil Expenditure by TVAAS ACT Composite Classification

$Ho2_5$: There is no relationship between Gateway English II scores and the three graded categories of TVAAS (Below, Above, No Detectable Difference).

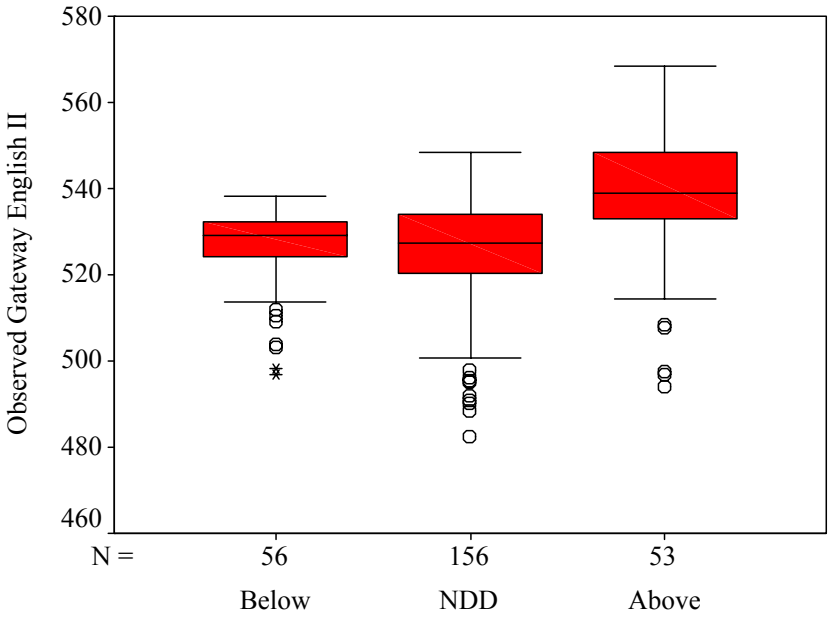
A one-way analysis of variance was conducted to evaluate the differences between the three graded categories of TVAAS and schools' Gateway English II scores, the test variable, measured as the schools' Gateway English II scores. The grouping variable was TVAAS which had three levels (Below, Above, No Detectable Difference). The ANOVA was significant, F (2, 262) = 18.49, $p < .01$. The null hypothesis was rejected. The measure of the strength of the relationship between the graded TVAAS classifications and Gateway English II scores as measured by η^2 was medium (.12). That is, 12% of the variance in the observed Gateway English II scores was accounted for by the graded TVAAS classifications.

Because the overall F test was significant, post hoc multiple comparisons were conducted to determine the pairwise differences in the means of the three TVAAS classifications. Because the Levene's test for equal variances was significant, F (2, 262) = 4.04, $p = .02$, the Dunnett post hoc test was used. The Dunnett does not assume equal variances. The Dunnett procedure showed that the mean for Gateway English II scores for schools that scored below on the TVAAS ACT grade was not significantly different from schools that had no detectable difference ($p = .90$) but was significantly different from schools that scored above ($p < .01$). There was also a significant difference between schools that had no detectable difference and schools that scored above ($p < .01$) regarding Gateway English II scores. As shown in Table 7, the mean Gateway English II scores for schools that scored below and no detectable difference on the TVAAS ACT Composite test were lower than the mean for schools that scored above on the TVAAS grades. The means and standard deviations for the three TVAAS groups are reported in Table 7. The boxplot for the distribution of observed English II scores by the three TVAAS classifications is shown in Figure 8.

Table 7

Means and Standard Deviations for Observed Gateway English II Scores by TVAAS ACT Composite Classification

TVAAS Act Composite Classification	*n*	*M*	*SD*
Below	56	526.03	9.76
No Detectable Difference	156	525.01	13.16
Above	53	537.60	16.33
Total	265	527.74	14.09



o = an observation between 1.5 times to 3.0 times the interquartile range

***** = an observation more than 3.0 times the interquartile range

Figure 8. Boxplot for Observed Gateway English II Scores by TVAAS ACT Composite Classification

$Ho2_6$: There is no relationship between Gateway Algebra I scores and the three graded categories of TVAAS (Below, Above, No Detectable Difference).

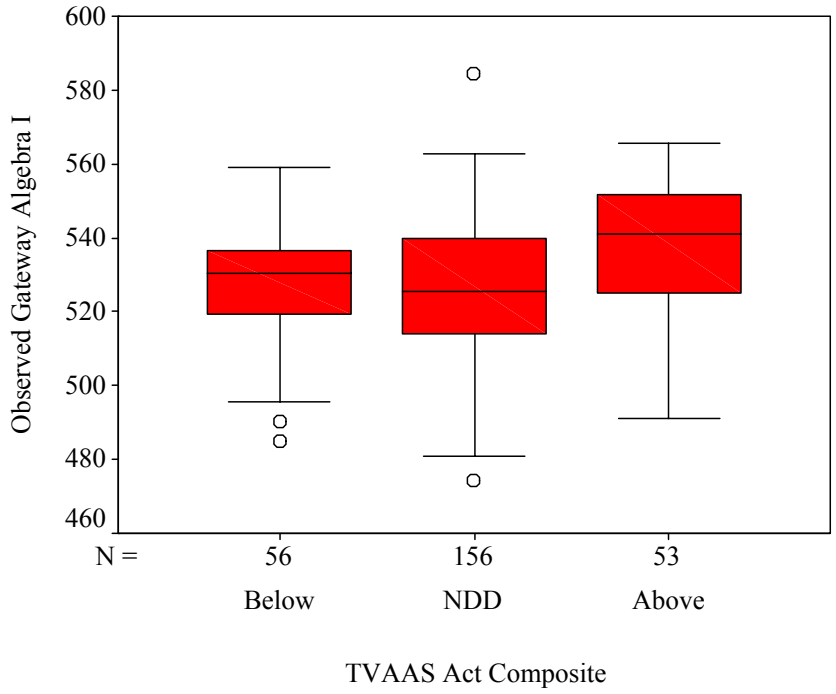
A one-way analysis of variance was conducted to evaluate the differences between the three graded categories of TVAAS and schools' Gateway Algebra I scores, the test variable, measured as the schools' Gateway Algebra I scores. The grouping variable was TVAAS which had three levels (below, no detectable difference and above). The ANOVA was significant, F (2, 262) = 7.14, $p < .01$. Therefore, the null hypothesis was rejected. The measure of the strength of the relationship between the graded TVAAS classifications and Gateway Algebra I scores as measured by $\eta^2$was small (.05). In other words, 5% of the variance in the observed Gateway Algebra I scores was accounted for by the TVAAS classifications.

Because the overall F test was significant, post hoc multiple comparisons were conducted to determine the pairwise differences in the means of the three TVAAS classifications. Because the Levene's test for equal variances was not significant, F (2, 262) = 2.61, p = .08, the Tukey post hoc test was used. The Tukey assumes equal variances. The Tukey procedure showed that the mean for Gateway Algebra I scores for schools that scored below on the TVAAS ACT grade was not significantly different from schools that had no detectable difference (p = .74) but was significantly different from schools that scored above (p = .03). In addition, there was a difference in the means on the Gateway Algebra I test between schools that had no detectable difference and the schools that scored above on the TVAAS ($p < .01$). As shown in Table 8, the mean Gateway Algebra I scores in schools that scored above on the TVAAS ACT Composite test was nine points higher than the mean for schools that scored below and 11 points higher than schools that had no detectable difference on the TVAAS ACT composite test. The means and standard deviations for the three TVAAS groups are reported in Table 8. The boxplot for the distribution of observed Gateway Algebra I scores by the three TVAAS classifications is shown in Figure 9.

Table 8

Means and Standard Deviations for Gateway Algebra I Scores by TVAAS ACT Composite Classification

TVAAS Act Composite Classification	*n*	*M*	*SD*
Below	56	527.31	15.31
No Detectable Difference	156	525.15	19.70
Above	53	536.33	18.69
Total	265	527.84	19.08



o = an observation between 1.5 times to 3.0 times the interquartile range

Figure 9. Boxplot for Observed Gateway Algebra I Scores by TVAAS ACT Composite Classification

CHAPTER 5

SUMMARY OF FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS

The purpose of this study was to determine whether a significant relationship exists between particular home, student, and school characteristics and ACT scores and the relationship these characteristics subsequently have with the TVAAS grades assigned to each high school's ACT scores. These home, student, and school variables were socioeconomic status, percentage of minority students, graduation rate, per-pupil expenditure, Gateway English II scores, and Gateway Algebra I scores. I began by establishing that a relationship exists between certain home, student, and school characteristics and the ACT scores from the same schools. After determining the predictive values of these characteristics, I then measured the relationship between those variables and the TVAAS grades as assigned by the state.

Summary of the Study

In recent years, high stakes testing has risen to the forefront as a means to measure student progress and provide accountability as to the performance of the districts and schools. Taxpaying citizens want to know that their taxes are being put to use in effective ways and the expectations of schools to produce results permeates public education. Tests are given annually to provide an analysis as to the effectiveness of the schools and districts and the subsequent results are distributed for all citizens to peruse to aid them in their evaluation of public schools. As a result, in Tennessee's high schools, students are tested in many subjects including English II and Algebra I. These tests are called Gateways and the state and federal government evaluates these results to ascertain the effectiveness of schools and districts.

A thorough review of the literature was conducted tracing the history of testing from centuries ago to modern days in American testing procedures. I also explored the evolution of achievement tests and intelligence tests and how they have an impact on modern testing

philosophy and practices. Tennessee's involvement in the growth model testing practices of Sanders (1998) was thoroughly explored, as was the lottery scholarship and the TVAAS system of school accountability.

Summary of Findings

This analysis focused on two research questions using a sample containing data from 265 Tennessee high schools. The sample included all Tennessee high schools except those that lacked information pertinent to the study and atypical high schools such as adult high schools, technology centers, and schools where students take college courses while in high school.

Research Questions

Research Question #1

Which home, student, and school variables have the strongest relationship with 2007 ACT scores for Tennessee high schools?

Before examining the finding of the regression analyses, I wanted to make certain that there were not violations of the assumptions of regression. The assumption of normality appeared to be met by looking at the histogram as shown in Figure 1 in Chapter 4. To further confirm that the assumptions of normality were met, I used the one-sample Kolmogorov-Smirnov Test. This test was used to test the distribution of the standardized residuals against a normal distribution. The null hypothesis that the distribution of the residuals from the regression model does not deviate from a normal distribution was retained ($p = .73$). Therefore, the assumption of normality was met. Second, there appeared to be no reason to question the assumption of linearity based on a visual examination of the normal probability plot (as shown in Figure 2) as indicated by the red dots falling very close to or on the green line. Likewise, the scatterplot of the standardized residuals and standardized predicted values (as shown in Figure 3) revealed no discernible pattern of nonlinearity or unequal variances. Based on the analyses of the residuals, I concluded that the assumptions of the regression model were met.

I then analyzed the six predictor variables in a regression model with observed ACT composite scores as the dependent variable: (a) socioeconomic status was measured as the percentage of students who participated in the free- or reduced-price meals program, (b) percentage of minority students, (c) graduation rate, (d) per-pupil expenditure, (e) observed Gateway English II scores, and (f) observed Gateway Algebra I scores. All of the variables except for per-pupil expenditure were school level variables. Per-pupil expenditure was a school district level variable. A multiple regression model with the six predictors entered as a set was used to evaluate the six hypotheses.

As shown in Table 2, each of the six zero-order correlations between the predictors and ACT Composite scores was significant. However, after controlling for the other variables in the regression model, only four predictors remained significant: (a) observed Gateway English II, (b) percentage of minority students, (c) socioeconomic status measured as the percentage of students participating in the free- or reduced-price meals program, and (d) per-pupil expenditure.

The findings of the regression analysis showed that the six predictor variables as a set were significantly related to observed ACT composite scores, F (6, 258) = 157.18, $p < .01$. The R^2 for the model was .79 meaning that 79% of the variance in observed ACT composite scores was accounted for by the six predictors. This is a strong relationship within these six predictor variables of home, student, and school characteristics. These six variables account for 79% of the variance on the ACT test that students take that is supposed to indicate their readiness for college.

Gateway English II scores significantly predicted ACT Composite scores, $\beta = .62$, t (258) = 11.70, $p < .01$. Therefore, the null hypothesis was rejected. The relationship between Gateway English II and ACT scores was positive. After controlling for the other variables in the model, the partial correlation between Gateway English II and ACT Composite scores was .59. The r^2 change for Gateway English II when entered into the model last was .11 indicating that Gateway English II scores contributed an additional 11% of the variance in ACT Composite scores over and above the variance accounted for by the five other predictors. It is not surprising to find that

the Gateway English II scores have such a high relationship with the ACT composite test. Reading proficiency is an important part of performing well on the ACT test because the entire battery of tests involves so much reading.

The percentage of minority students also significantly predicted ACT Composite scores after controlling for the other predictors, $\beta = -.25$, $t\ (258) = -5.44$, $p < .01$. Therefore, the null hypothesis was rejected. The relationship between percentage of minority students and ACT scores was negative. After controlling for the other variables in the model, the partial correlation between percentage of minority students and ACT Composite scores was -.321. The negative correlation reveals that the higher the percentage of minority students in a school, the lower the ACT composite score. The r^2 change for percentage of minority students when entered into the model last was .03 indicating that percentage of minority students accounted for an additional 3% of the variance in ACT composite scores over and above the variance accounted for by the other predictors.

Socioeconomic status as measured by the percentage of students participating in the free- or reduced-price meals program was also a significant predictor of ACT Composite scores, $\beta = -.17$, $t\ (258) = -4.30$, $p < .01$. Therefore, the null hypothesis was rejected. The relationship was negative, and like the percent of minority analyses, the schools with a higher percentage of students on free- or reduced-priced meals would score lower on the ACT composite scores. After controlling for the other variables in the model, the partial correlation between socioeconomic status and ACT composite scores was -.26. The r^2 change for socioeconomic status when entered into the model last was .02 meaning socioeconomic status contributed an additional 2% of the variance accounted for in ACT composite scores over and above the variance accounted for by the other predictors.

The regression showed that graduation rate was not a significant predictor of ACT Composite scores, $\beta = -.004$, $t\ (258) = -.09$, $p = .93$. This is understandable; since the passage of *No Child Left Behind* many schools are pushing hard for students to graduate. There are schools that have very low ACT composite scores that have high graduation rates. Cloudland High

School, for example has a graduation rate of 89.36% while its ACT composite average is 17.38. The mean of the graduation rate in this analysis is 83.04 and the ACT composite mean is 20.7. Oak Ridge, on the other hand, has a graduation rate of 81.38 and an ACT composite average of 23.53. Also, Algebra I was not a significant predictor of ACT composite, $\beta = .03$, $t\ (258) = .69$, $p = .49$. Therefore, the null hypotheses for both graduation rates and Gateway Algebra I scores were retained.

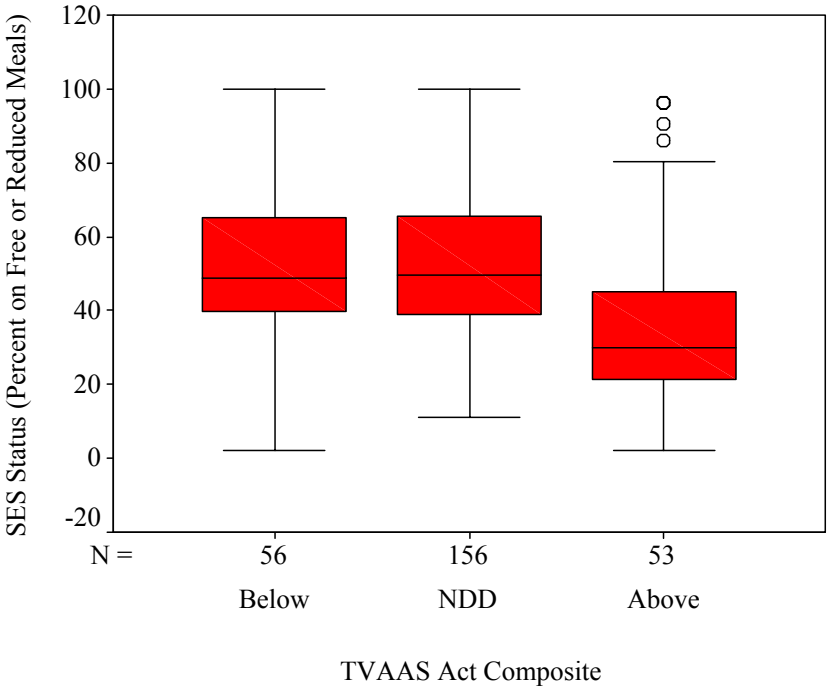
Research Question #2

Is there a relationship between the home, student, and school variables (socioeconomic status, percentage of minority, graduation rate, per-pupil expenditure, Gateway English II scores, and Gateway Algebra I scores) with the three graded categories of TVAAS (Below, Above, No Detectable Difference)?

The analyses of this research question also consisted of the six predictor variables and the hypotheses that there was no relationship between the variables and the three graded categories of TVAAS ACT composite scores. To ascertain if a relationship existed between the home, student, and school variables, I conducted a one-way analysis of variance to evaluate the differences between the three graded categories of TVAAS and socioeconomic status, the test variable, measured as the percentage of students who qualified for the free- or reduced-priced meals program. The grouping variable was TVAAS, which had three levels (Below, Above, No Detectable Difference). The ANOVA was significant, $F\ (2, 262) = 14.40$, $p < .01$. Therefore, the null hypothesis was rejected. The measure of the strength of the relationship between TVAAS classification and socioeconomic status as measured by η^2 was medium (.10). That is, 10% of the variance in TVAAS classifications was associated with socioeconomic status.

Because the overall F test was significant, post hoc multiple comparisons were conducted to determine the pairwise differences in the socioeconomic status means of the three TVAAS classifications. A Tukey procedure was used because equal variances could be assumed, $F\ (2, 262) = .41$, $p = .67$. The Tukey procedure showed that the mean for socioeconomic status for

schools that scored above on the TVAAS test was significantly different from both schools that scored below ($p < .01$) and schools that had no detectable difference ($p < .01$). In each case, the mean SES for schools that scored above on the TVAAS was lower (lower percentage of students on the free- or reduced-price meals program) than was the means for schools that scored below or those that had no detectable difference. There was no difference in the SES means among schools that scored below and those with no detectable difference on the TVAAS ACT Composite grades ($p = .86$). Figure 10 shows the three graded classifications of SES in a boxplot. Figure 10 reveals that the Above category has a lower SES percentage than the schools that have NDD or a Below designation on the 2007 State Report Card. This is significant given the claims that socioeconomics are filtered out and teacher effect is the dominant aspect of the schooling experience according to TVAAS claims.



o = an observation between 1.5 times to 3.0 times the interquartile range

Figure 10. Boxplot for Socioeconomic Status (Percentage of Students on the Free- or Reduced-Price Meals Program) by TVAAS ACT Composite Classification

The analysis of the SES category suggests that there is a relationship between the SES status of schools and the subsequent grades given to schools on the TVAAS ACT composite category. Another way of looking at these statistics is to rank the poorest schools to the richest schools. In looking at these rankings in my SPSS package, I discovered that the poorest 44 schools had only six "Above" scores. I also looked at the 44 wealthiest schools based on fewest percentage students receiving free- or reduced-price meals. These 44 schools revealed 26 schools receiving "Above" status on the state's Report Card. This would seem to indicate that the wealthier schools have an advantage in achieving the designation of "Above" status over the poorer schools and districts.

I then embarked to ascertain if the percentage of minorities in a school accounted for a significant difference in the TVAAS grades. To do this, I conducted an ANOVA to evaluate the differences between the three graded categories of TVAAS and schools' percent minority, the test variable, measured as the percentage of students who are not White. According to the statistical analysis, the measure of the strength of the relationship between TVAAS classification and percentage of minority as measured by η^2 was small (.05). In other words, 5% of the variance in schools' TVAAS classifications was associated with percentage of minority students.

The Dunnett procedure showed that the mean for percentage of minority students for schools that scored below on the TVAAS test was significantly different from schools that had no detectable difference ($p < .01$) and from schools that scored above on the TVAAS ($p < .01$). As shown in Table 4, the mean percentage of minority students in schools that scored below on the TVAAS ACT Composite test was much lower than the means for schools that scored no detectable difference and above on the TVAAS grades. The schools with no detectable difference and schools that scored above on the TVAAS were not statistically different ($p = .93$) regarding percentage of minority students in the schools. Because this established that there was a relationship between the percentage of minority students in schools and the TVAAS grades, I looked at the SPSS rankings of the schools in the percentage of minority students category. This was revealing. The 63 schools with the highest percentage of minorities students showed that

only five schools were below and 12 schools were above in the TVAAS graded category of ACT composite. The top 63 schools ranged from 100% minorities to 41.99%. I wondered what the other end of the percentage of minority students' spectrum would show. In the 64 schools with the least percentage of minority students, I discovered that 29 schools were categorized as being below the expected growth on ACT composite scores. There were only four schools that merited the above category. Is this suggesting that the grade for schools with a high percentage of minority students is not equitable with the schools with a small percentage of minorities? The schools with the lowest percentage of minority students ranged from 0% to 3.61% percent. Additionally, seeing this trend led me to looking deeper at the rankings and I discovered that when looking at the schools with the largest percentage of White students, out of the 113 schools with the greatest percentage of White students, there were only 9 above schools on the TVAAS ACT composite grades. Conversely, I looked at the 113 schools with the lowest percentage of White students and found that there were 32 schools with the above distinction.

The next home, school, and student variable I measured was the graduation rate of the schools. A one-way analysis of variance was conducted to evaluate the differences between the three graded categories of TVAAS and schools' graduation rate, the test variable, measured as the percentage of students that graduated on time. The measure of the strength of the relationship between TVAAS classification and graduation rate as measured by η^2 was small (.03). That is, only 3% of the variance in TVAAS classifications is accounted for by graduation rates.

The Dunnett procedure showed that the mean for graduation rate for schools that scored below on the TVAAS ACT grade was not significantly different from schools that showed no detectable difference (p = .51) and from schools that scored above (p =.35). As shown in Table 5, the mean graduation rate in schools that scored below on the TVAAS ACT Composite test was slightly higher than was schools that scored no detectable difference and slightly lower than the mean for schools that scored above on the TVAAS grades. However, there was a significant difference in the mean graduation rates of schools that scored no detectable difference and schools that scored above on the TVAAS (p = .03). As shown in the table, the mean graduation

rate in schools that scored no detectable difference was almost five percentage points lower than in those schools that scored above on the TVAAS grades. The boxplot for the distribution of graduation rates by the three TVAAS classifications is shown in Figure 6.

I then measured the home, school, and student, variable of per-pupil expenditure. Please remember that this was district level data, but each school in a given district would have the exact per-pupil expenditure that was reported from the district. A one-way analysis of variance was conducted to evaluate the differences between the three graded categories of TVAAS and schools' per-pupil expenditure, the test variable, measured as the districts' per-pupil expenditure. The measure of the strength of the relationship between the graded TVAAS classifications and per-pupil expenditure as measured by η^2 was small (.04). In other words, 4% of the variance in TVAAS classifications is associated with per-pupil expenditure. There would be a much greater correlation if some of the larger school systems such as Memphis City ($9,253 per pupil) and Davidson County ($9,299 per pupil) school systems were pulled from the equation. The average per-pupil expenditure in Tennessee is $7,715. These two systems alone account for 8.6% of the high schools in this study. These school systems spend a great deal of money per pupil but with poor results on achievement scores. This skews the relationship that has been analyzed here.

The Dunnett procedure showed that the mean for per-pupil expenditure for schools that scored below on the TVAAS ACT grade was significantly different from schools that scored no detectable difference ($p < .01$) and from schools that scored above on the TVAAS test ($p = .01$). The schools that scored no detectable difference and above were not significantly different ($p = .83$) regarding per pupil expenditures in the schools. As shown in Table 6, the mean per-pupil expenditures in schools that scored below on the TVAAS ACT Composite test was just over $400 less than was the mean for schools that scored no detectable difference and $523 less than schools that scored above on the TVAAS grades.

The Gateway English II scores were next examined to see if a relationship existed to the TVAAS grades assigned. A one-way analysis of variance was conducted to evaluate the differences between the three graded categories of TVAAS and schools' Gateway English II

scores, the test variable, measured as the schools' Gateway English II scores. The measure of the strength of the relationship between the graded TVAAS classifications and Gateway English II scores as measured by η^2 was medium (.12). That is, 12% of the variance in the graded TVAAS classifications was accounted for by the observed Gateway English II scores. As we saw in the earlier analyses, the Gateway English II test is a great predictor of future ACT scores. I also examined the rankings in a similar fashion examining the 64 schools with the highest grades in observed Gateway English II scores and the lowest grades and found that the schools with the highest grades on English II observed scores had 27 "above" schools while the schools with the lowest grades had 26 schools. This indicated to me that the grades in English II were fair and equitable. An ironic observation in this category was that the school (Hume-Fogg High Academic Magnet School) with the highest observed score in the state actually received a "below" grade in English II TVAAS category.

The Dunnett procedure showed that the mean for Gateway English II scores for schools that scored below on the TVAAS ACT grade was not significantly different from schools that had no detectable difference ($p = .90$) but was significantly different from schools that scored above ($p < .01$). There was also a significant difference between schools that had no detectable difference and schools that scored above ($p < .01$) regarding Gateway English II scores. As shown in Table 7, the mean Gateway English II scores for schools that scored below and no detectable difference on the TVAAS ACT Composite test were lower than the mean for schools that scored above on the TVAAS grades.

Finally, I examined the relationship between Gateway Algebra I scores to the TVAAS grades assigned. A one-way analysis of variance was conducted to evaluate the differences between the three graded categories of TVAAS and schools' Gateway Algebra I scores, the test variable, measured as the schools' Gateway Algebra I scores. The measure of the strength of the relationship between the graded TVAAS classifications and Gateway Algebra I scores as measured by η^2 was small (.05). In other words, 5% of the variance in the TVAAS classifications was accounted for by the observed Gateway Algebra I scores.

Because the overall *F* test was significant, post hoc multiple comparisons were conducted to determine the pairwise differences in the means of the three TVAAS classifications. Because the Levene's test for equal variances was not significant, $F(2, 262) = 2.61$, $p = .08$, the Tukey post hoc test was used. The Tukey assumes equal variances. The Tukey procedure showed that the mean for Gateway Algebra I scores for schools that scored below on the TVAAS ACT grade was not significantly different from schools that had no detectable difference ($p = .74$) but was significantly different from schools that scored above ($p = .03$). In addition, there was a difference in the means on the Gateway Algebra I test between schools that had no detectable difference and the schools that scored above on the TVAAS ($p < .01$). As shown in Table 8, the mean Gateway Algebra I scores in schools that scored above on the TVAAS ACT Composite test was 9 points higher than the mean for schools that scored below and 11 points higher than schools that had no detectable difference on the TVAAS ACT composite grades. The TVAAS grades for the Gateway Algebra I appear to be closely associated to the observed scores in Algebra I, as the top 62 schools garnered 54 above scores.

After examining the data, I wanted to do one more analysis to look at the scores that were given in the TVAAS ACT Composite category. I looked at the rankings of the ACT observed scores to see if the rankings would reveal any usable information. These observed scores are what the TVAAS grades should be based upon. I also looked at the schools that emerged in these rankings. The rankings of the ACT composite scores from top to bottom revealed some interesting information. Schools like Ravenwood, Farragut, Brentwood, and Maryville, all schools with high percentage of white students with 10% to 15% minority populations, stood out as having great test scores and Above ratings on the TVAAS composite grades. Schools like Tellico Plains, Greenback, and Campbell County, all rural predominantly white student bodies, received the Below ratings, despite having solid ACT composite scores. One school, White Station High School in Memphis, seemed to be the exception to this trend. I examined the School Improvement Plan for White Station High School and discovered that White Station High was a school in Memphis that took "high achieving" students from throughout Shelby County.

In fact, White Station offers a phenomenal selection of classes. Of the 23 Advanced Placement classes they offer, 307 students took AP exams and 93% scored 3 or more. These were outstanding scores by any measure. However, White Station's student body is not "naturally occurring" but rather a result of bright students from throughout Memphis making application to attend that high school. Another similar school is located in Nashville, and its scores are very impressive. Hume-Fogg Academic Magnet high school scores at the top of the state in almost everything, but students have to apply to be accepted.

Conclusions

Based on the analysis of the findings from this study, it appears there is still work to be done on the TVAAS ACT composite grading of high schools in Tennessee. With the coming of the Tennessee Diploma, to be implemented in 2009-10, there will be added public scrutiny on the grades assigned to the TVAAS ACT composite grades. The ACT test is taking on added importance, and communities across the state are looking to the public high schools to help their sons or daughters have every opportunity to compete in the ever emerging global economy. This only raises the stakes for public high schools in Tennessee and the necessity that grades from the state department be accurate and fair.

When I embarked on this journey, my curiosity was piqued by a previous study by Paul Webb (2005). In his study, he had serious criticisms of TVAAS in general. I am not near as critical of the growth model concept as his study was. However, I do feel that the ACT Composite portion of the TVAAS grading system needs to be modified. It seems that the rural school systems with the higher percentage of White students are not graded equitably with the TVAAS. The Gateway English II TVAAS and the Gateway Algebra I TVAAS grades seem to be equitable. Webb remarked that people should not purchase lottery tickets because the Hope Scholarship was based on money derived from the state lottery. I disagree with that (although I have never purchased one). Webb's premise was that the poor people in Tennessee were still being preyed upon by an unfair system such as TVAAS. It is ironic that since Webb's study in

2005, prekindergarten programs have been started throughout the state with money that was derived in a large degree from the purchase of lottery tickets. The prekindergarten programs that are now serving thousands of children throughout the state is aimed at students from low income families. Therefore, even with the assumption that Webb was right originally, today the poor people in Tennessee could be benefiting from the efforts of the lottery scholarship program.

In addition, Webb's (2005) study raised awareness on my part about the apparent inequities of the assignment of the TVAAS grades through the state department to schools. This study seems to indicate that the distributions of the grades need to be reevaluated because the results indicate that rural predominantly White schools have a much more difficult time rating an "above" under the current system. I would suggest that Sanders look at this issue and tweak the system to more fairly evaluate schools that are not from more metropolitan areas.

Whether his formula underestimates minority growth, or overestimates the growth of rural white students, the formula should be fair. Appropriate expectations for all students should be our goal. We must have high expectations for all groups and not allow underachievement to be rewarded with an "Above" designation, perhaps perpetuating future low expectations for our urban students.

Conclusion 1

There is a correlation between students' demographics and the achievement levels of students, schools, and districts. Findings from the study reaffirm what I know about the challenges of schooling students who are from impoverished areas in both urban areas or rural areas. To fairly evaluate and grade these students, schools, and districts, the state must strive to give an equal opportunity to all. I am not sure whether minority students are graded with expectations that are too low or rural schools with predominantly White populations have too high of expectations under the TVAAS growth model, but these expectations should be fair to all. The TVAAS composite grades, although purported to measure teacher effectiveness, continue to be a measurement of demographics as well as teacher effectiveness. The TVAAS

composite grades assigned to schools in Tennessee are not equitable in terms of percentage of minority students of schools. Findings from this study reveal that it is difficult for a heavily predominant White population in a school or district to score well in the TVAAS composite grade category of Above, No Detectable Difference, and Below.

Conclusion 2

Because of accountability, students, schools, and districts still need an appropriate grading system in place for the TVAAS composite grades assigned through the state department. A growth model is a good thing, but it must be equitable regardless of percentage of minority students, socioeconomics, or per-pupil expenditure in a school or district.

Conclusion 3

Per-pupil expenditure statistics can be deceiving based on large, underperforming school districts skewing the results, making it easy for public school detractors to minimize the effects of appropriate funding. These types of adversaries will call for vouchers and other programs to divert funding for public schools into private schools.

Conclusion 4

The strong correlation for high schools in the Gateway English II scores to the ACT composite suggests a need to have reading intervention with students to enhance their chances of making an ACT score that would allow them to qualify for the Hope Scholarship, avoid developmental classes, and graduate from college.

Recommendations for the Improvement of Practice

The following recommendations are based on the findings of this study:

1. The Tennessee Department of Education must examine the practices of assigning the ACT composite TVAAS grades. These grades, given the significance that the ACT scores of students, schools, and districts will acquire in the future, must be equitable. I recommend that we examine the practice of holding low expectations of growth for our inner city schools with a high percentage of minority students. Schools and districts need to understand how a school with a 15.05 can be "Above" on the ACT TVAAS grade while another school with 21.87 rates a "Below" distinction. The formula for the TVAAS grade on ACT composite must be examined to assure appropriate expectations as well as equitable assignment of grades to schools and districts.
2. Reading plays a vital role in improving ACT scores. ACT scores play an ever increasing importance with the coming Tennessee Diploma. Changing assessment practices to more closely match ACT's style of assessment may encourage the teaching of reading.
3. Our inner city schools are struggling with high per-pupil expenditures and low achievement and graduation rates. We must continue to seek ways to aid the growth and raise expectations for poor inner city youth.

Recommendations for Further Research

1. A deeper examination must be made into the formulas that Bill Sanders uses to determine the grade assigned to schools and districts enhancing equity for all schools and districts.
2. A deeper look at the "magnet schools" concept which, while being successful enterprises, may by their very presence, be dooming other schools because of the "brain drain" that results when magnet schools are created.

REFERENCES

Achieve. (2008). *Our goal*. Retrieved May 13, 2008, from http://www.achieve.org/

ACT. (2005). *About ACT: History of ACT.* Retrieved April 5, 2008, from http://www.act.org/aboutact/history.html

ACT. (2007a). *ACT high school profile report: High school graduating class 2007.* Retrieved April 5, 2008, from http://www.act.org/news/data/04/pdf/data.pdf

ACT. (2007b). *ACT national and state scores.* Retrieved April 14, 2008, from http://www.act.org/news/data.html

ACT. (2008a). *Facts about the ACT assessment.* Retrieved March 24, 2008, from http://www.act.org/aboutact/history.html

ACT. (2008b). *ACT's annual report*. Retrieved March 24, 2008, from http://www.act.org/aboutact/pdf/AnnualReport07.pdf

Adelman, C. 1999. *Answers in the toolbox: Academic intensity, attendance patterns, and bachelor's degree attainment*. Washington, DC: U.S. Department of Education.

Archer, J. (1999). Sanders 101. *Education Week. 18,* 34 [Online]. Retrieved May 8, 2008, from http://www.edweek.org/ew/vol-18/34sander.h18

Becker, K. A. (2003). *History of the Stanford-Binet intelligence scales: Content and psychometrics.* (Stanford-Binet Intelligence Scales, Fifth Edition Assessment Service Bulletin No. 1). Itasca, IL: Riverside.

Berliner, D. C., & Biddle, B. J. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools* . Cambridge, MA: Perseus.

Bianchi, A. B. (2003). A new look at accountability: Value-added assessment. *Forecast: Emerging issues in public education, 1,* 1–4.

Binet, A., & Simon, T. (1980). *The development of intelligence in children* (E. Kite, Trans.). Nashville, TN: Williams Printing. (Original work published 1905)

Bock, R. D., & Wolfe, R. (1996, March 15). *A review and analysis of the Tennessee value-added assessment system, part I: Audit and review of the Tennessee value-added assessment system (TVAAS): Final report.* Nashville, TN: Comptroller of the Treasury. Retrieved May 1, 2008, from http://www.coedu.usf.edu/~dorn/research/tvaas.html

Bracey, G. W. (2004). Serious questions about the Tennessee value-added assessment system. *Phi Delta Kappan, 85,* 716-717.

Brown v. Board of Education, 347 U.S. 483, 1954.

Buros Center for Testing. (2007). *Mental measurements yearbook*. Retrieved March 24, 2008, from http://www.nebraskapress.unl.edu/catalog/productinfo

Buros, O. K. (2008). *The 1938 mental measurements yearbook*. New Brunswick, NJ: Rutgers University Press.

Callender, J. (2004). Value-added student assessment. *Journal of Educational and Behavioral Statistics, 29,* 5.

Carey, K. (2004). The real value of teachers: Using new information about teacher effectiveness to close the achievement gap. *Thinking K–16, 8,* 1–42.

Ceperly, E., & Reel, K. (1997). *Grading teachers, grading schools: The impetus for the Tennessee Value-Added Accountability System*. Retrieved May 2, 2008, from http://library.plymouth.edu/read/297132

Coleman, J. S. (1990). *Equality and achievement in education*. Boulder, CO: Westview.

College Board. (2005). *SAT scores hold steady for college-bound seniors*. Retrieved April 3, 2008, from http://www.collegeboard.com/student/testing/sat/scores.html

Conant, J. (1943). Wanted: American radicals. *The Atlantic* [Online]. Retrieved March 26, 2008, from http://www.theatlantic.com/issues/95sep/ets/radical.htm

Cornwell, C., & Mustard, D. (2002). *Race and the effects of Georgia's HOPE scholarship*. Retrieved April 8, 2008, from http://www.civilrightsproject.harvard.edu/research/meritaid/6Cornwellch4.pdf

Crane, J. (2002). *The promise of value-added testing.* Washington, DC: Progressive Policy Institute. Retrieved on May 8, 2008, from http://www.ppionline.org/documents/Value_Added_Testing.pdf

Delisio, E. (2004, April). Would you switch schools for more money? *Education World*. Retrieved on May 9, 2008, from http://www.educationworld.com/a_admin/admin/admin349.shtml

Dorn, S. (2001). Tennessee value added assessment system commentary. Retrieved May 1, 2008, from http://www.coedu.usf.edu/~dorn/research/tvaas.html

Drury, D., & Doran, H. C. (2003). The value of value-added analysis. *National School Boards Association Policy Research Brief 3,* 1–4.

Dylan, B. (1964). *The times, they are a changin'*. Retrieved March 23, 2008, from http:/www.bobdylan.com/albums/times.html

Education Webmaster. (2005). *The GI bill*. Retrieved March 19, 2008, from http://www.gibill.va.gov/education/GI_Bill.htm

Educational Improvement Act. (1992). Chapter No. 353 enacted March 11, 1992 by the Tennessee Legislature. Retrieved May 1, 2008, from http://www.coedu.usf.edu/~dorn/research/tvaas.html

Elementary and Secondary Education Act. (1965). Section 201.

Fancher, R. E. (1985). *The intelligence men: Makers of the IQ controversy.* New York: W.W. Norton.

Fisher, T. (1996). *A review and analysis of the Tennessee value-added assessment system.* Retrieved April 2, 2005, from http://www.comptroller.state.tn.us./orea/reports/tvaascp2.pdf

Food & Nutrition Service. (2008). *School meals programs*. Retrieved March 21, 2008, from http://www.fns.usda.gov/cnd/governance/notices/iegs/iegs.htm

Francis Galton: The 1st modern attempt (late 1800s). (2003). Key players in the history & development of intelligence & testing. Retrieved March 21, 2007, from http://www.wilderdom.com/

Freed, S. A., & Freed, R. S. (1992). Clark Wissler. In P.H. Raven (Ed.), *National academy of sciences of the United States of America biographical memoirs* (pp. 468-497). Washington D.C.: National Academy Press.

Frontline PBS. (1999a). *Americans instrumental in establishing standardized tests*. Retrieved March 8, 2008, from http://www.pbs.org/wgbh/pages/frontline/shows/sats/where/three.html

Frontline PBS. (1999b). *A brief history of the SAT*. Retrieved March 19, 2008, from http://www.pbs.org/wgbh/pages/frontline/shows/sats/where/history.html

Frontline PBS. (1999c). *Secrets of the SAT*. Retrieved March 25, 2008, from www.pbs.org/wgbh/pages/frontline/shows/sats/interviews/**lemann**.html

Galton, F. (1879). Psychometric experiments. *Brain: A Journal of Neurology*, *II*, 149-162.

Gardner, H. (1985). *Frames of mind*. New York: Basic Books.

Gendren, S. (2007). *Maine state department of education news release on AYP*. Retrieved March 25, 2008, from http://www.ldame.org/docs/2007%20NEWS%20RELEASE%20DOE.pdf

Goertzel, T. (2004). *The myth of the bell curve*. Retrieved February 27, 2005, from http://www.crab.rutgers.edu/~goertzel/normalcurve.htm

Goldschmidt, P., Choi, K., & Martinez, F. (2004). *Using hierarchical growth models to monitor school performance: The effects of the model, metric and time on the validity of inferences.* Paper presented at the Council of Chief State School Officers annual conference on large-scale assessment, Boston.

Goldstein, S. R., Gee, E. G., & Daniel, P. T. (2000). *Law and public education: Cases and materials* (3rd ed.). Charlottsville, VA: Michie.

Goleman, D. (1995). *Emotional intelligence*. New York: Bantam Books.

Gould, S. J. (1981). *The mismeasure of man*. New York: W.W. Norton.

Guilford, J. P. (1954). *Psychometric methods* (2nd ed). NY: McGraw-Hill.

Heller, D. E., & Rasmussen, C. J. (2002). *Merit scholarships and college access: Evidence from Florida and Michigan*. Retrieved April 8, 2008, from http://www.civilrightsproject.harvard.edu/research/meritaid/4MIFLHellerch2.pdf

Herrnstein, R. J., & Murray, C. (1994). *The bell curve*. New York: Free Press.

Hershberg, T. (2005). *Value-added assessment and systemic reform: A response to America's human capital development challenge*. Retrieved May 1, 2008, from http://www.cgp.upenn.edu/pdf/aspen.pdf

Hershberg, T., Simon, V. A., & Lea-Kruger, B. (2004). *Measuring what matters*. National School Boards Association. Retrieved May 8, 2008, from http://www.asbj.com/2004/02/0204asbjhershberg.pdf

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1998). *Applied statistics for the behavioral sciences* (4th ed.). New York: Houghton Mifflin.

Human Intelligence. (2008). *The Wissler controversy*. Retrieved March 24, 2008, from http://www.indiana.edu/~intell/wisslers.shtml

Hume, D. (1998). *An enquiry concerning the principles of morals*. New York: Oxford.

Information Brief. (2004). *ACT and SAT scores in the south: The challenge to lead*. Retrieved April 7, 2008, from http://www.sreb.org/main/highschools/college/Information_Briefing.pdf

James, W. (1958). *Talks with teachers*. New York: W.W. Norton.

Jaspers, K. (1957). *Man in the modern age*. New York: Anchor Books.

Jencks, C., & Phillips, M. (Eds.). (1998). *The black-white test score gap*. Washington, DC: Brookings.

Lakoff, G. (1996). *Moral politics: How liberals & conservatives think* (2nd ed.). Chicago: University of Chicago.

Lemann, N. (1999). *The big test-The secret history of American meritocracy*. New York: Farrar, Straus, and Giroux.

Lewin, T. (2002). *College Board to revise SAT after criticism by university*. *New York Times* [Online]. Retrieved March 26, 2008, from http://query.nytimes.com/gst/fullpage.html

Lewin, T. (2005). *Strivers sharpen no. 2's for different college test*. *New York Times* [Online].. Retrieved March 25, 2008, from http://www.nytimes.com/2005/03/04/education/04SAT.html

Lippmann, W. (1922). *Debunking intelligence experts: Walter Lippmann speaks out*. Retrieved March 11, 2005, from http://historymatters.gmu.edu/search.php?function=print&id=5172

Lohninger, H. (1999). Teach/Me Data Analysis, Springer-Verlag, Berlin-New York-Tokyo. Retrieved October 16, 2008, from http://www.vias.org/tmdatanaleng/bio_pearson.html

Ludlow, L. H. (2008). *Galton: The first psychometrician*? Retrieved March 24, 2008, from http://www2.bc.edu/~ludlow/galton.html

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.

Math Works. (2005). Evaluating goodness of fit. Retrieved March 21, 2008, from http://www.mathworks.com/access/helpdesk/help/toolbox/curvefit/ch_fitt9.html

Matthews, J. (2006, November 16). Just whose idea was this testing? *Washington Post*, p. A06.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability.* Santa Monica, CA: Rand.

McCullough, D. (2001). *John Adams*. New York: Simon & Shuster.

Melville, H. (1952). *Great books of the western world: Moby Dick*. Chicago: University of Chicago. (Original work published 1851)

Meyer, R. H. (1996). Value-added indicators of school performance. In E. A. Hanushek & D. Jorgenson (Eds.), *Improving America's schools: The role of incentives* (pp.197-223). Washington, DC: National Academy Press.

Minton, H. L. (1988). *Lewis M. Terman: Pioneer in psychological testing*. New York: University Press.

National Commission on Excellence in Education. (2001). *A nation at risk.* Retrieved March 20, 2008, from http://www.ed.gov/pubs/NatAtRisk/risk.html

National Education Association. (2004). *Rankings & estimates; A report of school statistics.* Retrieved April 5, 2005, from http://www.nea.org/edstats/images/04rankings-update.pdf

No Child Left Behind Act of 2001, P.L. 107-110.

Olson, L. (2006). Growth models are still the rage**.** *Education Week* [Online]. Retrieved May 7, 2008, from http://www.truescores.com/2006/11/growth-models-are-still-rage.html

PBS. (1998). Eugenics movement reaches its height,1923. *A Science Odyssey* [Online]. Retrieved August 14, 2008, from http://www.pbs.org/wgbh/aso/databank/entries/dh23eu.html

Pearson, K. (1924). *The life, letters and labours of Francis Galton* (Vol. II). Cambridge, MA: Cambridge University Press.

Peterson, G. B. (2004). A day of great illumination: B.F. Skinner's discovery of shaping. *Journal of the Experimental Analysis of Behavior, 11,* 320.

Piaget, J. (1932). *The moral judgment of the child.* London: Routledge & Kegan Paul.

Plato, (1952). *Great books of the western world: The republic* (B. Jowett, Trans.). Chicago: University of Chicago. (Original work published 370 BC)

Plucker, J. (2004a). *Human Intelligence: Alfred Binet*. Retrieved February 26, 2008, from http://www.indiana.edu/~intell/binet.shtml

Plucker, J. (2004b). *Human intelligence: Arthur Jensen.* Retrieved April 3, 2008, from http://www.indiana.edu/~intell/jensen.shtml

Plucker, J. (2004c). *Human intelligence: David Wechsler.* Retrieved February 27, 2008, from http://www.indiana.edu/%7Eintell/wechsler.shtml

Pratt, C. C. (1917). Book review: The measurement of intelligence. *Journal of Applied Psychology, 1*,191–192.

Ravitch, D. (2000). *Left back*. New York: Simon & Schuster.

Ravitch, D. (2006). The Fall of the standard bearers. The *Chronicle of Higher Education* [Online]. Retrieved March 25, 2008, from http://chronicle.com/free/v52/i27/27b0f4401.htm

*Regents of the University of California v. Bakke***.** (1978)**.** 438 U.S. 265.

Robinson, D. (1997). *The great ideas of psychology: Part II, the great courses*. Retrieved April 2008, from http://www.teach12.com/

Robinson, D. (1998). *The mind*. New York: Oxford University Press.

Roid, G. H. (2003). *Stanford-Binet intelligence scales* (5th ed.). Itasca, IL: Riverside.

Rousseau, J. (1762). *Emile*. North Clarendon, VT: Everyman Library.

Rubin, D. B., Stuart, E. A., & Zanutto, E. L. (2004). A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics, 29,* 103–116.

Russell, B. (1950). *Unpopular essays*. New York: Simon & Schuster.

Sacks, P. (1999). *Standardized minds.* Cambridge, MA: Perseus.

Sanders, W. (1998). Value-added assessment. *The School Administrator, 12,* 24-32.

Sanders, W. L., & Horn, S. (1994). The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education, 8,* 299-311.

Seligman, M. E. (1998). *Learned helplessness*. Retrieved March 26, 2008, from http://www.noogenesis.com/malama/discouragement/helplessness.html

Shaughnessy, M. F., & Moore, J. (2005). *The wobegon intelligence test*. Retrieved April 30, 2008, from http://www.bouldertherapist.com/html/humor/MentalHealthHumor/wobegon_intelligence_test.htm

Sizer, T. R. (2004). *The red pencil*. New Haven, CT: Yale University Press.

Skinner, B. F. (1991). *The behavior of organisms*. Acton, MA: Copley. (Original work published 1938)

Spearman, C. (1937). *Psychology down the ages*. London: McMillan.

State Department of Education. (2008). *SDE directory*. Retrieved April 25, 2008, from http://www.k-12.state.tn.us/sde/Searches/SimpleSearchSchool.asp

State of Tennessee. (2008). *State of Tennessee statewide report card 2007*. Retrieved April 4, 2008, from http://www.k-12.state.tn.us/rptcrd07/state1pf.asp

Stern, W. (1914). *The psychological methods of testing intelligence* (G. Montrose, Trans.). Baltimore: Warwick & York.

Sternberg, R. J. (1990). *Metaphors of mind: Conceptions of the nature of intelligence.* Cambridge, MA: Cambridge University Press.

Summers, A. A. (2002, Summer). Expert measures. *Education Next* [Online]. Retrieved May 7, 2008, from http://www.educationnext.org/20022/16.html

Sweet, M. (2004). James Cattell. *Britannica Concise Encyclopedia* [Online]. Retrieved May 20, 2008, from http://www.encyclopedia.com/doc/1B1-360053.html

Symynet Educational Statistics Resources. (2004). *Normal distribution curve formula.* Retrieved March 20, 2008, from http://www.famousbelgians.net/quetelet.htm

Tennessee Advisory Commission on Intergovernmental Relations. (2000). *Changes in state formula improve equity in school funding*. Retrieved April 20, 2008, from http://www.state.tn.us/tacir/PDF_FILES/Education/educationbrief.pdf

Tennessee Department of Education. (2001). *Tennessee plan for implementing the teacher and paraprofessional quality provisions of the no child left behind act of 2001.* Retrieved April 14, 2008, from http://tennessee.gov/education/nclb/doc/NCLB_ImpPlan_08_18_05.pdf

Tennessee Department of Education. (2007a). *Gateway proficiency levels: Gateway and end-of-course proficiency levels*. Retrieved March 2008, from http://tennessee.gov/education/tsgatewayproflvl.htm

Tennessee Department of Education. (2007b). *Report card 2007*. Retrieved April 6, 2008, from http://www.k-12.state.tn.us/rptcrd07/

Tennessee Department of Education. (2007c). *Consolidated state accountability workbook.* Retrieved May 9, 2008, from http://www.ed.gov/admins/lead/account/stateplans03/tncsa.pdf

Tennessee Education Association. (2005). *Value-added assessment not popular in education: Committee senators question cost-exclusive arrangement with Sanders*. Retrieved April 9, 2008, from http://www.teateachers.org/legreports/currentreport.htm

Tennessee Education Association. (2006). *TEA differentiated pay guidelines*. Retrieved May 9, 2008, from http://www.teateachers.org/News.asp?nid=16

Tennessee Education Lottery Scholarship Program. (2007). Retrieved May 12, 2008, from http://www.state.tn.us/sos/rules/1640/1640-01-19.pdf

Tennessee Education Lottery Scholarship Program. (2008). *Annual report*. Retrieved May 12, 2008 from http://www.collegepaystn.com/mon_college/sch_app_pdfs/LotteryReport2008.pd

Tennessee Higher Education Commission. (2008). *Overview of fall 2007 recipients of the Tennessee education lottery scholarship program*. Retrieved April 8, 2008, from http://www.state.tn.us/thec/2007web/division_pages/pub_news_pages/current_publications/FINAL lottery report.pdf

Tennessee Report Card. (2007). *Report card terminology.* Retrieved March 16, 2008, from http://www.k-12.state.tn.us/rptcrd07/rptcrdterms.htm

Tennessee Small School Systems (TSSS) v. McWherter. (1993). 851 S.W. 2d 139 (Tenn. 1993).

Tennessee State Board of Education. (2007). *Guidelines for differentiated pay plans public chapter 376.* Retrieved May 9, 2008 from http://state.tn.us/sbe/Nov07/VB_Diff_Pay_Guidelines.pdf

Tennessee State Constitution. (2007). *Tennessee department of education history*. Retrieved April 10, 2008, from http://www.state.tn.us/sos/bluebook/online/

Tennessee Student Assistance. (2004). *Tennessee scholar dollar*. Retrieved April 8, 2008, from http://www.state.tn.us/tsac

Tennessee Value-Added Assessment System. (1997). *Graphical Summary of educational findings from the Tennessee Value-Added Assessment System*. Knoxville, TN: University Value-Added Research and Assessment Center.

Terman, L. (1916). *The measurement of intelligence*. Boston: Houghton Mifflin.

Terman, L. M., & Merrill, M. A. (1960). *Stanford-Binet Intelligence Scale: Manual for the Third Revision Form L-M*. Boston: Houghton Mifflin.

Terman, L. M., & Merrill, M. A. (1973). *Stanford-Binet Intelligence Scale: Manual for the Third Revision Form L-M* (1972 Norm Tables by R. L. Thorndike). Boston: Houghton Mifflin.

Thorndike, R. L. (1982). *Applied psychometrics*. Boston: Houghton.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford-Binet intelligence scale* (4th ed.). Itasca, IL: Riverside Publishing.

Touche, T. (1999). *THE BIG TEST: The Secret History of the American Meritocracy. - Review - book reviews.* Retrieved April 3, 2008 from http://findarticles.com/p/articles/mi_m1316/is_12_31/ai_58170294

U.S. Census Bureau. (2003). *Income 2003*. Retrieved April 3, 2008, from http://www.census.gov/hhes/income/income03/statemhi.html

Ulich, R. (1961). *Three thousand years of educational wisdom*. Cambridge, MA: Harvard University Press.

United States Department of Education. (2006). *Secretary Spellings approves Tennessee and North Carolina growth model pilots for 2005-2006*. Retrieved May 7, 2008, from http://www.ed.gov/news/pressreleases/2006/05/05172006a.html

United States Department of Education. (2007**).** *Secretary Spellings invites eligible states to submit innovative models for expanded growth model pilot.* Retrieved May 7, 2008, from http://www.ed.gov/news/pressreleases/2007/12/12072007.html

United States Supreme Court. (1978). *University of California regents v. Bakke*, 438 U.S. 265 (1978) 438 U.S. Retrieved May 20, 2008, from http://caselaw.lp.findlaw.com/scripts/getcase.pl?court=US&vol=438&invol=265

Vaughan, A. C. (2002). Standards, accountability, and the determination of success. *The Educational Forum, 66,* 206- 213.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Watson, J. B. (1924). *Psychology from the standpoint of a behaviorist* (2nd ed.). Philadelphia: J.B. Lippincott.

Watson, J. B. (1930). *Great books of the western world*: *The laws* (B. Jowett, Trans.). Chicago: University of Chicago. (Original work published 348 BC)

Webb, P. B. (2005). *The associations between the scores on the ACT test and Tennessee?s value-added assessment in 281 Tennessee high schools*. Retrieved November 1, 2008, from http://etd-review.etsu.edu/ETD-db/ETD-browse/browse?first_letter=E&browse_by=department

Wilson, E. O. (1998). *Consilience*. New York: Vintage Books.

Wissler, C. (2004). *The Wissler controversy*. Retrieved February 26, 2008, from http://www.indiana.edu/~intell/wisslers.shtml

Witte, J., & Witte, R. (2007). *Statistics*. Hoboken, NJ: John Wiley & Sons.

Yerkes, R. M. (Ed.). (1920). *The new world of science*. New York: Century.

VITA

PHILLIP L. SWANSON

Personal Data: Date of Birth: April 20, 1960
Place of Birth: Sweetwater, TN
Marital Status: Single

Education:

Tennessee Wesleyan College, Athens, TN;
B.S. in English Education;
1982

Lincoln Memorial University, Harrogate, TN;
M.S. in Educational Administration and Supervision
1989

Lincoln Memorial University, Harrogate, TN;
E.D.S. in Curriculum and Instruction,
1998

East Tennessee State University, Johnson City, TN;
Educational Leadership and Policy Analysis, Ed.D.;
2009

Professional Experience

Teacher/Coach
Monroe County School System, Madisonville, TN,
1983-90

Principal,
Tellico Plains High School;
1990-92

Asst. Principal/Principal,
McMinn High School/Vocational School, Athens, TN;
1992-2004

Supervisor of Secondary Education,
McMinn County School System, Athens, TN;
2004-2009