

STATE ACCOUNTABILITY RATINGS AS RELATED TO
DISTRICT SIZE AND DIVERSITY

Teresa M. Starrett

Dissertation Prepared for the Degree of
DOCTOR OF EDUCATION

UNIVERSITY OF NORTH TEXAS

May 2008

APPROVED:

Jane B. Huffman, Major Professor
James D. Laney, Committee Member
Frances van Tassell, Committee Member
Leslie Patterson, Chair of the Department of
Teacher Education and Administration
M. Jean Keller, Dean of the College of Education
Sandra L. Terrell, Dean of the Robert B. Toulouse
School of Graduate Studies

Starrett, Teresa M. State accountability ratings as related to district size and diversity.

Doctor of Education (Educational Administration), May 2008, 95 pp., 15 tables, 1 figure, reference list, 87 titles.

All Texas school districts were examined to determine the relationship of district size and diversity to the accountability ratings of selected Texas school districts and the implications of including all data in the accountability rating system. Eight large districts and 12 small districts were matched demographically utilizing data from the 2003-2004 school year.

Information from the Texas Education Agency was accessed over 2003-2004 and 2004-2005. The ratings were found to be lowered from Recognized to Academically Acceptable with the inclusion of these groups 6 out of 20 times. These findings indicate that the Texas accountability system, in its current structure, excludes certain students based upon race and economic status and is not in compliance with what the law intended.

This study should be replicated on a larger scale to assess its validity for a larger sample of small districts.. Equity among states should be examined to provide information for a nationwide accountability system.

Copyright 2008

by

Teresa M. Starrett

ACKNOWLEDGEMENTS

I would never have accomplished this goal without the unwavering support of countless people.

I would like to recognize and thank my advisor and committee chair, Dr. Jane B. Huffman, for her guidance, support and unwillingness to let me give up. I also wish to thank my committee members, Dr. James Laney and Dr. Frances van Tassell. They provided me with advice and invaluable input and for that I am grateful. Without the expertise of this talented team, I would not be successful.

I wish also to thank my friends and colleagues who have encouraged me along the way. The journey has been a long one, and I could not have done it without you. My school staff has been supportive and helped me wished me well both in my former district and now, in my present district. Teaching is a difficult job and I take my hat off to each of you for the job you do every day.

My friends have given me pep talks when needed and kicks when needed more. Among my dear ones, I count Alex Leavell and Diane McLean. I wish to especially thank Vicki Sargent for taking a chance so long ago on a very young and green me. Your friendship has allowed me to grow exponentially.

Finally, I want to thank my family. First, my father and mother, Mary and Frank Martin who encouraged me throughout the entire process. Without your support and assistance I couldn't have done it. My daughter Cate and son Cole were always understanding and proud of me even when it meant I was working on the computer and not playing. Thank you! Lastly, thank you to my dear husband, Bob. His love, encouragement and unwillingness to let me give up have allowed me to fulfill a dream.

I dedicate this dissertation to my dear family who has supported me for many, many years as I have sacrificed much time and energy to complete this degree. This does not solely belong to me; it belongs to each of you. Your encouragement and love gave me the strength I needed to finish the race. Without you, I am nothing.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
LIST OF TABLES AND FIGURES	vii
Chapter	
1. INTRODUCTION	1
Historical Background	1
Statement of Problem: District Size and Current Study	5
Research Questions	6
Purpose of the Study	6
Assumptions of the Study	6
Limitations of the Study.....	6
Definitions of Terms	7
Significance of Study	9
Summary	9
Organization of Study	10
2. REVIEW OF RELATED LITERATURE	11
Accountability	12
<i>A Nation at Risk</i>	12
Goals 2000	13
States' Responses to Accountability	15
California	17
Kentucky	21
New York.....	25
Testing in Texas.....	30
Texas Assessment of Basic Skills.....	32
Texas Educational Assessment of Minimal Skills.....	33
Texas Assessment of Academic Skills	34
Texas Assessment of Knowledge and Skills	35
Texas Accountability System	37
Texas Assessment of Knowledge and Skills	39
State Developed Alternative Assessment	40
Completion Rate	40
Annual Dropout Rate	41
Diversity and Testing.....	42
Poverty	44
Minorities.....	45
English Language Learners.....	49

Significance of Diversity and Testing.....	51
Summary	53
3. METHODOLOGY	55
Overview	55
Research Methodology	56
Procedures	56
Data Collection	58
Data Analysis	59
Summary	61
4. DATA PRESENTATION AND ANALYSIS	62
Corpus Christi and Ricardo.....	64
Conroe and Farmersville.....	65
Pharr San Juan-Alamo, Webb, and Los Fresnos.....	67
Spring Branch, Palacios, and Lometa	69
Spring and Fort Sam Houston.....	72
Pasadena and Brackett	73
Katy and Robinson.....	75
Irving and Springlake-Earth.....	77
Analysis.....	79
5. SUMMARY DISCUSSION AND RECOMMENDATIONS	82
Introduction.....	82
Statement of the Problem.....	82
Review of Methodology	82
Summary of Results and Discussion.....	83
Recommendations for Practical Application	85
Recommendations for Future Research	86
Implications.....	87
REFERENCE LIST	89

LIST OF TABLES AND FIGURES

Page

Tables

1. Average Targets required for AYP in California and Illinois.....17

2. Performance Levels and API Weight Factors in the California Department of Education (CDE)19

3. Subgroup Populations for the API Indicators as Defined by the CDE.....20

4. Kentucky Accountability System22

5. New York Performance Index (PI).....26

6. Percentage Population Increases in Texas43

7. Texas Population Projection43

8. Subgroup Gains in NAEP Mathematics Average Scale Scores (1996-2005).....48

9. Subgroup Gains on NAEP Depending on Basic or Proficient Level.....48

10. Recognized and Exemplary Districts with Non-Traditional Structure57

11. Texas District Comparability Study.....60

12. State Averages of Teacher Experience, Student-Teacher Ratio, Salary and Funding....63

13. Percentage of Students Who Met the TAKS Standard in 2003-2004 and 2004-2005....63

14. Ratings of Large and Small Districts 2004-2005.....79

15. District Accountability Changes81

Figure

1. Confidence Intervals of Annual Measurable Goals27

CHAPTER 1

INTRODUCTION

“There’s a growing consensus between policymakers and pundits that we must raise standards in our schools to ensure our students have the skills to compete in a world where what you know means far more than where you live” (Spellings, 2006). Secretary of Education Margaret Spellings made this remark in a speech to the National Conference of Editorial Writers in September 2006, where she highlighted the importance of education in the global community.

Accountability in public education is an issue in the 2008 presidential election and will continue to be integral in the future as the United States focuses on testing. The purpose of this dissertation is to examine the impact of district size and diversity on the accountability rating of selected Texas districts. This examination compares large districts comprised of 25,000 to 50,000 students with small districts which educate fewer than 8,000 students. Chapter 1 and Chapter 2 discuss, in detail, accountability as it relates to reauthorization of No Child Left Behind (NCLB), states’ responses to NCLB requirements, standardized testing including the Texas system, and diversity in testing.

Historical Background

The Elementary and Secondary Education Act (ESEA), crafted by Commissioner of Education Francis Keppel, was originally signed into law in 1965 by Lyndon B. Johnson less than three months after it was introduced. This piece of legislation was the main educational component of President Johnson’s War on Poverty. Since that time, ESEA has been reauthorized, at the very least, every five years and is the principal law affecting elementary and

secondary education throughout the United States (Linn, 2007). The original document included the following:

In recognition of the special educational needs of low-income families and the impact that concentrations of low-income families have on the ability of local educational agencies to support adequate educational programs, the Congress hereby declares it to be the policy of the United States to provide financial assistance...to local educational agencies serving areas with concentrations of children from low income families to expand and improve their educational programs by various means which contribute to meeting the special educational needs of educationally deprived children. (USDE, 2006, Section 201)

One of the central beliefs of the ESEA was that children from low-income homes required more educational services than children from affluent circumstances. At the time of the creation of ESEA, \$1,000,000,000 in funding was allocated, specifically for schools with high concentrations of low-income children. This was the beginning of Head Start, Follow Through, and Bilingual Education (Johnson, 1966).

Almost 40 years later, on January 8, 2002, President George W. Bush signed into law new revisions of ESEA, entitled the No Child Left Behind (NCLB) Act. According to the proposed budget for fiscal year 2008, President Bush expanded funding for NCLB from \$1.2 billion to \$24.5 billion. This represents an increase of 41% in funding since 2001 (Bush, 2007). In addition to funding for programs within the school day, there is increased federal support for before-school and after-school programs, school libraries, charter schools, and reading readiness programs for children in high-poverty neighborhoods.

In the revised 2002 edition of ESEA, specifically in the area of Title I, President George W. Bush proposed closing the achievement gap between disadvantaged students and their peers through accountability and high standards, annual academic assessments, and consequences for schools that fail to educate disadvantaged students. The intent of the NCLB Act was to close this achievement gap and to provide states additional assistance and flexibility in return for

implementing rigorous accountability standards. Schools that fail to make sufficient progress are promised special assistance, and students will not be forced to attend persistently failing schools.

The legislation:

- Sets high standards for reading, math, history and science
- Expects adequate yearly progress (AYP) for disadvantaged students
- Helps states with technical assistance funds for low-performing schools
- Requires progress reports on all student groups
- Increases flexibility for schools; lowers the poverty threshold from 50% to 40% for school-wide Title I funding
- Provides corrective action for low-performing schools and districts
- Rewards schools and states that narrow the achievement gap
- Puts in place consequences for failure (NCLB, 2002 pp. 20-26)

Most significantly, according to President George W. Bush (2000), NCLB requires that:

Students in at least grades three through eight must be tested on the basics of reading and math each year and those results, posted, by school, on the Internet. This will give parents the information to know if education is actually taking place and the leverage to demand reform. (p.122)

These assessments are intended to ensure that schools are able to implement measures to remediate and accelerate instruction, and to identify and correct problems quickly (Paige, 2002).

District scores are reported and the districts are rated as to whether students' needs are being met.

The percentage of schools meeting AYP targets increased in 2003-04 from the year before in most states, and the National Assessment of Educational Progress (NAEP) long-term trend scores have shown narrowing in achievement gaps. Despite the goal of NCLB that all students will be proficient in reading and math by 2014, based upon current research nearly all

schools in the U.S. will fail to meet AYP targets. Based upon standardization provided by the NAEP, in 2003 no state or large district had close to 100% of students performing at the basic NAEP level (Linn, 2005a).

According to data comparing the rigors of the NAEP and state assessments, there is a discrepancy from state to state. In looking at eighth grade mathematics for students in Missouri, 21% scored proficient on the NAEP while 16% scored proficient on the state assessment. In Tennessee, 26% of the students scored proficient on the NAEP while 87% scored proficient on the state assessment. The alignment of assessments and content standards is critical when looking at meeting AYP standards (Linn, 2007).

NCLB provides that school districts and schools that continually fail to make adequate progress will be identified for improvement and be subject to corrective action and restructuring by the federal government. Conversely, schools that consistently achieve higher standards will be honored with awards from a No Child Left Behind school bonus fund and an Achievement in Education state bonus fund (NCLB, 2002).

The long-term goal of NCLB is that 95% or more of all student groups will reach their state standards by the year 2014. One of the central tenets of NCLB is the requirement that test scores improve annually. To measure student success, each state is required to provide an assessment tool to measure progress and report to the federal government whether the state's educational program is effective.

Phelps (2005) reports the public is in favor of standardized testing and supports rigorous standards for all students; however, there are mixed reactions in regard to the frequency and depth of testing. One survey of teachers in California and Virginia showed that most of the teachers believe that a narrowing of the gap between subpopulations is occurring. According to a

recent report from the Center for Education Policy, the narrowing of the gap is widespread across the nation (Lee, 2006). Without assessment procedures and standards in place, little improvement can occur. A comprehensive accountability system must focus both on measurable elements of the process and on the results (Reeves, 2005).

Statement of the Problem: District Size and Current Study

One critical change that has been made in measuring performance under NCLB is that schools and districts must disaggregate data to ensure that all student groups are accounted for within the school and district rating. This includes a variety of demographic groups, such as students who are limited English proficient (LEP), low socioeconomic, and disabled.

Though data disaggregation and public reporting of student assessments is relatively new on the national level, Texas has been evaluating and assessing districts and campuses by utilizing standardized tests for many years. This particular investigation looks specifically at Texas and examines the impact of district size and diversity on the accountability ratings of selected Texas districts. In order to evaluate the effectiveness and equitability of the accountability system in relation to small and large districts, the Texas Education Agency's Academic Excellence Indicator System (AEIS) data and demographics are examined and tracked for the 2003-04 and 2004-05 school years to determine trends. Two groupings were studied: large school districts with student populations of 25,000-50,000, and small districts with fewer than 8,000 students. In addition, the study is limited to districts which have had an existing superintendent in position for at least three years prior to the original testing date. This provides a measure of district level continuity and allows for stability of programs at the upper administrative level.

Research Questions

Two research questions drive this study:

- What is the status of district size and diversity as related to the accountability rating of selected Texas school districts?
- If all size and diversity data were included in the accountability rating system, how might district ratings be affected?

Purpose of the Study

This descriptive study explores the impact of size and diversity on the accountability ratings of selected Texas school districts. It goes on to examine and interpret information to determine the consistency of the Texas system when the size of the subpopulation is not an issue.

Assumptions of the Study

All demographic and TAKS data analyzed are assumed to be accurate and true. Pearson Education Measurement is the data source that provides services to the Texas Education Agency and asserts that all reported data are correct.

Limitations of the Study

The difficulty of controlling for many variables such as principal retention, teacher quality, and curricular changes is a limitation on this study. Another uncontrollable variable is student ability. Those students who are most cognitively impaired are excluded from state standardized testing and are provided different testing opportunities through locally developed alternative assessments (LDAA). In addition, districts offer a wide variety of special educational services, enrolling many students who are eligible to take an alternative assessment [the state developed alternative assessment (SDAA)] rather than the TAKS.

This study is specific to Texas and the accountability system currently in place. Though some of the concepts and ideas can be generalized, the uniqueness of the Texas system may prevent general application of these conclusions. The Texas system of accountability is unlike any other state's system; therefore, the recommendations and conclusions may not apply to other accountability systems in the United States.

Definitions of Terms

Academic Excellence Indicator System (AEIS) – A comprehensive reporting system defined in Texas state statute. Local districts share responsibility for disseminating AEIS reports and for holding hearings for public discussion of the AEIS report content. All indicators for accountability are reported in the AEIS, with additional disaggregations depicting how each grade level and population performed. AEIS shows demographic information about students and staff, program information, and financial information, all of which provide context for interpreting accountability results (Texas Education Agency, 2005b).

Accountability – An obligation or willingness to accept responsibility or to account for one's actions (Merriam-Webster, 2007).

Adequate yearly progress (AYP) – A federal accountability program mandated under the No Child Left Behind (NCLB) Act. Under the accountability provisions in the No Child Left Behind (NCLB) Act, all public school campuses, school districts, and the state are evaluated for AYP. Districts, campuses, and the state are required to meet AYP criteria on three measures: reading/language arts, mathematics, and either graduation rate (for high schools and districts) or attendance rate (for elementary and middle/junior high schools).

Assessment – Testing (and other procedures) conducted to learn more about the competencies and weaknesses of the individual who is being tested. The purpose is to identify problem areas, needs and feedback (Merriam-Webster, 2007).

Criterion referenced – A type of test in which questions are written according to specific predetermined criteria. An individual's performance is compared to a specific learning objective or performance standard and not to the performance of other students.

English language learners (ELL) – A person who is in the process of acquiring English and has a first language other than English.

Limited English proficient (LEP) – A student for whom English is a second language.

Local education agency (LEA) – A term used to designate a school district or county office of education.

State Board of Education – The Commissioner of Education and the 15 elected members of the Texas State Board of Education (SBOE), who oversee the public education system of Texas in accordance with the Texas Education Code.

Student Success Initiative (SSI) – Grade advancement requirements, which apply to the Grade 3 reading test and the Grade 5 reading and mathematics tests. SSI was enacted by the 76th Texas Legislature in 1999, to ensure that all students receive the instruction and support they need to be academically successful in reading and mathematics.

Subgroups – Groups of students for which the Texas Education Agency disaggregates testing data when assigning accountability ratings to campuses and district. The named subgroups are African American, Hispanic, White, and economically disadvantaged.

Superintendent – The individual who has executive oversight and administration rights within the district.

Texas Essential Knowledge and Skills (TEKS) – Adopted by the State Board of Education in July 1997 and implemented in the 1998-99 school year. The TEKS replaced the Essential Elements as the statewide curriculum, and emphasizes the knowledge and skills that students are expected to learn rather than the delivery standard of the teacher (Texas Education Agency, 2007a).

Significance of Study

A review of pertinent literature revealed no other studies related to the impact of district size and diversity on the accountability ratings of districts. This study is timely, due to the current political climate emphasizing the importance of holding schools and districts responsible for student learning and the focus on accountability, as well as the likelihood of a reauthorization of No Child Left Behind. This study will provide additional information regarding the impact of district size and diversity on ratings. Examining the ways students and districts are assessed is essential to improving the system. Texas educational leaders are in a position to influence and direct the nation by creating a model accountability system.

Summary

With the upcoming presidential election in 2008 and historical data supporting the trend of reevaluating ESEA every five years, there is a strong possibility that NCLB will be reauthorized. With the intent of improving education, legislators and educators critically examine the system by which students are assessed, and evaluate the state's school districts. Therefore, until all students meet the goals of NCLB, school districts and superintendents will experience increasing pressure to achieve and perform at higher standards. Assessment and evaluation are essential parts of an educational system, as are high standards. Sirotnik (2004), in

Holding Accountability Accountable, asserts that a responsible accountability system must be sensitive to the complexity of the social, political, and economic circumstances within which it is to function. In addition, the system must also be focused on the students within its schools and provide equitable opportunities for student learning. This focus is as important as the attention on the curriculum. Sirotnik suggests there are many factors that affect a child's education and that the accountability system must look at all of these factors. Research addressing this issue should provide information that could impact decisions of policy-makers and practitioners.

Organization of Study

There are five chapters in this study. Chapter 1 presents an overview and introduction to the study. This examination of the Texas accountability system provides information regarding the impact of school district size and diversity on district ratings. In addition, the introductory chapter presents the background, statement of problem, and significance of findings, as well as the limitations of the study and definitions of terms.

Chapter 2 offers a review of literature and an examination of the state assessment systems of California, Kentucky, New York, and Texas. Chapter 3 describes the research design and methodology used in the study. Chapter 4 reports research findings and analysis of data. Chapter 5 provides a summary, discussion, and recommendations for future practice and research.

CHAPTER 2

REVIEW OF RELATED LITERATURE

This chapter examines the available research conducted on the topics of accountability and standardized testing and how diversity affects both. With the likely reauthorization of the No Child Left Behind Act in 2008, this study is of the utmost importance.

In order for lawmakers to provide a framework for fair and equitable district accountability, policy-makers must examine accountability systems and the equity involved on the state level, and then visualize application on the national level. Currently, accountability systems revolve solely around states' high-stakes testing. In this, the beginning stage of accountability in many states, the role of standardized testing and the balance between that and other variables must be addressed (Cizek, 2005). Critics of accountability and high-stakes testing note teachers and districts "teaching to the test." When curriculum and assessment are in alignment, however, teaching to the test is appropriate and is what should happen in order for content mastery to occur (Goodman & Hambelton, 2005).

The 1960s brought expanded opportunities for education for all students. Subsequent discussion in the 1970s addressed falling standards. National efforts to define standards date back to 1964 with public opinion in favor of more traditional discipline and minimum competency testing (Broadfoot, 1996). In preference to administration of tests on a single day, Cizek (2005) recommends an assessment system that includes general monitoring components, providing diagnostic and evaluation information for teachers and administration.

First, a review of various states' responses to accountability is reported below. Second, a chronological history of standardized testing in Texas is examined as it relates to the education of Texas students. Third, the impact of high-stakes testing is presented as it relates to special

populations such as minorities and subgroups. Also noted in this section is the impact of district size on achievement.

In any educational setting, evaluation and assessment play an important role in improving instruction. As Kohn (2000) suggests, the more a test is made to count in terms of being the basis for promoting or retaining students or for funding or closing down schools, the more anxiety is likely to rise and the less valid the scores become. In the state of Texas, standardized testing has been a part of the reality of districts, educators, and students for an entire generation.

Accountability

Requiring testing and accountability in education is not a new trend. High-stakes tests have been traced all the way back to 2,000 B.C. for military exams and 200 B.C. for civil service exams (Cizek, 2001). In more recent history, the public has asked two questions: (a) What counts? and (b) who is held accountable? Oftentimes, these two questions are answered in the narrowest of terms and do not account for the diversity of students and situations (Linn, 2003). Many of the high poverty schools that NCLB seeks to change have limited resources, poorly trained teachers, and instability of both student enrollment and staffing, making it difficult to accomplish goals without large increases in funding (Lee, 2006). This presents a challenge to districts that have large populations of economically disadvantaged or minority students, as well as those districts that do not enjoy a high per-pupil allocation.

A Nation at Risk

In 1983, *A Nation at Risk* was published. This report from the National Commission on Excellence in Education (NCEE) showed the need for reform on all levels, indicating that the United States was performing poorly in comparison to other countries. The 400,000 copies that

were distributed reached a readership of 5 million. The report suggested that minimum competency examinations fall short of what is needed as the minimum tends to become the maximum, thus lowering educational standards for all (Gardner, 1983). Rigorous assessments, high standards, and curriculum improvement were all recommended to hold schools responsible to meet these new standards. As a result of this report, every state but Iowa developed educational standards, and every state but Nebraska implemented assessment policies to evaluate their effectiveness.

Goals 2000

In September 1989, President George H. W. Bush and 50 state governors convened in Charlottesville, VA and agreed to work together to set education goals for the nation. In January 1991, the National Education Goals were announced by President George H. W. Bush and adopted by the governors. A panel was created in March 1991 to issue reports on the progress of the nation.

In 1991, President George H. W. Bush presented the following six America 2000 National Education Goals to the nation:

1. Ensure that children are ready to learn when beginning school
2. Maintain a 90% high school passing rate
3. Demonstrate competencies in core subjects at Grades 4, 8, and 12
4. Maintain a ranking of first in the world in science and math
5. Achieve complete adult literacy
6. Ensure that schools are free of drugs and violence (United States Congress, 1992)

After announcing these goals, President George H. W. Bush proposed national student testing. In the initial plan, the tests would be implemented at the 4th, 8th, and 12th grade levels and would be completely voluntary, providing incentives for students who passed (Phillips, 2004). With the support of the President, Congress established a National Council on Education Standards and Testing to explore and determine national education standards in regard to the six adopted goals.

Goals 2000 was signed into law by President Bill Clinton in 1994 with an increase in funding for education nationwide. His emphasis was on educational improvement with local control. Throughout President Clinton's proposal, there was support for alternative assessment; restrictions on standardized testing for promotion, graduation, or retention; and a mandated review of all standardized tests required for federal funding (United States Congress, 1992).

One study of 1,023 parents of school age children indicated that 83% believed tests provided important information regarding their children's progress, and 9 out of 10 wanted comparative data about their children and the schools they attend. Of these parents, two-thirds indicated they would like standardized test results for their children in every grade, with half of those parents indicating tests should be administered twice a year (Dreisler, 2001). The populace has historically supported testing for a number of reasons, including the relatively low cost, the quick changes that can be implemented, the visibility of testing results, and the testing that can create other changes that might be difficult to legislate.

According to Herman (2006), state assessments are focusing instruction, educators are working to align curriculum and instruction, administrators are becoming more attentive to data, and at-risk students are seeing new, augmented opportunities. Conversely, instruction is clearly aligned only with tests, not state standards; the curriculum can drastically narrow and become

boring; practice can become distorted and focused on triage only; and rhetoric outstrips quality of practice.

Performance standards vary from state to state, as is seen in an in-depth NAEP study by Linn (2005a). In 2003, he found, the state with the second-lowest average performance on the NAEP had more than 90% of its schools meet adequate yearly progress (AYP), while only one-half of the schools met AYP goals in the states with the highest performance on NAEP.

States' Responses to Accountability

As a response to the call for accountability, different states have addressed the issues with different standardized testing systems. Each state functions under two accountability systems: the state system and the federal accountability system embedded within the NCLB Act. All schools and districts are required to disaggregate data to examine subpopulations of students they serve. Each student group, including low socioeconomic, minority, limited English proficient (LEP), and students with disabilities are included. This disaggregation of scores, according to the United States Department of Education (USDE), no longer allows districts to hide students who were traditionally "left behind" and ensures that every student receives the quality education he or she deserves (USDE, 2006). By holding all states to a uniform proficiency target in reading and math, NCLB does not take into account differences in students' cognitive skills and ignores contributions that schools make to student learning. Consequently, districts that serve students with lower skills are at a disadvantage because they are required to improve at faster rates than other districts. Differences in the mean proficiency level among schools are often the result of differences in student skills and background characteristics before students enter school, not differences in rates of progress (Byrk & Raudenbush, 1988).

States have varying ways of addressing standardized testing and accountability systems. Legislators did not intend for NCLB to usurp states' authority or replace existing policy. The purpose of NCLB is to enhance state accountability systems. Those states that have had a strong accountability system in place are more prepared to meet the criteria set forth in the act.

There has been much praise for the emphasis on high goals for all children and the attention given to the groups that have been historically lower-performing. However, studies have found evidence that high-stakes testing that has resulted from NCLB has led to a narrowing of the instructional focus among teachers and principals. Teachers place greater emphasis on material that is covered on a high-stakes test than they do on other material. Students receive less instruction than previously on subjects not tested, and excessive pressure can actually detract highly qualified teachers from teaching at high-poverty schools.

Even though, at the beginning of the testing process, small gains were evident in mathematics, students have now reverted to the pre-existing patterns and these gains are no longer seen (Lee, 2006). Of the 44 possible AYP targets listed by the California Department of Education, large districts were required to meet, on average, 33 targets; medium districts averaged 20 possible targets; and small districts only 9 targets. In Illinois, there are 36 targets with large districts being required, on average, to meet 23 targets; medium-sized districts were required to meet 14; and small districts only had to meet an average of 8 targets. (See Table 1 below.) While disaggregation is essential, there is no real relevance for small schools with homogenous student bodies. Schools with multiple subgroups are at a relative disadvantage (Linn, 2007).

Table 1
Average Targets Required for AYP in California and Illinois

State	Possible Targets	Small Districts	Medium Districts	Large Districts
California	44	9	20	33
Illinois	36	8	14	23

The state systems examined below include California, Kentucky, New York, and Texas. These states are significant in the study because, according to Texas demographer Dr. Steve H. Murdock (2007), both California and New York will experience significant changes in their demographics over the next several years. California, New York, and Texas are the three largest states and are experiencing an influx of new, more diverse growth, especially in the area of Hispanic and English-language-learner populations.

California

California has a population of 36,132,147, according to the 2005 census data (United States Census Bureau, 2005). The demographics are increasingly diverse with the most current estimates from 2005 showing White 77%, Asian 12.2%, African American 6.7%, and American Indian 1.2%. Of the individuals who reported their racial group as White, 35.2% indicated they were Hispanic or of Latino origin while 43.8% indicated they were not of Hispanic origin (United States Census Bureau, 2005). California's 6,312,103 students attend 1,054 school districts where the average funding is \$7,658 per student, including local and federal dollars (California Department of Education, 2007a).

The California Academic Performance Index (API) is a numeric index ranging from a low of 200 to a high of 1000 that reflects a local education agency's (LEA's) performance level based upon the result of statewide testing. The API was established by the California Public

Schools Accountability Act (PSAA) of 1999. The PSAA has three main sections: the API, the Immediate Intervention/Underperforming Schools Program (II/USP), and the Governor's Performance Award (GPA) program. Used in calculating the API are results from the Standardized Testing and Reporting (STAR) Program and the California High School Exit Examination (CAHSEE). Statewide, the API performance target is 800. The two major purposes of API are to measure growth of school performance from one year to the next and to rank schools on an annual basis.

The Standardized Testing and Reporting (STAR) Program consists of the following:

- California Standards Tests (CSTs) in language arts, mathematics, history, social studies, and science in varying grades depending on content area
- California Alternate Performance Assessment (CAPA)
 - The CAPA in English language arts and mathematics is included for Grades 2 through 11.
- Norm-Referenced Test (NRT)
 - The California Achievement Test, Sixth Edition Survey, (CAT/ Survey) is included for all content areas at Grades 3 and 7 only. Content areas tested included reading, language, spelling, and mathematics.

The California High School Exit Examination (CAHSEE) is also used as a measure by which districts are assessed. Below are the performance levels and weighting factors used in the API, according to the California Department of Education (2007b) (See Table 2).

Table 2

Performance Levels and API Weight Factors in the California Department of Education (CDE)

CST or CAPA Performance level	NRT Performance Bands	CAHSEE Score	API Weighting Factors	Point Gain Movement
Advanced	80-99 th NPR	Pass	1000	1000-875=125
Proficient	60-79 th NPR	N/A	875	875-700=175
Basic	40-59 th NPR	N/A	700	700-500=200
Below Basic	20-39 th NPR	N/A	500	500-200=300
Far Below Basic	1-19 th NPR	N/A	200	N/A

The point-for-gain column shows points awarded to low-performing schools, allowing them to improve at an accelerated level when gains are made. The annual API growth target is 5% of the difference between the school's API and the statewide performance target, or a minimum of one point growth. Schools with an API base of 800 or above must maintain an API at 800 or above.

All numerically significant subgroups must show comparable improvement in meeting targets. However, the law does not specifically state what comparable improvement means. The subgroups that are recognized in reference to API are racial, socioeconomically disadvantaged, ELL, and students with disabilities. Each numerically significant student subgroup must achieve at least 80% of the school-wide annual growth target. (See Table 3.)

Table 3

Subgroup Populations for the API Indicators as Defined by the CDE

<p><i>A numerically significant subgroup</i> for the API is defined as:</p>	<p>OR</p>	<ul style="list-style-type: none"> • 100 or more students with valid Standardized Testing and Reporting (STAR) Program scores • 50 or more students with valid STAR Program scores who make up at least 15 % of the total valid STAR Program scores 			
<table border="0" style="width: 100%;"> <tr> <td style="vertical-align: top; padding-right: 20px;"> <p>Subgroups used in API calculations include</p> </td> <td style="vertical-align: top;"> <ul style="list-style-type: none"> • African American • American Indian or Alaskan Native • Asian • Filipino • Hispanic or Latin • Pacific Islander • White • Socioeconomically disadvantaged • English learners • Students with disabilities </td> </tr> </table>			<p>Subgroups used in API calculations include</p>	<ul style="list-style-type: none"> • African American • American Indian or Alaskan Native • Asian • Filipino • Hispanic or Latin • Pacific Islander • White • Socioeconomically disadvantaged • English learners • Students with disabilities 	
<p>Subgroups used in API calculations include</p>	<ul style="list-style-type: none"> • African American • American Indian or Alaskan Native • Asian • Filipino • Hispanic or Latin • Pacific Islander • White • Socioeconomically disadvantaged • English learners • Students with disabilities 				
<p><i>Socioeconomically disadvantaged</i> is defined as:</p>	<p>OR</p>	<ul style="list-style-type: none"> • A student whose parents both have not received a high school diploma • A student who participates in the free or reduced lunch program 			
<table border="0" style="width: 100%;"> <tr> <td style="vertical-align: top; padding-right: 20px;"> <p><i>English language learners</i> are defined as:</p> </td> <td style="vertical-align: top; padding-right: 20px;"> <p>OR</p> </td> <td style="vertical-align: top;"> <ul style="list-style-type: none"> • English learners • Reclassified fluent-English-proficient (RFEP) students who have not scored at the proficient level on the California Standards Test (CST) in English-language arts (ELA) for three years after being reclassified </td> </tr> </table>			<p><i>English language learners</i> are defined as:</p>	<p>OR</p>	<ul style="list-style-type: none"> • English learners • Reclassified fluent-English-proficient (RFEP) students who have not scored at the proficient level on the California Standards Test (CST) in English-language arts (ELA) for three years after being reclassified
<p><i>English language learners</i> are defined as:</p>	<p>OR</p>	<ul style="list-style-type: none"> • English learners • Reclassified fluent-English-proficient (RFEP) students who have not scored at the proficient level on the California Standards Test (CST) in English-language arts (ELA) for three years after being reclassified 			
<p><i>Students with disabilities</i> is defined as:</p>	<p>A student who receives special education services and has a valid disability code</p>				

(California Department of Education, 2007b)

API ranks are determined by matching districts with those that are similar demographically and by decile rank statewide. In order to ensure demographic similarity, a “similar school rank” is created. In order to match like districts, a number of factors are compared: pupil mobility, ethnicity, economic status, percentage of fully credentialed teachers, percentage of teachers who hold emergency credentials, average class size, structure of the school year, percentage of students in gifted and talented, percentage of students in special education, percentage of LEP students, percentage of migrant students, and percentage of students in a reduced class size for a full school day. According to the rankings associated with these results, districts may rank from a low of 1 to a high of 10. Therefore, districts are assessed based upon varying factors and using multiple measures (Education Data Partnership, 2007).

The California system is unique in that the base API, released each March, is calculated from state-wide test results to be used in providing continuing and new assessments from the system. The test results used in the API California’s accountability system measure the performance and progress of a school or LEA based on results of statewide tests, Grades 2 through 12 (California Department of Education, 2007b).

Kentucky

Kentucky was included in the study because of its established standardized testing system and because of its location in the United States. It provides a balance to New York on the East Coast and California on the West. Kentucky also enjoys relative neutrality, with the added caveat that the state rewards districts that are high-performing.

The Kentucky system has been in place in its current form since 1999. Kentucky was admitted to the Union as the 15th state in 1792, with Frankfort as its capitol. Its population in 2003 was 4,171,827, according to the 2003 census, with 89% of residents being White and 7%

African American, the remainder being Hispanic and Asian. The 653,248 students in the state attend 176 school districts where, in the school year 2002-2003, \$7,022 was spent per pupil and the average teacher salary was \$40,849 (Commonwealth of Kentucky, 2005).

The Commonwealth Accountability Testing System (CATS) is Kentucky’s system set in place for all schools to reach proficiency, as defined by the Kentucky Board of Education. The long-term accountability model adopted by the Kentucky Board of Education is a growth model where each school serves as its own baseline. School improvement plans are set in place, and students must show yearly improvement in all areas (Kentucky Department of Education, 2007).

Measuring the progress of districts is the Kentucky Accountability Index, a numeric composite score reflecting student performance with reference to Kentucky Performance Standards – Novice, Apprentice, Proficient, and Distinguished. (See Table 4.)

Table 4
Kentucky Accountability Index

Performance Standard	Criteria
Distinguished	<ul style="list-style-type: none"> • Student demonstrates an in-depth, extensive or comprehensive knowledge of content. • Student communication is complex, concise, and sophisticated with thorough support, explicit examples and justifications. • Student uses and consistently implements a variety of appropriate strategies.
Proficient	<ul style="list-style-type: none"> • Student demonstrates broad content knowledge and is able to apply it. • Student communication is accurate, clear and organized with relevant details and evidence. • Student uses appropriate strategies to solve problems and make decisions.

(table continues)

Table 4 (continued).

Performance Standard	Criteria
Apprentice	<ul style="list-style-type: none"> • Student demonstrates some basic content knowledge and reasoning ability. • Student communicates reasonably well but draws weak conclusions or only partially solves or describes. • Students attempts appropriate strategies with limited success.
Novice	<ul style="list-style-type: none"> • Student demonstrates minimal, limited, underdeveloped, and at all times inaccurate content knowledge and reasoning. • Student communication is ineffective and lacks detail with no evidence of connections within or between content areas. • Student uses inappropriate strategies.

(Kentucky Department of Education, 2007)

In addition, non-academic indicators are utilized in school and district ratings as well as a nationally norm-referenced test. Percentages of students scoring at each standard in reading, writing, mathematics, science, social studies, arts, humanities, practical living, and vocational studies at each tested grade level are converted to a composite index used to identify successful schools and those in need of improvement. For a school to be successful, it must meet a dropout criteria of less than or equal to 5.3% or reduce the percentage by 0.5% and reduce the percentage of Novice students so that by 2014, the school will have 5% or less of its students scoring at a Novice level. Schools are held accountable for students enrolled for a total of 100 instructional days in a school (Kentucky Department of Education, 2007).

The Kentucky Core Content Test (KCCT) is an open-response test whose results contribute most to the school accountability index. The KCCT is given at all grade levels where

reading, mathematics, science, and social studies are tested. Each student is given 7 open-response and 28 multiple-choice questions. According to the CATS Interpretive Guide, these open-response (OR) items are extremely important to the assessment because Proficient and Distinguished performance across the OR items depends on students having received high-quality instruction. Students who historically scored higher on the OR responses also achieved higher marks on the multiple-choice questions. The multiple-choice questions were added to the test to broaden the scope of material tested and to increase the reliability of scores (Kentucky Department of Education, 2007).

When reporting the scores, a *fairness margin* is noted. The fairness margin takes into account that testing is often imprecise. The CATS accountability model gives a cushion to schools if a school scores just below the goal line, but is within one standard error. If this is the case, the school is treated as though the school was at or above the goal line, allowing for some flexibility in determining ratings. The three levels of assistance determined by the state for schools failing to make progress are scholastic self-review, scholastic review, and scholastic audit (Kentucky Department of Education, 2007).

There are several points to remember in regard to the Kentucky assessment program. The accountability index measures the growth of successive cohort groups. Multiple measures are used and scores are weighted and combined utilizing both the non-academic and academic data. Schools and districts are recognized for performance, either by intervention or incentives. In 2003, 31% of students were proficient in math on Kentucky's assessment versus 24% that were proficient on the NAEP. This 1.3 discrepancy between the NAEP and the state assessment indicates that the standards for assessment have been set at a significantly lower level than the NAEP (Lee, 2006).

The strength of CATS is that each school and district serves as its own baseline, so growth is measured individually. School improvement plans are used to design interventions to improve academic progress. Campuses that do not meet their academic goals receive a scholastic audit and the services of a highly skilled educator to assist in their improvement efforts. In addition, these schools and districts are eligible to apply for additional grant funding. Incentive funds are awarded for schools that achieve award status. The Kentucky program offers funding for districts that meet and exceed the accountability requirements set forth. It also provides funds for districts that fail to meet their accountability targets (Kentucky Department of Education, 2007).

New York

New York's testing program offers a different perspective, as the state is educating an ever-increasing diverse population more than four times the size of Kentucky's public school enrollment. The 2006 census estimate showed 19,306,183 individuals residing in New York, including a White majority accounting for 73.8% of the population, African American 17.4%, Asian 6.7%, and the remainder identifying as "other." In 2006, 16.1% of individuals identified themselves as Hispanic or Latino, while many individuals reported ethnicity in multiple categories (United States Census Bureau, 2006). The New York educational system serves 2,710,000 students in 695 school districts. In the 2006-07 school year, the percentage of state to federal dollars spent was 44.3% with total expenditures of \$48,300,000,000.

Students in New York are tested in English/language arts, mathematics, and science in Grades 3-8. This test combines multiple-choice questions with short- and long-response essays based upon a listening selection and paired reading selection. Students who score below the state-designated performance level are required to receive academic intervention services the

semester following the administration of the test (New York State Department of Education, 2007b).

All students in public school, including those who have been retained, must take these assessments. Students who are classified as LEP are required to take the New York State English as a Second Language Achievement Test (NYSESLAT), provided they have been enrolled in United States schools for one year. Individual determinations are made regarding testing students with disabilities. For those for whom it is determined the test is not an appropriate measure, the New York State Alternative Assessment (NYSAA) is an option (New York State Department of Education, 2007b).

Student performance is measured on a four-point scale from basic to advanced proficiency. A performance index (PI) from 0-200 is then assigned to each accountability group. This PI is calculated as the number of continuously enrolled students scoring at 2, 3, or 4 plus the number scoring at 3 and 4 divided by the number of continuously enrolled students, once again times 100. (See Table 5.)

Table 5
New York Performance Index (PI)

Test Grade	Number of Students	Levels			
		1	2	3	4
3	35	1	7	10	6
4	43	3	6	20	14
5	30	6	10	10	4
Total	108	21	23	40	24

$$\text{Index} = [(23 + 40 + 24 + 40 + 24) / 108] \times 100 = 140$$

Note: The methodology is the same regardless of how many grade levels the school serves.
(New York State Department of Education, 2007a)

Annual measurable objective (AMO) is the PI value that New York uses to signify that each accountability group is making satisfactory progress toward meeting AYP. A confidence interval is provided around the AMO. For each accountability group, the four small squares represent four schools with the same PI, but with different numbers of tested students. The vertical lines represent the confidence intervals for each school based on the number of students tested. The more students tested, the smaller the confidence interval. (See Figure 1.)

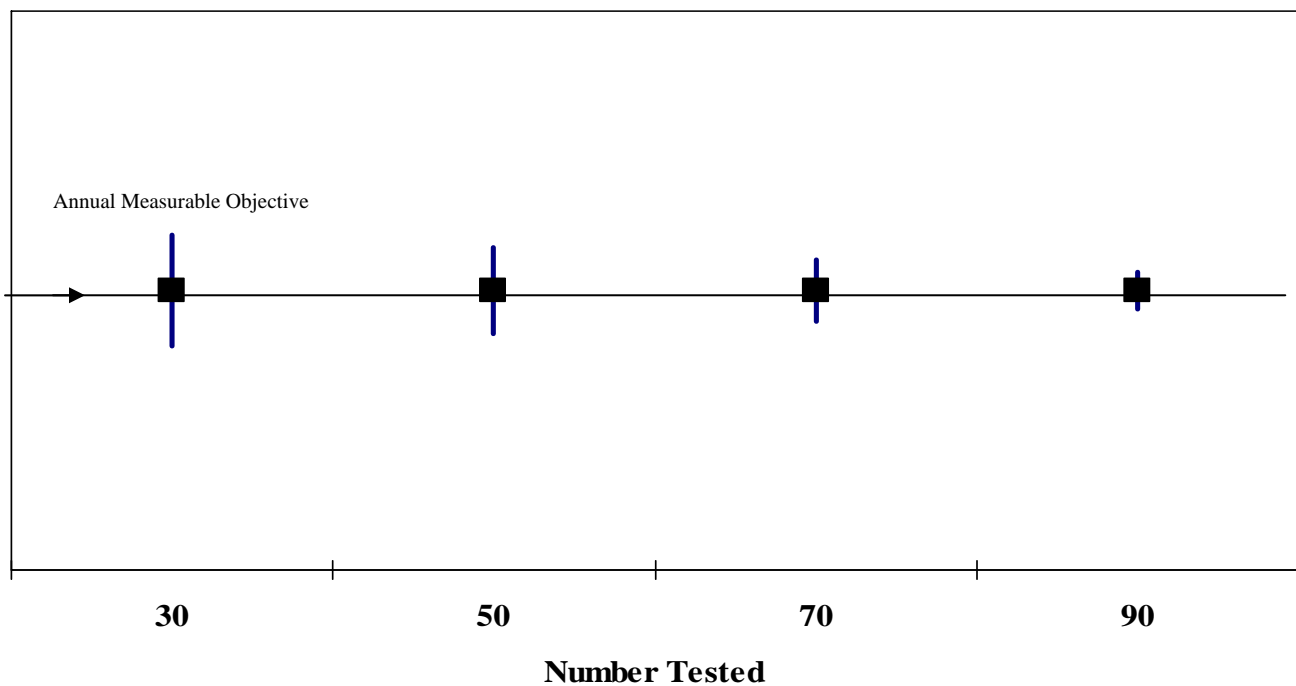


Figure 1. Confidence intervals of annual measurable objectives. (New York State Department of Education, 2007a).

An effective AMO is the lowest PI that an accountability group of a given size can achieve in a subject for the group’s PI not to be considered significantly different from the AMO for that subject. If an accountability group’s PI equals or exceeds the effective AMO, the group is considered to have made AYP (New York State Department of Education, 2007b).

If a campus fails to meet AYP for two consecutive years on the same measure, the campus is identified as needing improvement. To be removed from improvement status, a school

must make AYP on that measure for two consecutive years. The school may remain or be placed in improvement status on another measure for which it has not made AYP. District level results are disaggregated for all students attending school in the district as well as for students that are enrolled in approved private school placements. Accountability areas include English language arts (ELA), mathematics, science, and graduation rates with requirements for failure to meet AYP for two consecutive years in ELA or mathematics in both elementary and secondary or in science and graduation rate. Interestingly, a district may be identified for improvement even if no school in the district is identified for improvement. For small districts where there is only one school, the district and the school can have different accountability status scores because the district accountability groups include students placed outside the district (New York State Department of Education, 2007b).

If an elementary or middle school does not test 30 continuously enrolled students in ELA or mathematics, the scores over two years will be combined to determine the PI. If a school still does not have 30 students on which to base a decision, the school is subject to special procedures for determining AYP (New York State Education Department, 2007b).

One strength of the New York accountability system is the recognition that the groups of students change from year to year. The office of the Information and Reporting Services indicates that there are some areas that are within a school's control, such as curricular changes and personnel changes, that may reflect student performance increases or declines. However, there are three areas that are indicated as being subject to error beyond the district's control:

- A measurement error due to factors such as health, motivation, attention
- A sampling error caused by random variants in student ability, early preparedness, and motivation from grade to grade in the same school

- A change in the environment not under the school's control, for example the events of September 11 (New York State Department of Education, 2007b)

To account for some of these deviations, New York minimizes the chance that a district will be deemed to have not made AYP by utilizing a confidence interval to determine whether a group has met its AMO. This confidence interval recognizes the sampling error associated with an observed score and allows the examiner to analyze whether the difference between the observed performance index (PI) and the AMO falls within certain bounds or whether that difference falls outside of the margin of error and is not attributable to chance alone (New York State Department of Education, 2007a).

According to the New York Information and Reporting Services, because it is impossible to make statistical statements about the performance of a school with total accuracy, there is always a degree of error when deciding whether a group met the AMO. New York State's system minimizes the chance of falsely determining that a district did not meet AMO by allowing flexibility in the interpretation of the scores. What is reported on test day is a snapshot that represents a static picture of what students and districts have the opportunity to report on any given day. This is a decided strength of the New York testing program as it accounts for differences in testing situations, errors in grading, variances in student health, and many other factors that cannot be controlled. By building in a small margin of error, it lessens the pressure on students and staff. While the New York public education system educates nearly 3,000,000 students, the California system serves more than twice that number. The accountability structure incorporates a multifaceted approach to rank and evaluates school effectiveness (New York State Department of Education, 2007b).

Testing in Texas

Standardized testing, as the Texas Education Agency (TEA) has interpreted it, is defined as “an examination administered under strictly uniform conditions and interpreted in a consistent manner” (Kennedy, 2003, p. 7). In theory, the same test is administered to all grade level students under the same conditions and the data are interpreted in a standardized manner as well. This standardization, according to Wilgoren (2000), often drives good teachers out of the profession, can lead to widespread cheating, may turn teachers against students, and may cause teachers to look at students with special needs as a liability.

With NCLB came examination of the first- and second-generation testing states. Researchers have found that first-generation accountability states’ systems, as in Florida, North Carolina, and Texas, did not work. NCLB neither enhanced the first-generation states’ earlier academic improvement nor transferred the effects of an accountability system to states that adopted test-based accountability under NCLB, which are the second-generation accountability states. Data gathered showed that high-stakes testing in the first- and second-generation states failed to narrow the math and reading gaps after NCLB (Lee, 2006).

The mid-1980s began an era of change in Texas. Governor Mark White appointed a commission headed by Ross Perot to review Texas education and make recommendations for its improvement. The main recommendations of the Equal Education Opportunity Act of 1984, enacted under Texas House Bill 72, and of Perot’s Commission included reducing the monetary gap between funding for rich and poor school districts, lowering class size to 22 in kindergarten through second grade, moving to a full-day kindergarten, and providing subsidized pre-school education for low-income four-year-olds.

In addition, Perot proposed developing standardized achievement tests for all students. It was Perot who made popular the phrase that came to represent the thrust of his reforms: *No pass, no play*. This came to mean all students had to maintain passing grades in order to participate in extracurricular activities. In his book *United We Stand* (1997), he writes:

Establish national standards and measure results. We'll never fix this system until parents, as consumers, can plainly see how schools measure up against one another and against the world competition. This information is practically impossible to obtain today. Parents should be able to know how their elementary school performs against the nation's and the world's. Employers need to know how their local school districts perform against others in their state. Principals and teachers need to see where they are succeeding and where they need to concentrate their resources for improvement. Right now we have a \$185 billion enterprise operating essentially in the dark. We shouldn't be surprised that it doesn't work. We need to haul it out into the light of day, measure results, student by student, in a thorough, fair way and publish the results school by school for everyone to see. (p.79)

Perot began an era of accountability and change for Texas and Texas schools that, more than 20 years later, has reached to the national level. Critics say that the testing of students increasingly drives curriculum and compromises both teaching and the role of students in learning. This creates a new type of discrimination, as teaching to the fragmented and narrow information on the test begins to substitute for the curriculum in the schools of poor and minority youth. Disaggregating scores by race initially appears to promote equality, but the high stakes attached to the scores have forced many schools to replace the regular curriculum in minority students' classrooms with test-prep materials that have no real-world value. While the scores initially go up in these classrooms, academic quality goes down (McNeil, 2000). Between 1980 and 1998, the number of states that mandated student testing increased from 29 to 48 (Hoff, 1999). By 1998, 39 states were administering performance-based assessment while 24 states attached incentives, such as student recognition or promotion.

Texas Assessment of Basic Skills

The first formal standardized assessment occurred in Texas almost 30 years ago with the Texas Assessment of Basic Skills (TABS). In 1979, the Texas Legislature amended a bill that changed the Texas Education Code by linking student assessment with the statewide curriculum, thus creating the TABS test. This required the Texas Education Agency (TEA) to create and administer criterion-referenced tests designed to assess basic skills in the areas of mathematics, reading, and writing in Grades 3, 5, and 9. At this time, there was no mandated statewide curriculum, so the learning objectives represented on the TABS tests were indicative of a small sampling of the curriculum covered in the grades tested (Texas Education Agency, 2005b).

TEA worked with educator committees to review and revise assessment recommendations prior to formal approval by the Texas State Board of Education (SBOE). In 1983, the Texas Education Code was amended again to require that students in ninth grade who failed the TABS retake each year until passing. Students who failed the tests were not denied diplomas if they did not succeed, but the feeling among the Texas Education Agency was that the increased pressure on schools would result in higher academic standards and support. Furthermore, this was the first time the test results were released to the public (Texas Education Agency, 2005b).

Texas is considered a first-generation accountability system because of the presence of high-stakes testing and a strong accountability system prior to the 1990s. Strong accountability states, generally the first-generation systems, with larger achievement gaps have narrowed some of the gaps more than other states; however, there is no significant difference between the two groups once the initial difference in achievement level is considered. There is no indication that the gaps narrowed more or less in one group of states than the other after NCLB.

There was no connection between state accountability and pre-NCLB and post-NCLB changes in regard to racial and socioeconomic gaps. In terms of changes in achievement gaps – for African Americans, Hispanics, and economically disadvantaged students in both reading and math – few states narrowed the gaps significantly and there were no systemic differences between strong accountability states (such as Alabama, Florida, Illinois, Indiana, Kentucky, Louisiana, Maryland, New Jersey, New Mexico, New York, North Carolina, and Texas) and weak accountability states (such as Alaska, Arkansas, Colorado, Delaware, Idaho, Iowa, Maine, Massachusetts, Montana, Nebraska, New Hampshire, North Dakota, and Wyoming) (Lee, 2006).

Texas Educational Assessment of Minimal Skills

In 1984 the Legislature changed the wording in the Texas Education Code from *basic skills competencies* to *minimum basic skills*. The Texas Education Agency reacted by increasing the rigor of the assessments. In addition, this initiative, supported by State Board of Education rules and Texas state law, added individual student requirements for performance. The new assessment, titled the Texas Educational Assessment of Minimal Skills (TEAMS), was given to students in Grades 1, 3, 5, 7, 9, and 11. Students in Grade 11 were required to pass the exit-level test in order to graduate from high school, beginning with the graduating class of 1987. Students who did not pass the exit level test were denied their diplomas.

The most significant changes made were an increase in the number of grade levels tested and the mandatory remediation of students failing to master material at the exit level. This shift also evidenced an increased interest in educational reform. Campus and district summaries continued to be published in newspapers, thus highlighting the successes and struggles each district experienced (Texas Education Agency, 2005b). Though the public reporting of test scores is a way to inform the public, there should be clear explanations of legitimate and

potentially illegitimate interpretations of the results (Baker, 2002). Information should be provided to the general population regarding score interpretation and data disaggregation.

Texas Assessment of Academic Skills

Revisions to the assessment program were made by TEA under the direction of the State Board of Education in the late 1980s. These changes were based upon revisions of the Texas Education Code and the Texas Administrative Code rules. Changes included an expansion of content measured and a greater emphasis on the assessment of problem-solving skills. This new assessment, the Texas Assessment of Academic Skills (TAAS), was first implemented in 1990. TAAS was still a basic skills test but was more difficult and comprehensive than the TEAMS. In addition to revising the assessment, the Texas Education Agency set a new state curriculum, began to collect data for state accountability, and created a rating scale by which districts would be measured. With information gathered from the administration of TAAS, local districts could make instructional improvements in programs (Rhoten et al., 2003). TAAS was administered to students in Grades 3, 5, 7, 9, and 11, with Grade 11 serving as the exit level test. Untimed tests were administered in the areas of mathematics, reading, and writing.

Changes were made in 1994 to include Grades 4 and 8 for writing. Exit level testing was changed to 10th grade to provide for additional remediation and retesting. McNeil (2000) states:

The TAAS system of test-driven accountability masks the inequities that have for decades built unequal structures of schooling in Texas. Test score inflation, through concentrated test prep, gives the impression that teaching and learning are improving in minority schools when, in fact, teaching and learning may have been severely compromised in the attempt to raise scores. The investments in expensive systems of testing, test design, test contracts and subcontracts, training of teachers and administrators to implement the tests, test security, realignment of curricula with tests, and the production of test-prep materials serve a political function in centralizing control over education and linking public education to private commerce. (p. 259)

Texas Assessment of Knowledge and Skills

Sweeping changes were made by the 76th Texas Legislature to the standardized testing system with the passage of Senate Bill 103 in June 1999. The bill created the Texas Assessment of Knowledge and Skills (TAKS). The format of the test remained the same; however, content became more difficult due to the assertion that the test had not been sufficiently rigorous. Texas Education Code (TEC), Chapter 39, Subchapter B provides information regarding current assessment provisions (Texas Statutes, 2006).

In September 2001, requirements regarding the participation of English language learners (ELL) in the state assessment program were established. Regarding ELL participation, there are particular challenges these students encounter that regular education students do not. Historically, ELL students show slow improvement and score lower than their age-group peers, often 20 to 30 percentage points below. Because of the verbal demands of the test, ELLs are not measured accurately in either achievement or ability. Lastly, factors such as mobility affect the stability and integrity of student subgroup data (Abedi & Dietel, 2004). Regardless, English language learners continue to be evaluated as an accountability group, requiring these children to be assessed on a yearly basis with their peers.

In May 2002, guidelines regarding the Student Success Initiative (SSI) were published to support Senate Bill 4 testing requirements which had been passed by the 76th Texas Legislature. Amendments were made in the SBOE rules in the fall of 2004 with subsequent changes being made in the spring of 2005. The state developed alternative assessment (SDAA) for students in Grades 9 and 10 with disabilities was developed at that time (Texas Education Agency, 2005a, Texas Education Agency, 2007a).

Federal legislation required some clarification regarding the assessment of ELL student participation. NCLB requirements added English language proficiency for students in additional domains and grades. This was added to clarify exemptions provided for immigrant students and to require students who were merely limited English proficient to take the required assessment on grade level (Texas Education Agency, 2005a).

Conversely, SSI was amended to allow districts to consider past performance on standardized tests when considering retention. Also to be taken into account were extenuating circumstances and a student's ability to participate in regular instruction (Texas Education Agency, 2005a).

The TAKS was administered beginning in the 2002-2003 school year. At the present time, reading is tested in Grades 3-9; writing at Grades 4 and 7; English language arts at Grades 10 and 11; mathematics at Grades 3-11; science at Grades 5, 10, and 11; and social studies at Grades 8, 10, and 11. The Spanish TAKS is administered in Grades 3-6. Satisfactory performance on the TAKS at various grade levels and subjects, including the exit level, became a prerequisite for promotion (Texas Education Agency, 2007a).

According to Johnson and Johnson (2006), since the year 2000, seven more states have added retention policies based on a single test score. In Texas, test scores do not take into account student grades for the academic year. Students in Grade 3 are required to pass the reading TAKS and students in Grade 5 are required to pass both the reading and math TAKS. If students do not pass the test on the first attempt, parents are notified of the possibility of retention and an individual plan for acceleration is formulated. Students then have the opportunity to retake the test roughly six weeks later. This second administration determines the group of scores upon which district ratings are calculated. For students who do not pass the

second administration, school personnel convene a grade placement committee (GPC) meeting to discuss accelerated instruction and to prescribe the appropriate assessment for the third administration of testing. At this time parents may waive the third assessment opportunity and choose to automatically retain their child in grade. After the second administration, if the GPC decides retention is in the best interest of the child, the parent is given notification and the chance to appeal. The GPC may promote the student if it decides unanimously that the child will likely perform on grade level the following year, if given additional accelerated instruction. GPCs, however, can determine that a student will not be successful, even with the added intervention; that the student cannot move on to the next grade level; and that retention is the only option (Texas Education Agency,2007a).

The Texas Accountability System

Assessment and accountability work closely together in the state of Texas. Like the assessment system, the accountability system has evolved over the years. This study provides an analysis of data that examines the ratings of 2005 and 2006; therefore, an examination of the accountability system is included. Following is an account of how districts have been assigned ratings in the state and how their progress has been determined.

The state is divided into 1,046 districts that are governed by the Texas Education Agency. In 1993, the Texas Legislature enacted statutes that mandated the creation of the public school accountability system to rate school districts and evaluate campuses. With the TAKS, a new accountability system was introduced in the fall of 2004 (Texas Education Agency, 2005a).

The year 2006, the third year of the new system, introduced a number of changes:

- Significant increase in the rigor of the TAKS standards for all subject in order to achieve or maintain a rating of Academically Acceptable

- An increase in the rigor of the under-reported student indicator, which can prevent a district from being rated Exemplary or Academically Acceptable
- Completion of phasing in the passing standard on the TAKS
- the use of Completion Rate I, which does not count students who receive their GEDs as completers
- additional required improvement opportunities for SDAA II
- adjustments to the accountability subset as well as adjustments to ratings in situations where Hurricanes Katrina and Rita adversely affected schools and districts
- an increase in the rigor of the Recommended High School Program/Distinguished Achievement Program (RHSP/DAP) indicator for Gold Performance Acknowledgment (GPA)
- Replacement of the TAAS/TASP equivalency indicator with the Texas Success Initiative (TSI) – Higher Education Readiness Component indicator for GPA. (TEA, 2006)

According to the Texas Education Agency (2006), the 2006 accountability rating system uses the following four base indicators to assess districts:

- Spring 2006 performance on TAKS
- Spring 2006 performance on the State Developed Alternative Assessment II (SDAA-II)
- The completion rate I for the class of 2005, and
- The 2004-2005 annual dropout rate for Grades 7 and 8.

For ease of explanation, each indicator will be addressed separately below.

Texas Assessment of Knowledge and Skills

The TAKS indicator is the number students who took the TAKS who met the minimum standard, divided by the entire number of students who were tested. Students are evaluated in the areas of reading/ELA, writing, social studies, mathematics, and science. Performance in eighth grade science will not be used in accountability factoring until the 2008 school year (Texas Education Agency, 2005a).

The TAKS standards are evaluated for all student groups including subgroups such as African American, Hispanic, White, and economically disadvantaged. The scores are disaggregated and then placed in one of these following categories.

- Exemplary – For every subject, at least 90% of tested students passed the test.
- Recognized – For every subject, at least 70% of the tested students passed the test.
- Academically Acceptable – Varies by subject:
 - Reading/ELA – At least 60% pass
 - Writing – At least 60% pass
 - Social Studies – At least 60% pass
 - Mathematics – At least 40% pass
 - Science – At least 35% pass (Texas Education Agency, 2007)

In order for a particular subgroup's data to be calculated within the district's accountability score, minimum size requirements must be met. If a student group has fewer than 30 students, it is not evaluated. If there are 30-49 students within the student group and it comprises at least 10% of all students, the group is evaluated. If there are at least 50 students within the student group, it is evaluated. Student group size is calculated subject by subject. Therefore, grade levels

vary based upon the tests given. The student group size is integral to this study as group size and diversity are the focus of the study's comparisons.

State Developed Alternative Assessment (SDAA-II)

The SDAA-II is used to test students in special education in Grades 3-10 for whom the TAKS is not an appropriate measure of their progress. The SDAA-II standard is based upon Admission, Review, and Dismissal (ARD) Committee recommendation accuracy. There are no subgroup breakdowns for SDAA accountability ratings. Minimum group size is 30; however, depending on the grade level, one student may count as many as three times if he or she takes three different tests. Therefore, the minimum group size of 30 could potentially apply to a group as small as 10.

- Exemplary – Results on at least 90% of tests taken met ARD expectations
- Recognized – Results on at least 70% of tests taken met ARD expectations
- Academically Acceptable – Results on at least 50% of tests taken met ARD expectations (Texas Education Agency, 2005a)

Completion Rate

Districts who serve Grades 9-12 are evaluated by the longitudinal rate that shows the percentage of students who first attended ninth grade in the 2001-2002 school year and have completed their education four years later. For the 2006 accountability year, the definition of “completer” changed; a student who attains a GED is no longer considered a completer. Performance is evaluated for all students and for African American, Hispanic, White, and economically disadvantaged subgroups. In order for a subgroup to be evaluated, there must be at least 5 dropouts within the student group; there must be 30-49 students within the group and the

group must comprise at least 10% of the student population; or there must be at least 50 within the student group. The standards are as follow:

- Exemplary – Completion rate I of 95% or more
- Recognized – Completion rate I 85% or more
- Academically Acceptable – Completion rate I of 75% or more

Results are based on the original student cohort whether the student remains on grade level or not (Texas Education Agency, 2005a).

Annual Dropout Rate

The annual dropout rate is used to evaluate districts with students in seventh and eighth grade. This is a one-year measure calculated by summing up the number of dropouts across the two grades.

- Exemplary – Annual dropout rate of 0.2% or less
- Recognized – Annual dropout rate of 0.7% or less
- Academically Acceptable – Annual dropout rate of 1% or less (Texas Education Agency, 2007)

This four-pronged system provides the public with a great deal of information about the districts of Texas. Despite the advantages of informing the public, McNeil (2000, p. 231) cautions that linking the accountability system to cost-accounting actually limits public discourse on the nature and purpose of education, thus diminishing the role of the populus in education.

According to NCLB subgroup rules, each subgroup of students must reach the state-defined proficiency level in reading and mathematics. In addition, 95% of students in each subgroup must take the assessment. If any subgroup fails the minimum group size criteria, the subgroup is counted when calculating AYP. Large districts are at a disadvantage when

accounting for subgroups, since they are more likely than small districts to meet the minimum subgroup size required for reporting AYP data and therefore must report more subgroups than smaller districts. State accountability plans designate the minimum number of students that a subgroup must contain to be subject to the accountability provisions, For each additional subgroup that a district must report, the district must meet at least four additional indicators (for each subject, participation, and proficiency rate). If any one of the subgroups fails to meet one of these four indicators, the district fails to meet AYP. Therefore, a district with more subgroups has more students who are held accountable and the district is at a greater disadvantage (Tracey et al., 2005).

Diversity and Testing

As indicated in previous sections, the United States is becoming increasingly diverse economically, ethnically, and linguistically. According to the United States Bureau of the Census (2007), there is one birth every 8 seconds, one death every 13 seconds, and one international migrant every 27 seconds for a net gain of one person per 11 seconds.

According to Dr. Steve Murdock (2007), the population in Texas is changing significantly and will continue to do so over the next few decades. Table 6 shows a percentage population comparison between 2000 and 2005 with a population increase of 9.6%. The decrease in the White population and increase in Hispanic population should be noted.

Table 6
Percentage Population Increases in Texas

	2000	2005
White	53.1%	49.2%
African American	11.6%	11.2%
Hispanic	32.0%	35.1%
Other	3.3%	4.5%

(Murdock, 2007)

Table 7 shows percentage increases of estimated population through the year 2040, assuming that net migration is equal to that of the years 2000-2010.

Table 7
Texas Population Projection

	White	African American	Hispanic	Other	Total
2000-2010	2.7%	14.7%	47.3%	64.7%	20.4%
2010-2020	.8%	11.8%	41.7%	56.1%	20.5%
2020-2030	-1.4%	8.0%	36.7%	50.1%	20.1%
2030-2040	-3.5%	5.0%	33.1%	44.8%	20.1%
2000-2040	-1.6%	45.5%	279.7%	458.9%	109%

Murdock (2007) also indicates that for 2040, the total number of households living in poverty is expected to rise to 16.6% from 14.4%.

Some discrepancies have been noted regarding testing students who are in these minority subgroups. According to the American Evaluation Association (AEA) policy statement (2002), standardized testing has resulted in violations of AEA guidelines by encouraging, in either direct or indirect ways, that students who might not pass the test stay home on test day. Since 2004,

123 public schools in California have been identified as having engaged in cheating to improve standardized test scores, including encouraging students to stay home on test days (Gardner, 2007).

Poverty

Socioeconomic status is the most powerful indicator of academic success. Students from wealthier families enjoy more exposure to printed materials, watch more educational television programming, and benefit from a variety of experiences, such as museum visits, that translate into academic success (Jones et al., 2003). Much of the difficulty in equalizing test scores between groups such as African American and White, economically disadvantaged and wealthy; LEP and English-speaking are the factors that are beyond the school's control, such as the differences in financial resources (Ferguson, 1991).

Johnson and Johnson (2006) indicate poverty was a significant variable in testing and the impact of poverty increased steadily from 4th to 12th graders. Therefore, the longer a child stays in school, the greater the effect poverty has on test scores. Popham (1998) theorized there is a direct correlation between student success and parent income. If parent income is high, students usually score well. If parental income is low, a lower score is anticipated. The amount of poverty in communities where schools are located accounts for the great majority of difference in test scores from one school, or district, to the next (Kohn, 2000).

Districts failing to meet AYP enrolled a higher percentage of minority, ELL, and low-income students (Tracey et al., 2005). The subgroup rules mean that districts serving a diverse population of students will have a more difficult time making AYP, simply because they have more performance targets to make than more homogeneous districts. Large districts and districts serving minority and low-socioeconomic populations are more likely to be identified as not

meeting AYP. When looking at proficiency measures, it appears that accountability systems are biased against high-poverty schools. Despite the initial difference in scores, high-poverty schools generated annual learning gains that were similar to those of low-poverty schools. High-poverty schools typically started at a lower level and were forced to make gains at a faster rate (Kim & Sunderman, 2005).

Oftentimes, students who are from poor and minority communities receive less rigorous curriculum and are taught by teachers who are less prepared academically and less experienced. Those same students have access to fewer books and computers and often attend schools where they feel unsafe or where it is not cool to take schoolwork seriously (Shepperd, 2000).

Minorities

Equity related to testing is not a new issue. In the 1979 case of *Larry P. vs. Riles*, African American parents argued that the administration of culturally biased standardized intelligence tests resulted in the disproportionate identification of African American children as mentally retarded and their inappropriate placement in special education classes for the educable mentally retarded. The federal district court ruled against the use of intelligence tests for African American children for placement in these classes. In 1986, the injunction was extended to include the use of such tests for all special education purposes. In 1992, Judge Robert Peckham rescinded his 1986 ban which prevented the use of standardized IQ tests for all special education settings (Wyatt, 2003).

Language ability, learning style, student experiences, and motivation all affect student achievement. Cultural differences influence the scope of experiences a student might have. World knowledge, prior knowledge, and task attitudes have a strong effect on a student's ability to perform a task and interpret outcomes. In laboratory studies in the late 1960s, researchers

suggested that African American children possessed no language and were completely non-verbal. Further investigation found that these experiments were conducted by White investigators with whom the children had had no prior interactions. When asked to respond to verbal or written tasks, repeat phrases, and answer structured questions, the children often answered with monosyllables (Weener, 1969). When the experiment was then changed to a group setting conducted by African American researchers, the children's demonstrated verbal ability improved dramatically (Baratz, 1969).

With the introduction of the TAKS and the increased emphasis on accountability, it became essential that the Texas curriculum become standardized to ensure that all students receive the same instruction. McNeil (2000) argues that this creates inequities by widening the gaps between educational opportunities that are provided to poor and minority children and those that are offered to middle-class children. She continues the discussion of standardization, stating that the most negative effects fall upon the poorest children and minority children whose introduction to school experience is overshadowed by any and all attempts to raise test scores. Kohn (2000) maintains that high-stakes testing holds people accountable for factors over which they have little control, such as social and economic factors.

Overall, according to a study conducted by Lee in 2006, state assessments compared to NAEP underestimated the racial and socioeconomic achievement gap. Data show an achievement gap between White and African American Grade 4 math scores: the chance of a White student meeting the proficiency standard was 1.8 times that of an African American student. Additionally, the gap on the NAEP has shown that White students met the standard 4.3 times more often than their African American peers. This same disparity is shown in the White-Hispanic comparisons as well as the nonpoor-poor comparisons (Lee, 2006).

According to the Civil Rights Project at Harvard University, low-income and minority students are over-represented in low-track schools and programs. Because African American and Hispanic students have lower average test scores than White and Asian students, schools with either African American or Hispanic subgroups have a higher probability of failing to meet the AYP requirement. Though many of these schools were found in need of improvement, they were actually making progress comparable to or greater than those making AYP. For example, in New York, for Grade 4, the percentage of students in improvement schools who scored at or above the proficiency level in math increased by 4.80 percentage points, compared to an increase of 0.90 for students in AYP schools (FairTest, 2007).

Tables 8 and 9 show data from the National Center for Research on Evaluation, Standards and Student Testing (CRESST) gathered over a 10-year period. While the gains in NAEP scaled scores for African American and Hispanic subgroups are significant, it is important to note that often these students come to school behind their peers. Globally, when comparing achievement gains, there is no difference in the Grade 4 reading achievement scores before or after NCLB. Grade 8 showed a marked decline when comparing pre- and post-NCLB. In the area of mathematics, there was an increase both before and after NCLB with the post-NCLB growth holding steady (Lee, 2006).

According to a 2005 NAEP report, the percentage of African American and Hispanic students above the proficient level in mathematics is much lower than that of their White peers (47% for Whites vs. 13% for African American and 19% for Hispanics at Grade 4; 39% for Whites vs. 9 % for African Americans and 13% for Hispanics at Grade 8), but it also shows that a large majority of African American students fail to meet the proficiency standards. The achievement gap continues between the races after the implementation of NCLB. The gaps

between White and African American students, White and Hispanic, and poor and non-poor students remain unchanged in reading and mathematics in both Grades 4 and 8. There was a small decrease in the gap between White and Hispanic students in eighth grade mathematics (Table 8).

Table 8
Subgroup Gains in NAEP Mathematics Average Scale Scores (1996-2005)

Group	Grade 4	Grade 8
White	14	8
African American	22	15
Hispanic	19	11

Table 9 illustrates *basic* and *proficient* gains for White, African American and Hispanic students. Students in minority subgroups historically underperformed when compared to their same-age peers, showing larger gains in the basic level while their White counterparts achieved higher percentages in the proficient level.

Table 9
Subgroup Gains on NAEP Depending on Basic or Proficient Level

Group	Grade 4	Grade 4	Grade 8	Grade 8
	Basic	Proficient	Basic	Proficient
White	14	20	7	9
African American	33	10	17	5
Hispanic	28	12	13	5

Because African American and Hispanic students often belong to multiple subgroups as defined by NCLB, including economically disadvantaged and LEP, students in a minority subgroup

become even more integral to the success or failure of a district. These students' scores are often counted multiple times compared with the scores of students in the general population, thus skewing the results (Abedi & Dietel, 2004; Baker et al., 2002). The implications for underachieving minorities are numerous. Students, as a result of low test scores, receive more drill and practice in an effort to raise test scores, with the long-term consequence of higher drop-out rates (Jones et al., 2003).

According to an August 2006 report, Florida showed significant gains in closing the gap between White and minority students in the area of fourth-grade reading. Haney (2006) examined enrollment data and found that 10-12% of minority students were being forced to repeat third grade. The gains reflected on the fourth grade NAEP scores were a reflection of student retentions. Minority students were forced to repeat at a rate of two to three times those of their White counterparts.

According to Lee (2006), if these current trends in racial and socioeconomic achievement gaps continue, significant disparities in achievement rate between advantaged White and disadvantaged minority groups will result. Projections indicate that by 2014, 32-44% of Whites will reach the reading proficiency target with 40-78% reaching the math target. Conversely, 7-18% of African American students will achieve proficiency in reading with 25-55% meeting the math proficiency. The Hispanic population is projected to have a 14-21% reading proficiency level with 32-70% mastering math. Among non-poor students, 32-38% are projected to be proficient in reading while 47-81% will be proficient in math. Of the students who meet criteria for free and reduced lunch, 11-16% are projected to master reading while 20-76% are projected to be proficient in math. (Lee, 2006)

English Language Learners

The challenges that face the changing population are many, especially for students for whom English is a second language. State tests show that ELL students' school performance is far below that of other students, often by 20-30%. Additional challenges involved with state

testing of ELLs include: (a) The language demands of tests do not provide for an accurate measure of student ability; (b) mobility of the ELL-designated group provides for difficulty with continuity of services; and (c) factors outside the districts' control, such as parent education level or socioeconomic status, frequently exert negative influence on student performance. In a study of more than 30,000 students, both ELL and non-ELL, a gap of 15 percentile points on a standardized reading test was found between students whose parents had postgraduate education and those who had not graduated from high school (Abedi & Dietel, 2004).

Accurate ELL identification is crucial to providing services, and inconsistencies exist from district to district and state to state. While NCLB does not set a minimum subgroup size, there is an expectation that the subgroup be large enough to provide reliable data. Setting a higher minimum number for ELLs will produce less performance variation and provide more accurate data when examining long-term trends (Hill & DePascale, 2003). The rapidly growing ELL population is diverse, with many languages and cultures represented. Performance varies among these groups, with one study indicating substantial differences in performance between ELL students with a Chinese-speaking background and ELL students with a Spanish-speaking background. The students with a Chinese-speaking background had significantly higher performance on science and reading tests (Abedi & Dietel, 2004).

Though standardized tests have benefits and uses, there are limitations as well:

Standardized tests can't measure intuitive thinking, creativity, imagination, conceptual thinking, curiosity, effort, irony, judgment, commitment, nuance, goodwill, ethical reflection, or a host of other valuable dispositions and attributes. What they can measure and count are isolated skills, specific facts, and functions, the least interesting and least significant aspects of learning. (Ayers, 2001, p. 112)

Test questions are often biased against students whose first language is not English. The use of idiomatic terms with which they are not familiar puts them at a disadvantage.

Kane and Staiger (2003) show a decreasing probability that schools will achieve specific goals as the number of subgroups increases. Districts with large ELL subgroups tend to have higher numbers of subgroups overall, increasing the difficulty for those districts to meet AYP.

Significance of Diversity and Testing

The National Center for Fair and Open Testing (2005) states, “Flatline NAEP scores show the failure of test-driven school reform. No Child Left Behind has not improved academic performance. NAEP reading scores were essentially unchanged from 2002 to 2005 at grade 4 and declined markedly at grade 8” (p.1).

Higher AYP failure rates are present in high-poverty schools as well as in schools with large minority enrollments which are often responsible for meeting multiple subgroup targets. An examination of the racial composition of select schools in need of improvement (SINI) and schools meeting AYP found a disproportionate number of minority students at the SINI schools. Analyses of school-level ethnicity and free and reduced lunch data from 2003-2004 school year indicate that schools with large African American and Hispanic populations also have high poverty rates. This finding is significant for schools struggling to meet AYP because it indicates that in addition to the African American and Hispanic subpopulations, the economically disadvantaged subpopulation is counted in accountability ratings. Predominantly White schools, which usually have low poverty rates, are less likely to have this accountability group for low-income students (Kim & Sunderman, 2005).

A state that bases assessment of districts on a single state test, and then uses those scores for such high-stakes decisions such as grade promotion and high school retention, rules out the possibility of discussing student learning in the most basic of terms. Cognitive and intellectual developments, social awareness, social conscience, and social and emotional development have

often been ignored in regard to standardized testing (McNeil, 2000). Emphasis on individual scores of students ignores the social and collaborative aspect of learning. Oftentimes, when test scores are reported, children are depersonalized into aggregates where the numbers do not necessarily speak for the students who are involved.

NCLB provides great challenges to lower-performing schools, located most often in urban areas. These districts must show greater improvement in scores and tend to have a larger number of subgroups, thereby increasing the probability that at least one subgroup will not make adequate yearly progress. Lower-achieving schools are required to make immediate progress, while higher-performing schools do not face immediate sanctions (Linn, 2003b).

FairTest, a group devoted to ensuring that students are assessed accurately, warns that tests are not a straightforward indicator of achievement. The emphasis on basic skills in testing programs showed that children were left deficient in critical thinking and communication skills (Cooper, 1990).

In a 2006 speech, U. S. Secretary of Education Margaret Spellings credited NCLB for such gains in overall math and reading scores for fourth graders and gains in mathematics for eighth graders, claiming:

These results, like the long-term July data, confirm that we are on the right track with No Child Left Behind, particularly with younger students who have benefited from the core principles of annual assessment and disaggregation of data. (2006)

Obviously, state assessments can only test a small amount of information where students might have a much broader view of the curriculum. Rather than testing for higher order or complex thinking skills, the test may dictate the need to examine for a narrower range. This is more detrimental to low-performing students because there are more incentives for teachers to prioritize, increasing the number of students meeting proficiency (Owens & Sunderman, 2006).

Summary

There is no argument in education that ongoing evaluation and assessment are essential to ensure that students receive effective instruction during their years in public school. District size, diversity, and population are all important factors in student success. Tracey (2005) analyzed the accountability systems of six states: Arizona, California, Georgia, Illinois, New York, and Virginia. The districts identified for improvement enrolled a disproportionate share of most states' students. In four of the six states (Arizona, California, Illinois, and New York), the percentage of the students attending school in districts needing improvement exceeded the percentage of districts statewide that were identified for improvement. Of 752 districts in New York, 58 (7.7%) were identified as in need of improvement and these districts enrolled 27.3% of the state student population. In Illinois, of 868 total districts, 28.3% were identified as in need of improvement and the targeted districts enroll 61.7% of the student population. Districts that were identified for improvement were, on average, large districts. According to Tracey, in California, Illinois, and New York, districts which need improvement are four to five times larger than districts making AYP. Such disparities call for the examination of the accountability system. This study examines whether the specified districts are truly not meeting AYP or whether the scores are misinterpreted due to the influence of population size on scores to be reported.

The state legislature, the commissioner of education, and the Texas Education Agency have been given the difficult task of creating an accountability system for the state of Texas that takes into account differences in students and districts. Because the general public is so often eager to judge schools and districts on one label bestowed by the Texas Education Agency, the accuracy and equity of this classification must be above reproach. In order to further study these

trends and the current issues in education, the Texas accountability system needs further examination.

The effectiveness and equity of the system in regard to small versus large districts must be evaluated. The information presented in Chapter 2 has addressed the importance for all students of assessment, accountability, and diversity issues. This study provides valuable information, analysis, and research for future discussion regarding policy and practice about accountability on both state and national levels, especially when faced with factors such as diversity and varying group size.

Chapter 3 explains the research design and methodology of this study. In order to effectively compare and contrast large and small districts, demographics and leadership history of selected districts were analyzed. Detailed information on the process and procedures is provided.

CHAPTER 3

METHODOLOGY

Overview

This study examines the impact of district size and diversity on the accountability ratings of selected Texas school districts. The research design for this study uses a stratified random selection process to select school districts to participate. Descriptive research is used to describe the findings in various reports provided by the Texas Education Agency (TEA) (Glatthorn, 2005). District accountability data and profiles from 2004 and 2005 are examined, and test scores in the 25 accountability subpopulation groups are analyzed.

The study begins by examining all 1,046 Texas districts and separating them into groups based upon size during the 2003-2004 school year, utilizing data provided by TEA. Districts are separated into two predetermined categories: large districts, between 25,000 and 50,000 students; and small districts with fewer than 8,000 students. This work was accomplished by downloading all the Academic Excellence Indicator System (AEIS) information from TEA and examining each district individually to determine the student population. Those districts that fell within the 25,000 and 50,000 population range were placed in one spreadsheet and those districts that met the fewer than 8,000 student criteria were placed in another. The remaining school districts that do not meet these criteria according to population are excluded from this study. Once the districts were screened randomly according to size, the districts were matched according to demographic likeness. This information was placed on a spreadsheet to ensure close data matches and those districts which had discrepancies were not paired.

After this process was complete, Texas Education Agency data accessed via the electronic Texas Education Directory (askTed) were utilized to determine how long the

superintendent had been in position. District directories were accessed from 1999 to the present. In order for a district to be eligible for comparison of data for this research project, the superintendent must have been in place for at least three years prior to collection of data in the years 2003-2004. Originally, there were 36 large/small district matches prior to the superintendent longevity component. The district comparison group revealed a field of 20. In that group, there are five superintendents that continue to lead their districts. If that information was not available, the district was contacted and that information was recorded as it was a requirement of the study for the superintendent to be in place for three years prior. At this point in time, no information regarding their accountability ratings had been accessed to ensure that the sample population was entirely random.

Research Methodology

This descriptive methodology examines the frequency of change in accountability ratings when adding underreported groups (Glatthorn, 2005). According to Bartz (1979), descriptive statistics are used to report the basic features of data in a study. This provides simple summaries about the sample and the measures. Descriptive statistics are used to present quantitative descriptions in a manageable form. In this study, descriptive statistics are utilized to examine data related to 20 school districts gathered from multiple reports from the Texas Education Agency.

Procedures

To qualify for inclusion in the research, eligible districts were examined on the basis of several criteria. One criterion was that each district serves students kindergarten through 12th grade. If the district did not serve students kindergarten through Grade 12, the district was

considered ineligible regardless of district size. Of the small districts that achieved Recognized or Exemplary status, the following district configurations were noted (See Table 10).

Table 10
Recognized and Exemplary Districts with Non-traditional Structure

Configuration	Number of Districts
EE-8	3
PK-6	3
PK-7	1
PK-8	17
K-5	1
K-6	1
K-7	1
K-8	9
7-12	1

Once the districts were selected, demographic information was compared between large and small districts in the areas of African American, Hispanic, White, Native American, Asian, economically disadvantaged, and limited English proficient (LEP). Initially, there were 26 large districts identified and 309 small districts. However, if no sufficient matches were found, the districts were not utilized. The demographic analysis was completed by case-by-case document examination and comparison of data. Care was taken to ensure that student numbers were within one standard deviation of error in order to provide for data integrity.

After this process was complete, Texas Education Agency data accessed via the electronic Texas Education Directory (askTed) were utilized to determine how long the

superintendent had been in position. District directories were accessed from 1999 to the present. In order for a district to be eligible for comparison of data for this research project, the superintendent must have been in place for at least three years prior to collection of data in the years 2003-2004. Originally, there were 36 large/small district matches prior to the superintendent longevity component. The district comparison group revealed a field of 20. In that group, there are five superintendents that continue to lead their districts.

Data Collection

Accountability data for the years 2003-2004 and 2004-2005 were compiled and compared for each small- and large-district group. Data group comparisons were analyzed to determine which student groups counted toward accountability ratings. The field of 25 available accountability group areas, in all districts, focuses on the academic areas of reading/ELA, writing, social studies, mathematics, and science. Each accountability group is rated for all students, as well as four subgroups: African American, Hispanic, White and economically disadvantaged. In order for a subgroup to count for adequate yearly progress (AYP), a minimum percentage of students must qualify and take the exam offered.

A brief background about the district supplements the subgroup information that was integral to the study. Accountability ratings for the 2003-04 and 2004-05 school years and demographic data are highlighted as well.

The central premise of this dissertation focuses on the accountability system as it compares large districts with small districts. It is evident that districts and campuses with small numbers of students pose a special challenge to the accountability system. According to the Texas Education Agency, there are two types of small-numbers situations. One is small numbers

of students within a group, and these are handled by minimum size criteria. Though this information was referenced in Chapter 2, it is important to the structure of the study:

- Any student group with fewer than 30 students tested is not evaluated.
- If there are 30-49 students within the student group and the group comprises at least 10% of all students, it is evaluated.
- If there are at least 50 students within the student group, it is evaluated.
- Student group size is calculated subject by subject. Therefore, group size will vary. (Texas Education Agency, 2005a).

The second situation is small numbers of total students, that is, few students tested in the all-students category. The significance of these collection-size criteria is that the scores of entire subgroups within a district have the potential to affect accountability ratings, either positively or negatively. If the subgroup fails to meet the minimum size criteria and does not meet the minimum percentage standards for certain accountability rating thresholds, the district is not required to account for those students within its ratings.

Data Analysis

District accountability data tables were analyzed for each district. For the small districts, each analysis group not used to determine ratings because of size was examined to establish whether the subgroup had met standard. Once this analysis was complete, a determination was made regarding whether the accountability rating was affected (Texas Education Agency, 2005). In the final analysis, it was found that 30% of district ratings were affected for the smaller school districts over the course of the two-year study period.

Testing data were downloaded from the Texas Education Agency accountability website for each of the 20 participating districts. Each district's content area test, grade level, and sub-

group data were recorded, compiled, and compared. The percentage of students who met the minimum standard on the TAKS was examined by analyzing the 25 potential indicators for the five subject area tests: reading/ELA, writing, social studies, mathematics, and science. Furthermore, data were disaggregated by the subgroups: all students, African American, Hispanic, White, and economically disadvantaged. When analyzing the district data, TEA indicates with an “X” whether the group size was large enough for the data to count in the ratings. For those districts where any groups include subpopulations that did not count in the ratings, the data were examined to determine whether the inclusion of these students would have affected the rating of Recognized or Exemplary. As noted previously, ratings were negatively affected 30% of the time.

Table 11 shows the district comparison matches. There are two large districts, Pharr San Juan Alamo and Spring Branch, that had close demographic matches and which were compared to two smaller districts in the analysis.

Table 11
Texas District Comparability Study

Large District	Small District
Corpus Christi	Ricardo
Conroe	Farmersville
Pharr San Juan-Alamo	Webb Consolidated
	Los Fresnos Consolidated
Spring Branch	Lometa
	Palacios
Spring	Fort Sam Houston
Pasadena	Brackett
Katy	Robinson
Irving	Springlake-Earth

Summary

Data from the 2003-2004 school year were used to determine large and small district matches in terms of demographics. Referring to Table 11, the following district groupings were analyzed and compared over a two-year period: Corpus Christi and Ricardo; Conroe and Farmersville; Pharr San Juan-Alamo, Webb Consolidated, and Los Fresnos Consolidated; Spring Branch,, Lometa, and Palacios; Spring and Fort Sam Houston; Pasadena and Brackett; Katy and Robinson; Humble and Kennedale; and Irving and Springlake-Earth. The districts ranged in size from 322 to 46,002 students. Initially, analysis of district ratings included a broad overview as well as an analysis of subpopulations counted in the accountability ratings. For the purposes of this study, in the districts where any subpopulation was not originally counted in the ratings, the group was subsequently added. The study then reveals whether the subgroup being counted within the accountability rating system would affect the rating, and to what degree.

CHAPTER 4

DATA PRESENTATION AND ANALYSIS

As the literature review for this study has demonstrated, accountability is an integral part of the education system. Each time Texas Assessment of Knowledge and Skills (TAKS) scores are reported, districts have the opportunity to examine programs and curriculum and improve the organization. The difficulty with any system is determining whether the measures in place are truly accurate and are fair assessments of student and district success. In order to examine the effectiveness and equity of the Texas accountability system, districts that are similar in demographics but different in size were compared. The districts were matched through examination of state data utilizing the askTed system and the Texas Education Agency resources as well as census data, as appropriate. Careful consideration has been made to ensure that in the district comparisons district minority and socioeconomic percentages align in order to provide an accurate examination.

This chapter provides, first, a short background of each district, including the region as well as the population. Utilizing information from 2005, comparisons to state averages are reported in regards to teacher salary and experience, the district-wide average ratio of teacher to student, tax rate, and expenditures per pupil. Comparisons are also made to the state averages on the TAKS and all 25 subgroups are assessed. As noted previously, any student group with fewer than 30 students tested is not evaluated. If there are 30 to 49 students within the student group and the student group comprises at least 10% of all students, it is evaluated. If there are at least 50 students within the student group, it is evaluated. Then, data from the Academic Excellence Indicator System (AEIS) are analyzed to determine whether size and diversity have an impact on accountability ratings. Last, analysis of the study is presented.

For the ease of reference, Table 12 illustrates the 2005 state averages upon which comparisons will be drawn. This information is reported for each district and includes total teacher years experience, student teacher ratio, average salary and average funding.

Table 12
State Averages of Teacher Experience, Student-Teacher Ratio, Salary, and Funding

Years Experience	11.5
Student-Teacher Ratio	11.9
Salary	\$41,011
Funding	\$8,916

Table 13 shows the state percentages of students who met the TAKS standard for each test and subgroup for the year 2004 and 2005 on the TAKS test. This is provided for the reader as there will be data reported in reference to the state averages for each district indicating variances in performance. This percentage indicates the overall passing rate, according to the state, for that particular group.

Table 13
Percentage of Students who Met the TAKS Standard in 2003-2004 and 2004-2005

	State	African American	Hispanic	White	Economically Disadvantaged
Reading 04	80%	71%	72%	89%	70%
Mathematics 04	66%	49%	57%	78%	55%
Writing 04	89%	84%	85%	93%	84%
Science 04	56%	38%	41%	73%	39%
Social Studies 04	84%	77%	76%	92%	74%
Reading 05	83%	76%	77%	91%	76%

(table continues)

Table 13 (continued).

	State	African American	Hispanic	White	Economically Disadvantaged
Mathematics 05	71%	55%	63%	83%	61%
Writing 05	90%	86%	87%	94%	85%
Science 05	63%	45%	50%	79%	48%
Social Studies 05	87%	81%	80%	94%	79%

Corpus Christi and Ricardo

Corpus Christi ISD is located in the Gulf region of Texas. It was established in 1909 and encompasses 68 square miles. Of the more than 39,000 students who attended school in the district in 2005, a majority (57.5%) qualified for free and reduced lunch under federal guidelines. Corpus Christi had a minority majority, with Hispanic students comprising 72.5% of the student population. Teachers, on average, had 1.9 years more experience than the state average, and the teacher-student ratio was 0.9 below state average. Both salaries and tax rate were above state average as well. Expenditures per pupil fell below the state average by \$1,026.

Because of its size, all indicators in the accountability system were taken into consideration for Corpus Christi ISD. In looking at data comparisons for the school years 2004 and 2005, Corpus Christi ISD performed at or above state average on most of the tests given in 2004; however, in 2005 the averages dropped significantly. Of particular concern were the social studies, mathematics, and science tests, as all groups performed below state average, some by as much as 7%. The district's rating for both 2004 and 2005 was Academically Acceptable.

Ricardo ISD is located in Kingsville, Texas. In 2005, this small district served 527 students, 61% of whom qualified as economically disadvantaged. At the time, the majority of

the students enrolled (74.4%) were Hispanic students. Teacher experience was 2.2 years greater than the average state while class size was 0.6 below. Salaries for teachers were \$10 above the state average while tax rates were below. The state average for expenditures per pupil was higher than at Ricardo by \$280.

Ricardo ISD had a rating of Recognized in 2004 and Academically Acceptable in 2005. In examining the data, students in Ricardo achieved higher than the state average on all areas the TAKS test in the year 2004. In 2005, there were several areas of deficiency, including White students in social studies, mathematics, and science and Hispanic students in science. Because of the numbers of students in the various subgroups, only 16 of the 25 performance results groups counted toward the district's rating in 2004. However, in 2004, even if those groups had been counted, the rating would not have been affected. In 2005, 18 out of 25 performance results groups counted toward the district's rating. Once again, even with the inclusion of these scores, the district's rating of Academically Acceptable would not have been affected.

Conroe and Farmersville

Conroe is located close to Houston, Texas and encompasses 348 square miles. In 2005, the district served more than 39,000 students with projections that, by 2012, enrollment will exceed 52,000. More than 70% of the students were White while the remaining student body was made up of Hispanics (21%) and African Americans (5.4%). Of those students enrolled, 28.5% were economically disadvantaged. Student-teacher ratio was 1.2 above the state average while teacher experience was shown as average. Teacher salaries were \$265 lower than the state average in 2005 while property taxes were higher. Per pupil expenditures were \$307 greater than the state average of \$8,916.

In 2004, students performed above state average on each portion of the TAKS test. Scores dropped in 2005 when most of the test scores were equal to state average, with science being the strongest area. Because of the size of the district, all accountability groups were assessed and had an impact on ratings. For 2004 and 2005, Conroe ISD achieved a rating of Academically Acceptable due to a lower passing rate for African American students in mathematics. African American students represented 5% of the student group tested.

Farmersville ISD, a district that was established in 1850, encompasses 89 square miles of land located in North Texas. Of the almost 1,500 students in 2005, 73% were White, 19.4% were Hispanic, and 6.6% were African American. One-third of the students qualified for free and reduced lunch. Teacher salary was below state average by \$3,375 with teacher experience being 0.2 years greater than state average. Teacher-student ratios were slightly below average. Tax rate and per-pupil expenditures fell below state averages as well.

On the 2004 TAKS, all student subgroups achieved at or above state average except in reading and mathematics for African American students. The district achieved a rating of Recognized based upon these scores. However, when examining the data, 69% of African American students met the TAKS standard in mathematics. Because of the group size, this was one of the 6 out of 25 accountability areas that did not count in Farmersville's ratings. If these scores had been counted, the district would have achieved a rating of Academically Acceptable. In 2005, the same trend occurred with several additions. All students, especially Hispanics, performed well below state average in social studies. African American students met the TAKS mathematics standards at a rate of 60%, but the group size was large enough to qualify for inclusion in the ratings. The district achieved a rating of Academically Acceptable with 22 out of 25 indicators reporting.

Pharr San Juan-Alamo, Webb, and Los Fresnos

Pharr San Juan-Alamo is located in the Rio Grande Valley of South Texas. Its demographics are not very diverse, with more than 26,400 students who are over 98% Hispanic and the remainder African American and White. Roughly 90% of students who attended school in Pharr San Juan-Alamo in 2005 met the federal criteria for free and reduced lunch. Teacher experience and salary fell below the state average for the year 2005 while class sizes were larger by an average of 1.3 pupils. Individuals in the community experienced a slightly higher than average tax rate at \$1.614. Per student expenditures were below state average by \$680.

In 2004, White students who attended this district performed below state average in reading, writing, and mathematics. African American students performed well above state average in the area of science. Scores in mathematics and science dropped below state average in all subject areas (reading, mathematics, writing, science, and social studies) in 2005. Though the district is a large district, only 19 of the 25 accountability groups counted toward the rating of Academically Acceptable in the year 2004 due to the lack of diversity. After examining the data, inclusion of the additional groups would not affect the rating. In the year 2005, the same number of groups was counted with the same rating result. If these groups were included, the rating would not be changed.

Los Fresnos Consolidated Independent School District is located in Cameron County near the southernmost tip of the state. The district encompasses over 460 square miles and serves Arroyo City, the town of Rancho Viejo, Indian Lake, the City of Bayview, Olmito, Las Yescas Laureles, and portions of Brownsville (Los Fresnos CISD, 2007). With over 93% of its 7,000 students of Hispanic origin, the lack of diversity that Pharr San Juan-Alamo experienced is mirrored here. In 2005, nearly 86% of the students were eligible for free and reduced lunch.

Teachers, on an average, had 2.4 fewer years experience, and their salaries, on average, were more than \$2,200 less than those statewide while the student-teacher ratio was .8 greater. The tax rate was nearly average (\$0.0017 above state average) with per pupil expenditures above average by \$884.

Los Fresnos achieved a Recognized rating in 2004 with achievement in mathematics for African American students falling significantly below state average by 18%. For this accountability cycle, the district was rated on 20 out of 25 indicators. If mathematics for African American students had been included, the rating would have dropped to Academically Acceptable as the passing rate was 40%. The 2005 school year brought the same number of accountability groups; however, the district's rating dropped to Academically Acceptable. The district continued to show low passing rates in the African American subgroups: reading at 67% and mathematics at 44%. Writing, social studies, and science were not reported due to small populations and confidentiality rules.

Webb Consolidated Independent School District is located 40 miles east of Laredo in Bruni, Texas. It consists of three schools that served a total of 333 students in 2005. According to the Texas Education Agency, teacher salaries were more than \$5,000 over state average and the teachers had, on average, 4.9 years more experience than their colleagues elsewhere in the state. Student-teacher ratio in WCISD was lower by 7.2 students. Per pupil expenditures were well over average by \$9,260 while the tax rate was lower than average.

TAKS scores in 2004 show most students achieving well above state average on every section of the test with one exception: White students in the area of reading scored two percentage points below the state average. There were also several student groups for which data were not available due to small group size: African American and White students in all subjects

and economically disadvantaged students in writing. (As stated previously, any student group with fewer than 30 students tested is not evaluated. If there are 30 to 49 students within the student group and the student group comprises at least 10% of all students, it is evaluated. If there are at least 50 students within the student group, it is evaluated.) The district received a rating of Recognized with 14 of 25 groups reporting data. Utilizing the data that are available, the rating would not have been affected if additional information had been included. Due to small group size, there were still eight non-reporting groups for which data were not available.

For the year 2005, most areas were at or above state average, though science was an area of difficulty for more than one group. In that year, there were 15 of the 25 accountability groups that were counted for the rating of Academically Acceptable. If all the groups with sufficient group size had been counted within the ratings, this rating would not have changed.

Spring Branch, Palacios, and Lometa

Spring Branch ISD is located in the western part of Houston and encompasses approximately 44 square miles. In 2005, the district served 32,920 students with slightly more than half the student population being Hispanic, 35% White, 6% African American, and 6% Asian. Economically disadvantaged students made up 54% of the population. Teacher experience was slightly below the state average by 0.3 years while salaries were well above average by over \$3,000. Student-teacher ratios were slightly below average. Tax rates in Spring Branch were above the state-wide average while student expenditures were \$3 below average.

Overall, students in Spring Branch achieved at or above state average in 2004 in all areas of the TAKS test. Math and science were strong areas for all demographic groups, with averages between 14 and 21% higher than the state. In 2004, the district achieved a rating of Academically Acceptable. In 2005, scores dropped across the board, but most significantly in the African

American subgroup in mathematics, social studies, and science. A rating of Academically Acceptable was assigned to the district for the 2005 school year as well.

Palacios ISD is a small district located on Tres Palacios Bay, centrally located between Houston and Corpus Christi. Among the 1,681 students in attendance in 2005, 55% were Hispanic, 27% White, 4% African American, and 13% Asian. Students qualified for free and reduced lunch at a rate of 65% under federal guidelines. Teacher experience was, on average, almost five years greater than other districts while salaries were just slightly above average by \$377. Tax rates were \$0.182 below average and student expenditures lagged behind by \$567.

On the TAKS test in 2004, students in all subgroups performed at or above state average in all subjects with one exception: African American students in the area of social studies performed 1% below. The district received a rating of Recognized. Of the possible 25 reporting groups, 20 groups were counted within this rating. If all groups had been counted, the accountability rating would have been lowered to Academically Acceptable due to performance by African American students in mathematics and science, with percent mastery of 67 and 58 respectively. As noted previously, for a district to receive a Recognized rating, the minimum passing rate for any subject and subpopulation is 70%. In 2005, scores dropped in the areas of reading and social studies across the board. In addition, there were scattered drops in subpopulations for mathematics, social studies, and science. The district's rating for this year was Academically Acceptable with 20 of the 25 potential performance results areas reporting. If each of the non-included scores had been incorporated, the district rating would not have changed.

Lometa ISD is located in Lampasas County in central Texas. This district served 322 students in 2005. A little more than 50% of the students who attend school in Lometa were

White while 42% were Hispanic and 5% African American. Economically disadvantaged students numbered 202 or 69%. Teachers in Lometa had, on average, 2.5 more years of experience than the state average while the teacher-student ratio was lower by 3.6 students. Salaries were nearly \$6,000 less than the statewide average. Individuals in the community paid \$0.259 less per \$100 valuation in taxes than the state average and student expenditures were behind by \$601.

Test scores in 2004 showed Lometa students at or above average on most of the areas of TAKS with the exception of writing, where all students and White subgroups achieved below the state average by 1% and 10% respectively. Science was a strong area for students, especially those who were Hispanic, who achieved 44% over the average state passing rate, and economically disadvantaged students, who exceeded the average by 50%. Recognized was the rating that Lometa was assigned for 2004 with 11 of the 25 groups included in the evaluation. If the 14 missing subgroup scores were included, the rating would not have been affected as 82% was the lowest score noted.

In Lometa, scores went down in 2005 when seven of the scores were reported as below state average. Most significant was writing where all students, Hispanics, and economically disadvantaged students achieved 15-34% below state average. Eleven of the 25 accountability groups were examined in the 2004-2005 school year to assign the rating of Recognized. If the additional reporting groups had been included, the label would have dropped to Academically Acceptable. Four scores were reported below 70%, the threshold for the Recognized rating. In writing, Hispanics scored at 53% and economically disadvantaged students met the standard at 63%. The same two groups scored below standard in the area of science, with a 52% for Hispanics and 59% for economically disadvantaged students.

Spring and Fort Sam Houston

On the surface, Spring ISD and Fort Sam Houston ISD have little in common; however, they mirror each other demographically. Spring is a thriving community in northern Harris County near Houston that, in 2005, had 26,664 students who were diverse in many different ways. The district had 31% African Americans, 31% Hispanics, 31% Whites, and 5% Asians. More than 45% of those students qualified for the free and reduced lunch per federal guidelines. Teacher salaries were nearly \$1,400 above state average, while teacher experience fell below by 0.1 year. Student-teacher ratio was higher by 0.6. The levied tax rate for Spring ISD was higher as were expenditures per student.

TAKS achievement in 2004 was at or above state average except in the areas of mathematics and writing for White students. These students ranked one percentage point below average in each area. For the 2004 school year, the district was assigned a rating of Academically Acceptable due to the achievement of African American students in the areas of mathematics and science. The year 2005 brought comparable results in these curricular areas. The district received a rating of Academically Acceptable with all 25 accountability groups reporting. Particular weaknesses were in the area of mathematics for all students, African Americans, Hispanics, and economically disadvantaged. In addition, science for the same groups showed similar challenges.

Fort Sam Houston ISD is located in San Antonio and its boundaries are coterminous with Fort Sam Houston, Lackland Air Force Base, and Randolph Air Force Base (Fort Sam Houston ISD, 2007). The district served 1,249 students in 2005, 32% of whom were African American, 20% Hispanic, and 42% White. On the average, teacher experience was 5.5 years beyond that of colleagues in the state while the teacher-student ratio was almost four below average. Salaries were well above average by more than \$8,000. Because the district is established on federal

property and has no taxing or bonding authority, it is jointly funded by the Department of Education Impact Aid Funds and the State of Texas Public School Foundation.

The year 2004 brought above-average scores for all student groups on each section of the TAKS test with science being an area of strength. The district received a rating of Recognized with all 25 groups reporting. Fort Sam Houston is a unique example of a small district for which every accountability group has a large enough population to count toward the district rating. This is the only small district group in the study that displays this characteristic. In 2005, the district's economically disadvantaged writing scores dropped below state average; however, the district still maintained the Recognized rating. During this rating period, the state assessed 24 of 25 accountability groups, with Hispanic students in writing excluded due to small numbers of students tested. The inclusion of these students would not have impacted the rating of the district in 2005.

Pasadena and Brackett

In 2005, Pasadena ISD served 46,000 students in sections of South Houston, Pearland, and Pasadena. The majority of these students, 70%, were Hispanic, while nearly 20% were White, and 6.4% were African American. Of these students, 63% met the federal guidelines as economically disadvantaged. Teacher experience was a little more than one year below state average with salaries \$1,034 above average. Student-teacher ratio was slightly above average by 0.8. The tax rate in the community was above the state average; however, student expenditures lagged behind by \$985.

Students in the year 2004 achieved at or above state average on all sections of the TAKS test. During this period of time, the district received a rating of Recognized. Science was an area of strength for all subgroups, as the lowest percentage rate was for African American students at

66%. This group met the required improvement standard with an actual change of 14%, thereby achieving the Recognized status. Students in 2005 achieved at a lower rate. All students as well as African American students in mathematics were below the state average. In addition, the overall science score fell below the average mark. These scores impacted the ratings with an Academically Acceptable label for the 2005 school year.

Brackett ISD is a small district located in Brackettville in Kinney County in southwest Texas. This district of 629 students served 64% Hispanic, 33% White, and 2% African American students in 2005. Economically disadvantaged students accounted for 63% of the student population. During the 2005 school year, teachers in the Brackett school district had nearly three years greater experience on average yet were paid \$2,400 less. Student-teacher ratios were slightly below average, by 1.6 students while per pupil expenditures exceeded the state average by \$637.

The majority of student groups achieved at or above state average in the year 2004 on the TAKS test with several groups non-reporting due to group size. For this year, TEA reported a rating of Recognized; however, of the 25 standard groups, only 19 reported data. Further examination showed that the African American subgroup in mathematics achieved a 50% passing rate. If this score had been included in the calculations, the ratings would have been affected and downgraded to Academically Acceptable as the passing rate should be, at a minimum, 70% to be Recognized.

The year 2005 brought changes in levels of scores for students in many different areas. Students in the African American subgroup fell below the state average in reading by 9%, social studies by 22%, and mathematics by 15%. Five other groups were below state average by as much as eight percentage points during this assessment year. Examination of the accountability

data table, for which the district was assigned an Academically Acceptable rating, revealed that of the 25 accountability groups, 19 were included in the calculations. The lowest of the available published scores, due to group size, was that of African American students in mathematics, at 42%. If this score were included in the rating system, it would not affect the assigned rating of Academically Acceptable.

Katy and Robinson

Katy ISD is a suburban school district that is located in East Texas. It encompasses 181 square miles and its eastern boundary stretches to within 16 miles of downtown Houston (Katy ISD, 2007). Nearly 64% of the students attending Katy's schools are White, 22% Hispanic, 6% African American, and 7% Asian. Of the 41,690 in attendance in 2005, 17.3% qualified for free and reduced lunch. Teacher experience was 0.4 years below average in 2005 while the teacher-student ratio was above state average by 0.3. Salaries for teachers were above state average by more than \$2,700. Per student expenditures exceeded the average by nearly \$3,000, while the tax rate was \$0.441 above average.

Students in 2004 performed at or above state average on all sections of the TAKS test. Among different demographic groups, African Americans continually performed well above the average. Students in this subgroup achieved 24% above average in mathematics and 29% above in science. Because of the large student population, all 25 group sizes were integral to the Academically Acceptable rating from the Texas Education Agency. The difficulty the district had was in the areas of Hispanic and economically disadvantaged students in science. Nearly 1,500 Hispanic students took the science TAKS test, while 652 economically disadvantaged students completed it as well.

In 2005, the district again achieved a rating of Academically Acceptable. That year, students who were African American, Hispanic, and economically disadvantaged met the standard at 64%, 61%, and 56%, ensuring the district a rating of Academically Acceptable.

Robinson ISD is located near Waco and served 2,050 students in 2005. The majority of these fell within the category of White, while Hispanic students were at 13% and African American students at 4%. Nearly 22% of the students were eligible for free and reduced lunch. Teacher-student ratio was slightly above average by 0.6 and teacher experience was below average by 1.8 years. Salaries were reported above state average by almost \$1,400 while taxes exceeded the average by \$0.311 as did expenditures. Per student expenditure exceeded state average by more than \$800.

Scores on all tests in 2004 showed students achieving above state average for all TAKS tests. Students were well above their peers in each subgroup. The Texas Education Agency rated Robinson ISD a Recognized district, taking into account scores from 20 of 25 possible demographic groups. The African American group size for every test was not sufficient to be reported and incorporated. If these additional five scores were taken into account, the district's rating would have been affected due to a 59% TAKS Met Standard passing rate for African American students in science. This was the only area in which the required improvement was a negative, showing a drop of 8% from the previous year.

Comparisons in the year 2005 showed drastic drops for Hispanic, economically disadvantaged, and White subpopulations. In social studies, all three of these groups achieved below the state average. White students achieved below average in social studies, mathematics, and science. These scores affected the accountability rating from the state, lowering the district to Academically Acceptable. As in 2004, of the possible 25 accountability indicators, 20 were

included in the ratings due to small numbers of African American students. If these five additional groups had been included, the rating would not have been affected.

Irving and Springlake-Earth

Irving ISD is located in the Dallas-Fort Worth Metroplex covering 48.5 square miles. It serves most of the city of Irving except for Valley Ranch and a portion of Las Colinas (Irving ISD, 2007). The 31,200 students who attended Irving ISD in 2005 were 58% Hispanic, 24% White, 13% African American, and 5% Asian with a total of 67 languages spoken by students. Students qualified for free and reduced lunch at a rate of 61%. Student-teacher average in Irving ISD was slightly above average by 0.3 students while teacher experience lagged behind by two full years. Salaries were above average by \$1,181 and taxes for the community were \$0.278 above average. Expenditures per student fell slightly below the state average by \$71.

Scores in the year 2004 showed that Irving students performing below state average on 17 out of 25 indicators. Students in the all student and Hispanic categories performed below average on each test. The area of the greatest difficulty was science, where scores ranged from 2% below for African American students to 12% below for all students. Because of the large district size, on the 2004 district accountability data table, each of the 25 accountability groups was influential when it came to Irving's rating. African American, Hispanic, and economically disadvantaged subgroups scored below 70% in mathematics. In science, all students, African American, Hispanic, and economically disadvantaged subgroups scored 62%, 53%, 50%, and 50% respectively. The district received a rating of Academically Acceptable.

The year 2005 showed scores that were improving slightly in most areas compared to others in the state. Reading scores went up overall except for students in the White

subpopulation, which decreased by one percentage point. Scores for writing improved for all subgroups except for those students in the White subpopulation, who dropped three percentage points below state average. Academically Acceptable was the rating assigned for this school year due to performance in mathematics and science by African American, Hispanic, and economically disadvantaged subgroups.

Springlake-Earth ISD was created through consolidation in 1923 and is located in northwest Texas in Lamb County. This district served 399 students in 2005, the majority (58%) of whom were Hispanic, 39% White, and the remainder African American. Of these students, 62% were identified as economically disadvantaged. Teachers in Springlake-Earth averaged 0.1 year less experience than their counterparts, but their class sizes were smaller by 3.9 students. Salaries were lower by more than \$3,300, tax rates were slightly higher, and expenditures per student were \$63 over the state average.

In the year 2004, students in Springlake-Earth performed well above the state average. For that year, the district had 16 of 25 groups reporting. African American students in reading and mathematics scored significantly beyond the average; however, the subgroup was not large enough to count for the district's rating of Recognized. Even if these non-reporting groups had been included, there were some additional groups (such as African American in writing, social studies, and science) that would not have been reported. When including the 22 of 25 groups, the rating of Recognized that the district achieved for 2004 have not changed.

The Texas Education Agency considered 12 of 23 accountability groups in assigning accountability ratings for Springlake-Earth. District data comparisons show high scores for all subgroups and no scores that fell below the state averages. For this year, the district achieved an Academically Acceptable rating due to lower scores in the special education assessment division.

Based specifically on TAKS, if all excluded subgroups had been included, the district’s rating would not have changed. The lowest score was that of Hispanic students in science showing a passing rate of 69%. The rating of Academically Acceptable supports these scores.

Analysis

Eighteen different districts were examined in this analysis. In establishing the study, eight large districts were included as comparison groups to illustrate the accountability system as it is currently in place. In addition, 10 small districts were selected for comparison. Both the large and small districts were examined over two years.

In looking at the eight large districts over the two-year period, the rating assigned to them was Academically Acceptable 15 out of 16 times, with the single exception being Pasadena ISD in the year 2004. The districts were rated Recognized only 1 time out of 16. (See Table 14).

Table 14
Ratings of Large and Small Districts 2004-2005

District	2004 Rating	2005 Rating
Corpus Christi	Academically Acceptable	Academically Acceptable
Ricardo	Recognized	Academically Acceptable
Conroe	Academically Acceptable	Academically Acceptable
Farmersville	Recognized	Academically Acceptable
Pharr San Juan-Alamo	Academically Acceptable	Academically Acceptable
Webb Consolidated	Recognized	Academically Acceptable
Los Fresnos	Recognized	Academically Acceptable
Spring Branch	Academically Acceptable	Academically Acceptable

(table continues)

Table 14 (*continued*).

District	2004 Rating	2005 Rating
Lometa	Recognized	Academically Acceptable
Palacios	Recognized	Academically Acceptable
Spring	Academically Acceptable	Academically Acceptable
Fort Sam Houston	Recognized	Recognized
Pasadena	Recognized	Academically Acceptable
Brackett	Recognized	Academically Acceptable
Katy	Academically Acceptable	Academically Acceptable
Robinson	Recognized	Academically Acceptable
Irving	Academically Acceptable	Academically Acceptable
Springlake-Earth	Recognized	Academically Acceptable

The ten small comparison districts experienced different results, receiving Recognized status 12 out of 20 times.

If all student scores had been counted in the state's accountability system and the aforementioned unreported scores had been counted, the percentages of Recognized districts would drop significantly. Table 15 illustrates how small districts that would have been affected in the 2004 or 2005 school year if all students had been accounted for. When including these groups, the subgroups were underreported at a rate of 6 out of 20. However, even with the districts whose ratings did not change, there were still group sizes too small to report. The state requires that when a group size is so small that confidentiality might be breached, the scores are not reported. Therefore, it is impossible to determine whether additional districts would be affected making the ratio even higher.

Table 15

District Accountability Changes

District	Year	Prior Rating	New Rating
Farmersville	2004	Recognized	Academically Acceptable
Los Fresnos	2004	Recognized	Academically Acceptable
Palacios	2004	Recognized	Academically Acceptable
Lometa	2005	Recognized	Academically Acceptable
Brackett	2004	Recognized	Academically Acceptable
Robinson	2004	Recognized	Academically Acceptable

As a result of this sampling of the 1,046 Texas school districts over the 2004 and 2005 school years, there is evidence that the accountability system has some built-in inequities involving small and large districts. Large districts, by virtue of their size, are generally more diverse and are held to more accountability standards, while smaller districts are essentially held to less rigorous standards through the implementation of minimum-group-size exclusions. Chapter 5 summarizes the findings, includes discussion, provides recommendations for future research, and examines practical applications for additional study.

CHAPTER 5

SUMMARY, DISCUSSION AND RECOMMENDATIONS

Introduction

For clarity, this final chapter reiterates the research problem and reviews the major methods used in the study. The chapter also summarizes the results, provides discussion, includes areas for practical application, and suggests ideas for future research.

Statement of the Problem

Two main questions were examined in this study: What is the status of district size and diversity as related to the accountability rating of selected Texas school districts? If all size and diversity data were included in the accountability rating system, how would district ratings be impacted? These two main points were the focus of the state-wide district analysis.

Review of Methodology

This descriptive statistical study began by determining which of Texas' 1,046 public school districts met the criteria for the study as large or small districts. Districts were considered to be a large district if they served between 25,000-50,000 students. Districts that served fewer than 8,000 students were considered small districts. Once districts were identified as falling into large and small categories, their demographics were compared to determine their similarities. Congruency in the areas of minority and economically disadvantaged and limited English proficient (LEP) populations were considered. If the districts proved to be comparable, they were eligible for the study and were included in one of the matched pairs or triads that were examined to determine what possible relationship district size and diversity have with ratings on the Texas accountability system.

Summary of Results and Discussion

Linn (2005 b) observes that reporting more groups increases the number of ways schools can fail to meet adequate yearly progress (AYP). While only one group missing its goal will prevent a district from meeting AYP, those schools and districts (usually urban, with large numbers of Latino, African American, English language learners, or students with special needs) have multiple groups reporting. In a study of AYP by Tracey and Sunderman (2005), it was found that a higher percentage of large districts met the minimum subgroup requirements than medium-sized or small districts. In fact, large districts were four times more likely to report a black or Hispanic subgroup. Regarding English language learners, large districts were almost 10 times as likely to have reported the subgroup. For the Texas state accountability system, there are five possible subgroup targets for each of the five tests given, totaling 25 reporting scores. According to Tracy and Sunderman, (2005), there is a correlation between districts failing to meet AYP and those who enrolled a higher percentage of minority or English language learners (ELL) and low-income students. Districts that are larger and more diverse are assessed based upon all the subpopulations and are being held accountable for teaching all groups, but smaller districts are not held to the same standards due to different expectations. These gains are similar, however, but the larger districts face larger gaps to make up. Therefore, their deficits force them to show improvement at a faster rate (Kim & Sunderman, 2005).

The current study examined the effect of including the demographic groups that had been originally omitted based upon small group size. After analysis, these groups were included back into the district assessment and the ratings were reexamined. When the accountability groups that had not originally been counted in were included, the original ratings were reduced to a lower rating 6 out of 18 times. This potentially places large districts at a distinct disadvantage because,

due to large student group size, each of their 25 reporting groups is always used to determine ratings. There is the potential for smaller districts to take advantage of the system, calculating at what point each student group would be large enough to have their scores included, and serving only students who *count*. The accountability system in the state of Texas has become so high-stakes that each year there are stories of cheating and dishonesty. The worst-case scenario is that schools and districts could feel compelled to encourage students to stay home on testing days so that the district would not meet the criteria for group size.

For example, the study examines a small heterogeneous school district in north Texas with roughly 7% socioeconomically disadvantaged students and over 93% White students. This district earned the Academically Acceptable rating in 2004 due to low scores for economically disadvantaged students in mathematics. The following year (2005) the district's rating increased to Recognized with a drastic increase in this subpopulation's score. In 2006, this same subgroup did not meet the criteria for group size by two students in the area of English language arts, and the score failed to meet the standard as well. Economically disadvantaged students met TAKS expectations at 69%; therefore, if the score had been counted the district would have achieved Academically Acceptable another year. Because this score was not included, the district was labeled as Recognized.

The question arises as to what affected or influenced these children who were not tested? It is important to determine whether the cause of this shift in the number of students tested was mobility, special education, absence, choice, or another factor. There is also question regarding what factors might influence the numbers of children in attendance on the day of testing.

Recommendations for Practical Application

Nearly one-third of the examined districts' ratings would have been changed with the inclusion of additional accountability groups. There are certain minimum standards that a district must achieve in order for the district to obtain Recognized status. Are these districts actually meeting these standards? As shown in the data, there were instances where student groups achieved at a passing rate as low as 40%, and the district still received a rating of Recognized. Thus, it is clear that Texas needs an equitable system that accounts for all children, whether in large or small districts, in order to provide accurate reporting.

This study indicates a need for a system that allows students to correctly demonstrate their abilities and skills. This may not be accomplished in a one-day, criterion-referenced test. As noted, there are certain sections of the population, based upon location and numbers, that are not included in the reporting in the accountability rating reports. If there are not enough students in a subgroup to *count*, the district could potentially choose not to serve them with a high-quality, equitable education. This does not mirror the goals of No Child Left Behind in that NCLB seeks to narrow class and racial gaps in school performance by creating common expectations for all. NCLB also requires schools and districts to focus their attention on the academic advancement of underserved groups of children, such as low income and minority students. Based upon the current system in Texas, districts are able to opt out of serving these students if their numbers do not meet subgroup qualifications.

The system must be revisited. By examining the California, Kentucky, and New York systems, this study illustrates that there are ultimately 50 different systems in as many different states. As the trend for nationwide testing continues, the onus is on lawmakers to determine a uniform testing system for all states to serve all the nation's children. Until that is in place,

critical changes must be made in the testing and accountability system in order to provide a clearer picture of the Texas school system. Linn (2002b) recommends that accountability systems should employ different types of data from multiple sources, such as observation, portfolios, or performance tasks that allow for interpretations of student performance. In addition, they should include the performance of all students, including those subgroups that have historically been difficult to assess. Though this would be time consuming, the additional data would provide a more multidimensional view of each student, rather than relying on a one-day standardized test.

All Texas schools and districts should be assessed at least annually. That same expectation applies nationwide as we address assessment on the national level with a system structured so that all children are included. This ideal is embodied in the central tenet of the policy and procedures of No Child Left Behind. The current Texas accountability system allows districts to leave children behind if they fall within certain subpopulations and the numbers fail to match up. Therefore, the system must be revised utilizing a multidimensional, developmentally appropriate assessment that will measure student achievement and allow instruction to improve (Linn, 2002a). Governing bodies need to consider student scores counting for only one subgroup while still providing valuable information about subpopulation performance (Tracey & Sunderman, 2005). Districts would still be responsible for the performance of all students, but the disadvantage of students who account for multiple subpopulations would be lessened.

Recommendations for Future Research

Further research should focus on two areas. The first area is individual district accountability in considering each student's performance. A study focused on this issue would

accurately measure each subgroup's data and compile the information regardless of how small the data sets. This study should be replicated on a larger scale to determine whether this affects all smaller districts' accountability ratings in a similar way.

The second recommended area of study would be a comparison of different states' systems to determine equity within the systems. Systems such as New York's should be examined for variance factors based upon extenuating circumstances, such as the "September 11th effect" and student fatigue. Kentucky's system should be analyzed based upon districts' ability to be their own benchmarks by providing initial data points for district growth and determining how effective teaching has been implemented. Such documentation could lay the foundation for designing a nation-wide, cohesive accountability system at both state and national levels that would match stated objectives with AYP guidelines. This information would contribute to informing policy-making and would provide a more equitable education for all students.

Implications

This is the challenge educators must confront. The recommendation for accountability recording is that all students, regardless of their background or color of their skin, count in the determination of ratings for public schools. Since the origination of the Elementary and Secondary Education Act of 1965, upon which NCLB is based, the goal has been to reach the children who are most disadvantaged. The central tenet of No Child Left Behind is that all children learn to their maximum potential and reach grade level by the year 2014. The Texas accountability system, in its current structure, excludes certain students based upon race and economic status and is not in compliance with what the law intended. The Texas Legislature must examine the system to determine other options, rather than exclude entire sectors of the

population based upon the number of students reporting. The Texas system, as related to the national AYP system, also has discrepancies reflected at the district and state levels. Until that inequity is addressed, districts with higher percentages of ELL, minority, and low income students will continue their inconsistency in meeting AYP under the current structure (Tracy & Sunderman, 2005). The data provided in this study may serve to increase awareness and stimulate further discussion regarding how most effectively to provide school, district, and state accountability.

REFERENCE LIST

- Abedi, J., & Dietel, R. (2004). Challenges in the No Child Left Behind Act for English language learners. Policy Brief 7, Winter. CRESST: National Center for Research on Evaluation, Standards, and Student Testing. Retrieved on December 12, 2006, from <http://www.cse.ucla.edu/products/policy.html>
- American Evaluation Association (2002). Position statement on high stakes testing. Retrieved December 14, 2007, from <http://www.eval.org/hst3.htm>
- Ayers, W. (2001). *To teach: The journey of a leader*. New York: Teachers College Press.
- Baker, E., Linn, R., Herman, J. & Koretz, D. (2002). Standards for educational accountability systems. Policy Brief 5, Winter. CRESST: National Center for Research on Evaluation, Standards, and Student Testing, Retrieved December 12, 2006 from <http://www.cse.ucla.edu/products/policy.html>
- Baker, E., & O'Neil, H. (1995). Diversity, assessment, and equity in educational reform. In M. T. Nettles & A. L. Nettles (Eds.), *Equity and excellence in educational testing and assessment* (pp. 69-88). Boston, MA: Kluwer Academic Publishers.
- Baratz, J. (1969). A bidialectical task for determining language proficiency in economically disadvantaged Negro children. *Child Development*, 40, 889-901.
- Bartz, A. (1979). *Descriptive statistics for education and the behavioral sciences*. Minneapolis, MN: Burgess Publishing Company.
- Broadfoot, P. (1996). *Education, assessment, and society*. Philadelphia, PA: Open University Press.
- Bush, G. (2000). George W. Bush's plans for education in America. *Phi Delta Kappan*, 82, 122.
- Bush, G. (2007). President's FY 2008 education budget: Building on progress. Retrieved February 12, 2007 from <http://www.ed.gov/about/overview/budget/budget08/budget08-factsheet.doc>
- Byrk, A., & Raudenbush, S. (1988). Heterogeneity of variance in experimental studies: A challenge to conventional interpretations. *Psychological Bulletin*, 104, 396-404.
- California Department of Education. (2007a). Data and statistics. Retrieved January 13, 2007, from <http://www.cde.ca.gov/ds/>
- California Department of Education. (2007b). Overview of California's 2006-2007 accountability progress reporting system. Retrieved March 17, 2007, from <http://www.cde.ca.gov/ta/ac/ay/documents/overview07.pdf>

- Cizek, G. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice*, 20, 19-27.
- Cizek, G. (2005). High stakes testing: Contexts, characteristics, critiques, and consequences. In R. P. Phelps (Ed.), *Defending standardizing testing* (pp.23-54). Mahwah, NJ: Lawrence Erlbaum Associates.
- Commonwealth of Kentucky (2005). Kentucky facts. Retrieved January 13, 2007, from <http://www.thinkkentucky.com/kyedc/pdfs/kyfactsbk.pdf>
- Cooper, K. (1990, April 8). Educator's tests show school reforms take time. *The Post*, p. A4.
- Driesler, S. D. Whiplash about backlash: The truth about public support for testing. *NCME Newsletter*, 9, 2-5.
- Education Data Partnership (2007). Fiscal, demographic, and performance data on California's K-12 schools. Retrieved January 13, 2007, from <http://www.ed-data.k12.ca.us/welcome.asp>
- FairTest (2005). Flatline NAEP scores show failure of test driven school reform. *Fair Test Examiner*. Retrieved March 12, 2007, from <http://www.fairtest.org/flatline-naep-scores-show-failure-test-driven-scho>
- FairTest (2007, January). NCLB neither spurs improvement nor prevents failure. *FairTest Examiner*. Retrieved February 20, 2007, from <http://www.fairtest.org/examarts/2007%20January/NCLBNeitherSpursImprovement.html>
- Ferguson, R. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation*, 28, 465-498.
- Gardner, D. (1983). *A nation at risk*. Retrieved October 14, 2006, from <http://www.ed.gov/pubs/NatAtRisk/index.html>
- Gardner, W. (2007, July 15). Cheating on standardized tests isn't fleeting – it's predictable. *San Francisco Chronicle*, p. D4.
- Garner, A. (1974). *The performance based campus: How to build a total system*. Paper presented at the National Conference on Higher Education. (ERIC Document Reproduction Service No: ED 093 411).
- Glatthorn, A. (2005). *Writing the winning thesis or dissertation: A step-by-step guide*. Thousand Oaks, CA: Corwin Press.

- Goodman, D. & Hambelton, R. (2005). Some misconceptions about large-scale educational assessments. In R. P. Phelps (Ed.), *Defending standardizing testing* (pp. 91-110). Mahwah, NJ: Lawrence Erlbaum Associates.
- Haney, W. (2006, September). *Evidence on education under NCLB and how Florida boosted NAEP scores and reduced the race gap*. Paper presented at the 2006 Hechinger Institute: Broad seminar for K-12 reporters, New York, New York. Retrieved April 22, 2007, from <http://www.bc.edu/research/nbetpp/statements/nbr6.pdf>
- Herman, J. (2006, April). *Accountability and assessment in the service of learning: Is public interest being served?* Paper presented at AERA Annual Meeting, San Francisco, CA.
- Hill, R. & DePascale, C. (2003). *Reliability of No Child Left Behind accountability designs*. Dover, NH: National Center for the Improvement of Educational Assessment.
- Hoff, D. (1999, June 16). Made to measure. *Education Week on the Web*. Retrieved May 4, 2006, from <http://www.edweek.org/ew/articles>
- Johnson, D. & Johnson, B. (2006). *High stakes: Poverty, testing and failure in American schools*. Lanham, MD: Rowman & Littlefield Publishing.
- Johnson, L. (1966). *Public papers of the presidents of the United States, 1965* (Book 1). Washington, DC: US Government Printing Office, pp. 407-408.
- Jones, M. G., Jones, B. D., & Hargrove, T. (2003). *The unintended consequences of standardized testing*. Lanham, MD: Rowman & Littlefield Publishers, Inc.
- Kane, T., & Staiger, D. (2003). Unintended consequences of racial subgroups rules. In P. E. Peterson (Ed.), *No Child Left Behind? The politics and practice of school accountability*. Washington, DC: Brookings Institution Press.
- Katy Independent School District (2007). About Katy ISD. Retrieved September 25, 2007 http://katyisd.org/about_kisd.htm
- Kennedy, E. (2003). *Raising test scores for all students: An administrator's guide to standardized test performance*. Thousand Oaks, CA: Corwin Press.
- Kentucky Department of Education (2007). Accountability system. Retrieved January 13, 2007, from <http://education.ky.gov/KDE/Administrative+Resources/Testing+and+Reporting+/CATS/Accountability+System/>
- Kim, J. & Sunderman, G. (2005). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equity. *Educational Researcher*, 34, 3-13.

- Kohn, A. (2000). *The case against standardized testing: Raising the scores, ruining our schools*. Portsmouth, N.H.:Heinemann.
- Lee, J. (2006). *Tracking the achievement gaps and assessing the impact of NCLB on the gaps: An in-depth look into national and state reading and math outcome trends*. Cambridge, MA: Civil Rights Project at Harvard University.
- Linn, R. (2002). Standards for educational accountability systems. Policy Brief 5, Winter. CRESST: National Center for Research on Evaluation, Standards, and Student Testing. Retrieved May 16, 2006 from http://www.cse.ucla.edu/products/policy/cresst_policy5.pdf
- Linn, R. (2003a). Accountability: Responsibility and reasonable expectations. *Educational Researcher*, 32, 3-13.
- Linn, R. (2003b). Requirements for measuring adequate yearly progress. Policy Brief 6, Winter. CRESST: National Center for Research on Evaluation, Standards, and Student Testing. Retrieved December 12, 2006 from <http://www.cse.ucla.edu/products/policy.html>
- Linn, R. (2005a). Fixing the NCLB accountability system. *CRESST Policy Brief 8*. National Center for Research on Evaluation, Standards, and Student Testing.
- Linn, R. (2005b). Test based educational accountability in the era of No Child Left Behind. *CSE Report 651*, Retrieved December 10, 2006, <http://www.cse.ucla.edu/products/reports/r651.pdf>.
- Linn, R. (2007, January). *Educational accountability systems*. Paper Presented at CRESST Conference: The future of test based educational accountability, Los Angeles, CA.
- Linn, R., Baker, E. & Herman, J. (2002). Minimum group size for measuring adequate yearly progress. *CRESST Line*. Retrieved October 4, 2006 from http://www.cse.ucla.edu/products/newsletters/cresst_cl2002_4.pdf
- Los Fresnos Consolidated Independent School District (2007). Retrieved September 18, 2007 from <http://www.lfcisd.net/1558105118552490/site/default.asp>
- McNeil, L. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.
- Merriam-Webster (2007). *Merriam Webster on-line*. Retrieved April 22, 2007, from <http://www.m-w.com/>
- Murdock, S. (2007). *Presentation to the House Ways and Means Committee*. Retrieved March 7, 2007, from http://txsdc.utsa.edu/download/pdf/presentations/2007_03_07_House_Ways_and_Means_Committee_Austin.pdf

- National Council on Educational Standards and Testing (1992). *Raising standards for American education*.
- New York State Department of Education (2001). *Report on implementation of a system of accountability for student success*. Retrieved January 20, 2007, from <http://www.emsc.nysed.gov/nyc/PDFs/brd6.pdf>
- New York State Department of Education (2007a). *Confidence intervals*. Retrieved January 20, 2007, from <http://www.emsc.nysed.gov/irts/school-accountability/confidence-intervals.htm>
- New York State Department of Education (2007b). *School and district accountability rules: Implementing No Child Left Behind*. Retrieved February 8, 2007, from <http://www.emsc.nysed.gov/irts/accountability/2006-07/accountability-rules-Feb2007.ppt#31>
- New York State Department of Education (2007c). *Table of contents: Elementary and secondary demographic information*. Retrieved January 20, 2007, from <http://www.emsc.nysed.gov/irts/educationstats/edstats-07/toc.htm>
- No Child Left Behind Act (2002). Retrieved March 2, 2006, from <http://www.ed.gov/policy/elsec/leg/esea02/107-110.pdf>
- Owens, A. & Sunderman, G. (2006). *School accountability under NCLB: Aid or obstacle for measuring racial equity?* National Center for Research on Evaluation, Standards, and Student Testing [Policy Brief].
- Paige, R. (2002). An overview of America's education agenda. *Phi Delta Kappan*, 83, 708-713.
- Phelps, R. (2005). *Defending standardized testing*. Lawrence Erlbaum Associates: Mahwah, New Jersey.
- Phillips, K. (2004). *Testing controversy: A rhetoric of educational reform*. Cresskill, NJ: Hampton Press, Inc.
- Perot, R. (1997). *United we stand: How we can take back our country*. New York: Hyperion.
- Popham, W. (2003, January 8). The debasement of student proficiency. *Education Week*. Retrieved September 14 from www.edweek.org/ew/westory.cfm?slug=16popham.h22
- Rhoten, D., Carnoy, M., Chabran, M., & Elmore, R. (2003). The new accountability: High schools and high stakes testing. In M. Carnoy, R. Elmore, & L. Siskin (Eds.), *The conditions and characteristics of assessment and accountability*. New York: Routledge Falmer.

- Reeves, D. (2005). *Accountability in action: A blueprint for learning organizations*. Englewood, CO: Advanced Learning Press.
- Robinson, G., & Brandon, D. (1994). *NAEP test scores: Should they be used to compare and rank state educational quality?* Arlington, VA: Educational Research Service.
- Shepperd, L. (2000). *The role of classroom assessment in teaching and learning*. Santa Cruz, CA: Center for Research on Education Diversity and Excellence.
- Sirotnik, K. (2004). Introduction: Critical concerns about accountability. In K. A. Sirotnik (Ed.), *Holding accountability accountable* (pp.1-17). New York: Teachers College Press.
- Spellings, M. (2006). National Conference of Editorial Writers. Retrieved July 13, 2006 from <http://www.ed.gov/news/speeches/2006/09/09142006a.html>
- Texas Education Agency (2005a). *Accountability manual*. Retrieved July 15, 2006 from <http://www.tea.state.tx.us/perfreport/account/2005/manual/index.html>
- Texas Education Agency (2005b). *2005 AEIS Report*. Retrieved August 18, 2007 from <http://www.tea.state.tx.us/perfreport/aeis/2005/state.html>
- Texas Education Agency (2007a). *Grade placement committee manual*. Retrieved August 18, 2007 from <http://www.tea.state.tx.us/student.assessment/resources/ssi/gpcmanual07.pdf>
- Texas Education Agency (2007b). *Student enrollment reports*. Retrieved April 22, 2007, from http://www.tea.state.tx.us/cgi/sas/broker?_service=marykay&_program=adhoc.addispatch.sas&major=st&minor=e&endyear=07&format=W&linespg=60&charsln=120&selsum=m=ss&key=TYPE+HERE&grouping=g
- Texas Statutes (2006). *Texas education code*. Retrieved September 1, 2006, from <http://tlo2.tlc.state.tx.us/statutes/edtoc.html>
- Tracey, C. A., Sunderman, G. L., & Orfield, G. (2005). *Changing NCLB district accountability standards: Implications for racial equity*. Cambridge, MA: Civil Rights Project at Harvard University.
- United States Census Bureau (2005). State and county quick facts – California. Retrieved April 22, 2007 from <http://quickfacts.census.gov/qfd/states/06000.html>
- United States Census Bureau (2006). State and county quick facts – New York. Retrieved May 13, 2007 from <http://quickfacts.census.gov/qfd/states/36000.html>
- United States Census Bureau (2007). American factfinder. Retrieved May 13, 2007 from http://factfinder.census.gov/home/saff/main.html?_lang=en

- United States Congress (1992). *National education goals*. Retrieved September 1, 2006, from <http://www.ed.gov/legislation/GOALS2000/TheAct/sec102.html>
- United States Department of Education. (2006). Elementary and Secondary Education Act of 1965. Retrieved September 1, 2006 from <http://www.ed.gov/policy/elsec/leg/esea02/beginning.html#sec6>
- Weener, P.D. (1969). Social dialect differences and the recall of verbal messages. *Journal of Educational Psychology*, 60, 194-199.
- Wilde, S. (2002). *Testing and standards: A brief encyclopedia*. Portsmouth, NH: Heinemann.
- Wilgoren, J. (2000, March 14). Florida's vouchers a spur to two schools left behind. *New York Times*, pp. A1, A18.
- Wyatt, T. (2003). The assessment of African American children: An update on Larry P. Retrieved September 1, 2006, from <http://www.csha.org/Positionpapers/FinalLarryPTaskForce03.pdf>