

Summer 2008

Standardized Testing: Are the Benefits Really there?

Alan D. Foster
Regis University

Follow this and additional works at: <http://epublications.regis.edu/theses>



Part of the [Arts and Humanities Commons](#)

Recommended Citation

Foster, Alan D., "Standardized Testing: Are the Benefits Really there?" (2008). *All Regis University Theses*. Paper 455.

This Thesis - Open Access is brought to you for free and open access by ePublications at Regis University. It has been accepted for inclusion in All Regis University Theses by an authorized administrator of ePublications at Regis University. For more information, please contact repository@regis.edu.

Regis University
College for Professional Studies Graduate Programs
Final Project/Thesis

Disclaimer

Use of the materials available in the Regis University Thesis Collection ("Collection") is limited and restricted to those users who agree to comply with the following terms of use. Regis University reserves the right to deny access to the Collection to any person who violates these terms of use or who seeks to or does alter, avoid or supersede the functional conditions, restrictions and limitations of the Collection.

The site may be used only for lawful purposes. The user is solely responsible for knowing and adhering to any and all applicable laws, rules, and regulations relating or pertaining to use of the Collection.

All content in this Collection is owned by and subject to the exclusive control of Regis University and the authors of the materials. It is available only for research purposes and may not be used in violation of copyright laws or for unlawful purposes. The materials may not be downloaded in whole or in part without permission of the copyright holder or as otherwise authorized in the "fair use" standards of the U.S. copyright laws and regulations.

STANDARDIZED TESTING;
ARE THE BENEFITS REALLY THERE?

by

Alan D. Foster

A Research Project Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Arts in Liberal Studies
Adult Education / Statistics

Regis University
May 2008

STANDARDIZED TESTING;
ARE THE BENEFITS REALLY THERE?

by

Alan D. Foster

has been approved

May 2008

APPROVED:

. Course Facilitator

.Faculty Advisor

ABSTRACT

STANDARDIZE TESTING; ARE THE BENEFITS REALLY THERE?

The proponents of standardized high-stakes testing argue that this type of assessment will and does promote increased knowledge acquisition by students. The opponents of this testing claim any apparent knowledge increase as measured by improvements in tests scores are mainly due to teaching to the test. This study was done to determine whether high-stakes testing does in fact improve the knowledge of the students. Data from two national college entrance examinations were collected and reviewed changes over time as reference to knowledge improvements. Test scores from three state high-stakes tests were collected (California, Texas, and Colorado). The scores from the national tests both the national averages (composite scores) and the individual state averages were tabulated and plotted. The relative change over time was determined for the national examinations and the state examinations.

The results of this study call into question the hypothesis that high-stakes testing improves results of education. The national examinations improved at approximately the same rate for the period examined (12-14 years) while the state test scores show greater changes either positively or negatively for the final 5-6 years examined. One conclusion that can be drawn by that is that the state test scores improved based on changes in what is being taught in the classroom, i.e. teaching to the test.

The conclusion of this work is that the hypothesis that high-stakes testing will improve the knowledge of student must be rejected. Because of that the use and value of high-stakes testing is called into question and should be reevaluated and if not just eliminated then these tests need to be revised both in content and emphasis.

TABLE OF CONTENTS

| | |
|---------------------------------------------|----|
| 1. INTRODUCTION..... | 1 |
| Statement of the Problem | 2 |
| Background of the Problem..... | 3 |
| Purpose of the Project..... | 4 |
| Chapter Summary..... | 5 |
| 2. REVIEW OF LITERATURE | 7 |
| Chapter Summary | 19 |
| 3. METHOD | 20 |
| Chapter Summary | 22 |
| 4. RESULTS..... | 23 |
| Chapter Summary..... | 35 |
| 5. DISCUSSION..... | 36 |
| REFERENCES | 40 |
| Appendices | |
| A. Goals 2000 | 42 |
| B. No Child Left Behind Summary..... | 46 |
| C. California Testing Program (STAR)..... | 48 |
| D. Texas Testing Programs (TAAS, TAKS)..... | 51 |
| E. Colorado Testing Program (CSAP)..... | 52 |

LIST OF TABLES

| | |
|----------------------------------------|----|
| 1. ACT composite scores..... | 26 |
| 2. Combined SAT Scores..... | 28 |
| 3. California STAR Scores..... | 30 |
| 4. Colorado CSAP Scores..... | 32 |
| 5. Texas State Tests (TAAS, TAKS)..... | 34 |

LIST OF FIGURES

| | |
|----------------------------------------|--------|
| 1. ACT composite scores..... | 26 |
| 2. Combined SAT Scores..... | 28 |
| 3. California STAR Scores..... | 30 |
| 4. Colorado CSAP Scores..... | 32 |
| 5. Texas State Tests (TAAS, TAKS)..... | 34, 35 |

Chapter 1

INTRODUCTION

Beginning in the early 1980's, there was a movement underway to establish standards or goals for the United States educational system. In part, this was driven by the 1983 report, *A Nation at Risk* from the National Commission on Excellence in Education (NCEE; 1983). The Commission members were appointed by Secretary of Education Bell, in 1981, in response to his concern about "the widespread public perception that something is seriously remiss in our educational system" (p. 1). The development of standards was undertaken following the publication of the Commission's report as a way to assure that U.S. students would be prepared to compete in the global market. The primary focus was to establish standards for mathematics, science, English, and social studies.

The initial work on the standards was conducted at the national level driven in part by the results of the Education Summit held by President Bush and the U.S. governors in 1989, but the development of standards was passed onto the States for implementation (Gordon, 2003). The need for assessments and the type of assessments needed to monitor progress toward attainment of the goals put forth as standards has become a source of debate (Fetler, 1994; Neill, 2003; Sanders & Horn, 1995; Seidman, 1996). Currently, in some states the scores from the assessments are used to determine grade progression and high school graduation (Albrecht & Joles, 2003; Neill, 2003; Wright, 2002). This has led to discussions as to whether the use of *high-stakes* testing has a positive effect on progress in education (Albrecht & Joles, 2003; Chudowsky &

Pellegrino, 2003; Hillcocks, 2002; Neill, 2003; R.Valencia & Villarreal, 2003; Schemo, 2003; Schemo & Fessenden, 2003).

The next major education legislation from the Federal government was The Goals 2000: Educate America Act (United States Department of Education, 1994), which became law in 1994 and was amended in 1996. Goals 2000 was the legislative out come to the Bush Education Summit. This legislation represents a larger scale approach for “improving student learning through long-term, broad-based effort to promote coherent and coordinated improvements in the system of education throughout the Nation at the State and local levels. Goals 2000 was followed in 2001 by the No Child Left Behind Act (NCLB) (U.S. Department of Education, 2005). These acts and the federal education funding tied to them have acted as the major forces driving the implementation of standards based reform at the state and local levels.

Terms

Terms and concepts which will be used as the issue of Standardized Testing is explored further are: (Albrecht & Joles, 2003)

- Standardized tests are assessments, which are written to measure knowledge of students at a given grade, level and are given to all students at that grade level. These tests are usually written to assess predefined standards for that grade level.
- Either high-Stakes tests are tests whose results have consequences for the student or the school or both, these consequences can involve, but are not limited to grade progression or graduation for the student and funding for the school. The consequences are usually negative.

- Academic performance is a measure of performance, which as it relates to performance of the students relative to standards set for his or her grade level.

Statement of the Problem

The problem that was be explored in this project is whether the use of standardized assessments, as they are currently being used, results in improved academic performance. The current thinking in some academic circles (Berliner, 1993; Fetler, 1994; Sanders & Horn, 1995) and much of the general population is that the use of the scores from standardized tests to drive curriculum and teaching methods will improve the academic achievement of students in Grades K-12 by focusing the teaching on defined standards. There is a body of literature on the topic of assessments and improved academic performance, but as Berliner pointed out, much of the legislation that supports high stakes testing is based on opinion, perceptions, and misperceptions put forth to meet a political goal. In the report, *A Nation at Risk* (N. C. o. E. i. Education, 1983), which seemed to be the catalyst for much of the current thought, the Commission members recommended that higher standards be developed and that the students, teachers, and schools needed to be held accountable to those standards. However, assessments were not suggested as the primary means to accomplish the goals. Most of the current work based on longitudinal student data from standardized assessments does not fully support the use of high stakes testing as it is being used in many states (Abrams, Pedulla, & Madaus, 2003; Albrecht & Joles, 2003; Chudowsky & Pellegrino, 2003; Hillcocks, 2002; Neill, 2003; Schemo, 2003; Schemo & Fessenden, 2003; Tretter & Jones, 2003; Valencia & Villarreal, 2003). Only recently have studies been published indicating that these policies

in some cases have driven improvement in student achievement both at the state and national level ((Greene, Winters, & Forster, 2003).

Background of the Problem

Standardized testing is the result of the movement to establish content standards for the United States public schools. Once the content standards were formalized and agreed upon, an assessment to measure the results from the use of the standards was the next logical step. The assessments have been developed by individual states for use in their public schools and, to date, 49 states have adopted some form of standardized testing for their public school students (Officers, 2002).

This author's initial review of the key literature about standardized testing was that much of the original legislation and policy that supports high-stakes testing was based more on opinion than fact, while most of the current field research does not support the use of high-stakes testing in its present form. Because of this, opinions and theory rather than peer reviewed results seem to be major factors that have influenced policy makers (e.g., state legislators, school boards, etc.). The logical inference from this would be that many legislative decisions about the use of standardized testing have been and are being made based on opinion, political expediency and/or theory more than on well-designed research using measurable results.

Purpose of Project

The purpose of this project was to analyze available data with respect to the outcomes of standardized testing, over time, and use that data to test the premise that high-stakes testing will improve academic knowledge (i.e., performance). Various standardized tests have been developed as an attempt to measure student performance at discrete points in the public schools, Grades K-12. What seems to be missing to date is sufficient review of the testing results against an established standard or standards so that the premise of high-stakes testing can be assessed or even whether the assessments themselves are a valid measure of what is being assessed.

Chapter Summary

Current educational policy being put forth by all of the States is that the use of assessments will advance the progress of their students toward meeting current educational standards. Different states and educational entities have taken these testing policies to different levels and directions ranging from a means of tracking progress to defining grade progression in the school systems.

The use of “high stakes” assessments is and has been questioned by a large portion of the current educational research. There is evidence that high-stakes testing can have a negative effect on actual educational progress. All of this information brings into question types of assessments that are currently being used and what the motivating factors are that are driving their use. Much of the support for “high stakes” testing has been from the political arena.

Chapter 2 provides a review of the current published literature on educational assessments. Chapter 3, describes the methodology, which will be used to evaluate the

data obtained from selected standardized testing. Chapter 3 will also define the statistical methods for evaluation of the data that were used in an attempt to answer the question, does high stakes testing lead to improved student performance (i.e., increased knowledge) as measured by standardized college entrance examinations or standardized national tests.

Chapter 2

REVIEW OF LITERATURE

In reviewing assessments that are in place at this time and how they have evolved to the current level of high stakes testing a look back at U. S. educational policies is required. Gryphon and Meyer (2003) in a review of U. S. educational policy for the Cato Institute attempted to put today's government mandated assessments in historical context. The authors' research indicates that schools in the U. S. were slowly being moved toward some form of government control beginning in the mid 19th century. Prior to this time schools were mainly under local parental (community) control, with little higher-level government involvement. This movement toward more government control was accelerated at the beginning of the 20th century. Many factors combined and drove this shift toward increased government control of education, some of which were high numbers of immigrants from non-Protestant Europe and Asia and the political ramifications related to World War I. The prejudice involving these immigrant groups and the war lead to education being more fully moved under government control as much for political reasons as anything else. This control allowed the government agencies to take control of curriculum, teacher training, etc., which has lead progressively to the current need by government to assess the schools and students under the guise of accountability.

The current focus began to take shape in the mid 20th century (Hillocks, 2003; Gryphon and Meyer, 2003) with the public's belief that the U.S. educational system was not performing well. At the time the use of college entrance exams had become the norm and the scores for these exams (Scholastic Aptitude Test (SAT) and American College

Testing Program (ACT)) had begun a steady decline in the 1960's. This decline and the perception that the U.S. education system was responsible lead to the 1983 report by the National Commission on Excellence in Education titled a *Nation at Risk*. This report was very critical of the educational system and policies in place at the time and made several recommendations (NCEE, 1983) to address the problems the commission believed existed.

The NCEE's report (*A Nation at Risk*, 1983) recommended changes in the following areas: a) Content, that all high school graduation requirements be strengthened; b) Standards and Expectations, all schools, colleges, and universities should adopt more rigorous and measurable standards and higher expectations; c) Time, significantly more time should be devoted to learning, i.e. more effective use of the school day, lengthening of the school day, and lengthening of the school year; d) Teaching, this recommendation consists of seven parts, all of which require higher standards for all persons involved in the educational system; e) Leadership and Fiscal Support, all levels of leadership must be held accountable for the status of the educational system and fiscal support must be increased at all levels (local, state, and federal) of government.

In 1989, the nation's Governors and President Bush met for an Education Summit. From these meetings the groundwork was laid for National Education Goals, which became part of the *Goals 2000 Education Program*. With the passing of the Goals 2000: Educate America Act on March 31, 1994 the federal government was given a new and more involved role in U. S. education policy (United States Department of Education, 1994). This new role allows the federal government to promote a more comprehensive approach toward education.

The Goals 2000: Educate America Act (United States Department of Education, 1994) created the National Educational Standards and Improvement Council (NESIC), which gave standards a greater role in national education policy. In a review (Mulcahy, 1994) of the Goals 2000 act Mulcahy identifies five basic duties that the NESIC was given; 1. to identify areas in which standards ought to be developed; 2. to certify the content and performance standards; 3. to identify and develop the criteria for certifying such standards; 4. to develop criteria for certify, and assist in the development of exemplary nation opportunity-to-learn standards; and 5. to certify State assessments.

The Goals 2000 Act (United States Department of Education, 1994) under Title 1 – National Education Goals, Sec. 102. The following is a summary of the National Education Goals (United States Department of Education, 1994), the complete goals are in Appendix A:

1. *SCHOOL READINESS.--*
 - 1a. *By the year 2000, all children in America will start school ready to learn.*
2. *SCHOOL COMPLETION.--*
 - 2a. *By the year 2000, the high school graduation rate will increase to at least 90 percent.*
3. *STUDENT ACHIEVEMENT AND CITIZENSHIP.—*
 - 3a. *by the year 2000, all students will leave grades 4, 8, and 12 having demonstrated competency over challenging subject matter including English, mathematics, science, foreign languages, civics and government, economics, arts, history, and geography, and every school in America will ensure that all students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment in our Nation's modern economy.*
4. *TEACHER EDUCATION AND PROFESSIONAL DEVELOPMENT.—*
 - 4a. *By the year 2000, the Nation's teaching force will have access to programs for the continued improvement of their professional skills and the opportunity to acquire the knowledge and skills needed to instruct and prepare all American students for the next century.*
5. *MATHEMATICS AND SCIENCE--*
 - 5a. *By the year 2000, United States students will be first in the world in mathematics and science achievement.*

6. *ADULT LITERACY AND LIFELONG LEARNING.--*
 - 6a. *By the year 2000, every adult American will be literate and will possess the knowledge and skills necessary to complete in a global economy and exercise the rights and responsibilities of citizenship.*

7. *SAFE, DISCIPLINED, AND ALCOHOL AND DRUG-FREE SCHOOLS—*
 - 7a. *By the year 2000, every school in the United States will be free of drugs, violence, and the unauthorized presence of firearms and alcohol and will offer a disciplined environment conducive to learning.*

8. *PARENTAL PARTICIPATION.—*
 - 8a. *By the year 2000, every school will promote partnerships that will increase parental involvement and participation in the social, emotional, and academic growth of children.*

In 2001 President George W. Bush, just after taking office, introduced the No Child Left Behind (NCLB) legislation which he described as the “cornerstone of my administration” (U.S. Department of Education, 2005). Less than a year later the legislation was passed and the NCLB became law. It reauthorized the ESEA and added strategies and principles defined by President Bush. These include; increased accountability for States, school districts, and schools; greater choice for parents and students, particularly those attending low-performing schools; more flexibility for States and local educational agencies (LEA) in the use of Federal education dollars; and a stronger emphasis on reading, especially for our youngest children (Secretary, 2002).

The NCLB Act (Secretary, 2002) places accountability at the forefront with its requirements that all states implement accountability systems for public schools and students under Title I programs. States and schools must implement annual assessments for grades 3 – 8 with the goal to be that all students meet proficiency levels by the end of 12th grade. Schools must also establish annual proficiency goals and make annual yearly progress (ALP) toward those goals or face funding consequences.

The Act also provides for greater school choice by requiring local education authorities (LEA) to provide students attending under performing schools the opportunity

to attend another school of choice. The LEA is required to set aside up to 20% of their Title I funds to pay for these needs.

Current Research

The use of standardized assessments in schools is currently being driven by the belief that their use will improve academic achievement for students Grades K-12. Many of the arguments for the use of standardized assessments are based more on personal and political beliefs than on solid scientifically assembled data (Berliner, 1993). This leaves open the possibility that the use of standardized assessments may have very little effect on improving the academic performance of students (Abrams et al., 2003; Albrecht & Joles, 2003; Chudowsky & Pellegrino, 2003; Hillcocks, 2002; Neill, 2003; Schemo, 2003; Schemo & Fessenden, 2003; Tretter & Jones, 2003).

Berliner (1993) cited several cases where misinformation was used to influence educational policy; this was being done using media reports and published books. One such piece of misinformation was the reported shift in the Scholastic Aptitude Test scores, as cited by Berliner (1993) Carson, Huelskamp, and Woodall (1991) had shown a marked decrease in mean score over a 25 year period (1965 –1990); the actual drop had been only 3.3 % of the raw score total. This is a small and perhaps not a significant shift when one examines the population of students who took the test. More students from a wider range of academic levels took the test at the end of the 25-year period than at the beginning. This holds true to an even greater extent in 2005, 12 years after this report. What continues to be omitted or minimized is the shifts in demographics of the populations being tested and the resulting influence that has on the testing results.

Chudowsky and Pellegrino (2003) argued that in most current assessments, students are measured incorrectly because the assessments are not designed to measure and/or support student learning. According to these authors, most current and

past assessments are designed to measure aptitude and not learning ability. Because of this, the use of most current assessments cannot provide data, which can be used to measure learning progression. They believe that new assessments must be designed which can provide useful feedback about learning progression. The authors also contended that most of the current assessments being used were put into place too quickly and since they cannot assess student learning they are not a useful tool for the measurement of the effectiveness of instruction. Because of this, the measurement of student proficiency, relative to standards, is not currently being done. This point is also brought up in the study by (Greene et al., 2003), but is not fully explored as part of the reason behind the shift in the ranges of the data while the scores remain somewhat steady.

Valencia and Villerreal (2003) analyzed the Texas high-stakes testing plan for Texas public school students. Their review of the Texas data suggested that continued use of the current assessments would have a negative outcome on student performance rather than improve student performance. In the current Texas plan, high-stakes testing results for reading are used to determine progression of third grade students into the next grade. The authors believe that other alternate plans, for example a universal preschool education, would have a higher degree of success. They cited the Donovan and Cross (2002) study of Head Start, in which it was indicated that the early learning of print, sounds, and writing made it more likely that the students would be successful readers in elementary school. Further, they recommended the use of early screening and intervention in reading as a means to maximize the probability of students being able to pass the third grade reading test. The Valencia and Villereal cited an earlier study in which, Valencia (2002a) supported the use of high quality K-3 Spanish reading programs as a means to improve student success for those students whose first language is Spanish. Finally, they cited Valencia, (2000a) in which it was indicated that

certified, highly qualified teachers were as vitally important in student success. This latter conclusion was also one of the conclusions in the report, *A Nation at Risk* (N. C. o. E. i. Education, 1983).

Schemo and Fessenden (2003) examined the Houston Texas School system in a study looking at the effects of the Texas high stakes test. They interviewed graduates and parents of the Houston schools and compared data from the Texas tests and the Stanford Achievement Test, a national examination. The testing data indicated that Houston students who took the Stanford exam in 2002 and 1999 did not show improvement in reading or mathematics when compared to other students nationally. However, the tests scores for the Texas test did show improvement. The authors suggested that the gains shown on the Texas tests are unusually high when compared to the scores of the same group of students on the Stanford exam. This discrepancy suggests that teaching to the test may have been occurring, which would inflate the test results but not improve the students actual performance or knowledge. They also interviewed several people outside of the schools as a means to gather information on the public view of the testing and apparent improvement. The overall response from the interviews was that some education material was omitted by the teachers, because the focus was on teaching to the test and this resulted in sometimes omitting some general knowledge. One example, which Schemo and Fessenden used, was a college student who had graduated from a Houston school with a B average. This student struggled in college because she had never learned to take notes in high school. Another example was that her brother was frustrated with the SAT college entrance test because of the vocabulary; the Houston high school, which he attended, did not place emphasis on vocabulary.

Tretter and Jones(2003) used a case study to investigate the relationship between inquiry based instructional style and student scores on the North Carolina standardized tests. During the first two years of the study, the students were instructed with a more traditional lecture style with low levels of inquiry. The following two years the students were taught by the use of inquiry methods. The findings from this study suggested that the use of inquiry techniques did not significantly alter students' achievement, as measured by the North Carolina standardized test for physical science. However, use of the inquiry methods did increase student participation and higher grades within the classroom. In addition, the use of inquiry-based instruction resulted in more uniform achievement of the students than did the traditional instruction, in both the classroom measures and in more objective standardized tests.

Tretter and Jones (2003) cited Costenson and Lawson (1986) who interviewed teachers and reported that their 10 top reasons for not using inquiry based instruction. The top two reasons were “too much time and energy are required” and “too slow content coverage” (p. 345). In the authors' study classroom grades improved with the inquiry-based instruction; however, the end of course test showed a lower mean, though not statistically significant, than the traditional lecture based instruction. This all points to a conflict, which is, that while the added effort required for inquiry based instruction improved classroom participation, students scores on the North Carolina standardized test did not reflect what appears to be an improvement in student proficiency as measured by in class performance.

In another study (Jorgenson & Vanosdall, 2002) the authors reviewed the effects of standardized testing on inquiry based methods for science education. They concluded that the increased use of standardized testing, i.e. high-stakes testing, will ultimately put an end to inquiry based science education which has been shown to be

“an effective strategy...that has been shown to improve science learning, promote critical thinking and problem solving, and even improve reading, writing, and math achievement” (p.4).

Hillcocks (2002) reported that curriculum is impacted more by testing than testing is by curriculum. His point was that generally teaching to the test follows assessments if those assessments are politically mandated and carry consequences such as graduation, grade progressing and/or school funding. Hillcocks argued that the use of high stakes testing has a negative impact on overall learning. He believes that these assessments markedly influence the content of the material that teachers present. He observed that when high stakes testing is used content is restricted to what is specifically being assessed rather than the instructor presenting a broader range of material.

In an article written for Phi Delta Kappan (Meier, 2002) the author examined the effects of the current application of standards to education is having. The author points out that there is great appeal to using standards and the tests that go with them to measure educational progress. One of the arguments used by standards based reformers is that it is possible to design tests that will measure educational progress. This is in direct conflict with the work of (Chudowsky & Pellegrino, 2003) in which they argue that current assessments are unable to measure a student's learning. The appeal for the concept that a test can be used to measure everyone's knowledge against an education standard is the apparent simplicity that this would bring to evaluation of the educational system. This type of testing would allow not only the measurement of a individuals relative placement against a standard, but it will allow for ranking schools themselves against the same standard. However much of the current research does not support this concept.

Meier (2002) argues that the development of a test of this nature leads to a test that pushes the education system to have all participants “pass” it with a high score. This concept leads to teaching to the test in order for the desired outcome to occur, this is called curriculum and test alignment. While there is not much difference in these assessments from the traditional norm-referenced test without the use of distributions (bell curve) it can be difficult to determine what should be in the test and how to set expectation, scores, and cutoffs. This leads to a greater influence of political pressures on the scoring and less influence of what can actually be expected from a group of students at a particular level in the educational process.

In addition, the use of testing to measure students against the standards has led several states to use the test scores to determine grade progression and high school graduation. Meier (2002) and several others (Butler, 2003; Chudowsky & Pellegrino, 2003; Gordon, 2003; Hillcocks, 2002; Neill, 2003; Schemo & Fessenden, 2003; Valencia & Villarreal, 2003) have noted that this is contrary to current research, which has shown that holding students back has a negative impact on their likelihood of success later in their educational career, it has also been found that this policy has a negative impact on graduation rates. Another effect of this has been that in order to appear more successful, schools and teachers have begun teaching to the test by modifying curriculum to match what is being tested.

Meier (2002) further concludes that rather than standardize teaching to meet a test educators should be setting and using real standards and exercising personal and professional judgment as to how individuals are meeting those standards. This will require more varied approaches by educators, students, and parents as to how to best meet the standards. Standardized testing cannot be the only tool, which is used to measure educational progress in our school systems.

In another study (Moon, Brighton, & Callahan, 2003) to evaluate the effect of high-stakes testing on gifted students reported several disturbing trends. Elementary teachers and students were surveyed in Texas, California, and Virginia and survey included schools in urban, suburban, and rural areas. The study involved instructional practices and perceptions of the influences that these tests had on curricula, instructional process, and student attitudes toward school.

The general trend reported from all of the schools by (Abrams et al., 2003; Moon et al., 2003) was that the tests had reduced the time allotted for other material, which was not being included in the high-stakes tests. The preparation for the tests includes increase use of worksheets, test-taking strategies, and review and practice of state released test material. Another effect is that the teaching to the test reduces the amount of higher-level material covered. The effect of this is that gifted students reported less engagement in learning. This also leads to the result that none of the students (gifted or not) are challenged beyond the level of the testing material.

Teachers (Abrams et al., 2003; Moon et al., 2003) reported that they focused almost entirely on topics included on the tests and gave little time or attention to other areas, i.e. fine and performing arts. They reported that most of the classroom assessments were designed as drills for the state tests. Other types of classroom assessments were only use after the state tests were completed.

Abrams and associates (2003) after reviewing published literature and a nationwide survey of teachers in which the participated. Their conclusion was that the survey supported the published literature in that high-stakes testing focuses more of the time in the classroom on the test and less time on general knowledge.

According to Moon et. al. (2003) a perception held by both students and teachers was that the gifted students had to perform very well on the tests in order to “pull up” the scores of the less talented students. This put undue pressure on those gifted students,

to which they responded in various ways and not always positively. This pressure and the unwillingness of teachers to provide more than one approach to test preparation are likely to yield lower scores from these students than might have been expected.

Overall, this study (Moon et al.,2003) suggests that high-stakes testing by the states is having a negative impact on the quality of education that not only the gifted students are getting but all of the others students. Due to the limiting effect, the preparation for the tests has on the quality and quantity of material that is covered in the classroom. Teachers reported that regardless of academic ability they spent substantially greater time preparing for state mandated tests prior to the testing than afterward. They report that they attempted to simulate the testing experience and consequently the classroom activities tended to focus on isolated skills and to emphasize facts and rules. The majority of teachers when asked about how the testing affects their instructional practices indicated that they teach to the test more than they would without the tests. Greater than 26% of all the teachers, again regardless of the student's academic level, indicated that they omit material they feel would enhance the students learning and focus only on things related to the tests.

Greene et.al (2003) conducted a review of test results from two states and seven school districts. This review compared results from both high-stakes and low-stakes test as well as the results from national tests, which were given to same group of students. Their conclusion was that when the high stakes test was well designed the results correlated well with national tests and that yearly progress correlated well. They believe that lack of correlation test results was more the effect of test design that of any other major effect. The authors concluded that given well designed assessments that high-stakes testing can give good accurate information as to the students' progress and the schools influence on that progress. In addition, that well designed assessments can be reliable policy tools. They also acknowledged that most of the high stakes tests that you

looked at did not perform well; Florida's program was used as the example of a successful high-stakes testing program

Chapter Summary

Much of the current research on the effects of high-stakes assessments tends to go counter to the idea that use of these types of assessments will improve the student's knowledge. Most of the authors reviewed here found flaws with use of high-stakes testing. They included, but were not limited to: (a) teaching to the test, (b) no improvement of tests scores when the students were given national standardized test, (c) a lack of basic skills, which are need for success after high school, and (d) increased risk of failing to graduate from high school.

Current Educational policy under both states and federal agencies is promoting and in most areas, requiring expanded use of high-stakes testing as a means to improve the public education system. If the use of well designed high stakes test can promote and measure student achievement as suggested by Greene et.al. (2003) then the problems identified by most researchers is the assessments and not the use of such assessments.

Chapter 3 defines the methodology which were used to examine the data in an attempt to determine if high-stakes testing as it is currently being used is meeting the goals of improving student performance or not.

Chapter 3

METHOD

The purpose of this project was to evaluate the impact of high stakes standardized testing. The current educational policy of the United States Department of Education and most state education departments is that the use of high stakes testing will improve student outcomes (i.e., improved test score equal increased knowledge). This author examined state test scores and national examination scores, in an attempt to determine whether increases in state test scores equate to increased student knowledge as determined by national examination scores. The national examinations were chosen because of the large quantity of data over a long period of time that is associated with them; these national examinations included in this study are; SAT and ACT.

Data Collection

The data for this project was collected from published standardized testing data; the states chosen for this project are California, Texas, and Colorado. These states were chosen because of the number of years standardized tests they have administered to their K-12 public school population. The goal of this project was to assess the test results from California, Texas, and Colorado over a period of time that most of the students will have been involved in the testing cycle for most of their primary and their entire secondary career in the public schools. Also the test data for students from those states for the national examinations (i.e., ACT and SAT) was collected (ACT) (Board, 2007; Service, 2007). Results were evaluated based on whether statistically significant changes occurred in the national examinations scores prior to and following the beginning of high-stakes testing by the states. A quantitative comparison of the scores, before testing

and after testing, was done to determine if the use of standardized assessments can be isolated as a factor in the changes in college entrance examination scores if these scores did in fact shift and if so in which direction.

Data for the national tests was collected from the College Entrance Examination Board (SAT) (Board, 2007), and the American College Testing Service (ACT) (Service, 2007). Individual state education data will be collected from the National Center for Educational Statistics, the College Entrance Examination Board, the American College Testing Service, and from the individual states (California, Texas, and Colorado) departments of education (Agency, 2007; Board, 2007; California Department of Education, 2007; Colorado Department of Education, 2007; Service, 2007).

The hypothesis that the use of high-stakes testing improves student's knowledge was tested by comparing scores from national college entrance examinations with scores from state high-stakes examination at the end point of the students high school career, generally the final test is given during or at the end of 11th grade.

Method of Analysis

The data was compiled in a tabular format and the results over time will be compared. The mean scores of the college entrance examinations were compared beginning before state tests were initiated and through the period of this study, 2006. State testing data was collected from the initiation of testing in each of the states through the academic year ending in 2006.

The data was compared using the student t-test for unequal variances to determine significance of change and the percent change over time was calculated to order to compare the changes in national and state test scores.

Chapter Summary

Assessment data from official state department of education and national college entrance examination sources was gathered and compared over time to test the hypothesis that the use of high-stakes testing will improve students' tests scores on all levels of testing (e.g. local and national). This was done using summary statistics, students t-test for significance of differences and percent change in the mean composite scores.

Chapter 4

RESULTS

The underlining premise of High-Stakes testing is that this type of testing will improve the knowledge of the students. With this in mind the hypothesis of this research is that, high-stakes testing as is currently being done in all 50 states will improve learning, thereby increasing knowledge, and that this will result in improved test scores. If this is correct then not only will the tests administered by the states show improving scores, but also national test scores should show similar improvement. This hypothesis was tested by comparing results from state tests with results from national tests, i.e. SAT and ACT college entrance exams. Results from both the state and national tests were compared over a period of several years.

Three states were chosen California (CA), Texas (TX), and Colorado (CO). These states were picked due to the period during which they had administered their own state tests. The goal was to examine national scores and state exit exams for high school students who had been in the testing cycle of the states for most their primary and their entire secondary school career before they took both the state and national tests. The final or exit exam for each state test was compared to the national tests.

SAT and ACT test scores for the three states were collected for the years 1993 to 2006 for the SAT examination (Board, 2007) and 1995 to 2006 for the ACT examination (Service, 2007), the composite national scores were compare to the composite scores of each of the states. State test scores were collected over a shorter time frame, beginning in 2002-2003 school years for California (California Department of Education, 2007), 1998-2006 school years for Texas (Agency, 2007), and 2001-2006 school years for Colorado (Colorado Department of Education, 2007) . The scores on the state tests used were from

the final high-stakes test given to the students in their high school career. The changes of the state scores were compared to the changes of that state for the national tests. The changes tracked were the shift in the national scores, the shift of the state on the national test and the shift of the state test scores. If high-stakes testing will in fact improve overall knowledge as measured by the college entrance exams (SAT and ACT) then the positive or negative changes in the state's scores on the national test should follow the changes in scores on the state test.

The California high stakes-testing program is STAR (California Standardized Testing and Reporting) (Appendix C) it was established in 1997. All students in grades two through eleven were tested in Reading, language, and mathematics. Students in grades two through eight were also tested in spelling, and students in grades nine through eleven were tested in science and social science.

The Colorado testing program CSAP (Colorado School Assessment Program) began in 2001 (Appendix D). Students in grades three through tenth, in grade eleven, the final exam is the ACT test. They are tested in reading, writing, and mathematics for grades three - ten, and for science in grade eight.

The Texas testing program began as the TAAS (Texas Assessment of Academic Skills) in 1998 and remained the exit exam for students enrolled in 9th grade on January 1, 2001. Following their graduation, the TAKS (Texas Assessment of Knowledge and Skill) became the state's high-stakes test. This test (TAKS) is the revised test for the state and was introduced in the 2002-2003 school year. (Appendix E).

The ACT scores from were tracked from 1995 to 2006 for each of the states (California, Texas, and Colorado). Table 1 is a compilation of the composites scores for each of the states and the national average. The ACT examination consists of four parts

with a total possible score of 36, based on this percent changes were calculated for each of the states. The largest change over the 12 years tracked was in Colorado and that was a negative change of 1.1 points or -3.1%, with California having a 0.7 shift (+1.9%) and Texas having a 0.2 shift (+0.6%) the national composite scores moved 0.3 points (+0.8%). It should be noted that Colorado requires all eleventh grade students to take the ACT examination and that the shift of 1.1 occurred at that time (2003) before that the change was 0.1 (+0.3). This shift was most likely a result of the change in demographics of the students taking the examination. Each state composite score was compared to the National composite scores using the t-test for unequal variances. The results in all cases indicated that there has been no significant changes in any of the state's composite ACT scores (Table 1). This is further illustrated when the composite scores are plotted (Figure 1). California and Texas scores changed in parallel with the National scores while Colorado has a dip in 2003 when they started testing 100% of the students. The Colorado scores moved in the same direction after that one time drop. This all supports the statistical results that there have been no significant changes in the ACT composite scores in the 12 years from 1995 to 2006 or during the time since 2000 which corresponds with the state tests.

Table 1

ACT composite scores

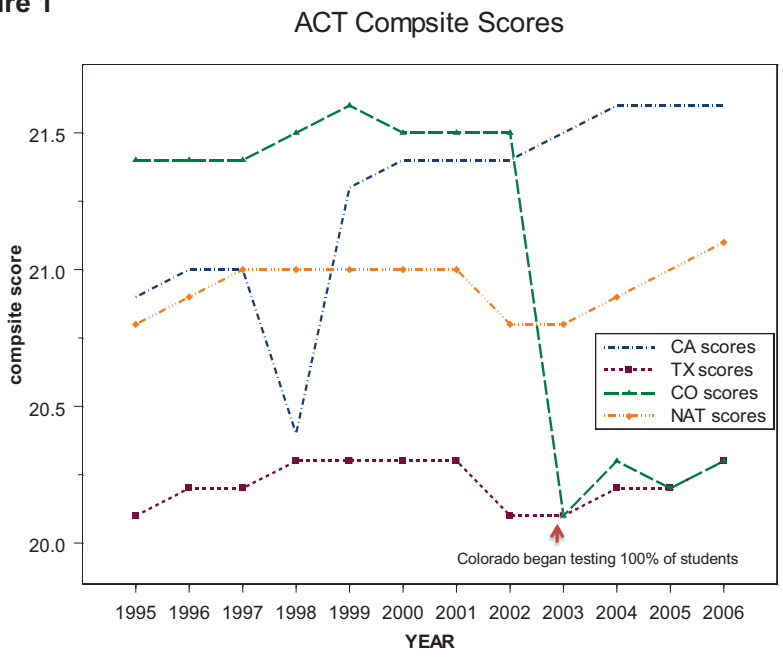
| Year | California (% students) | CA avg score | Texas (% students) | TX avg score | Colorado (% students) | CO avg score | National (% students) | National avg score |
|------|-------------------------|--------------|--------------------|--------------|-----------------------|--------------|-----------------------|--------------------|
| 1995 | 12.00 | 20.9 | 33.00 | 20.1 | 63.00 | 21.4 | 37.00 | 20.8 |
| 1996 | 11.00 | 21 | 33.00 | 20.2 | 60.00 | 21.4 | 35.00 | 20.9 |
| 1997 | 11.00 | 21 | 30.00 | 20.2 | 62.00 | 21.4 | 38.00 | 21.0 |
| 1998 | 12.00 | 20.4 | 32.00 | 20.3 | 63.00 | 21.5 | 37.00 | 21.0 |
| 1999 | 12.00 | 21.3 | 31.00 | 20.3 | 62.00 | 21.6 | 36.00 | 21.0 |
| 2000 | 12.00 | 21.4 | 32.00 | 20.3 | 64.00 | 21.5 | 38.00 | 21.0 |
| 2001 | 12.00 | 21.4 | 33.00 | 20.3 | 62.00 | 21.5 | 38.00 | 21.0 |
| 2002 | 13.00 | 21.4 | 32.00 | 20.1 | 99.00 | 21.5 | 39.00 | 20.8 |
| 2003 | 15.00 | 21.5 | 33.00 | 20.1 | 100.00 | 20.1 | 40.00 | 20.8 |
| 2004 | 14.00 | 21.6 | 29.00 | 20.2 | 100.00 | 20.3 | 40.00 | 20.9 |
| 2005 | na | 21.6 | na | 20.2 | 100.00 | 20.2 | na | 20.9 |
| 2006 | 14.00 | 21.6 | 14.00 | 20.3 | 100.00 | 20.3 | 40.00 | 21.1 |

| | CA avg score | TX avg score | CO avg score | Nat. avg score |
|--------------------|--------------|--------------|--------------|----------------|
| Mean | 21.25833 | 20.21667 | 21.05833 | 20.93636 |
| Standard Error | 0.105489 | 0.0241 | 0.178995 | 0.030963 |
| Standard Deviation | 0.365459 | 0.083485 | 0.620056 | 0.102691 |
| Sample Variance | 0.133561 | 0.00697 | 0.38447 | 0.010545 |
| Range | 1.2 | 0.2 | 1.5 | 0.3 |
| Minimum | 20.4 | 20.1 | 20.1 | 20.8 |
| Maximum | 21.6 | 20.3 | 21.6 | 21.1 |
| Count | 12 | 12 | 12 | 11 |

t-Test: Two-Sample Assuming Unequal Variances

| | CA | Nat. | TX | Nat. | CO | Nat. |
|------------------------------|----------|----------|----------|----------|----------|----------|
| Mean | 21.25833 | 20.93636 | 20.21667 | 20.93636 | 21.05833 | 20.93636 |
| Variance | 0.133561 | 0.010545 | 0.00697 | 0.010545 | 0.38447 | 0.010545 |
| Observations | 12 | 11 | 12 | 11 | 12 | 11 |
| Hypothesized Mean Difference | 0 | | 0 | | 0 | |
| df | 13 | | 19 | | 12 | |
| t Stat | 2.928362 | | -18.3426 | | 0.671443 | |
| P(T<=t) one-tail | 0.005875 | | 7.63E-14 | | 0.257328 | |
| t Critical one-tail | 1.770933 | | 1.729133 | | 1.392286 | |
| P(T<=t) two-tail | 0.01175 | | 1.53E-13 | | 0.514656 | |
| t Critical two-tail | 2.160369 | | 2.093024 | | 2.178813 | |

Figure 1



The SAT combined scores (Table 2) were examined in the same manner as the ACT composite scores. The SAT scores were examined for the period beginning with the 1993-1994 academic year through the 2005-2006 academic years. In 2005-2006, a writing section was added to the SAT examination, this increased the final combined score to 2400 from the previous 1600. To remain consistent for the 2005-2006 year only the mathematics and verbal components were used to determine the combine scores, this resulted in a possible score of 1600 for all of the years that were used.

Over this time frame (1994-2006) there was a +0.81% shift in the national scores, while the individual states (CO, TX, CA) had changes of +0.13%, +1.13%, +0.56% respectively. Using the t-test the results of each state was compared to the national results; again, there was no significant difference. This confirms that although there were small positive changes in the states scores that they did not improve at a greater rate than the national averages.

Table 2
Combined SAT Scores

| Academic Year | combined National | combined Colorado | combined California | combined Texas |
|---------------|-------------------|-------------------|---------------------|----------------|
| 2005-2006** | 1021 | 1021 | 1021 | 1021 |
| 2004-2005 | 1028 | 1028 | 1026 | 1028 |
| 2003-2004 | 1026 | 1026 | 1019 | 1026 |
| 2002-2003 | 1026 | 1026 | 1018 | 1026 |
| 2001-2002 | 1020 | 1020 | 1013 | 1020 |
| 2000-2001 | 1020 | 1020 | 1015 | 1020 |
| 1999-2000 | 1019 | 1019 | 1015 | 1019 |
| 1998-1999 | 1016 | 1016 | 1011 | 1016 |
| 1997-1998 | 1017 | 1017 | 1013 | 1017 |
| 1996-1997 | 1022 | 1016 | 1028 | 1016 |
| 1995-1997 | 1013 | 1013 | 1006 | 1013 |
| 1994-1995 | 1010 | 1010 | 1001 | 1010 |
| 1993-1994 | 1008 | 1003 | 1012 | 1003 |

** 2005-2006 a writing section was added to the SAT examination, these scores only include the reading and math scores

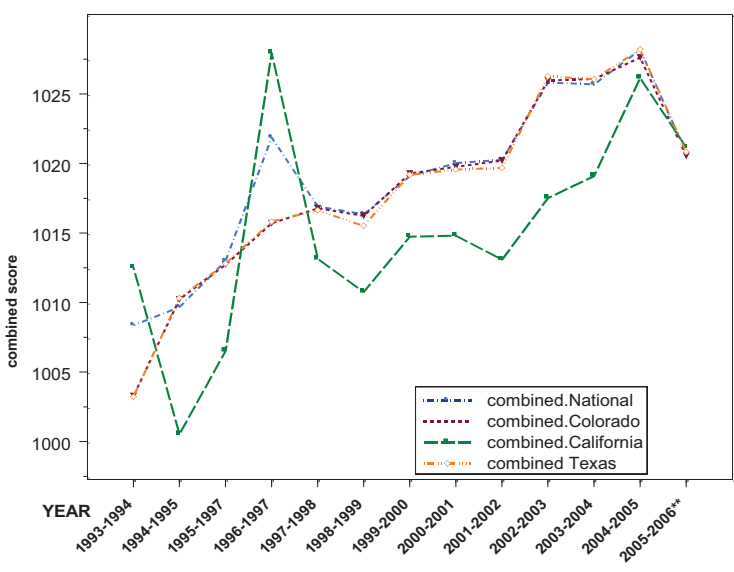
| | combined National | combined Colorado | combined California | combined Texas |
|--------------------|-------------------|-------------------|---------------------|----------------|
| Mean | 1018.923 | 1018.077 | 1015.231 | 1018.077 |
| Standard Error | 1.889 | 1.910 | 2.057 | 1.910 |
| Standard Deviation | 6.089 | 6.886 | 7.418 | 6.886 |
| Sample Variance | 37.077 | 47.410 | 55.026 | 47.410 |
| Range | 20 | 25 | 27 | 25 |
| Minimum | 1008 | 1003 | 1001 | 1003 |
| Maximum | 1028 | 1028 | 1028 | 1028 |
| Count | 13 | 13 | 13 | 13 |

t-Test: Two-Sample Assuming Unequal Variances

| | | | | | | |
|------------------------------|----------|----------|----------|----------|----------|----------|
| Mean | 1018.077 | 1018.923 | 1015.231 | 1018.923 | 1018.077 | 1018.923 |
| Variance | 47.410 | 37.077 | 55.026 | 37.077 | 47.410 | 37.077 |
| Observations | 13 | 13 | 13 | 13 | 13 | 13 |
| Hypothesized Mean Difference | 0 | | 0 | | 0 | |
| df | 24 | | 23 | | 24 | |
| t Stat | -0.3319 | | -1.3872 | | -0.3319 | |
| P(T<=t) one-tail | 0.3714 | | 0.0893 | | 0.3714 | |
| t Critical one-tail | 1.7109 | | 1.7139 | | 1.7109 | |
| P(T<=t) two-tail | 0.7428 | | 0.1787 | | 0.7428 | |
| t Critical two-tail | 2.0639 | | 2.0687 | | 2.0639 | |

Figure 2

SAT combined scores



The next step was to examine the individual states using the results from each of their state examinations. Beginning with California the data used was from the state examination (STAR) for the academic years 2002-2003 to 2005-2006 (Table 3) (Figure 3). The STAR examination tests English/Language Arts and Mathematics, data is reported as % not proficient, % proficient, and % advanced. I used the data for students who met or exceeded the state standards as the data set to compare changes in the scores, i.e. % proficient + % advanced (Table 3). These scores were then plotted and the percent changes compared. The data shows a change from 2003-2006 of 2% for the English/Language Arts and 8% for the Mathematics examinations. When these changes were compared to the changes for the SAT and ACT examinations, the improvements for the STAR examination did not carry over to either of the national examinations.

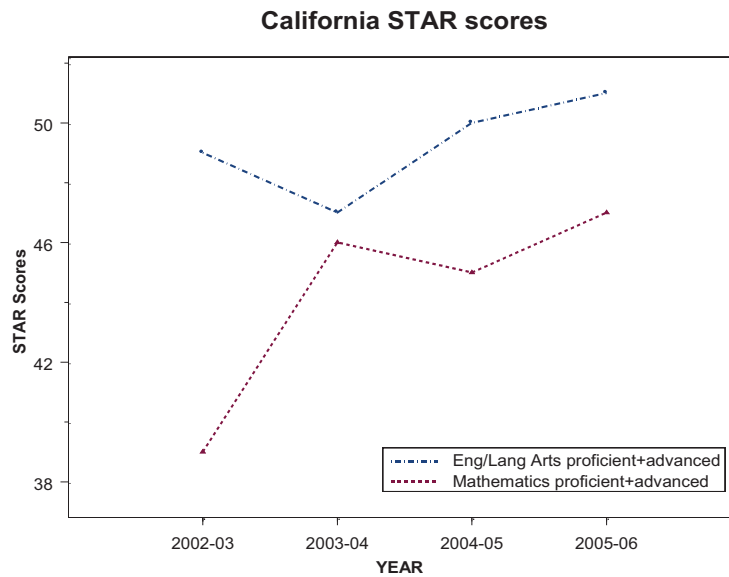
Table 3

California STAR scores

| year | % tested Eng/Lang Arts | % not proficient | % proficient | % advanced | % proficient + advanced | % tested Mathematics | % not proficient | % proficient | % advanced | % proficient + advanced |
|---------|------------------------|------------------|--------------|------------|-------------------------|----------------------|------------------|--------------|------------|-------------------------|
| 2002-03 | 89 | 52 | 26 | 23 | 49 | 91 | 61 | 27 | 12 | 39 |
| 2003-04 | 96 | 51 | 21 | 26 | 47 | 96 | 54 | 32 | 14 | 46 |
| 2004-05 | 96 | 51 | 22 | 28 | 50 | 96 | 54 | 31 | 14 | 45 |
| 2005-06 | 96 | 49 | 22 | 29 | 51 | 96 | 53 | 31 | 16 | 47 |

| | % tested Eng/Lang Arts | % not proficient | % proficient | % advanced | % tested Mathematics | % not proficient | % proficient | % advanced |
|--------------------|------------------------|------------------|--------------|------------|----------------------|------------------|--------------|------------|
| Mean | 94.25 | 50.75 | 22.75 | 26.5 | 94.75 | 55.5 | 30.25 | 14 |
| Standard Error | 1.750 | 0.629 | 1.109 | 1.323 | 1.250 | 1.848 | 1.109 | 0.816 |
| Standard Deviation | 3.500 | 1.258 | 2.217 | 2.646 | 2.500 | 3.697 | 2.217 | 1.633 |
| Sample Variance | 12.250 | 1.583 | 4.917 | 7.000 | 6.250 | 13.667 | 4.917 | 2.667 |
| Range | 7 | 3 | 5 | 6 | 5 | 8 | 5 | 4 |
| Minimum | 89 | 49 | 21 | 23 | 91 | 53 | 27 | 12 |
| Maximum | 96 | 52 | 26 | 29 | 96 | 61 | 32 | 16 |

Figure 3



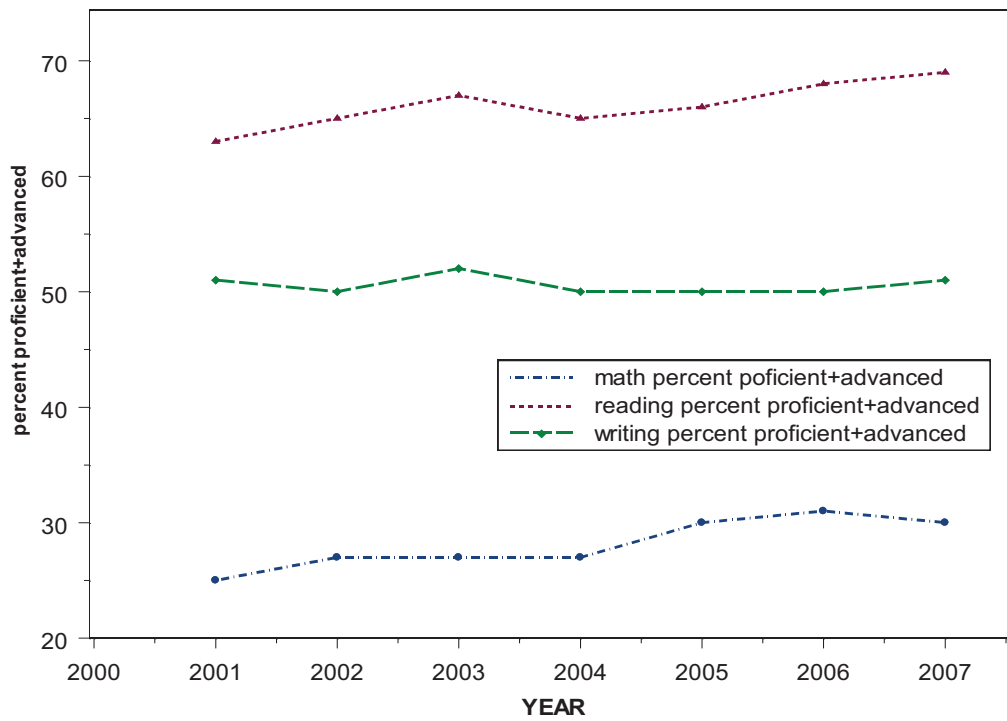
The Colorado state examination (CSAP) (Table 4) (Figure 4) tests reading writing and mathematics at the tenth grade level, the eleventh grade students are required to take the ACT examination as their final examination. When the scores for those student scoring proficient or above were examined for the academic years between 2001 and 2006 changes of +5% for mathematics, 6% for reading, and 0.0% for writing were observed. Again, these improvements did not follow the shifts for either the ACT or SAT examinations. Both the ACT and the SAT scores for Colorado had a less than 1% change over the same period.

Table 4 Colorado CSAP scores

| YEAR | Grade Level | % Proficient & Above Mathematics | % Proficient & Above Reading | % Proficient & Above Writing |
|------|-------------|----------------------------------|------------------------------|------------------------------|
| 2001 | 10 | 25 | 63 | 51 |
| 2002 | 10 | 27 | 65 | 50 |
| 2003 | 10 | 27 | 67 | 52 |
| 2004 | 10 | 27 | 65 | 50 |
| 2005 | 10 | 30 | 66 | 50 |
| 2006 | 10 | 31 | 68 | 50 |
| 2007 | 10 | 30 | 69 | 51 |

| | % Prof+Adv | | |
|--------------------|------------|--------|--------|
| Mean | 28.105 | 66.143 | 50.627 |
| Standard Error | 0.853 | 0.769 | 0.315 |
| Standard Deviation | 2.258 | 2.035 | 0.835 |
| Sample Variance | 5.099 | 4.143 | 0.697 |
| Range | 6.266 | 6 | 2 |
| Minimum | 24.733 | 63 | 50 |
| Maximum | 31 | 69 | 52 |
| Count | 7 | 463 | 7 |

Figure 4 Colorado CSAP Scores



Texas began high-stakes testing at the eleventh grade level began in 1998 using the TAAS examination and this was replaced in 2003 by a revised examination TAKS. Again, I examined the Texas results through the 2006 academic year and compared the results to the SAT and ACT examination results for Texas. The TAAS examination tested reading, writing and mathematics. Using the TAAS data for students who met all objectives (state standards) (Table 5) (Figure 5) for the years 1998-2002, the % students scoring at that level in mathematics had a shift of -8%, a -1% shift for reading, and 0.0% shift for writing. These declines did not follow the shifts in the national tests, which trended positively over the same period.

The TAKS examination tests English/Language Arts, Mathematics, Social Studies, and Science. Again the % students who met objectives (state standards) was used (Table 5) (Figure 6) for the years 2003-2006 for the comparisons. The first year of the examination (2003) all of the results were low. However, the next 3 years (2004-2006) they increased from the 2003 results but were either flat or declining from the 2004 results. Students who met objectives for the time frame 2004 to 2006 for English/Language Arts had a +1% change, for Mathematics had a -8% change, for Social Studies had a -3% change and Science had a -10% change. When this performance was compared to the national tests (ACT and SAT) the changes did not track. The state examination showed negative trend while the national examinations almost 1% positive trend for the same period.

Table 5
Texas State Tests (TAAS, TAKS)

| TEST / YEAR | | | | | | |
|--------------------|--------------------------|-------------------------|--------------------------|-------------------------|--------------------------|-------------------------|
| *TAAS | reading comprehension | | mathematics | | writing | |
| | met minimum expectations | mastered all objectives | met minimum expectations | mastered all objectives | met minimum expectations | mastered all objectives |
| 1998 | 86 | 54 | 75 | 27 | 87 | 40 |
| 1999 | 88 | 61 | 81 | 31 | 90 | 46 |
| 2000 | 90 | 53 | 86 | 26 | 90 | 47 |
| 2001 | 90 | 50 | 89 | 15 | 89 | 33 |
| 2002 | 94 | 53 | 92 | 19 | 91 | 40 |
| Mean | 89.6 | 54.2 | 84.6 | 23.6 | 89.4 | 41.2 |
| Standard Error | 1.327 | 1.828 | 3.010 | 2.891 | 0.678 | 2.518 |
| Standard Deviation | 2.966 | 4.087 | 6.731 | 6.465 | 1.517 | 5.630 |
| Sample Variance | 8.8 | 16.7 | 45.3 | 41.8 | 2.3 | 31.7 |
| Range | 8 | 11 | 17 | 16 | 4 | 14 |
| Minimum | 86 | 50 | 75 | 15 | 87 | 33 |
| Maximum | 94 | 61 | 92 | 31 | 91 | 47 |
| Count | 5 | 5 | 5 | 5 | 5 | 5 |

*TAAS - Texas Assessment of Academic Skills, grade 10 exit level

| **TAKS | English Language Arts | | Mathematics | | Social Studies | | Science | |
|--------------------|-----------------------|-----------------------|--------------|-----------------------|----------------|-----------------------|--------------|-----------------------|
| | met standard | commended performance | met standard | commended performance | met standard | commended performance | met standard | commended performance |
| 2003 | 69 | 5 | 68 | 6 | 90 | 9 | 67 | 2 |
| 2004 | 87 | 10 | 85 | 15 | 97 | 20 | 85 | 5 |
| 2005 | 88 | 20 | 81 | 18 | 94 | 25 | 80 | 5 |
| 2006 | 88 | 21 | 77 | 18 | 94 | 29 | 75 | 9 |
| Mean | 83 | 14 | 77.75 | 13.75 | 93.75 | 20.75 | 76.75 | 5.25 |
| Standard Error | 4.673 | 3.894 | 3.637 | 2.658 | 1.436 | 4.328 | 3.838 | 1.436 |
| Standard Deviation | 9.345 | 7.789 | 7.274 | 5.315 | 2.872 | 8.655 | 7.676 | 2.872 |
| Sample Variance | 87.333 | 60.667 | 52.917 | 28.250 | 8.250 | 74.917 | 58.917 | 8.250 |
| Range | 19 | 16 | 17 | 12 | 7 | 20 | 18 | 7 |
| Minimum | 69 | 5 | 68 | 6 | 90 | 9 | 67 | 2 |
| Maximum | 88 | 21 | 85 | 18 | 97 | 29 | 85 | 9 |
| Count | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |

**TAKS - Texas Assessment of Knowledge and Skills, grade 11 exit level

Figure 5
Texas TAAS Scores

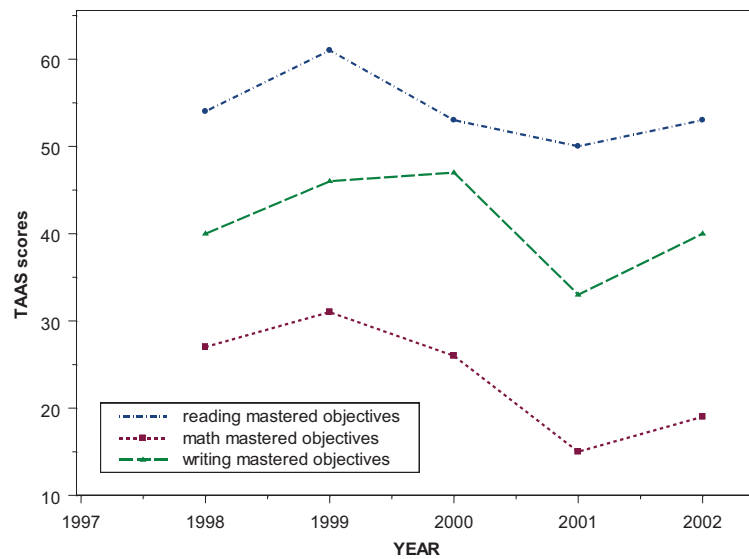
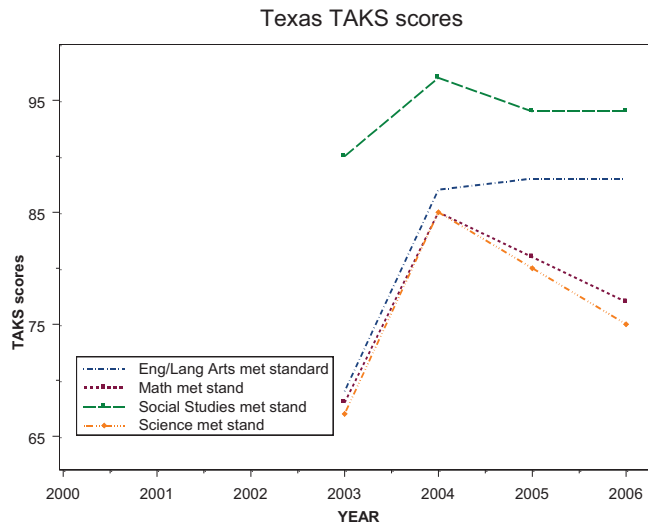


Figure 6



Chapter Summary

In summary, the data presented here indicates that while the scores for the ACT and SAT examinations have slowly improved over the past 12 years the scores of the high-stakes tests given by the states of California, Texas, and Colorado do not show that they have been a major factor in this gain. This is demonstrated by the larger shifts in state test scores compared with the changes in the national tests scores for each of the states examined in this study. The changes in the national test scores of each of the three states are not significantly different from the changes of the all of the states while the changes in the state tests score exceed the national test scores.

Chapter 5

Discussion

This study set out to examine the premise that High-Stakes testing as it is practiced by the individual states of the United States and promoted by the United States Department of Education will improve the learning of students. Proponents of high-stakes testing point to the increased scores of most of the state's tests, while opponents of high-stakes testing as it is being practiced believe that any improvement of scores on state tests can be generally attributed to teaching to the test and not real overall increases in knowledge.

In an attempt to try and answer the question, *does high-stakes testing improve knowledge of students*, three states (Colorado, California, & Texas) were chosen and the scores from the two major college entrance examinations (SAT, ACT) were collected for each of these states. Then the national averages for a period starting before the individual state tests and continuing over the period used for the state tests were collected. The scores for each of high-stakes tests given in each of the three states were collected and progress over time were evaluated. The data were examined for changes over time, significance of the changes, and whether the change on the state tests was comparable to the change on the national tests.

Scores of the ACT examination were tracked for an eleven-year period beginning with the academic year ending in 1995 to the academic year ending in 2006. The results indicate that there was a small positive shift of 0.3 points or +0.83%. When the results from each of the three states were compared to the national average composite score using the t-test there was no significant difference in the scores, in other words the states

had not improved more than the entire population of students in the nation taking the ACT examination.

The scores for the SAT examination were tracked a thirteen-year period beginning the academic year ending in 1994 and ending with the academic year ending in 2006. The data over this period again demonstrate a small positive change, +0.81%, which follows closely the shifts in the ACT examination. Again, there was no significant difference between the shifts seen in the states examined and the national average.

When individual state high-stakes test scores were examined there were greater shifts in these tests scores than the national tests (ACT, SAT) for the students in the individual states examined here (CA, CO, TX). Beginning with California, the state test scores improved by 2% (English/Language Arts) and 8% (Mathematics) for the four-year period between 2003 and 2006. These shifts in state test scores do not reflect the shifts seen in the national examinations of +0.7 (ACT) and +0.56 (SAT). Colorado had similar results, the state examination showed improvement of +5.0% (mathematics), +6.0% (Reading), and 0.0% (writing) while both of the national tests (ACT and SAT) had a positive change of just under 1% for the same time frame (2001-2006). I should be noted that when Colorado began requiring all eleventh grade students(2003) to take the ACT examination there was a -3.1% drop in scores, most likely due to the shift in demographics of the test taking population. After that initial drop, scores on the ACT examination moved in concert with the national average. Texas demonstrated another twist on the state vs. national trends. Texas has used two versions of their state test since 1998, the TAAS examination (1998-2002) and the TAKS examination (2003-2006). The state scores declined -.8% mathematics, -1% reading, and 0% writing for the TAAS examination (1998-2002) and +1% English/Language Arts, -7% mathematics, -3% social

studies, and -10% science for the TAKS (2003-2006). The scores on the national tests (ACT and SAT) had an almost +1% shift and tracked with the rest of the states for the same period.

The most obvious conclusion of the data from these 3 states when compared with the data from the national examinations is that results either positive or negative from the state tests do not translate to changes in the national examinations. The positive trends in CA and CO can probably be attributed to some degree of “teaching to the test”, while the negative trend in TX of the state examinations and positive trend for the national examinations in TX indicate some kind of disconnect between what is being taught and the examination. All of this points to being unable to support the original hypothesis that the use of high-stakes testing should improve student’s knowledge. Because of this, I have to reject the hypothesis and conclude that high-stakes testing as it is currently being done is not accomplishing the desired result of improving student’s knowledge during their primary and secondary school years as measure at the end of their high school career.

The test results seen in this study tend to support the opponents of high-stakes testing as it is currently being practiced in the public schools. While the data examined for this study is limited to three states it is in line with the educational research studies, which are critical of high-stakes testing. Because of this I believe that if public educational institutions are going to be tracked because of Federal legislation, such as NCLB, then the means by which progress is measured needs to be modified and become more broad based than current high-stakes testing allows. It can easily be argued that some means of monitoring is necessary but this monitoring would probably be better accomplished with a more general national test than with the current group of 50 different

high-stakes tests. Because the current approach makes it is very difficult to compare one states tests to the others, this in turn makes determining whether there is progress by the public schools very difficult and confusing.

REFERENCES

- Abrams, L. M., Pedulla, J. J., & Madaus, G. F. (2003). Views from the Classroom: Teachers' Opinions of Statewide Testing Programs. *Theory Into Practice, 42*(1), 12.
- ACT. (Publication., from American College Testing Service: Agency, T. E. (2007). *TAKS*. Retrieved 3/8/2007. from <http://www.tea.state.tx.us/assessment.html>.
- Albrecht, S. F., & Joles, C. (2003). Accountability and access to opportunity: Mutually exclusive tenets under a high-stakes testing mandate. *Preventing School Failure, 47*(2), 86-92.
- Berliner, D. C. (1993, February 2, 1993). Educational reform in an era of disinformation. *Educational Policy Analysis Archives* Retrieved December 19, 2004, 2004
- Board, C. E. E. (2007). SAT (Publication. Retrieved 3/9/2007, from College Entrance Examination Board: <http://www.professionals.collegeboard.com/data-reports-research/sat/archived>
- Butler, P. A. (2003). Achievement outcomes in Baltimore City schools. *Journal of Education for Students Placed at Risk, 80*(1), 33-60.
- Chudowsky, N., & Pellegrino, J. W. (2003). Large-scale assessments that support learning: What Will It Take. *Theory and Practice, 42*(1), 75-83.
- Education, C. D. o. (2007). *About STAR*. Retrieved. from <http://www.cde.ca.gov>.
- Education, C. D. o. (2007). *Title 22 - Education*. Retrieved 9/14/2007. from http://www.cde.co.us/index_stdn-access.htm.
- Education, N. C. o. E. i. (1983). *A Nation at Risk: The imperative for educational reform*. Washington D.C.: United States Department of Education. Document Number)
- Goals 2000: Educate America Act, (1994).
- Education, U. S. D. o. (2005). Executive summary of the No Child Left Behind Act of 2001. Retrieved 11/27/2005, 2005
- Fetler, M. E. (1994, 1994). Carrot or Stick? How do School Performance Reports Work? V2N13. Retrieved 6/15/2005, 2005, from <http://epaa.asu.edu/epaa/v2n13.html>
- Gordon, D. T. (Ed.). (2003). *A Nation Reformed? American education 20 years after A Nation at Risk* (1 ed.). Cambridge MA: Harvard Education Press.
- Greene, J. P., Winters, M. A., & Forster, G. (2003). Testing High Stakes Tests: Can We Believe the Results of Accountability Tests? *Center for Civic Innovation at the Manhattan Institute, 33*(February 2003), 19.
- Hillcocks, G. (2002). *The testing trap: how state writing assessments control learning*. New York: Teachers College Press.
- Jorgenson, O., & Vanosdall, R. (2002). Death of Science. *Phi Delta Kappan, 83*(8), 5.
- Meier, D. (2002). Standardization versus standards. *Phi Delta Kappan, 84*(3), 190-198.
- Moon, T. R., Brighton, C. M., & Callahan, C. m. (2003). State standardized testing programs: friend or foe of gifted education? (On gifted students in school). *Roeper Review, 25*(2), 12.
- Neill, M. (2003). The dangers of testing. *Educational Leadership, 60*(5), 43-46.

- Officers, C. o. C. S. S. (2002). *State education accountability reports and indicator reports: Status of reports across the States - 2002*. Washington, DC: Office of Educational Research and Improvement. Document Number)
- R.Valencia, R., & Villarreal, B. J. (2003). Improveing students' reading performance via standards-based school reform: A critique. *The Reading Teacher*, 56(7), 612-621.
- Sanders, W. L., & Horn, S. P. (1995, March 3, 1995). Educational Assessment Reassessed: The Usefulness of Standardized and Alternative Measures of Student Achievement as indicators for the Assessment of Educational Outcomes. Retrieved 6/15/2005, 2005, from <http://epass.asu.edu/epaa/v3n6.html>
- Schemo, D. J. (2003, December 14, 2003). Students' scores rise in math, not in reading. *The New York Times*,
- Schemo, D. J., & Fessenden, F. (2003, December 3, 2003). Gains in Houston Schools: How are they. *The New York Times*,
- Secretary, W. H. O. o. t. P. (2002, 1/8/2002). Fact Sheet: No Child Left Behind Act. Retrieved 05/02/2006, 2006, from <http://www.whitehouse.gov/news/releases/2002/01/print/20020108.html>
- Seidman, R. H. (1996, 1996). National Education 'Goals 2000': Some Disastrous Unintended Consequences. v4n11. Retrieved 6/15/2005, 2005, from <http://epaa.asu.edu/epaa/v4n11/>
- Service, A. C. T. ACT Scores (Publication., from American College Testing Service: Service, A. C. T. (2007). ACT National and State Scores [Electronic Version]. Retrieved 3/9/2007, from <http://www.act.org/news/data>
- Tretter, T. R., & Jones, G. M. (2003). Relationships between inquiry-based teaching and Physical science standardized test scores. *School Science and Mathematics*, 103(7), 345.
- Valencia, R. R., & Villarreal, B. J. (2003). Improveing students' reading performance via standards-based school reform: A critique. *The Reading Teacher*, 56(7), 612-621.
- Wright, W. E. (2002). The effects of high Stakes testing in inner-city elementary school: the curriculum, the teachers, and the English language learners. *Current Issues in Education*, 5(5), 36.

Appendix A

The Goals 2000 Act (Goals 2000:....., 1994) under Title 1 – National Education

Goals, Sec. 102. National Education Goals lists the following goals:

9. *SCHOOL READINESS.--*

9a. By the year 2000, all children in America will start school ready to learn.

9b. The objectives for this goal are that—

- i. All children will have access to high-quality and developmentally appropriate preschool programs that help prepare children for school;*
- ii. Every parent in the United States will be a child's first teacher and devote time each day to helping such parent's preschool child learn, and parents will have access to the training and support parents need; and*
- iii. Children will receive the nutrition, physical activity experiences, and health care needed to arrive at school with healthy minds and bodies, and to maintain the mental alertness necessary to be prepared to learn, and the number of low-birth weight babies will be significantly reduced through enhanced prenatal health systems.*

10. *SCHOOL COMPLETION.--*

10a. By the year 2000, the high school graduation rate will increase to at least 90 percent.

10b. The objectives for this goal are that--

- i. the Nation must dramatically reduce its school dropout rate, and 75 percent of the students who do drop out will successfully complete a high school degree or its equivalent: and*
- ii. the gap in the high school graduation rates between American students from minority backgrounds and their non-minority counterparts will be eliminated.*

11. *STUDENT ACHIEVEMENT AND CITIZENSHIP.—*

11a. by the year 2000, all students will leave grades 4, 8, and 12 having demonstrated competency over challenging subject matter including English, mathematics, science, foreign languages, civics and government, economics, arts, history, and geography, and every school in America will ensure that all students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment in our Nation's modern economy.

11b. The objectives for this goal are that—

- i. the academic performance of all students at the elementary and secondary level will increase significantly in every quartile, and the distribution of minority students in each quartile will more closely reflect the student population as a whole;*
- ii. the percentage of all students who demonstrate the ability to reason, solve problems, apply knowledge, and write and communicate effectively will increase substantially;*

- iii. *all students will be involved in activities that promote and demonstrate good citizenship, good health, community service, and personal responsibility;*
- iv. *all students will have access to physical education and health education to ensure they are healthy and fit;*
- v. *the percentage of all students who are competent in more than one language will substantially increase; and*
- vi. *all students will be knowledgeable about the diverse cultural heritage of this Nation and about the world community.*

12. *TEACHER EDUCATION AND PROFESSIONAL DEVELOPMENT.—*

12a. *By the year 2000, the Nation's teaching force will have access to programs for the continued improvement of their professional skills and the opportunity to acquire the knowledge and skills needed to instruct and prepare all American students for the next century.*

12b. *The objectives for this goal are that—*

- i. *all teachers will have access to preservice teacher education and continuing professional development activities that will provide school teachers with the knowledge and skills needed to teach to an increasingly diverse student population with a variety of educational, social, and health needs;*
- ii. *all teachers will have continuing opportunities to acquire additional knowledge and skills needed to teach challenging subject matter and to use emerging new methods, forms of assessment, and technologies;*
- iii. *States and school districts will create integrated strategies to attract, recruit, prepare, retain, and support the continued professional development of teachers, administrators, and other educators, so that there is a highly talented work force of professional educators to teach challenging subject matter; and*
- iv. *partnerships will be established, whenever possible, among local educational agencies, institutions of higher education, parents, and local labor, business, and professional associations to provide and support programs for the professional development of educators.*

13. *MATHEMATICS AND SCIENCE--*

13a. *By the year 2000, United States students will be first in the world in mathematics and science achievement.*

13b. *The objectives for this goal are that—*

- i. *mathematics and science education, including the metric system of measurement, will be strengthened throughout the system, especially in the early grades;*
- ii. *the number of teachers with a substantive background in mathematics and science, including the metric system of measurement will increase by 50 percent; and*
- iii. *The number of United States undergraduate and graduate students, especially women and minorities, who complete degrees in mathematics, science, and engineering will increase significantly.*

14. *ADULT LITERACY AND LIFELONG LEARNING.--*

14a. *By the year 2000, every adult American will be literate and will possess the knowledge and skills necessary to complete in a global economy and exercise the rights and responsibilities of citizenship.*

14b. *The objectives for this goal are that—*

- i. *every major American business will be involved in strengthening the connection between education and work;*
- ii. *all workers will have the opportunity to acquire the knowledge and skills, from basic to highly technical, needed to adapt to emerging new technologies, work methods, and markets through public and private educational, vocational, technical, workplace, or other programs;*
- iii. *the number of quality programs including those at libraries, that are designed to serve more effectively the needs of the growing number of part-time and mid-career students will increase substantially;*
- iv. *the proportion of the qualified students, especially minorities, who enter college, who complete their degree programs will increase substantially;*
- v. *the proportion of college graduates who demonstrate an advanced ability to think critically, communicate effectively, and solve problems will increase substantially; and*
- vi. *schools, in implementing comprehensive parent involvement programs, will offer more adult literacy, parent training and life-long learning opportunities to improve the ties between home and school, and enhance parents' work and home lives.*

15. *SAFE, DISCIPLINED, AND ALCOHOL AND DRUG-FREE SCHOOLS—*

15a. *By the year 2000, every school in the United States will be free of drugs, violence, and the unauthorized presence of firearms and alcohol and will offer a disciplined environment conducive to learning.*

15b. *The objectives for this goal are that—*

- i. *every school will implement a firm and fair policy on use, possession, and distribution of drugs and alcohol;*
- ii. *parents, businesses, governmental and community organizations will work together to ensure the rights of students to study in a safe and secure environment that is free of drugs and crime, and that the schools provide a healthy environment and are a safe haven for all children;*
- iii. *every local educational agency will develop and implement a policy to ensure that all schools are free of violence and the unauthorized presence of weapons;*
- iv. *every local educational agency will develop a sequential, comprehensive kindergarten through twelfth grade drug and alcohol prevention education program;*
- v. *drug and alcohol curriculum should be taught as an integral part of sequential, comprehensive health education;*
- vi. *community-based teams should be organized to provide students and teachers with needed support; and*
- vii. *every school should work to eliminate sexual harassment.*

16. *PARENTAL PARTICIPATION.—*

- 16a. *By the year 2000, every school will promote partnerships that will increase parental involvement and participation in the social, emotional, and academic growth of children.*
- 16b. *The objectives for this Goal are that—*
- i. *every state will develop policies to assist local schools and local educational agencies to establish programs for increasing partnerships that respond to the varying needs of parents and the home, including parents of children who are disadvantaged or bilingual, or parents of children with disabilities;*
 - ii. *every school will actively engage parents and families in a partnership which supports the academic work of children at home and shared educational decision making at school; and*
 - iii. *parents and families will help to ensure that schools are adequately supported and will hold schools and teachers to high standards of accountability.*

Appendix B

The No Child Left Behind Act (NCLB) was signed into law on January 8, 2002, a summary of the major points of the Act (fact sheet cite) follows;

Accountability for Results: Creates strong standards in each state for what every child should know and learn in reading and math in grades 3-8. Student progress and achievement will be measured for every child, every year.

- *Results from these tests will be made available in annual report cards parents can measure school performance and statewide progress, evaluate the quality of their child's school, the qualifications of teachers, and their child's progress in key subjects.*
- *Statewide reports will show progress of all student groups in closing achievement gaps between disadvantaged students and other groups of students.*
- *Schools will be held accountable for improving performance of all student groups, so every school will be performing at proficient levels within 12 years.*

Unprecedented State & Local Flexibility & Reduced Red Tape: Provides new flexibility for all 50 states and every local school district in America in the use of federal education funds.

- *Every local school district in America and all 50 states will receive the freedom to target up to 50 percent of federal non-Title 1 dollars to programs that will have the most positive impact on the students they serve.*
- *The new law consolidates and streamlines programs and targets resources to existing programs that serve poor students, reducing the overall number of ESEA programs from 55 to 45.*

Focusing Resources on Proven Educational Methods: Focuses educational dollars on proven. Research-based approaches the will most help children to learn.

- *Implements President Bush's Reading First initiative by increasing federal funding for reading programs from \$300 million in FY 2001 to more than \$900 million in FY 2002, and tying federal funding to the use of scientifically-proven methods of reading instruction.*
- *Implements a new Early Reading First program to support early language, literacy, and pre-reading development of preschool-age children, particularly those from low-income families.*
- *Strengthens teacher quality by providing \$2.8 billion for teacher quality programs and allowing local school districts to use additional federal funds to hire new teachers, increase teacher pay, improve teacher training and development or other uses.*

Expanded Choices for Parents: Enhances options for parents with children in chronically failing schools – and makes these options available immediately in the 2002-03 school year for students in thousands of schools already identified as failing under current law.

- *Public/Charter Schools Choice: Once a school is identified as failing, parents will be allowed to transfer their child to a better-performing public or charter school.*
- *Supplemental Services: For the first time, Federal Title I funds (approximately \$500 to \$1,000 per child) can be used to provide supplemental educational services – including tutoring, after school services, and summer school programs – for children in failing schools. Services can be provided by faith- and community-based organizations.*
- *Charter Schools: Expand the charter school initiative, creating more opportunities for parents, educators and interested community leaders to create schools outside the education establishment.*

Appendix C

California Testing Program (STAR)

About STAR (*Standardized Testing and Reporting*)

Program Background

The governor signed Senate Bill 376 authorizing the Standardized Testing and Reporting (STAR) Program in October 1997. The State Board of Education, as required by statute, designated the Stanford Achievement Test Series, Ninth Edition (Stanford 9) as the national norm-referenced achievement tests for the Program. These tests were first administered to students in grades two through eleven in California public schools during spring 1998 and were last administered as part of the STAR Program during spring 2002. School districts were required to administer the tests to all students in grades two through eleven except for:

- *Students who were receiving special education services with individualized education programs (IEPs) that specified that the students were to have an alternate assessment, and*
- *Students whose parents/guardians submitted written requests to exempt the students from testing.*

Students in grades two through eleven were tested in reading, language, and mathematics. Students in grades two through eight were also tested in spelling, and students in grades nine through eleven were tested in science and social science. All questions on the tests were multiple choice. The purpose of the Stanford 9 was to compare each student's achievement of general skills taught throughout the United States to the achievement of a national sample of students tested in the same grade at the same time of the school year.

In 1998, the State Board of Education designated the Spanish Assessment of Basic Education, Second Edition (SABE/2) as the primary language test for the Program. Beginning in spring 1999, Spanish-speaking English learners who were enrolled in California public schools less than 12 months when testing began were required to take the SABE/2, as well as taking the Stanford 9 and the Stanford 9 Augmentation/California Standards Tests (CSTs). Districts were given the option of also testing Spanish-speaking English learners enrolled in California public schools 12 months or more with the SABE/2.

During the 1998–99 school year, multiple-choice questions were developed specifically to assess the California English-Language Arts and Mathematics Content Standards. These questions, initially referred to as the Stanford 9 Augmentation, were administered for the first time during spring 1999. Students received CST scores based on questions selected from the Stanford 9 tests and the California-specific questions. The CSTs then evolved during the next several years.

The purpose of the CSTs is to determine students' achievement of the California Content Standards for each grade or course. Students' scores are compared to preset criteria to determine if the students' performance on the test is advanced, proficient, basic, below basic, or

far below basic. The state target is for all students to score at the proficient and advanced levels.

The legislature reauthorized the STAR Program during 2002, and the State Board of Education selected the California Achievement Tests, Sixth Edition Survey (CAT/6 Survey) to replace the Stanford 9 as the national norm-referenced test for the Program beginning with the spring 2003 test administration. The State Board also authorized the development of the California Alternate Performance Assessment (CAPA), an individually administered assessment for students with significant cognitive disabilities whose disabilities preclude them from taking the CSTs and CAT/6 Survey even with modifications. The CAPA assesses the California English-Language Arts and Mathematics Content Standards that were identified as appropriate for students with significant cognitive disabilities. The CAPA was first administered during spring 2003.

In August 2004, the governor signed legislation reauthorizing the STAR Program through 2011. The reauthorized program reduced the CAT/6 Survey to grades three and seven.

During 2005, the State Board of Education designated the Aprenda: La prueba de logros en español, Tercera edición (Aprenda 3) to replace the SABE/2 as the designated primary language test (DPLT) for the STAR Program. In 2006, Spanish-speaking English learners who were receiving instruction in Spanish were required to take the DPLT as well as English learners who had been enrolled in school in the United States less than 12 months when testing began. Districts had the option of administering the Aprenda 3 to Spanish-speaking English learners who had been enrolled in school in the United States 12 months or more who were not receiving instruction in Spanish. The change was from new state law that became effective on January 1, 2006.

Senate Bill 1448, which reauthorized the STAR Program, included the development of assessments for reading-language arts and mathematics in the state's dominant primary language. The legislation required that the assessments be aligned to state academic content standards. The Standards-based Tests in Spanish (STS) will replace the DPLT as it is developed. The STS is required for the same population of students who take the DPLT. The STS was first administered in the spring of 2007 to students in grades two through four. Tests for grades five through eleven will be administered as they are developed. Students taking the STS are also required to take the CSTs and the CAT/6 Survey.

The 2007 STAR Program included five components:

- *California Standards Tests (CSTs)*
- *Standards-based Tests in Spanish (STS)—grades 2, 3, and 4*
- *California Alternate Performance Assessment (CAPA)*
- *California Achievement Tests, Sixth Edition Survey (CAT/6 Survey)—grades 3 and 7*
- *Aprenda: La prueba de logros en español, Tercera edición (Aprenda 3)—grades 5 through 11*

The CSTs are a major component of California's accountability system for schools and districts. CST and CAPA results are the major component used for calculating each school's Academic Performance Index (API). These results are also used for determining if elementary and middle schools are making adequate yearly progress (AYP) in helping all students become proficient on

the state's content standards as required by the federal No Child Left Behind (NCLB) Act of 2001. Schools use CST results to identify seniors eligible for the California Golden State Seal Merit Diploma. The eligibility requirements for the diploma are posted at <http://www.cde.ca.gov/ta/tg/sr/documents/meriteligibility.doc>.

Milestones for the STAR Program

- 2000 *Stand-alone mathematics CSTs that used no Stanford 9 questions developed for grades 8–11*
- 2001 *Stand-alone history-social science and science CSTs developed and administered in grades 9–11
Writing component added to the grade 4 and 7 English-Language Arts CSTs
Performance levels reported for English-Language Arts CSTs*
- 2002 *Performance levels reported for all CSTs
The grade 4 and 7 writing components combined with the multiple-choice components to produce the English-Language Arts CST scores*
- 2003 *Grade 9 History-Social Science CST moved to grade 8
All CSTs administered as stand-alone tests
CAPA first administered*
- 2004 *Grade 5 Science CST added*
- 2006 *Grade 8 Science and Grade 10 Life Science CSTs added*
- 2007 *Standards-based Tests in Spanish (STS) in Reading-Language Arts and Mathematics for grades 2, 3, and 4 added
Students in Grade 7 were allowed to take the Algebra I CST if they had completed the course
Students in grades 9, 10, and 11 were allowed to take the CST for World History if they had completed the course*

California Department of Education

©2008 California Department of Education

Appendix D

Texas Testing Programs

Tests Administered by TEA

Texas Assessment of Knowledge and Skills (TAKS)

As mandated by the 76th Texas Legislature in 1999, the TAKS will be administered beginning in the 2002-2003 school year. The TAKS measures the statewide curriculum in reading at Grades 3-9; in writing at Grades 4 and 7; in English Language Arts at Grades 10 and 11; in mathematics at Grades 3-11; in science at Grades 5, 10, and 11; and social studies at Grades 8, 10, and 11. The Spanish TAKS is administered at Grades 3 through 6. Satisfactory performance on the TAKS at Grade 11 is prerequisite to a high school diploma.

Reading Proficiency Tests in English (RPTE)

The RPTE are designed to measure annual growth in the English reading proficiency of second language learners, and are used along with English and Spanish TAKS to provide a comprehensive assessment system for limited English proficient (LEP) students. LEP students in Grades 3-12 are required to take the RPTE until they achieve a rating of advanced.

State-Developed Alternative Assessment (SDAA)

The SDAA assesses special education students in Grades 3-8 who are receiving instruction in the Texas Essential Knowledge and Skills (TEKS) but for whom TAKS is an inappropriate measure of their academic progress. SDAA becomes a part of the school accountability system in the 2002-2003 school year.

Texas Assessment of Academic Skills (TAAS)

The TAAS measures the statewide curriculum in reading, mathematics and writing at the exit level. TAAS will remain the graduation requirement for students who were enrolled in Grade 9 or higher on January 1, 2001.

Appendix E

Colorado Testing Program (CSAP)

Title 22 - EDUCATION

- *Article 7 - Educational Accountability*
 - *PART 1 - Educational Accountability (22-7-102)*
 - *PART 4 - Education Reform (22-7-406 and 22-7-409)*

Colorado Revised Statute (C.R.S.) 22-7-102.

Legislative Declaration (Educational Accountability)

- *Declares that the purpose of (Educational Reform related to Standards and Assessments) in Colorado is to institute an accountability system to define and measure academic quality in education and thus to help public schools of Colorado to achieve such quality and to expand the life opportunities and options of the students of this state.*
- *Further, a stated purpose is to provide to local school boards and local schools assistance in helping their patrons to determine the relative value of their school program as compared to its cost.*

C.R.S. 22-7-102 (1)

- *Further declares that the educational accountability program developed should be designed to measure objectively the quality and efficiency of the educational programs offered by the public schools. C.R.S. 22-7-102 (2)*

The program should:

- *develop broad goals*
- *identify the activities of schools that can advance students toward these goals*
- *develop a means for evaluating the performance of students*

It is the belief of the general assembly that developing the evaluation mechanisms will provide for:

- *means for determining whether decisions affecting the educational process are advancing or impeding student achievement,*
- *means for reporting to students, parents and the general public on the educational performance of the public schools, and*
- *the collection and provision of performance information that could help school districts to increase their efficiency in using available financial resources.*

Colorado Revised Statute (R.S) 22-7-406.

Standards and Assessments Development and Implementation

- *Requires the state to adopt and revise Colorado Model Content Standards in first priority areas:*

C.R.S. 22-7-406 (1) (a)

- *Reading*
- *Writing*
- *Math, and*
- *Science*

Requires the state to adopt assessments in the areas of Reading, Writing, Math, and Science which are aligned with the state model content standards and to specify an acceptable performance level on each state assessment.

C.R.S. 22-7-406 (3)

- *Requires districts to adopt first and second priority content standards according to the timeline in C.R.S. 22-7-407.*

Colorado Revised Statute 22-7-409.

Assessments

Development of the CSAP

- *Requires the state to implement the Colorado Student Assessment Program (CSAP) in the first priority areas of Reading, Writing, Mathematics and Science. C.R.S. 22-7-409 (1)*
 - *Requires the tests be administered in English*
 - *Allows the state to administer tests in other languages except that any student who has participated in the English language proficiency program (created pursuant to 22-24-104) for more than a total of three school years must take the English version of the statewide assessments.*
- *Determined the following implementation schedule for the CSAP C.R.S. 22-7-409 (1) (a-f)*
:

| Beginning in.... | Grade level... | Content Areas... | | | |
|------------------------------------|-------------------------|-------------------------|---------|------|---------|
| Spring 1997 | -- Grade 4 | Reading | Writing | | |
| Spring 1998 | -- Grade 3 | Reading | | | |
| Fall 1999 (2001 move to spring) | -- Grade 5 | | | Math | |
| Spring 1999 | -- Grade 7 | Reading | Writing | | |
| Spring 2000 | -- Grade 8 | | | Math | Science |
| Spring 2001 | -- Grades 5, 6, 8, 9 | Reading | | | |
| | Grade 10 | Reading | Writing | Math | |
| Spring 2002 | -- Grades 3, 5, 6, 8, 9 | | Writing | | |
| | -- Grades 6,7,9 | | | Math | |

Calendar for Test Administration

1. Requires the assessments to be conducted during the period beginning second Monday in March and ending on the third Monday in April of each year. C.R.S. 22-7-409 (1.2) (a) (I)

Who must take the tests?

2. Requires every student enrolled in a public school to take the assessments in the grade level in which the student is enrolled (1.2)(d)(I)
 - Excepting those students taking the grade level assessment available in the alternative test - CSAPA.
C.R.S. 22-7-409 (1.2) (d) (I) (A)

Accountability Exclusions

3. Defines which test scores will not be used for purposes of calculating school academic performance ratings (SAR) or accreditation:
 - Any student who transfers into a school after October 1
C.R.S. 22-7-409 (1.2) (a) (I) (B).
 - Any student whose dominant language is not English for the first three years enrolled in Colorado public school
C.R.S. 22-7-409 (1.2) (a) (I) (C).
 - Any student taking the CSAPA
C.R.S. 22-7-409 (1.2) (a) (I) (A).

Note: ALL students must TAKE the assessments, even when their scores are excluded from accountability reports.

Colorado ACT

4. Requires all students enrolled in the eleventh grade in a Colorado public school to take the Colorado ACT. The ACT is the standardized, curriculum-based, achievement, college entrance examination selected by the department of education pursuant to this statute meeting the following criteria outlined in the statute C.R.S. 22-7-409 (1.5) (a) :
 - selected by the Colorado Department of Education
 - administered throughout the United States
 - relied upon by institutions of higher education that at a minimum test in the areas of reading, writing, mathematics, and science

Report Cards and Transcripts

5. Requires the results of the Colorado Student Assessment Program (CSAP) be included on each student's final report card for that school year and shall be part of the student's permanent academic record. C.R.S. 22-7-409 (1.9)
 - This means including a hard copy of the test results in the students' cumulative files.
 - Districts are including the results on the first report card after receiving the test results - fall of the next school year.
6. Requires the results of the Colorado ACT (COACT) be included on each student's transcript. C.R.S. 22-7-409 (1.9)

Assessments in the Spanish Language

7. *Requires the state to administer reading assessments in Spanish for students enrolled in the third and fourth grades and a writing assessment in Spanish for students enrolled in the fourth grade. C.R.S. 22-7-409 (3.5) (a)*

8. *Requires the state to administer writing assessments in Spanish for students enrolled in the third grades if NCLB funds are sufficient to pay for this assessment. C.R.S. 22-7-409 (3.5) (b)*