DISSERTATION

POVERTY, SCHOOL SIZE AND CHARTER DESIGNATION AS PREDICTORS OF SCHOOL ACHIEVEMENT ON THE COLORADO STUDENT ASSESSMENT PROGRAM (CSAP) IN 2001 AND 2004

Submitted by

James Lauritz Chamberlin

School of Education

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2006

COLORADO STATE UNIVERSITY

May 5, 2006

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY JAMES CHAMBERLIN ENTITLED POVERTY, SCHOOL SIZE AND CHARTER DESIGNATION AS PREDICTORS OF SCHOOL ACHIEVEMENT ON THE COLORADO STUDENT ASSESSMENT PROGRAM (CSAP) IN 2001 AND 2004 BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Ray Yang Ph.D.

Joel Sheridan Ph.D.

Donna Cooner Ed.D.

George Morgan Ph.D. Advisor

Jean Lehman Director, School of Education

ABSTRACT OF DISSERTATION

POVERTY, SCHOOL SIZE AND CHARTER DESIGNATION AS PREDICTORS OF SCHOOL ACHIEVEMENT ON THE COLORADO STUDENT ASSESSMENT PROGRAM (CSAP) IN 2001 AND 2004

The purpose of this ex-post-facto research study was to investigate the relationship of poverty, school size and charter designation to middle school scores on the Colorado Student Assessment Program (CSAP) in 2001 and 2004. A second purpose was to investigate whether there was a change in middle school CSAP achievement between 2001 and 2004. The study also determined if there was an association between the 2001 and 2004 Colorado middle school CSAP results. Finally, the study investigated whether changes from 2001 to 2004 in individual school's CSAP performance could be predicted from the school characteristics of poverty, school size, and charter designation.

The researcher used the Overall Academic Performance Index (OAPI) score based on school CSAP performance for (N=357) Colorado middle schools in 2001 and 2004. The researcher also utilized the percent of students on free and reduced meals (FARM) as a proxy for poverty. The same schools were studied in both 2001 and 2004.

The results of this study found a greater than typical effect size relationship between high poverty and low achievement. Charter school designation alone indicated a small to medium positive effect, and school size only in combination with poverty showed a small negative effect. Charter designation in combination with poverty in 2004

showed a small negative effect. This research study found no correlation between school change scores and school poverty, school size or designation as a charter school. While schools generally improved their OAPI score from 2001 to 2004, individual school scores were highly correlated between the two years studied.

In conclusion, results from this study call into question the current use of the CSAP as a state-wide high-stake testing system intended to hold schools accountable for student achievement because of the much greater than typical negative effect of poverty on student achievement. Recommendations for improving the accountability model to include test results that would provide a clearer picture of student achievement are recommended.

James Chamberlin
School of Education
Colorado State University
Fort Collins, CO 80523
Summer, 2006

ACKNOWLEDGEMENTS

The author wishes to thank Dr. George Morgan for his patience and guidance through out the planning, research and writing process. His expertise and support were invaluable. I would also like to thank my committee; Dr. Yang, Dr. Sheridan and Dr. Cooner for their expertise and time. I am especially grateful for the ongoing support of my wife, Amy, and son Mason. Amy has been instrumental in helping me focus and complete this dissertation by showing unconditional love and tolerance. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

Table

CHAPTER 1

INTRODUCTION

Since the 1983 publication of "A Nation at Risk," state policy makers have taken a larger role in directing school reform initiatives. The demand for K-12 public education accountability has increased legislation at both the federal and state level to improve public education. In 2001 the U.S. Congress passed the "No Child Left Behind Act" (NCLB). This legislation requires public schools to show "Adequate Yearly Progress" (AYP) and to utilize "scientifically based" reform efforts (Bracey, 2003). Numerous state legislatures across the country have passed reform initiatives centering on high-stakes state-wide testing and accountability systems. The tests have become yardsticks that state policymakers use to compare schools and to mandate school reform through public and legislative pressure (Jones et al., 1999).

With all fifty states moving towards some form of high-stakes accountability system, the need to understand the impact school characteristics have on school test performance is increasing (Clark & Clark, 2000). Although each state has developed its own assessments, most have followed a model recommended by the National Research Council. The model focuses on the alignment of expectations (often called standards), curriculum, instruction, and assessment (DeVito & Koening, 2001). Colorado, Texas, Maryland, North Carolina, and Massachusetts have all implemented statewide standards based testing programs.

Although each state's testing program is different, all fifty states are moving to some form of testing aligned with state content standards. Forty-five states issue school report cards and in eighteen states high school graduation is contingent on state test performance (Johnson & Johnson, 2002). Colorado implemented yearly testing as part of a high-stakes testing and accountability system in 1997. The current Colorado Student Assessment Program (CSAP) includes standards-based testing of students in grades three through ten in reading, writing, math, and fifth, eighth and, tenth grade science. Student test results are used to hold school systems accountable. School districts are accredited by the Colorado Department of Education based partly on the CSAP performance of schools in their district. Individual schools are then rated by the Colorado Department of Education based upon the school's CSAP performance. The Colorado Department of Education (CDE) publishes the School Accountability Reports (SARs) annually. The school accountability reports or SARs compare school ratings with ten other closest "nearby schools." The SAR rating is determined by an Overall Academic Performance Index score or (OAPI). The OAPI is determined by each student's test results on the CSAP test. The CSAP program and SAR are similar to legislative accountability measures in many other states. Most states also include a rating for improvement or decline in student achievement over time (Linn, 2002).

As the call for "scientifically-based research" reform efforts in public education increases, the need for empirical research on student achievement on statewide tests is imperative. High-stakes testing and the accountability systems being implemented across the country have been met with resistance and praise. The use of school testing results to imply school quality and make comparisons to other schools and school years continues

to grow as a popular legislative lever to hold schools accountable for student learning. However, some researchers contend that these comparisons reveal little about the quality of schools, only the varying difficulty of educating different populations of students (Cooley, 1993). If the goal of the reform movement is to truly "Leave No Child Behind," then what influences school achievement needs to be identified. The influence of poverty, school size, and designation as a charter school on student achievement is of critical importance when evaluating state policies mandating school accountability programs.

Purpose Statement

One purpose of this study was to confirm the hypothesis that poverty, as indicated by the percentage of students on free and reduced meals, was a major predictor of student achievement on the Colorado middle school CSAP in both 2001 and 2004. This study determined whether a school's designation as a charter school and the school's size would add significantly to the prediction from poverty of middle school CSAP achievement in both 2001 and 2004.

A second purpose was to investigate whether there was a change in middle school CSAP achievement between 2001 and 2004. The study also determined if there was an association between the 2001 and 2004 student CSAP results.

Finally, the study investigated whether changes from 2001 to 2004 in individual school's CSAP performance could be predicted from school characteristics. By considering if a relationship exists between school characteristics and school improvement, then school improvement efforts, best teaching practices, and public policy can benefit from research to practice.

Research Questions

1. Is there a combination of 2001 percent of students on free and reduced meals, 2001 school size, and a school's designation as a charter school that predicts 2001 Colorado Middle School CSAP testing performance better than 2001 free and reduced meals alone?
2. Is there a combination of 2004 percent of students on free and reduced meals, 2004 school size, and a school's designation as a charter school that predicts 2004 Colorado middle school CSAP testing performance better than 2004 free and reduced meals alone?
3. Is there a difference in 2001 and 2004 OAPI scores over all of Colorado middle schools?
    a. Did scores change over time?
    b. Is there a correlation between 2001 and 2004 Overall Academic Performance Index scores?
4. What factors studied predict the direction and amount of change?
    a. Does 2001 school percent on free and reduced meals predict change in OAPI (2001-2004)?
    b. Does 2001 school size predict change in OAPI (2001-2004)?
    c. Does 2001 school designation as charter or not predict change in OAPI scores in (2001-2004)?
    d. Does the 2004 percent of students on free and reduced meals predict a change in OAPI (2001-2004)?
    e. Does the 2004 school size predict a change in OAPI (2001-2004)?

5. a. Is there a combination of 2001 percent of students on free and reduced meals, 2001 school size, and school designation as a charter school that predicts change in the 2001- 2004 Colorado middle school CSAP testing performance?

   b. Is there a combination of 2004 percent of students on free and reduced meals, 2004 school size, and school designation as a charter school that predicts change in the 2001- 2004 Colorado middle school CSAP testing performance?

6. After initial analysis of the research questions, the researcher investigated the 30 most improved schools and the 30 schools that declined the most on their OAPI scores from 2001 to 2004. From the analysis the researcher investigated the following questions:

   a. Do the top 30 and bottom 30 schools differ on percent of students on free and reduced meals, school size, and charter designation from the 357 middle schools included in this study?

   b. Do the top 30 and bottom 30 schools differ from each other on percent of students on free and reduced meals, school size, and charter designation?

Definition of Terms

Adequate Yearly Progress or AYP. A nationwide measurement of school performance created as part of the federal No Child Left Behind Act of 2001. Schools, and districts must be able to demonstrate that students are making academic progress as evidenced by annual gains in test scores at grade levels no individual students. AYP is a minimum level of progress that must be achieved each year (CDE, 2005).

Charter School. This term refers to a public school chartered by the local school board to operate as a public school within the district. The charter school is publicly

funded but operates independently of other schools within the district. Charter schools are generally small and have a focus or theme to the curriculum of the school. For example, Edison Schools, Core Knowledge, Back to Basics, or Expeditionary Learning. Charter schools serve a wide variety of purposes and functions in the public education system.

Colorado Student Assessment Program or CSAP. Created by the state legislature in 1997. Students in third grade through tenth grade are tested annually on a standardized assessment in reading, writing, math, and science (depending on grade level).

Middle School Size. This term was defined as the number of students enrolled in the middle school as of October 1st, 2000 and October 1st, 2003 as reported by the Colorado Department of Education.

No Child Left Behind Act or NCLB. This term refers to legislation passed by the U.S. Congress in 2001. The law directs each state to determine whether schools are meeting federal goals regarding student achievement, highly qualified teachers, safe schools, and other important education policies (CDE, 2005).

Overall Academic Performance Index or OAPI. The numerical score calculated for every Colorado middle school. The calculation is based on student CSAP results in a statistically reliable procedure.

Poverty. Defined as the percent of students on Free and Reduced Meals (FARM). This term was defined as the percentage of students who qualified for a federal program called free and reduced meals as reported by the Colorado Department of Education. (FARM) is used as a proxy for poverty in this study because qualifications for FARM are based on parent income and widely used in the literature as a proxy for poverty.

School Accountability Report or SAR. Created by Senate Bill 186 (CRS 22-7-601). The Colorado Department of Education publishes the Colorado School Accountability Report annually. The report rates all schools in the state as either excellent, high, average, low, or unsatisfactory based on annual CSAP scores. The SAR also reports a comparison of ratings of ten neighboring schools.

State-wide High-Stakes Testing. A standardized achievement test administered to all students in certain grades and in certain subjects for the purpose of school accountability. For example, the Colorado Student Assessment Program or CSAP, The TASS or Texas assessment of Student Success, The WASL or Washington Assessment of Student Learning.

Value Added Test Results. A term used to indicate test results that show a comparison of one test to another on the same student over time, for the purpose of showing growth or decline in where the student began and where the student ended up.

Limitations/Delimitations

The research was limited by the use of a non-experimental research approach. There was a limitation with the method of data collection and the maintenance of data records. Because this was an ex-post-facto study, there is no guarantee of the condition or accuracy of the existing data. There was an assumption that the CSAP tests were administered in a standardized manner as was required by the Colorado Department of Education and that the test results reflect students' best efforts.

This study was limited to defining student achievement based on the Colorado Student Assessment Program. The Colorado Student Assessment Program is unique to Colorado and is a standards-based criterion referenced test that is a point in time

measurement used in Colorado as part of a legislative mandated school accountability system.

This study is delimited to the CSAP results of Colorado middle schools. The school performance indicator OAPI was calculated by the Colorado Department of Education. For the 2000-2001 school year sixth graders were given the reading test; seventh graders were given reading, writing, and math while eighth graders were given reading and math assessments.

During the 2004 administration, all sixth and seventh graders were given the reading, writing, and math assessments; while all eighth graders were administered the reading, writing, math, and science assessments. The school's cumulative performance indicator composite score was calculated in 2001 and 2004 using a statistically valid and comparable measure.

This study was limited by using the number of students receiving free and reduced meals reported to the Colorado Department of Education as a proxy for the school's poverty level. This study was also limited in the definition of a charter school, which is unique to Colorado legislation.

Significance of the Study

This study examined the influence of poverty, school size, and charter designation on student achievement as measured by the Colorado Student Assessment Program (CSAP) in 2001 and 2004. The results of this research will add to a body of knowledge related to student achievement on statewide high-stakes testing. Educators and policy makers have a significant interest in information concerning the influences on student achievement.

An extraordinary amount of literature has emerged, both in support of and against the use of high-stakes testing as a school accountability and reform measure. However, there seems to be limited quantitative, correlational research on the CSAP results and a school's demographic and characteristic makeup. There also seems to be limited quantitative, longitudinal correlations that compare school achievement results over a three-year period.

Scholarly Research

Although a number of studies have examined the influence of poverty and school size on student achievement, only a few have specifically looked at statewide testing programs and made statewide correlations over time. The quantitative research is clear about the strong correlation between high poverty and low student achievement. The quantitative research is mixed on the influence of school size on student achievement. The quantitative research is also mixed on the influence of charter designation on student achievement.

Improving Educational Practice

This research study potentially benefits public education practice. A strong emphasis has been put on educators to improve the effectiveness of their practice by accountability measures. By examining the influence of school characteristics on student achievement, educational practitioners can better understand the factors influencing student achievement.

Improving Educational Policy

After five years of full implementation, the Colorado School Accountability Program (CSAP) is being challenged by opponents and praised by supporters. This

research study potentially benefits policymakers by examining the influences of school characteristics on student achievement. The use of the School Accountability Reports (SAR's) to communicate academic performance ratings based on middle school CSAP tests results was the focus of this study. Another potential benefit of this study was the longitudinal nature of the research design.

This research study examined the question of whether there has been a change in student achievement over time. The findings may reveal the effectiveness of the CSAP and School Accountability Report program in Colorado over the past five years.

Researcher's Perspective

The researcher has a professional and personal commitment to the field of educational research. As a teacher and principal over the past fourteen years the researcher has a vested interest in understanding the influences of poverty, school size and charter designation on student achievement. The researcher also has a vested interest in school reform policies that impact the public education system. The current use of the CSAP and School Accountability Reports as catalysts for school improvement are just beginning to be understood. As one middle school in the Denver Public School system is being converted to a charter school because the school failed to improve CSAP scores over time, the stakes on school accountability are heating up. The need for this level of data analysis on the CSAP test is timely and critical to helping shape school improvement efforts. The more that can be learned about the influences of student achievement the better public education systems can authentically improve student learning.

CHAPTER 2

LITERATURE REVIEW

One purpose of this study was to confirm the hypothesis that poverty, as indicated by the percentage of students on free and reduced meals, was a major predictor of student achievement on the middle school Colorado Student Assessment Program (CSAP) in both 2001 and 2004. This study determined whether a school's size and a school's designation as a charter school added significantly to the prediction of the middle school CSAP achievement in both 2001 and 2004.

The primary purpose of this literature review is to present research on the relationship between student achievement on statewide high-stakes testing and school poverty, school size, and charter designation. Through this literature review, the researcher examined the theoretical framework for the relationship between school poverty levels, school size, charter school designation, and student achievement on statewide, high-stakes tests.

A second purpose was to investigate whether there was a change in middle school CSAP achievement between 2001 and 2004. The study also determined if there was an association between the 2001 and 2004 Colorado middle school CSAP results.

Finally, the study investigated whether changes from 2001 to 2004 in individual school's CSAP performance could be predicted from school characteristics of poverty levels, school size, and charter designation. By considering if a relationship exists

between school characteristics and school improvement, then school improvement efforts and public policy can benefit from research to practice.

The first section of the literature review focuses on the variables of school poverty, school size, and charter school designation. The researcher presents literature that supports the hypothesis that school poverty, school size, and charter designation predict school achievement and improvement or decline on statewide high-stakes tests. The review examines the school poverty variable, school size variable, and school designation as a charter school variable as it relates to student achievement in the literature.

The second section of this literature review is to present research associated with statewide high-stakes public school testing and accountability programs. The researcher used the Colorado Student Assessment Program to answer the research questions in this study. In order for the results to be related to student achievement outside of Colorado, the researcher compares the Colorado Student Assessment Program to assessment programs in Texas, Massachusetts, North Carolina, and Maryland.

Relationship Between Student Achievement and School Characteristics

This section of the literature review examines the relationship that exists in the literature between student achievement and school poverty, school size, and designation as a charter school. Because this research study focuses on statewide high-stakes tests the researcher tried to find empirical studies that used statewide tests to indicate student achievement. Since the statewide high-stakes testing movement is less than ten years old there were not many empirical studies found. Therefore, the researcher expanded student

achievement to include other national assessments like the National Assessment of Educational Progress (NAEP) and ACT.

School Poverty

Research has demonstrated that family background variables like socio-economic status are highly related to student achievement (Ainsworth-Darnell & Roscigno, 1999). The question that is often asked by researchers is how strong is the relationship between poverty and student achievement? The answer depends on how the researcher defines both poverty level and student achievement. The following studies have attempted to show a relationship between the variables of school poverty and student achievement.

National studies. Regarded as the beginning of outcome-based assessment, the 1966 Equality of Educational Opportunity Survey headed by James Coleman sought to determine how to provide equality of education to all American pupils (Lissitz & Schafer, 2002). The survey sample included 645,000 pupils, 60,000 teachers and over 4,000 schools across America. What was drawn from the survey was that:

> Schools bring little influence to bear on a child's achievement that is independent of his background and general social context...this very lack of independent effect means that inequalities imposed on children by their home, neighborhood and peer environment are carried along to become the inequalities with which they confront adult life at the end of school. (Lissitz & Schafer, 2002 p. 52).

The results of the Coleman survey weakened the belief in the presumed direct connection between school inputs and school outputs and led to a national debate over the effectiveness of schools (Lissitz & Schafer, 2002). In the literature the Coleman survey was regarded as one of the first studies to strongly link high poverty with low student achievement.

The National Assessment of Educational Progress (NAEP) as a measure of student achievement is the only test using nationally representative samples of U.S. students that can be used to track long-term national trends and accurately measure differences between states (Grissmer, et. al, 2000). The NAEP test consists of multiple choice as well as constructed response questions. The state results have been compared to both census data and student reported data on family characteristics.

> Research that attempts to explain variance in NAEP test scores across populations of diverse groups of students shows that family and demographic variables explain the largest part of total explained variance. Among commonly collected family characteristics the strongest associations with test scores are parental education level, family income and race/ ethnicity (Grissmer et al., 2000 p. 15).

A 1990 state-by-state comparison of NAEP scores found the states that scored the highest were small northern states with the highest level of parental income and parental education while the lowest scoring were southern states with the lowest levels of income and education. The family characteristic variables on the NAEP analysis accounted for about 75 percent of the variance in average state achievement scores (Grissmer et al., 2000). From the literature, poverty appears to be strongly related to student achievement on the National Evaluation of Educational Progress.

Although state-wide high-stakes assessments differ from the NAEP in some aspects, many state tests are based on national standards and a common curriculum similar to the NAEP. State tests include both multiple choice and constructed response questions. Some state tests are norm-referenced like the NAEP and some are not. Both the NAEP and statewide high-stakes assessments are used by policy makers to determine school quality based on student achievement data.

State-wide studies. One of the earliest statewide high-stakes testing programs implemented was in Kentucky. Although it has now been scrapped and changed many ways, it drew research attention in the early 1990's. The state compiled an accountability index score on every school in the state. The score was based on student performance on portfolios of student work, scores from on-demand assessments, and student performance events in the areas of mathematics, science, social studies, arts and humanities, practical living, and vocational skills (Guskey, 1997).

Kentucky also collected the percent of students qualifying for free and reduced lunch at the school level. The free and reduced lunch program is a federal program with parent income as the main criteria for consideration. The study used the free and reduced lunch variable as a proxy for socio-economic status. Guskey (1997) used the composite index results from 49 schools in Kentucky from 1993-1995 and correlated them with the percent of students on free and reduced lunch. Guskey found for middle schools "the correlations between socio-economic indices and accountability index scores were negative and large in magnitude (r= -.60 to -.96)" (p. 10). This research suggested that as the percent of students qualifying for free and reduced lunch increased index scores generally decreased. Interestingly, no relationship was found between gain scores and socio-economic variables over the three years studied. Guskey (1997) also concluded that "a single socio-economic variable, percent of students qualifying for free and reduced lunch benefits, provided the most parsimonious model for predicting accountability index scores" (p.11).

Another empirical study in Alabama examined the State Education Report Card. The study used the school performance grade as the dependent variable and the percent of students on free and reduced lunch, school size, and other school characteristics as predictor variables. The study found that the percent of students on free and reduced lunch was the strongest predictor of student achievement followed by average daily attendance, school size, and percent of teachers with masters' degrees (Miller-Whitehead, 2001).

While many states also collect information on the percent of minority students, student mobility rates, percent of students in special education, and retention rates, the data were found to be highly correlated, with the percent of students qualifying for free and reduced lunch for middle schools at r=.92 (Guskey, 1997). The literature appears to support using the percent of students qualifying for free and reduced lunch as a proxy for socio-economic status in educational research. The studies found seem to indicate that parent income, which free and reduced lunch numbers are based on, has a strong relationship to student achievement.

The strong relationship between a school's poverty level and student achievement is not a new research finding. Since the Coleman report in 1966, educational researchers have debated what schools can do to overcome the impact of poverty on student achievement. The current literature on the issue is framed by the term "closing the achievement gap." Federal legislation of No Child Left Behind and the statewide high-stakes testing movement is directly aimed at helping poor children achieve more and holding schools accountable for the improved learning of disadvantaged students.

The achievement gap. A wide variety of educational interventions emerged in the literature as ways schools can try to overcome the strong relationships between high poverty and low student achievement. Student achievement on statewide high-stakes testing programs has shown consistently, that certain groups of students score far below students in other groups (Hertert & Teague, 2003). The "achievement gap" is a term used in the literature to describe the achievement differences between certain groups of students. Hertert and Teague (2003) assert that, "Poverty is the single best explanation research has found for why children differ in the ways that affect school performance, both before they enter school and once they are enrolled" (p.3).

Achievement in schools with a high percent of students in poverty was the focus of a recent Colorado newspaper study. The study conducted by Mitchell and Hubbard of the Rocky Mountain News (July 16, 2005, p. 2a) reports that after eight years of CSAP testing, progress by poor kids is mixed. The study analyzed fourth grade reading results from 1997-2004 in the twelve largest and most diverse school districts in the state. Five districts had improved ie., reduced the achievement gap and seven had either no improvement or declined. With such a strong correlation between poverty and poor educational achievement, researchers are searching for ways to overcome the obstacles for poorer students. " ...state and local educators need better data and definitive research about how k-12 students are doing and what practices make a difference in their performance" (Hertert & Teague, 2003, p. 4).

While state-wide high-stakes accountability systems are exposing the issues of school poverty and student achievement, practitioners are struggling to find answers to the achievement problems that prevail. The literature emerged with numerous

intervention suggestions all the way from single sex schools and back to basics curriculum to total privatization of public schools and place-based education curriculum. While the literature is saturated with theoretical opinion articles, no empirical studies revealed a silver bullet approach to help educators overcome the influence of poverty on student achievement.

The purpose of this dissertation study was to investigate the prediction of school achievement on the Colorado Student Assessment Program in 2001 and 2004 from the percent of students on free and reduced meals, school size, and charter designation. The literature strongly supports the hypothesis that the percent of students on free and reduced meals is a strong predictor of student achievement. The follow-up question is what can educators do to overcome the impact of poverty on student achievement? By exploring school size and charter designation, this researcher hopes to add to the limited literature about what school characteristics help overcome the influence of high poverty on low student achievement.

From the literature, both charter schools (school choice movement) and school size emerged as frequently studied school characteristics that policy makers had used to try and overcome the influence of high poverty on low achievement. School size and charter designation were identified as variables worth studying and are described later in this literature review.

School Size

School accountability programs continue to grow in popularity as legislative programs to improve public schools. Improving student achievement often is the goal of statewide, high-stakes testing programs. The size of a school and its relationship to

student achievement is often cited in the research. The concept that school size matters comes from a variety of empirical studies. "Effective-schools research supports the notion that all children can learn and that schools control the factors necessary to assure student mastery of a rigorous core curriculum" (Hertert & Teague, 2003 p. 4). School size is one factor that policy makers can and do control fairly easily. "The effective-schools research counters the interpretations of Coleman's earlier work that schools could do little to compensate for the differences in family backgrounds that so strongly influenced children's school performance" (Hertert & Teague, 2003 p. 15).

The literature on school size has historically focused on optimizing the administration of schools for financial efficiency (Howley, 1996). The literature about school and district size dates back to 1915 with the work of Cubberly, who asserted that pupil teacher ratios could be increased in consolidated schools, longer terms could be held, and transportation could be provided (Howley, 1996). His research guided the consolidation of rural schools across the United States.

Today, the school size question is more associated with the need for school academic effectiveness than for the cost efficiency. The early research studies on effectiveness looked to determine an optimal size for elementary, middle, and high schools. This research found that the optimum size for a school to be effective achievement wise was different than for financial efficiency. Basically, the schools needed to be twice as large for financial efficiency as for academic effectiveness, based on student achievement data (Howley, 1989).

Research studies have uncovered a generally negative relationship between school or district size and student achievement. The literature indicates a range of maximum

school size for both effectiveness and efficiency, to be between 300 and 900 students depending on the demographic make-up of students (Ishmer, 1997). Most researchers agree that school size alone does not improve student achievement. Small school advocate and educational researcher, Debra Meier (1996) explains that small schools have better governance, respect, simplicity, safety, parent involvement, accountability, and belonging.

The most recent research on school size involved looking at the results of student achievement on state-wide assessments. The emerging literature seems to elaborate on the school size research to show that optimal school size depends highly on the socio-economic background of the community. School size researchers Howley and Bickel (2000) conducted a study analyzing the relationship between school level performance on tests, school size, and community poverty level in Georgia, Ohio, Montana, and Texas. This empirical study found that the correlation between poverty and low achievement was as much as ten times stronger in larger schools than in smaller ones in all four states.

A second empirical study conducted on Arkansas statewide student achievement data by Johnson, Howley and Howley (2002) indicated that the higher the poverty level in a school, the more negative influence large school size held on student achievement. They also concluded that the achievement gap between more affluent and less affluent students was narrowed in smaller schools and widened in larger schools.

A third empirical study conducted in Alabama investigated the 2000 school report card. The study used the school report card ratings to compare student achievement with school socio-economic status as well as school size. The research found that after controlling for socio-economic status, the variables of school size, and average daily

attendance were the two strongest predictors of student achievement (Miller-Whitehead, 2001). The literature on school size has evolved from recommending school consolidation in the 1920s to the National Association of School Principals recommending high schools no larger than 600 in the 1990s to current research which shows the benefits of smaller schools, especially for poorer students. The issue of school size appears to have a strong relationship to student achievement; but the relationship is not necessarily linear and varies greatly by the level of school poverty.

The variable of school size was identified by the researcher as a possible predictor of student achievement on state-wide high-stakes assessments. The relationship between student achievement and school size is one of the main research questions in this study.

Charter School Designation

Charter schools fall under a larger category of school reform initiatives called "school choice". According to Hadderman, (2002) school choice has become a popular strategy for reforming American education and attempting to equalize educational opportunities. Hadderman, (2002) describes the school choice movement to include magnet schools, open enrollment schools, charter schools, contracted schools, voucher programs, tax credit programs, scholarship programs, alternative schools, and home schooling.

Leading charter school researchers Buckley and Fischer (2002) report that charter schools are an institutional innovation where a particular learning approach or curriculum is not endorsed. The charter school is not a kind of school but rather an empty institutional structure allowed by law. Charter school teaching approaches and curriculum

come from what the organizers of a charter school put into it. Therefore, charter schools have been developed to serve a wide variety of purposes across the country.

Charter schools are public schools that are created through a contract or charter with a state education agency or local school board. The schools are publicly funded and are held to the same accountability standards as regular public schools. As of 2002, 34 states had charter schools and statewide testing programs (Loveless, 2002). The number of Charter schools operating in the United States has soared over the past decade to serve 15-20 percent of the public school population in this country (Hadderman, 2002). Because charter school laws are created by state legislation there is a wide variety of charter schools across the country. The purpose and process for operating a charter school depend upon the state and district in which it is located.

In a review of the literature on charter school legislation, there really are no "typical" charter schools (Goldhaber, 1999). There are charter schools that are college preparatory, back-to-basics, alternative schools, at-risk schools, and private-corporation run schools, schools for dropouts, and schools for second language learners. Charter school legislation seems to have grown out of both the accountability movement and school choice movement (Loveless, 2002). If a charter school is not accountable, its charter is revoked and parents have a choice of sending their students somewhere else. The popularity of charter schools continues to grow despite the lack of conclusive research on the positive effects on student achievement. In 1992 there were two charter schools in the United States. In 2002 there were an estimated 2,500 charter schools across the country (Loveless, 2002). Loveless explains, "research has been slow to accumulate because most charters, even today, are relatively new they lack standardized test score

data on which to compare achievement, and the test data from very small charters, which make up a large portion of the total number of charter schools operating, is sketchy at best" (p. 2).

One aspect of charter schools that was consistent was charter school size. Almost all charter schools were found to be small. The median enrollment was 137 students per school compared to non-charter schools where the median was 475 students per school (Nelson et al., 2000).

Research comparing student achievement in charter schools and non-charter schools on statewide tests is beginning to emerge with mixed results. In 2001, Miron and Nelson conducted a meta-analysis of 15 studies on charter school achievement. They found overall results to be mixed or very slightly positive. Some states like Arizona showed positive findings for charter schools, while Michigan had strong negative findings for charter schools.

A study by Loveless (2001) examined statewide testing information on 638 charter schools and non-charter schools from 10 states from 1999-2001. Loveless found that performance varied by state. Charter schools in Texas, Pennsylvania, Minnesota, Michigan, and Florida scored approximately one standard deviation or more below the state means. However, he found that charter students in Arizona and California were achieving comparable to students in average non-charter schools. Only in Colorado are charter schools test results significantly above the average of non-charter schools (Loveless, 2001).

Slovacek, Kunnan, and Kim, (2001) conducted an in-depth analysis of California charter schools serving low socio-economic students. The researchers used the academic

performance index scores for schools in California to compare achievement in 1999 and 2000. They only made comparisons between charter and non-charter schools with either greater than 50 percent receiving free and reduced lunch or greater than 75% receiving free and reduced lunch. They found that at the 50% or higher-level charter schools showed a 22% gain in student achievement while non-charters showed a 19.4% gain. When comparing schools at the 75% or more level, they found charter schools gained 28.1% while non-charters gained 23.8% (2002). Slovacek et. al., (2001) concluded, "In general, the results showed that charter schools are doing a more effective job of improving academic achievement of California's most difficult to serve students, those students from low-income families" (p. 11).

In Colorado, a 2003 study concluded that the distribution of how charter schools performed compared to non-charters on the CSAP assessment found that in 2002, 20% of the charter schools scored in the "excellent category" while 7% of non-charter schools scored in the excellent category (Fitzgerald, 2003). The study also found that 81% of charter schools in Colorado showed stable or improved performance from 2001-2002, while 87% of non-charter schools showed stable or improved performance (Fitzgerald, 2003). The Colorado report also stated:

> While noting performance differences between charter school students and non-charter schools, the report did not attribute such differences to the distinctive educational programs offered by charter schools and non-charter schools. Disparities in student performance may reflect other differences among charter school and non-charter school students. (Fitzgerald, 2003 p. 27.)

In a comprehensive review of the research on charter schools (Buckley & Fisler, 2002) reviewed 25 studies on charter school achievement. They did not find any conclusive data to indicate that charter schools on the whole were failing students

compared to non-charter schools or out performing non-charter schools. Rand researchers also concluded that the "evidence on the academic effectiveness of charter schools is mixed" (Gill, Timpane, Ross, & Brewer, 2001, p.95).

Drawing overall conclusions about charter school effectiveness is problematic due to the wide discrepancy among state charter school legislation. Colorado charters have shown greater achievement than the general population, other states have shown less achievement. Charter schools are a large part of the statewide high-stakes testing initiative. They are also part of the school choice and public school competition and accountability initiative. Colorado's accountability program mandates that failing schools be converted to charter schools. Policy makers and educators need more research on student achievement in charter schools. The literature overall seems mixed on the effectiveness of charter schools as an accountability and school choice reform movement.

State-wide High-stakes Testing Programs

The second purpose of this literature review was to present research about statewide high-stakes testing and accountability programs. The first part of this literature review focused on student achievement as it related to school poverty, school size, and designation as a charter school. The second part will report on the theoretical concepts of statewide high-stakes testing programs and their effectiveness. Additionally, the researcher will compare the Colorado testing programs to the Texas, Massachusetts, North Carolina, and Maryland testing programs. Through this comparison, the researcher will provide a theoretical framework to generalize the research findings beyond Colorado.

High-stakes Testing

As of 2002, 27 states used high-stakes testing programs to measure school quality and hold schools accountable for student achievement (Amrein & Berliner, 2002). Since most statewide testing programs have been implemented over the past five years, the research is emerging through conference papers, state department of education reports, and presentations. The topic of high-stakes statewide testing is a controversial topic. The controversy has garnered a lot of opinion papers about the perceived effectiveness or lack there of on statewide high-stakes testing programs as a tool to reform public education. While the empirical research on the success of statewide high-stakes testing is limited, the impact on public school systems across the country has been enormous. Linn (2000) explains:

> Student achievement is being used not only to single out schools that require special assistance, but also provide cash incentives for improvements in performance. Yet several fundamental questions remain about the student assessments, the accountability model, and the validity, impact, and credibility of the system (p. 3).

Statewide high-stakes accountability programs have been playing an increasingly active role in educational reform. However, it is unclear what effects these reform efforts have on educational processes and student achievement (Muller & Schiller, 2000).

The use of statewide high-stakes testing programs for the purpose of public school accountability stems from the standards-based educational reform movement. The philosophy of the current standards-based reform movement asserts that all children can learn and that schools control the factors necessary to assure student mastery of a rigorous core curriculum (Hertert & Teague, 2003). The use of statewide tests to measure student mastery of core curriculum becomes high-stakes when the test results are used to

hold public schools accountable for the students' achievement. For the purpose of this literature review and research study, statewide high-stakes tests are defined as tests from which results are used to make significant educational decisions about schools, teachers, administrators, and students.

The following model has been used by most states in the process of implementing statewide high-stakes assessments: First, states adopt content standards for the subject matter and grade levels to be tested. Second, from the content standards and grade level benchmarks, the state works with a testing company to design a testing framework. Third, the testing company uses the testing framework to develop the assessment. Fourth, the cut scores or performance standards are developed. Fifth, the scores are reported and used to make significant educational decisions about the quality of schools' educational programs.

While the process of implementation has been similar among states, the definition of content standards and level of expected performance vary greatly. Linn (2003), president of the American Educational Research Association explains, "Almost every state has adopted content standards for the areas of reading/ language arts and mathematics. The statewide assessments vary greatly in their specificity and arguably how challenging they are" (p.4).

Another area of difference among statewide testing programs is the high-stakes component of the testing program. Some state legislatures have set performance standards for schools and districts to be accredited, while other states have required the students to be proficient in order to graduate. With the passage of federal legislation in 2001 of No Child Left Behind or NCLB, schools are required to show steady gains in

student achievement called Adequate Yearly Progress or AYP. Adequate yearly progress is expected for all students especially those of minority, special education, and English language learners (Linn, Baker, & Betebemner, 2002). The use of statewide testing programs to meet federal requirements has standardized some components of the testing program across states. For example, the subjects tested and what grades are tested has become more uniform.

The term "high-stakes" goes beyond publishing test results publicly. Amrein and Berliner (2002) of Arizona State University's Education and Policy Studies Unit describe the stakes in the following categories:

Stakes affecting schools

1. Authority to close or revoke a school's accreditation, and or take over or re-constitute low-scoring schools.
2. Monetary awards are given to high performing or improving schools.

Stakes affecting administrators or teachers

1. Monetary awards can be use for teacher bonuses.
2. The state has the authority to replace principals or teachers due to low-test scores.

Stakes affecting students

1. Grade to grade promotion is contingent upon a promotion exam.
2. The state permits students in failing schools to enroll elsewhere.
3. Monetary awards of scholarships for college tuition are given to high performing students (p. 7).

Table1 details five states with consequences written into state K-12 testing policies in states with high-stakes testing programs.

Table 1

Consequences Written Into State K-12 Testing Policies in States with High-Stakes Testing Programs

		Consequences in Policy					
State	Year high - stakes implemented	State take over of school	Monetary awards	Replace principals or teachers	Grade promotion	Permits students in failing schools to enroll elsewhere	Scholarships for high performing students
Colorado	2000	Y	Y	Y	N	Y	N
Texas	1993	Y	Y	Y	N	Y	N
Massachusetts	1999	Y	N	Y	N	N	Y
North Carolina	1997	Y	Y	Y	Y	N	N
Maryland	1993	Y	Y	Y	N	Y	N

Note. Table 1 excerpted from (Amrein & Berliner, 2002, Table 1, p.7)

Results of High-stakes Testing

The reported successes and failures of statewide high-stakes testing programs often depend upon who or what organization is doing the reporting. The proponents of statewide high-stakes testing programs believe that if the state tests the content standards, the teachers will teach the standards, and students will learn the standards. The assessment results can guide the educational system to be more effective and productive when teachers and administrators are held accountable for student learning (Firestone et al., 1998). The opponents of statewide high-stakes testing programs contend that, "The current obsession with standardized testing that is driving education of all kinds in the United States is unhelpful, unfair, and illegitimate. It is invidious to the pursuit of real

learning and helps to legitimate the deep inequalities that characterize access to educational opportunity" (Broadfoot, 2000, p.1).

According to Darling-Hammond (2000), "A body of research has emerged that suggests that accountability strategies that have relied on primarily high-stakes testing programs have not always improved the quality of instruction or the outcomes for educationally needy students" (p. 5).

Although student achievement on statewide high-stakes tests has shown improvement, many professional organizations have issued position statements concerning the use of statewide high-stakes assessments. The International Reading Association (1999) voiced the concern that test results now control instruction rather than inform it. The American Educational Research Association (2000) expressed concern over the possibility that more students would be placed at increased risk of failure and eventually dropout of school (Zeller & Jenkins, 2001). Darling-Hammond (2000) contends that, "ultimately, accountability is not only about measuring student learning but actually improving it. Consequently, genuine accountability involves supporting changes in teaching and schooling that can heighten the probability that students meet standards" (p.46).

Amrein and Berliner (2000) conducted a study on the impact of high-stakes tests on student performance. The researchers analyzed the NAEP results in 28 states with high-stakes tests and ACT, SAT, and AP test results. The study found that although test scores on state-administered tests usually increased after high-stakes testing policies were implemented the researchers found no evidence of similar gains on the NAEP, ACT, SAT or AP test. They concluded that students are learning only the content and item

forms of the state-administered test and there doesn't seem to be a carryover effect when student achievement is measured by independent measures on the NAEP, ACT, SAT or AP test.

Another study conducted in Maine and Maryland by Firestone et al. (1998) concluded that statewide high-stakes assessments made an impact in the classroom by aligning content with what is covered on the test; however, accountability policies are less successful in changing basic instructional strategies in the classroom. The instructional strategies and school characteristics recommended to improve student achievement are mixed in their ability to replicate success. Hertert and Teague (2003) explain that, "the findings from much of this work are mixed and provide little guidance in designing discrete solutions with reliable and predictable results. To the extent conclusions have been drawn, they indicate that no silver bullet exists" (p. 4).

Regardless of the conflicting opinions on the use of statewide high-stakes tests, there is agreement that the use of assessment is a powerful lever for shaping instruction (Firestone et al., 1998). The use of high-stakes testing and accountability measures continue to be implemented across the nation. The implementation and design of the high-stakes testing programs continue without a clear picture of their effectiveness. Loveless (2001) claims, "Although it may seem intuitive that holding educators accountable for student achievement will improve the likelihood of its attainment, there is a dearth of empirical evidence that accountability systems actually raise test scores" (p. 3).

States With High-stakes Testing

This section of the literature review describes testing programs in five states: Colorado, Texas, Massachusetts, North Carolina, and Maryland. The purpose for the in-depth review is to illustrate similarities and differences in the way state legislatures have used high-stakes testing as an educational reform tool. This section of the literature review will also summarize literature on the successes and failures of statewide high-stakes testing programs as a tool to improve student achievement.

Colorado. Colorado's high-stakes testing program has taken shape over the past ten years. Similar to other states, Colorado has had to make adjustments to meet federal mandates of NCLB and AYP. The CSAP is described as "the longest-standing, standards-based accountability assessment program in the United States and is often praised as a model assessment program nationwide" (CDE Feb. 2005, p.19). Similar to other states, Colorado adopted statewide standards, required assessments on those standards, and implemented accountability measures based on the results of the assessments. The Colorado Department of Education organizational commitment specifies increasing student achievement levels for all students through three interlocking components:

1. High Standards for what student must know and be able to do
2. Tough Assessments that honestly measure whether or not students meet standards and tell citizens the truth about how well our schools serve children
3. Rigorous Accountability Measures that tie the accreditation of school districts to high student achievement (Fitzgerald, 2003, p3.).

The accountability program involves the Colorado Department of Education using student achievement data to publish a school accountability report annually. The Colorado Department of Education is also responsible for accrediting school districts based on student assessment data.

The CSAP started in 1997 with two tests, and in 2004 included 27 tests. Students are tested in reading grades 3-10, in writing grades 4-10, and in math grades 4-10. Student performance on the Colorado Student Assessment Program is at the core of Colorado's school reform initiative.

The Colorado Student Assessment Program is the basis for the Colorado Department of Education's accountability measures. CSAP results are used to determine the school's rating on the school accountability report and accreditation status. CSAP is used to determine disaggregated results of student groups mandated by NCLB and AYP. CSAP is used in measuring longitudinal growth of a school's annual progress.

The CSAP test was designed collaboratively with Colorado teachers and CTB-McGraw Hill. The test is a standards-based assessment and criterion referenced. The reading and writing tests include multiple choice, short answer, constructed response questions, and an essay. The math test includes multiple choice, short answer, and constructed response. Students' raw scores are converted to scale scores, and performance is reported as unsatisfactory, partially proficient, proficient, and advanced. Students in each category score points used for the calculation of the school's weighted index score. The weighted index score is used to determine the rating of the School Accountability Report.

The School Accountability Report uses the overall academic performance index or OAPI to rate schools. Schools are rated as Excellent, High, Average, Low, or Unsatisfactory. A school that rates Unsatisfactory for three consecutive years is converted to a charter school by the Colorado Department of Education. Schools are also compared to ten other schools within a seventy-five mile radius on the accountability reports. Since the OAPI is the dependent variable in this study, a more in-depth explanation is in Chapter 3.

The Colorado legislature, along with a handful of other states, has led the way in implementing state-wide high-stakes assessments. In Colorado, the high-stakes are placed on teachers and administrators because of the ability of CDE to convert the school to a charter for "unsatisfactory" performance. Students in schools labeled "Unsatisfactory" schools are also given vouchers to attend another public or private school of their choice (CDE, 2005).

Colorado is among one of the first states in the nation to convert a public school into a charter school for failing to improve on the statewide high-stakes testing program. Overall, the publication, School Accountability Reports: A Five Year Review of Progress 2000-2005, released by the Colorado Department of Education in February 2005, indicates that CSAP scores have improved in most tested areas and for all population groups. While improvement has been steady, the progress has been slow. Table 2 indicates the number of schools in Colorado being categorized as unsatisfactory, low, average, high, or excellent from 2000-2004.

Table 2

Colorado School Accountability Report (SAR) Distribution of Ratings 2000-2004

	Years			
Ratings	2000-2001	2001-2002	2002-2003	2003-2004*
Excellent	129	119	187	200
High	403	451	507	507
Average	631	665	634	679
Low	398	421	394	358
Unsatisfactory	30	49	32	13
No Rating	154**	16	21	85
Totals	1745	1721	1775	1842

**2003-2004 data includes alternative schools as defined by S.B. 04-83*
***This figure included programs, not schools, that were later incorporated into other SARs.*
Note: All data is excerpted from the School Accountability Reports: A Five Year Review of Progress 2000-2005 released by the Colorado Department of Education in February 2005, p.20.

Texas. The state of Texas has received significant media attention for its high-stakes testing program because President Bush cites the program's success as the foundation of his federal "No Child Left Behind" legislation. According to the Texas Education Agency, which is the branch of Texas government that oversees public education in Texas, high-stakes testing and the accountability system has been in place since 1993 (TEA, 2005).

> This system depends upon the existence of a statewide curriculum, a statewide assessment system based on that curriculum, a comprehensive data collection system, and an essentially stable set of statutes over the past eight years. The system's stability permits analysis of statewide change in performance on the academic excellence indicators as well as on the accountability system outcomes (TEA, 2005).

The components of the Texas accountability system are similar to those in Colorado. The Texas Education Agency along with the State School Board of education implemented a system of accountability which rates schools and districts based on statewide test performance. The district rating levels are Exemplary, Recognized, Academically Acceptable, and Academically Unacceptable. The school ratings are Exemplary, Recognized, Acceptable, and Low Performing (TEA, 2005). The criteria considered and cut scores for obtaining the ratings have changed since 1993 due to changing legislation. The incentives and sanctions for performance have also changed depending upon the financial allocations.

The Texas Successful School Awards System provides financial awards to schools rated Exemplary, Recognized, and Acceptable that have shown gains in students' performance. A peer review team visits the schools in Texas that score the lowest rating and specifies improvement plans that must be put in place. Schools that fail to improve for two or more years are subject to a higher level of intervention by the state. In addition, any parent who has a child in a low performing school may apply for a grant from the state to help pay for him/her to attend a higher performing school.

From 2003 to 2005 The Texas Education Agency phased in the following criterion referenced assessments for the public school accountability system: reading tests for grades three through nine; English/language arts for grades ten and eleven; math for grades three through eleven; writing in grades four and seven; social studies in grades eight, ten, and eleven; and science in grades five, ten, and eleven.

In Texas, the Academic Excellence Indicator System used to rate all schools and school district was based upon the percent of students passing the tests along with

dropout rates. These components make up the rating part of the school report card. All public schools, charter schools, and alternative schools participate in the testing program.

The high-stakes testing and school accountability program in Texas holds the schools accountable for student performance on the Texas Academic Assessment System or TAAS. At this time the TAAS is not used for students' grade promotion or retention. The Texas Education Agency does collect data on school size, percent of students on free and reduced lunch, and charter schools. All of these components compare closely with the Colorado accountability system.

Massachusetts. The Massachusetts Comprehensive Assessment Program or (MCAS) was adopted by the state legislature in 1999. Similar to other states, changes to the program were made to comply with federal legislation, the "No Child Left Behind" Act, in 2002. Currently, MCAS includes performance and accountability ratings for all public schools and districts in the Commonwealth (MCAS, 2004).

All public schools in Massachusetts are rated based on a composite performance index score derived from two years of student testing data. The rating includes current performance and improvement from the baseline years. All tests are criterion referenced to state standards and include multiple choice, constructed response, short answer, and essay. Harcourt Educational Measurement has been contracted to construct and grade the tests.

The MCAS program tests in English and Language Arts in grades 3, 4, 7, and 10. Math is tested in grades 4, 6, 8, and 10. Students' raw scores are converted to scale scores, which are used to determine the Composite Index Score for the school and

district. Students' scores are categorized into Advanced, Proficient, Needs Improvement, and Warning/ Failing (MCAS 2005).

The MCAS school and district accountability reports are based on the requirements of NCLB and AYP. Schools that fail to meet AYP for more than two consecutive years are identified for improvement. Schools that fail to meet AYP beyond four consecutive years are subject to corrective action and restructuring by the state. The actual accountability for Massachusetts's schools identifies which sub-groups are meeting the goals of AYP. For example, what percent of special education students are proficient on the MCAS tests? The entire MCAS accountability report focuses on how well a school is reaching the 2014 goal of 100% proficiency of all students on the MCAS assessment. The target goals for each cycle depend on the baseline for that school and subject tested. The goal is calculated based on the improvement needed to reach 100% by 2014.

Maryland. Maryland's State Department of Education revamped its state accountability system formerly known as MSPAP in 2003 to meet federal mandates in NCLB. The new testing program is called Maryland School Assessments or MSA. The Maryland State Department of Education describes their accountability system as Maryland holds schools and school systems accountable for improvement by reporting achievement data, recognizing progress, and taking action when schools are not improving (MSDE, 2005).

The new test is a hybrid of both norm-referenced components along with criterion-referenced components. Students receive multiple-choice questions along with writing and short answer sections. For accountability purposes, only the criterion

referenced scores are used. The state contracted with CTB McGraw Hill in 2002 to develop and grade the assessments. The company is the same testing company Colorado contracts with to develop and grade the CSAP test.

The Maryland School Assessment program tests all students in reading and math at grades three, five, and eight. It also tests tenth graders in math. The results from the tests are used by the Maryland State Department of Education to determine if schools' improvement goals have been met. School improvement goals are closely aligned with the federal mandates of NCLB. For example, if the state department of education sets a school's performance goal for a specific grade and subject tested, the Maryland School Performance Report shows the goal as either being met or not. The report also indicates if the school has met Annual Yearly Progress or AYP. When a school does not meet AYP students may transfer to a higher performing school. If a school fails to meet AYP and its improvement goals for four years the state may take over the school.

The MSA test reports student scores on a scale score as well as on three levels. Students' scores are reported as basic, proficient, or advanced. Improvement goals are based upon the number of students performing proficient or advanced. The Maryland Department of Education does disaggregate the data for students on free and reduced lunch as well as school size. The MSA is not used for promotion or retention of students individually. The Maryland Department of Education does not rate schools based on their students' performance but does provide the testing results as part of the Maryland School Performance Report. The testing component is very similar to the tests used in Colorado; however, the accountability measures are different than Colorado's because each school's

performance goals are different based on their baseline testing results and longitudinal growth.

North Carolina. The North Carolina State Board of Education (SBE) developed the ABCs of Public Education in response to the School-Based Management and Accountability Program enacted by the General Assembly in June 1996 (NCSBE, 2005). The program, like Colorado's, focuses on strong accountability for school performance on high-stakes statewide tests. The program was also updated in 2003 to comply with NCLB and AYP.

Currently, North Carolina public schools administer End of Grade or EOG tests in reading and math to students in grades 3-8. Students in grades 9-12 take the appropriate End of Course or EOC tests in Algebra, geometry, chemistry, English, physical science, and physics. All tests are multiple choice and norm referenced. Scores are reported both as a comparison to the mean and comparison to expected growth using the previous year's test as a pre-test and the following year's as a post test. Students are also scored using the contrasting group method of standard setting (NCSBE, 2005). Teachers in North Carolina set the standards for performance on the end of grade and end of course tests into four levels. The target level is level III, which indicates that students consistently demonstrate mastery of knowledge for that subject and grade level.

The North Carolina School Report Cards give a rating to every school in the state based on the EOG and EOC tests administered. The rating is based on three weighted composite scores: the school's expected growth, high growth, and overall performance composite. Because the test is norm referenced, overall ratings are calculated comparing actual mean scores for the school to the state mean controlling for effect size. The overall

performance composite is reported as the percent of students achieving a level III score on the EOG and EOC tests.

The 2003-2004 North Carolina Report Card also presented disaggregated data on for all schools in the state. The report card indicates that 68% of economically disadvantaged students perform at level III or higher where as the not economically disadvantaged students perform 90% at level III or higher.

The state department of education rates all schools as: Schools of Excellence, Schools of Distinction, Schools of Progress, a Priority School, or a Low-Performing School. The department of education also rewards high growth schools and expected growth schools with direct $1,500 and $750 awards to the teachers in those schools. The sanctions for schools failing to perform with significantly less than 50% of the students scoring below level III make a school a low performing school and over time the Department of Education may take over the school.

The students in North Carolina are held accountable for the results of their high school end of course tests. The tests must be passed in order to graduate from high school. The demographic data of school size, charter designation, and performance of economically disadvantaged students are all included in the NC State Report Card.

North Carolina's ABCs of education program differs from Colorado's Student Assessment Program in many components. All tests are norm-referenced and multiple choice. Students are held accountable for their performance and teachers are directly rewarded for high student growth monetarily.

Summary of Literature Findings

In general, the literature strongly supported the hypothesis that the percent of students on free and reduced meals is a strong predictor of student achievement. The literature was not as conclusive about the relationship between school size and student achievement. It did emerge that for the highest poverty students, smaller schools seemed to lead to greater gains in achievement. Also, in general, students with lower poverty levels seemed to achieve better in larger schools. The overall conclusions about school size and achievement were not conclusive.

Charter school designation was mixed on the relationships to student achievement. Since charter schools serve a wide variety of purposes and differ greatly from state to state, the research on the impact on student achievement was just emerging from the literature. The one finding that was consistent for charter schools was that charter school size was about half the size of the average school.

The impact of high-stakes testing programs on student achievement was not conclusive in the literature. While most states that have implemented high-stakes testing programs have seen increases in student achievement over time, the same improvements have not been recognized on national assessments by the same states. The debate over statewide high-stakes accountability systems basically has two sides. One side contends that high expectations, rigorous testing and tough accountability measures will force public education to reform and service students better. The other side contends that the high-stakes testing programs have unfairly identified schools with the greatest number of poor students and labeled them as failing. Another contention is that there does not seem

to be conclusive empirical evidence that high-stakes testing has improved student achievement.

The Colorado Student Assessment Program is similar to other statewide high-stakes testing systems. The overall quality and quantity of the tests administered in Colorado seem better than most states. The Overall Academic Performance Index however is a unique formula only used in Colorado to determine school ratings and improvement ratings. Because of the consistency of the CSAP program over the past seven years Colorado is a good model for longitudinal study.

CHAPTER 3

METHODS

This chapter presents an overview of the research methodology, proposed analysis, and process details needed for replication of the study. The sample for both 2001 and 2004 are explained. The Colorado Student Assessment Program (CSAP) school index scores or Overall Academic Performance Index (OAPI) are described as the dependent variable. The attribute variables of percent of students on free and reduced lunch, school size, and charter school designation are also described. The purpose of this study was to investigate the relationship between student achievement and the percent of students on free and reduced meals, school size, and charter school designation. The researcher used the results of the Colorado Student Assessment Program in 2001 and 2004 as the dependent variable.

Research Approach

The research methodology will be shown as appropriate to answer the following research questions:

1. Is there a combination of 2001 percent of students on free and reduced meals, 2001 school size, and school designation as a charter school that predicts 2001 Colorado middle school CSAP testing performance better than 2001 percent of students on free and reduced meals alone?

2. Is there a combination of 2004 percent of students on free and reduced meals, 2004 school size, and school designation as a charter school that predicts 2004 Colorado middle school CSAP testing performance better than 2004 percent of students on free and reduced meals alone?

3. Is there a difference between 2001 and 2004 OAPI scores for Colorado middle schools? Specifically:

   a. Did mean scores change over time?

   b. Is there a correlation between 2001 and 2004 OAPI scores?

4. What factors studied predict the direction and amount of change?

   a. Does 2001 school percent on free and reduced meals predict change in OAPI (2001-2004)?

   b. Does 2001 school size predict change in OAPI (2001-2004)

   c. Does school designation as a charter school or not a charter school predict a change in OAPI (2001-2004)?

   d. Does the 2004 percent of students on free and reduced lunch predict a change in OAPI (2001-2004)?

   e. Does the 2004 in school size predict a change in OAPI (2001-2004)?

5. Is there a combination of 2001 percent of students on free and reduced lunch, 2001 school size, and school designation as a charter school that predicts change in the 2001- 2004 Colorado middle school CSAP testing performance?

6. After initial analysis of the research questions, the researcher investigated the 30 most improved schools and the 30 schools that declined the most on their OAPI

scores from 2001 to 2004. From the analysis the researcher investigated the following questions:

a. Do the top 30 and bottom 30 schools differ on percent of student on free and reduced lunch, school size, and charter designation from the 357 middle schools included in this study?
b. Do the top 30 and bottom 30 schools differ from each other on percent of student on free and reduced lunch, school size, and charter designation?

Research Design

The research design for this study was quantitative. The research questions ask what degree of relationship exists between the dependent variable, Colorado middle school (CSAP) performance (OAPI), and the school characteristics (attribute independent variables) of school percent of students on free and reduced meals, school size, and a school's designation as a charter school? The research questions also asked what degree of relationship existed between the dependent variable and the attribute independent variables over time, in 2001 and 2004.

In order to investigate if there was a combination of percent of students qualifying for free and reduced meals, school size, and a school's designation as a charter school, that predicts a school's CSAP performance better than any one-predictor variable alone, a simultaneous multiple regression analysis was used. A multiple regression analysis using the change score as the dependent variable was also appropriate in order to investigate if there was a combination of percent of students on free and reduced meals, school size, and a school's designation as a charter school that predicted change in the 2001- 2004

Colorado middle school CSAP testing performance better than any one-predictor variable alone.

Based on the research questions, a quantitative paradigm using a non-experimental approach was used. This approach allowed for the examination of relationships between attribute independent variables and a dependent variable (Gliner & Morgan, 2000). The non-experimental research approach was employed based on the types of variables investigated. There was no manipulation of the variables in this study, as the variables were attribute variables. The researcher used existing data collected by the Colorado Department of Education in 2001 and 2004, thus the study can be further described as comparative and ex-post-facto.

The advantage of the research design was that it was practical and feasible. The strength of this study was the examination of the possible influence of attribute variables on middle school CSAP performance in 2001 and 2004. The amount of time between comparison years was deliberate in the research design because the students in the school that made up the sample in 2001 were completely different than the students who made up the sample for the 2004 data; thus, theoretically, the three school attributes are more independent than a year to year comparison design. This helped the researcher in interpreting the relationship between the school attributes and student achievement.

The sample is similar to the theoretical population in many aspects. Accordingly, the study has a medium to high external validity (Gliner & Morgan, 2000). The results from this study can help policy makers evaluate current policy and make decisions related to the variables in this study.

Participants

The theoretical population for this study was sixth, seventh, eighth, and ninth grade public school students in the United States. The accessible population was all middle school students in Colorado public schools who participated in the 2001 and 2004 CSAP program. The researcher used a composite score derived from individual student scores and assigned as the Overall Academic Performance Index or (OAPI) for the entire school in both 2001 and 2004. From the 393 Colorado middle schools that participated in the CSAP, 357 were found to have accurate data in all of the variables studied. The selected school sample (N=357) included 58% with grades 6-8, 38% with grades 7-8, 3% with grades 6-7, and 2% with grades 8-9. Only schools with the same grade configuration in 2001 and 2004 were included in the study.

Data Collection

The data for this study were collected from the Colorado Department of Education for the test years of 2001 and 2004. The data for this study were post-hoc in nature. The dependent variable of the school index score was provided by the Colorado Department of Education. The attribute variables of percent of students on free and reduced meals, school size, and charter school designation were provided by the Colorado Department of Education. The data was available in Excel and *PDF* formats and then hand-entered by the researcher into SPSS for analysis.

Measures

The researcher initially collected data from 357 middle schools on school location teacher experience, teacher pay, student teacher ratios, percent of students on

free and reduced meals, school size, charter designation, and the Overall Academic Performance Index score.

Through initial analysis the researcher determined that there was multi-collinearity between teacher pay, teacher experience, and the percent of students on free and reduced meals. The researcher decided to use the percent of students on free and reduced meals variable because the literature described it as a strong proxy for poverty. The researcher also found multi-collinearity between school size, location, and student to teacher ratios. Since size was numerical and normally distributed the researcher chose to study school size. Student to teacher ratios were correlated with school size and percent of students on free and reduced meals. Student to teacher ratios were also problematic because the available data turned out to be the number of adults in the building divided by the student population; therefore, they were not exactly class size indicators. The predictor variables of percent of students on free and reduced meals (FARM), school size, and designation as a charter school emerged as the best available variables for study. The dependent variable of the Overall Academic Performance Index (OAPI) score emerged as the best available indicator of overall school achievement on the Colorado Student Assessment Program.

Free and Reduced Meals (FARM)

The variable was collected as a numerical percent of students in the school who qualified for free and reduced meals. The Colorado Department of Education collected the data from the individual schools. Individual schools used the same federal regulations to qualify students as eligible for free and reduced meals. Parent income level was the main criteria for determining a student as eligible. The reliability and validity of the data

was dependent upon the collection and reporting by the Colorado Department of Education. Since the data were audited and tied to federal funding, they were assumed to be reasonably reliable and valid.

School Size

The school size variable was collected by the Colorado Department of Education as the October 1st enrollment-count in both 2001 and 2004. The variable was used by CDE to determine school funding. The criteria that individual schools used for calculating student enrollment were consistent among schools and between the years of 2001 and 2004. Since the data were audited and tied directly to school funding, they were assumed to be reasonably reliable and valid.

Charter Designation

The school names and identification numbers for all charter schools in Colorado serving middle school students were collected from CDE. Colorado Charter School Law prescribed the criteria for being recognized as a charter school. The criteria was consistent between schools and between 2001 and 2004. Only charter schools that participated in the CSAP program in both 2001 and 2004 were included in the study. Since the data are audited and tied directly to Colorado Charter School law they were assumed to be reasonably reliable and valid.

Colorado Student Assessment Program (CSAP)

The dependent variable of student achievement used the CSAP scores from individual student results on the CSAP tests. The Colorado Student Testing Program results were the main measurement instruments for this study. The CSAP and OAPI will

be described in great detail since the use of CSAP results were the only criteria used for measuring student achievement in both 2001 and 2004.

The Colorado Student Assessment Program was created by the state legislature in 1999. Students in Colorado public schools grades 3-10 are tested in reading, writing, and math annually (CDE, 1999). For the purpose of this study, the test results used depended upon the school's grade configuration. The selected schools (N=357) included 58% with grades 6-8, 38% with grades 7-8, 3% with grades 6-7, and 2% with grades 8-9.

The CSAP assessment was developed by the Colorado Department of Education and the CTB (California Testing Bureau). The test is a standards-based assessment. All 176 Colorado School Districts have adopted the Colorado Model Content Standards as blueprints for curriculum in their district. From the Content Model Standards, CTB created a testing framework. The tests include 25% constructed response questions and 75% multiple choice. The constructed response questions are weighted and can make up 50% of the points possible on the test. From the assessment framework CTB created an item bank. From the item bank CTB pulled the actual test questions. Each year, 25% of the questions are released from the test, never to be put back in the item bank (CDE, 1999).

The CSAP tests are administered in a standardized way across the state. All test administrators must be trained and all students must access the test in the same way and at the same time within a school. Schools have a four-week window in March and April to complete the testing program.

CSAP tests are scanned into computers for scoring purposes. Constructed response questions are graded by trained readers using rubrics for evaluation. The results

are used to determine a scale score. The scale scores are then used to determine the "book mark" procedure. Colorado teachers and CTB "book mark" the cut points for the scale score that fall into the proficiency categories of unsatisfactory, partially proficient, proficient, and advanced. All students receive a proficiency rating level and a scale score on the reading, writing, and math CSAP. For the purpose of this study the researcher used the proficiency rating levels as the indicators of student achievement.

The Overall Academic Performance Index (OAPI) for each middle school was calculated by the Colorado Department of Education. The calculation procedure is described by CDE in the following six steps:

Step 1. CDE removed all excluded students from the tested group. This included students new to the district after October 1st and students who qualified to take the CSAP Alternate test. Then CDE recalculated the percentages in each of the five proficiency levels and used the following weighted factors: -.05 for Unsatisfactory and No Scores, .05 for partially proficient, 1.0 for proficient, and 1.5 for advanced. .

Step 2. CDE obtained the standardized weighted total scores for each grade level on each test. The state mean was subtracted from the district mean and then divided by the standard deviation.

Step 3. CDE calculated a weighted average of all the standardized, weighted total scores (SWTS) for each content area in 6th-9th grade, based on the number of student scores in each content area

Step 4. CDE computed an overall standardized weighted total score based on the total number of student scores. See table 3 for an example.

Step 5. CDE rank-ordered the OAPI for every public middle school in the state.

Step 6. CDE labeled the top 8% of OAPI schools with an "excellent" rating; the next 25% received a "high" rating, the next 40% received an "average" rating, the next 25% received a "Low" rating, and the bottom 2% received an "unsatisfactory" rating. Because the distributions are not re-standardized each year, it is possible for schools to improve their performance rating over time (CDE, 2005).

Table 3

Example Computation of the Overall Performance Index (OAPI) for a School

Content area	Content area student count	Z- content area	Content area x Z--content	Total student count for school	School OAPI
Reading	400	-1.27	-508.194		
Writing	400	-1.46	-585.585		
Math	269	-0.95	-255.550		
Total	1069		-1349.33	1069 =	-1.26

For this study the OAPI for 357 Middle Schools that participated in the CSAP program in 2001 and 2004 were used. The researcher did not use the sixth step of the rating label since the OAPI was available on every school and the scores were normally distributed.

The test reliability of the Colorado Student Assessment Program scores were measured each year using Cronbach's alpha. Cronbach's alpha was used as a measure of internal consistency (CDE, 2005). The split-half coefficient also was applied and adjusted using the Spearman-Brown formula. Total score reliability coefficients are all greater than .85. The reliability coefficients indicate that the CSAP has strong internal consistency and produces relatively stable scores (CDE, 2005).

Content validity of the CSAP test was ensured by correlations to the sub-content standards in test development, and then the results were tested with Differential Item Function Analyses (CDE, 2005). The standard errors of measurement indicate, based on item response theory, that scores closer to the lowest and highest possible score for a particular grade and test have higher measurement errors than those closer to the mean (CDE, 2005).

Procedure

This ex-post-facto research study involved the collection and analysis of middle school testing data and school demographic information from the Colorado Department of Education. The Colorado Student Testing Program or CSAP is a statewide school accountability program enacted in 1997 with the goal to improve public education in Colorado.

The researcher collected the OAPI or Overall Academic Performance Index score for 357 Colorado middle schools in the 2001 and 2004 school years. The researcher also collected demographic data on the 357 middle schools in the 2001 and 2004 school years. The demographic data included; the percent of students on free and reduced lunch, school size and designation as a charter school. The researcher requested data from the Colorado Department of Education data request department. The data were in both Excel and PDF format then hand entered by the researcher into the SPSS program.

Data Analysis

The data analysis proposed for this research study was based on the research questions posed in Chapter 1. Associational and comparative statistics were used to

analyze the following research questions. After each research question the statistical analysis procedures are described.

1. Is there a combination of 2001 percent of students on free and reduced meals, 2001 school size, and school designation as a charter school that predicts 2001 Colorado middle school CSAP testing performance better than 2001 free and reduced meals alone? The researcher used simultaneous multiple regression analysis.
2. Is there a combination of 2004 percent of students on free and reduced meals, 2004 school size, and school designation as a charter school that predicts 2004 Colorado middle school CSAP testing performance better than 2004 free and reduced meals alone? The researcher used simultaneous multiple regression analysis.
3. Is there a difference between 2001 and 2004 OAPI scores for Colorado middle schools? Specifically, the researcher used a paired sample *t*-test for analysis.
   a. Did mean scores change over time?
   b. Is there a correlation between 2001 and 2004 OAPI scores?
4. What factors studied predict the direction and amount of change? The researcher utilized regression analysis.
   a. Does 2001 school percent on free and reduced lunch predict change in OAPI (2001-2004)?
   b. Does 2001 school size predict change in OAPI (2001-2004)
   c. Does school designation as charter school or not predict change in OAPI (2001-2004)?

d. Does the 2004 percent of students on free and reduced meals predict a change in OAPI (2001-2004)?

e. Does the 2004 school size predict a change in OAPI (2001-2004)?

5. a. Is there a combination of 2001 percent of students on free and reduced meals, 2001 school size, and school designation as a charter school that predicts change in the 2001- 2004 Colorado Middle School CSAP testing performance?

b. Is there a combination of 2004 percent of students on free and reduced meals, 2004 school size, and school designation as a charter school that predicts change in the 2001- 2004 Colorado Middle School CSAP testing performance? The researcher used simultaneous multiple regression analysis.

6. After initial analysis of the research questions, the researcher investigated the 30 most improved schools and the 30 schools that declined the most on their OAPI scores from 2001 to 2004. From the analysis the researcher investigated the following questions:

a. Do the top 30 and bottom 30 schools differ on percent of students on free and reduced meals, school size, and charter designation from the 357 middle schools included in this study? The researcher used a one-sample *t*-test analysis.

b. Do the top 30 and bottom 30 schools differ from each other on percent of students on free and reduced lunch, school size, and charter designation? The researcher used an independent samples *t*- test analysis.

CHAPTER 4

RESULTS

One purpose of this study was to confirm the hypothesis that poverty, as indicated by the percentage of students qualifying for free and reduced meals, was a major predictor of student achievement on the middle school Colorado Student Assessment Program (CSAP) in both 2001 and 2004. This study determined whether a school's size or a school's designation as a charter or non charter would add significantly to the prediction of the middle school CSAP achievement in both 2001 and 2004.

A second purpose was to investigate whether there was a change in middle school CSAP achievement between 2001 and 2004. The study also determined if there was a relationship between the 2001 and 2004 student CSAP results.

Finally, the study investigated whether changes from 2001 to 2004 in an individual school's CSAP performance can be predicted from the school characteristics of percent of students qualifying for free and reduced meals, schools size, and charter designation. Additionally, the researcher examined the thirty middle schools that improved the most from the 2001-2004 CSAP and the thirty middle schools that declined the most to investigate their relationship to the original 357 middle schools. By considering if a relationship exists between school characteristics and school improvement, then school improvement efforts, best teaching practices, and public policy can benefit from research to practice.

Description of the Sample

The sample was composed of 357 Colorado public middle schools. Included in the sample were the schools that received an overall academic performance (OAPI) rating in both 2001 and 2004. Excluded from the sample were schools with invalid free and reduced lunch data and school size data. Schools that did not offer a lunch program in both 2001 and 2004 and schools that had changed their grade configuration from 2001 to 2004 were not used in answering the research questions. Because schools were identified and compared by the Colorado Department of Education school code, schools that had changed codes were also excluded from the study. Schools change codes when re-configuring grades that a school serves or when closing one school and opening another due to construction or demographic changes.

Overall Academic Performance Index (OAPI)

The Overall Academic Performance Index for each middle school was calculated by the Colorado Department of Education for 2001 and 2004. The calculation procedure was described by CDE in the following six steps:

Step 1. CDE Removed all excluded students from the tested group. This includes students new to the district after October 1st and students who qualified to take the CSAP Alternate test. Then CDE calculated the percentages in each of the five proficiency levels and used the following weighting factors: -.05 for Unsatisfactory and No Scores, .05 for partially proficient, 1.0 for proficient, 1.5 for advanced.

Step 2. CDE obtained the standardized weighted total scores for each grade level on each test. The state mean was subtracted from the district mean and divided by the standard deviation.

Step 3. CDE calculated a weighted average of all the standardized, weighted total scores (SWTS) for each content area in 6^{th}-9^{th} grade, based on the number of student scores.

Step 4. CDE computed an overall standardized weighted total score based on the total number of student scores.

Step 5. CDE rank-ordered the OAPI based on the SWTS for every public middle school in the state.

Step 6. CDE labeled the top 8% of OAPI schools with an "excellent" rating, the next 25% received a "high" rating, the next 40% received an "average" rating, and the next 25% received a "Low" rating, and the bottom 2% received an "unsatisfactory" rating. Because the distributions were not re-standardized each year it is possible for schools to improve their performance rating over time (CDE, 2003).

The overall academic performance index scores for 357 middle schools that participated in the CSAP program in 2001 and 2004 were used as the dependent or outcome variable for this study. The 2001 range of scores was from –2.96 to 2.28 with a mean of .018. The 2004 range of scores was from –2.67 to 2.51 with a mean of .086.

Charter Schools and Non-charter Schools

Charter schools are public schools that are chartered by the local school district. Charter institutes are schools chartered by the state board of education. This study only included charter schools since charter institutes were invoked after 2001. Charter schools

in Colorado are typically organized around a particular curriculum approach and instructional approach. Charter schools operate with public tax dollars flowing though the local school district to the charter school, and they are open to all students in a district's attendance boundary. Of the 357 middle schools studied, 48 schools were identified by the Colorado Department of Education as charter schools serving middle school students. Included in this study were charter schools that received an OAPI score in both 2001 and 2004. 86.6% of the schools were non-charter while 13.4% were charter schools.

Free and Reduced Meals (FARM)

Each of 357 Colorado middle schools reported to the Colorado Department of Education the percent of students in the school qualifying for free and reduced meals in both 2001 and 2004. In 2001 the range was from 0 to 97% with a mean of 31.43% and standard deviation of 23.37. In 2004 the range was from 0 to 95.4% with a mean of 36.56% and standard deviation of 24.54. Each of the 357 schools offered some form of lunch program. The researcher used the percent of free and reduced meals as a proxy for poverty.

School Size

The sample included 357 middle schools with student populations in 6^{th}-9^{th} grades. School configurations varied from 6-8, 7-9, 7-8 & 6-7. The sample included only those schools with the same grade configurations from 2001 and 2004. The range of school size in 2001 was from 24 to 1,531 with a mean of 464.8 and a standard deviation of 327.77. In 2004 the range of school size was from 28 to 1,576 with a mean of 452.45 and with a standard deviation of 315.11.

Summary of Sample Descriptions

The sample included 357 public Colorado middle schools. Schools that did not have data available on their OPAI, size, charter status or percent of students on free and reduced lunch were not included in the sample. Schools that reorganized their grade configurations from 2001-2004 were also not included in the sample. The descriptive statistics are presented in Table 4.

Table 4

Means and Standard Deviations on the Variables of OAPI, Charter Designation, Free and Reduced Meal (FARM) and School Size in 2001 and 2004

Variable	N	Minimum	Maximum	M	*SD*
OAPI01	357	-2.96	2.28	.017	.88
OAPI04	357	-2.67	2.51	.087	.89
Charter	357	0	1	.13	.34
FARM01	357	0	97.00	31.44	23.40
FARM04	357	0	95.40	36.58	24.55
Size01	357	24	1531.00	464.55	327.99
Size04	357	28	1576.00	453.52	314.91

Examination of the Research Questions

Research Question 1

The first research question examined the data from 2001. Is there a combination of percent of students on free and reduced meals, school size, and school designation as a charter school or not that predicts 2001 Colorado middle school CSAP testing performance better than the 2001 percent of students on free and reduced meals alone? Simultaneous multiple regression was conducted to investigate the best predictors of 2001 Colorado Middle School CSAP achievement. The means and standard deviations

can be found in Table 4. The intercorrelations for CSAP achievement and predictor variables are presented in Table 5.

The percentage of students qualifying for free and reduced meals in 2001 was found to be negatively correlated with the school's 2001 overall academic performance index (OAPI) score. With an N=357 middle schools, the correlation coefficient was - .77 with a significance level of p<.01. According to Gliner and Morgan (2000) this is a much greater than typical effect in social science research. The 2001 percentage of students on free and reduced meals (FARM01) variable was a strong predictor of school performance on CSAP performance in 2001. Generally, a school with a low percentage of students on free and reduced lunch scored higher on the CSAP in 2001.

Table 5

Intercorrelations for CSAP Achievement (OAPI) and Predictor Variables in 2001

	Predictor Variables		
Variable	FARM01	School Size01	Charter01
OAPI01	-.77***	-.05	.18**
FARM01	--	.07	-.27**
School Size 01		--	-.23**

*p < .05; **p < .01, ***p<.001

Note. FARM01 = Percentage of students on free and reduced meals in 2001
OAPI = Overall academic performance index score

School size in 2001 was not correlated with the 2001 overall academic performance index (OAPI) scores with a correlation coefficient of -.05. A school's designation as a charter school in 2001 showed a small but significant correlation with

2001 overall academic performance index score (OAPI). The correlation coefficient was .18 with a $p<.01$. Gliner and Morgan (2000) describes this as a small effect.

When the combination of variables to predict CSAP performance in 2001 included free and reduced meals (FARM01), school size in 2001, and charter designation, $F(3,354)=187.1$, $p<.001$, indicating that in combination these three variables are highly significant predictors of overall 2001 CSAP achievement. The beta coefficients are presented in Table 6. Charter designation in 2001 when combined with school size in 2001 and percentage of students on free and reduced meals in 2001 (FARM01) was not a significant predictor with a beta of -.06. Note that both free and reduced meal percentage in 2001 (FARM01) and school size in 2001 significantly predict CSAP achievement when all three variables are included. The FARM01 beta was -.80 and School Size in 2001 beta was -.12, both significant at the $p<.01$ level. Generally, schools with a low percentage of students on free and reduced lunch along with a small school size scored higher on the 2001 CSAP. The adjusted R squared value was .61. This indicates that 61% of the variance in CSAP achievement was explained by the model. According to Gliner and Morgan (2000), this is a very large effect. Note that charter designation did not add significantly to the prediction even though there was a significant bivariate relationship between charter designation and CSAP performance. Conversely, size did add significantly to the prediction with poverty even though there was not a significant bivariate correlation of size and CSAP performance. Note also that the beta for percentage of students on free and reduced meals is at least as large (-.80) as the bivariate correlation (-.77), indicating that it is a very powerful predictor even in combination with other variables.

Research Question 2

The second research question examined the data from 2004. Is there a combination of percent of students on free and reduced meals, school size, and a school's designation as a charter school or not that predicts 2004 Colorado Middle School CSAP testing performance better than the 2004 percentage on free and reduced meals alone? Simultaneous multiple regression was conducted to investigate the best predictors of 2004 Colorado middle school CSAP achievement. The means and standard deviations can be found in Table 4. The intercorrelations for CSAP achievement and predictor variables are presented in Table 7.

Table 6

Simultaneous Multiple Regression Analysis for 2001 Percentage on Free and Reduced Meals (FARM), School Size, and Charter Designation Predicting CSAP Achievement (OAPI)(N=357)

Variable	*B*	*SEB*	*B*
FARM01	-.03	.01	-.80***
School Size in 2001	-.00	.00	-.12**
Charter	-.17	.09	-.06

*$p<.01$; **$p<.001$, ***$p<.001$

Note. $F(3,353)=186.39$, $p<.001$. Adjusted $R^2 = .61$
FARM01 = Percentage of students on free and reduced meals in 2001
OAPI = Overall academic performance index score

The percentage of students qualifying for free and reduced meals in 2004 was negatively correlated with the schools 2004 OAPI score. With an N=357 middle schools, the correlation coefficient was - .75 with a $p<.01$. According to Gliner and Morgan (2000) this is a much greater than typical effect in social science research. The percent of students qualifying for free and reduced meals in 2004 was a strong predictor of school

performance on CSAP performance in 2004. Generally, a school with a low percentage of students on free and reduced meals scored higher on the CSAP.

School size in 2004 was not correlated with the 2004 OAPI scores with a correlation coefficient of .02. A school's designation as a charter school in 2004 showed a small but significant correlation with 2004 OAPI. The correlation coefficient was .17 with a $p<.01$. Gliner and Morgan (2000) describes this as a small effect.

When the combination of variables to predict CSAP performance in 2004 included the free and reduced meals (FARM) in 2004, school size in 2004, and charter designation in 2004, $F(3,354)=157.2$, $p<.001$, indicating that this combination of three variables was a highly significant predictor of overall 2004 CSAP achievement. The beta coefficients are presented in Table 8. Both school size in 2004 and charter designation in 2004 when combined with the free and reduced meal variable were significant predictors, with betas of -.10 and -.08 respectively. Note that the percent of students on free and reduced meals (FARM04) was a highly significant predictor of 2004 middle school CSAP achievement when all three variables were included. The FARM04 beta was -.78 and is significant at the $p<.01$ level. Schools that were small, non-charter with a low percentage of students on free and reduced lunch scored higher on the 2004 CSAP. The adjusted R squared value was .57. This indicates that 57% of the variance in CSAP achievement was explained by the model. According to Gliner and Morgan (2000), this is a very large effect. Note that even though the bivariate correlation of charter designation and CSAP was positive (charter performed better), in combination with poverty and school size, non-charter status was predictive of higher performance.

Table 7

Intercorrelations for CSAP Achievement (OAPI) and Predictor Variables in 2004

	Predictor Variables		
Variable	FARM04	School Size04	Charter04
OAPI04	-.75***	-.02	.17**
FARM04	--	-.12*	-.29**
Size04		--	-.20**

*$p < .05$; **$p < .01$, *** $p<.001$

Table 8

Simultaneous Multiple Regression Analysis for Percentage on Free and Reduced Meals (FARM)2004, School Size04, and Charter Designation Predicting 2004 CSAP Achievement (OAPI)(N=357)

Variable	*B*	*SEB*	*B*
FARM04	-.03	.001	-.78***
School Size in 2004	-.00	.000	-.10**
Charter	-.20	.098	-.08*

*$p<.05$; **$p<.001$, ***$p<.001$
Note. $F(3,354)=158.1$, $p<.001$. Adjusted $R^2 = .57$

Research Question 3

The third research question examined the OAPI index scores from 357 Colorado middle schools in 2001 and 2004. The researcher examined whether there was a difference between 2001 and 2004 OAPI scores, whether mean scores changed over time, and whether there was correlation between 2001 index scores and 2004 index scores. The mean scores are reported in Table 4 with a 2001 mean score of .018 and a 2004 mean score of .086. A paired sample *t*-test indicated that overall OAPI scores improved from 2001 to 2004. With a t (357)=2.89, p=.004, d=.15, the difference, although statistically

significant was small using Gliner and Morgan's (2000) guidelines. The paired sample correlations found a correlation of r =.88 between 2001 and 2004 OAPI scores. This indicated that, the relative ranking of the school was similar in both years. The 2001 scores were highly correlated with 2004 scores with approximately 80% of the variance accounted for by the 2001 score.

Research Question 4

The fourth research question examined what factors studied predicted the direction and amount of change in CSAP scores between 2001 and 2004. Using bivariate regression analysis, the results indicated the percent of (FARM) , size, and charter designation do not significantly correlate with a school's change in CSAP scores. Table 9 illustrates the findings. Thus, a change in CSAP scores cannot be predicted from school size in 2001 or 2004, percent of students on free and reduced meals in 2001 or 2004, or charter designation. Note in Table 9 the very high correlations between the two free and reduced meal predictors and the two school size measures, indicating strong consistency in these measures.

Research Question 5

The fifth research question examined if there was a combination of school size, percent of students on free and reduced lunch, and a school's designation as a charter school or not that predicts a change in the 2001- 2004 Colorado Middle School CSAP testing performance better than any one predictor alone. The researcher conducted two simultaneous regression analyses on 357 Colorado middle schools that received an OAPI index score in both 2001 and 2004. The researcher found none of the individual variables

combined to predict the change score from 2001 to 2004. Tables 10 and 11 illustrate the regression analyses.

Table 9

Intercorrelations for CSAP (OAPI) Change from 2001-2004 and Predictor Variables in 2001, 2004

				Predictor Variables	
Variable	FARM01	School Size01	FARM 04	School Size 04	Charter
OAPI's change 01-04	.08	-.01	.05	.00	-.03
FARM 01	--	-.074	.96**	-.15*	-.27**
School Size01		--	-.05	.97**	-.23**
FARM 04			--	-.12*	-.29**
School Size 04				--	-.20**

*$p < .05$; **$p < .01$

Table 10

Simultaneous Multiple Regression Analysis for Percentage on Free and Reduced Meals (FARM)2001, School Size01, and Charter Designation Predicting 2001-2004 CSAP Change (OAPI)(N=357)

Variable	*B*	*SEB*	*B*
FARM01	.001	.001	.079
School Size01	-.000	.000	.009
Charter	-.010	.074	-.008

Note. $F(3,353)=.808$, $p>.49$

Table 11

Simultaneous Multiple Regression Analysis for Percentage on Free and Reduced Meals (FARM)2004, School Size04, and Charter Designation Predicting 2001-2004 CSAP Change [OAPI]

Variable	*B*	*SEB*	*B*
FARM04	.001	.001	.044
School Size04	-.000	.000	-.004
Charter	-.003	.074	-.015

Note. $F(3,353)=.308$, $p>.820$

Research Question 6

The sixth research question investigated the data for patterns of school improvement or decline. To analyze this statistically, the researcher selected the top 30 schools that had shown the most improvement and the bottom 30 schools that had shown the greatest decline in their Overall Academic Performance Index (OAPI) score from 2001 to 2004 to see what predictor variable and change variable means differed between these groups as well as from the original 357 norm group.

Table 12 illustrates a comparison of means for the thirty greatest gaining schools and the thirty greatest declining schools with the overall means of the 357 middle schools on the predictor variables of free and reduced meals, school size, charter designation, and change scores from 2001 to 2004, using 14 one- sample *t*-tests. Table 13 shows the comparison of the 30 most improved and 30 most declined schools using independent samples *t*-tests. .

Table 12 shows that the means of the top 30 and bottom 30 schools do differ from the means of all the middle schools in the original sample (N=357). The greatest difference is that the mean school size of the top and bottom 30 schools in 2004 is smaller (244.36 and 275.86) compared to a mean of 453.52 of all 357 middle schools. The researcher used a one sample *t*- test to compare the means. The means of school size in 2001 were significantly smaller for both the top 30 (p<.001), and for the bottom 30 (p=.002). In 2004 the same pattern was found. Also significant was the pattern that the top 30 schools differed from the N=357 mean in the number of students qualifying for free and reduced meals.

Interestingly, the top schools had a somewhat higher percent than the original 357, 42.79% compared to 31.44% in 2001 and 48.33% compared to 36.57% in 04. In 2001 p<.010 and in 2004 it was p<.006. The possible explanations for this finding are discussed further in Chapter 5. No other significant differences were found using the one sample *t*-test.

Table 12

Comparison of Means for the Top 30 Improving and Bottom 30 Declining Schools With the Overall Means for the 357 Colorado Middle Schools on the Predictor and Change Variables

	Predictor Variables				
Variable	*M*	*SD*	*T*	*df*	*p*
Size01 *N*=357	465.55	327.99			
Top 30	248.56	245.54	-4.82[a]	29	<.001*
Bottom 30	262.90	280.27	-3.40[b]	29	.002*
Size04 *N*=357	453.52	314.91			
Top 30	244.40	235.89	-4.84	29	<.001*
Bottom 30	275.90	281.86	-3.44	29	.002*
SizeCh. *N*=357	-12.1	78.14			
Top 30	-4.2	37.65	1.17	29	.240
Bottom 30	-14.7	63.60	-.210	29	.835
Charter *N*=357	13%	.34			
Top 30	16%	.38	.53	29	.600
Bottom 30	13%	.35	.053	29	.958
Farm01 *N*=357	31.44	23.40			
Top 30	42.79	22.38	2.77	29	.010*
Bottom 30	33.22	20.64	.472	29	.640
Farm04 *N*=357	36.57	36.57			
Top 30	48.33	21.80	2.90	29	.006*
Bottom 30	39.83	20.95	.854	29	.400
FarmCh. *N*=357	5.12	6.80			
Top 30	5.54	5.55	1.17	29	.248
Bottom 30	6.42	6.61	1.34	29	.188

*Note: *p<.05, [a]Top 30 schools in a one-sample t-test compared to the original 357.*
[b]Bottom 30 schools in a one-sample t-test compared to the original 357.

Table 13 shows that none of the comparisons between the top 30 and bottom 30 schools were significantly different. The comparison of means for the percent of students on free and reduced meals in 2001 came the closest to being significant. The top 30 (42.33%) was somewhat higher than the (33.49%) of the bottom 30 schools; however, the difference was not statistically significant (p=.09).

Supplemental Research Question

In the initial analysis, a pattern of student enrollment loss was noted by the researcher between 2001 and 2004. The original sample of 357 Colorado middle schools in 2001 and 2004 excluded 32 schools because their characteristics of grade configuration had changed or the school had opened after 2001. The 32 excluded schools were included for an analysis of 2004 data only to determine whether their exclusion significantly impacted the second research question illustrated in Table 8. The comparison of the original N=357 for 2004 and N=389 is shown in Table 14.

The comparison of the 2004 data with an N=357 and N=389 in the correlations and multiple regressions did not show much difference, adjusted R^2 =.57 vs. R^2 =.51 Of the 32 schools included in the N=389 the majority (17) were charter schools. The researcher will discuss the implications of the increase in the number of charter schools and their performance distribution in chapter 5.

Table 13

Comparison of 30 Most Improved Schools and 30 Schools that Declined the Most on School Size in 01 and 04 and Change in School Size, Charter Designation and Percent of Students on Free and Reduced Meals in 01 and 04 and Change

Variable	*M*	*SD*	*t*	*df*	*p*
School Size 01			-.62	58	.54
Top 30	248.56	245.54			
Bottom 30	262.90	280.27			
School Size 04			-.47[a]	56.25[a]	.64
Top 30	244.40	235.89			
Bottom 30	275.90	281.86			
Schl Size Chg.			.78	58	.44
Top 30	-4.2	37.65			
Bottom 30	-14.7	63.60			
Charter			.36	58	.72
Top 30	16%	.38			
Bottom 30	13%	.35			
FARM01			1.72	58	.09
Top 30	42.79	22.38			
Bottom 30	33.22	20.64			
FARM04			1.54	58	.13
Top 30	48.33	21.80			
Bottom 30	39.83	20.95			
FARM Change			-.55	58	.58
Top 30	5.54	5.55			
Bottom 30	6.42	6.61			

Note. [a] The *t* and *df* were adjusted because variances were not equal.

Table 14

Simultaneous Multiple Regression Analysis and Comparison for Percentage on Free and Reduced Meals (FARM)2004, School Size04, and Charter Designation Predicting 2004 CSAP Achievement (OAPI)(N=357)and (N=389)

Variable	*B*	*SEB*	*B*
FARM04 N=357	-.03	.001	-.78***
FARM04 N=389	-.03	.001	-.75***
Size 04 N=357	-.03	.000	-.10*
Size 04 N=389	-.03	.000	-.09*
Charter N=48	-.20	.099	-.08*
Charter N=65	-.30	.090	-.11*

Note. *(N= 357) $F(3,354)=157.1$, $p<.001$. Adjusted $R^2 = .57$
**(N=389) $F(3,385)=137.3$, $p<.001$. Adjusted $R^2 = .51$
*$p<.01$; **$p<.001$, ***$p<.001$

Summary of Research Findings

First of all, through the research model, the findings confirmed the hypothesis that poverty, as indicated by the percent of students qualifying for free and reduced meals, was a very strong predictor of student achievement on the Colorado Student Assessment Program in both 2001 and 2004. In 2001 the free and reduce meal variable negatively correlated r= .-77 with CSAP achievement. In 2004 the free and reduced meal variable negatively correlated r=-.75 with CSAP achievement. In both years, the higher the percent of students on free and reduced meals the lower the CSAP achievement scores.

Secondly, the researcher determined that in 2001 the combination of small school size (beta -.12) and low FARM01 (beta-.80) significantly predicted higher CSAP achievement when all three variables were included. Charter designation did not add

statistically to the 2001 prediction model. However there was a significant positive bivariate relation between charter designation and CSAP performance. Note also that the beta for percentage of students on free and reduced meals is at least as large (-.80) as the bivariate correlation (-.77), indicating that it is a very powerful predictor even in combination with other variables.

The researcher determined that in 2004 school size was not correlated with the 2004 CSAP performance with a correlation coefficient of .02. A school's designation as a charter school in 2004 showed a small but significant positive correlation with CSAP performance.

Both school size in 2004 and charter designation in 2004 when combined with the free and reduced meal variable were significant negative predictors, with betas of -.10 and -.08 respectively. Note that the percent of students on free and reduced meals (FARM04) was a highly significant predictor (beta=-.78). The combination of low poverty, small size, and *not* charter predicted higher CSAP performance. Table 15 illustrates the findings of the correlations and betas for all three predictor variables in both 2001 and 2004.

Third, the researcher determined that there was a change in middle school CSAP achievement between 2001 and 2004. The study also determined there was a relationship between the 2001 and 2004 student CSAP results. A paired sample *t*- test indicated that overall OAPI scores improved from 2001 to 2004. The difference, although statistically significant, was small using Cohen's (1988) guidelines. The paired sample correlations found an r =.88 between 2001 and 2004 OAPI scores. This indicated that schools that

scored low were very likely to have scored low in 2004 and the schools that scored high in 2001 scored high in 2004.

Table 15

Summary Findings of FARM, School Size and Charter Designation, with OAPI in 01, 04 and the Change OAPI (N=357)

	01 OAPI correlation	01 OAPI Beta	04 OAPI correlation	04 OAPI Beta	OAPI Change Correlations	OAPI Change Beta
FARM	VL-	VL-	VL-	VL-	0	0
Size	0	S-	0	S-	0	0
Charter	SM/ME +	0	SM/ME +	S-	0	0

Note. *(0)= no significant statistical relationship*
(VL-) = very large negative effect
(S-)= small negative effect
(SM/ME+)= small to medium positive effect

Fourth, the researcher investigated whether changes from 2001 to 2004 CSAP performance could be predicted from the school characteristics. Using correlation and multiple regression analyses, the results indicated that the percent of students on free and reduced meals, school size, and charter designation were not significantly correlated with a school's change in CSAP scores and the combination of these variables was not a significant predictor. This is depicted in Table 15 as well.

Lastly, the researcher investigated the data for patterns in the characteristics of the 30 most improved and the 30 schools that declined the most. None of the characteristics of school size, change in size, free and reduced meals, change in free and reduced meals or charter designation were statistically significant between the top 30 and bottom 30. However in 2004 the mean school size of the original *N*=357 was 453.52 while the top 30 was 244.40 and the bottom 30 was 275.90. Generally, the schools that experienced the

greatest gain and decline in CSAP achievement were only about half the size of the total 357 schools. Statistical reasons of this finding will be discussed in chapter 5. Interestingly, the top schools had a somewhat higher percent of students on free and reduced meals than the total 357, 42.79% compared to 31.44% in 2001 and 48.33% compared to 36.57% in 2004. The possible explanations for this finding are discussed further in Chapter 5.

Additionally, the researcher included a comparison and analysis of the 2004 data with 32 new or reorganized schools added. The researcher determined that the exclusion of the 32 schools did not make much of a difference in the correlations and multiple regressions even though 17 of the 32 new schools were charter schools.

CHAPTER 5

DISCUSSION

Overview of the Study

High-stakes testing and public school accountability systems have increased dramatically over the past ten years. With the inception of the federal legislation "No Child Left Behind" in 2002, all states were required to implement high-stakes testing and hold schools accountable for student learning. However, since most high-stakes testing systems have been implemented by state legislatures, testing programs differ among states and have changed from year to year. Because of the constant changing, the ability to empirically study the high-stakes testing systems over time is fairly limited. The literature is limited in regards to longitudinal empirical studies of high-stakes, statewide accountability systems.

In this research study the researcher investigated the relationship between student achievement and various demographic variables that were reported on the School Accountability Reports or SAR's in Colorado. The researcher used the Overall Academic Performance Index (OAPI) score for 357 middle schools in both 2001 and 2004 as the dependent variable. The OAPI was based on individual student results on the high-stakes statewide testing program, CSAP. This ex-post-facto research study involved 357 Colorado middle schools that had the same grade configuration in both 2001 and 2004. The student achievement and demographic variables were consistent, measured the same ways in 2001 and 2004.

The researcher initially collected data from 357 middle schools on school location, teacher experience, teacher pay, student to teacher ratios, percent of students on free and reduced meals, school size, charter designation and the Overall Academic Performance Index score. Through initial analysis the researcher determined that there was multi-collinearity between teacher pay, teacher experience, and the percent of students on free and reduced meals. The researcher then decided to use the percent of students on free and reduced meal variable because the literature described it as a strong proxy for poverty. The researcher also found multi-collinearity between school size, location, and student to teacher ratios. Since size was numerical and normally distributed, the researcher chose to study school size. Student teacher ratios were correlated with school size and percent of students on free and reduced meals. Student teacher ratios were also found to be problematic because the data was actually the number of adults in the building divided by the student population; therefore, the data was not exactly class size indicators. The predictor variables of percent of students on free and reduced meals (FARM), school size, and designation as a charter school emerged as the best available variables for study. The dependent variable of the Overall Academic Performance Index (OAPI) score emerged as the best available indicator of overall school achievement on the Colorado Student Assessment Program. Because of the multi-collinearity and data validity problems, the researcher narrowed the focus of the study to the following purpose.

The purpose of this study was to investigate the results of middle school student performance on the Colorado Student Assessment Program (CSAP) in both 2001 and 2004 and to see if student performance could be predicted based on the percent of

students qualifying for free and reduced meals (FARM), school size, and designation as a charter school. The free and reduced meal variable was used as a proxy for poverty. The purpose was also to investigate if there was a relationship between a school's change in performance and the percent of students on free and reduced meals, school size, and designation as a charter school. After initial analysis, the researcher investigated the 30 most improved and the 30 schools that declined the most on their CSAP scores from 2001 to 2004 for patterns of predictability.

Data from both 2001 and 2004 were collected by the researcher and checked for accuracy against databases at the Colorado Department of Education. Six general research questions were developed by the researcher to investigate the relationship between student achievement and school poverty, school size, and charter designation. The data analyses included descriptive statistical methods, *t*-test analyses, Pearson product-moment correlations, multiple regressions, and bivariate regressions.

Overview of the Findings and Relationship to the Literature

The data analyses of this study were based on six general research questions. Five questions were identified prior to the complete analysis of data, and one additional question was posed after the initial data analysis.

The five original questions focused on the relationship between the percent of students qualifying for free and reduced meals, school size, designation as a charter school and Colorado middle school CSAP performance in 2001, 2004, and the changes in performance between years. Research questions one and two looked at the relationship between the predictor variables and the outcome variable (OAPI) in 2001 and 2004, respectively. Research question three looked at the difference in OAPI scores from 2001

and 2004. Research questions four and five examined the relationship between the attribute variables (FARM, school size, and charter designation) in 2001 and 2004 and the outcome variable of the school's change in performance on CSAP between 2001 and 2004.

The additional research questions focused on looking for patterns of relationships in the data. Research question six analyzed the 30 most improved schools and 30 schools that declined the most on the OAPI scores from 2001 and 2004. The analyses examined whether the three predictor characteristics were different between the top and bottom 30 and whether there were differences between these groups of 30 and the original sample of 357 schools.

Prediction of CSAP Performance in 2001 and 2004

Poverty. The major purpose of this study was to confirm the hypothesis that a strong relationship exists between CSAP performance and poverty. Using the percent of students on free and reduced meals as a proxy for poverty and the OAPI as the outcome variable, the research confirmed earlier studies in the literature on statewide high stakes testing. Guskey (1997) found for middle schools "the correlations between socio-economic indices and accountability index scores were negative and large in magnitude (r= -.60 to -.96)" (p. 10). In another study by the National Assessment of Educational Progress (NAEP), a national comparison of student achievement, the family characteristic variables accounted for about 75 percent of the variance in average state achievement scores (Grissmer et al., 2000). In addition, an empirical study of Alabama's high-stakes statewide assessment program found that the percent of students on free and reduced lunch was the strongest predictor of student achievement. Miller-Whitehead (2001) found

that the percent of students on free and reduced meals along with school size, daily attendance rates, drop out rates, teacher education, and per pupil expenditure combined to have *r*-'s ranging from .76 to .81 and adjusted *r* squared from .57 to .64.

The results of this study confirm the strong relationship that exists between poverty and student achievement in the literature and first brought to light in the 1966 Coleman report which concluded that:

> Schools bring little influence to bear on a child's achievement that is independent of his background and general social context...this very lack of independent effect means that inequalities imposed on children by their home, neighborhood and peer environment are carried along to become the inequalities with which they confront adult life at the end of school. (Lissitz & Schafer, 2002 p. 52).

For Colorado middle schools in 2001, the free and reduced meal variable was negatively correlated $r = -.77$ with CSAP achievement. In 2004, the free and reduced meal variable was negatively correlated $r = -.75$ with CSAP achievement. According to Cohen (1988) these are a much greater than typical effect sizes in social science research.

School size. Another purpose of this study was to confirm the hypothesis that school size was a predictor of student achievement on statewide high-stakes tests. One empirical study in the literature reported that after controlling for socio-economic status, school size and average daily attendance were the two strongest predictors of student achievement (Miller-Whitehead, 2001). Howley and Bickel (2000) conducted another study analyzing the relationship between school-level academic achievement, school size, and community poverty levels in Georgia, Ohio, Montana, and Texas. The empirical study found that the correlation of poverty and low achievement is almost ten times stronger in large schools than it is in small schools. In another study in Arkansas it was found that the achievement gap between more affluent and less affluent students was

narrowed in smaller schools and widened in larger schools (Johnson, Howley & Howley, 2002).

Although the Pearson correlations of school size and CSAP in this study were essentially zero in 2001 and 2004, using multiple regression analysis, the researcher found that smaller school size added to the prediction of CSAP performance from poverty in both 2001 and 2004. Thus, like the studies cited above, there is some evidence that school size in combination, but not by itself, is a predictor of student achievement on the Colorado middle school CSAP.

The researcher did, however, find that the schools that had shown the greatest gain and the greatest loss in CSAP performance compared to all schools were smaller on average. This may be because the mean of a small group is easier to change than the mean of a large group. This is discussed in the Recommendations for Research section of this chapter.

School designation as a charter school. A third purpose of this study was to determine the relationship between student achievement and charter schools. The literature on charter schools indicates that student achievement in charter schools compared to student achievement in non-charter schools varied by state. Charter schools in Texas, Pennsylvania, Minnesota, Michigan, and Florida scored approximately one standard deviation or more below the state means. However, Loveless (2001) found that charter students in Arizona and California were achieving comparable to students in non-charter schools. Only in Colorado were charter school test results significantly above the average of non-charter schools. (Loveless, 2001). A meta-analysis of 15 studies on charter school achievement found overall results to be mixed or very slightly positive

(Miron & Nelson, 2001). Fitzgerald, (2000) found in Colorado that in 1998-99 charter schools had performed better than the state average and their authorizing districts. Fitzgerald, (2000) also concluded that charter schools in Colorado performed better than other public schools with similar socio-economic levels. One finding in the literature was consistent in that charter schools tended to be smaller in size than non-charter schools (Nelson et al., 2000).

In the current study, designation as a charter middle school in 2001 and 2004 had a small but significant positive correlation with the 2001 and 2004 overall academic performance index score (OAPI). Thus, in 2001 and 2004 the average charter school scored higher on the CSAP than the average non-charter schools. Also, it was found that charter middle schools in Colorado tended to be smaller than the average school, 278 compared to 494, and had fewer students on free and reduced meals 15.4% compared to 33.9%.

However, the multiple regression analyses indicated that charter designation was not a significant contributor to the prediction of performance in 2001 but made a small *negative* contribution to the prediction of performance, on CSAP in 2004 in combination with the FARM and school size variables.

Prediction of Change in Student Achievement Over Time

Another purpose of this study was to investigate the relationship between school characteristics and student achievement over time. The literature findings for student achievement over time on statewide high-stakes tests were very limited because the statewide testing programs have changed dramatically in most states. Guskey (1997) found the FARM variable the most parsimonious model for predicting accountability

index scores; however, he found no relationship with gain scores over the three years studied.

Another empirical study compared California statewide high-stakes tests from 1999 to 2000. Slovacek et al. (2001) found that in California charter schools serving students with 75 percent or more students on free and reduced lunch, student achievement improved 28.1 percent while similar non-charters improved 23.8 percent. They concluded that high poverty students improved their performance more in a charter school setting than in a non- charter school setting. A longitudinal study of Colorado fourth grade reading results on the CSAP from 1997-2004 in twelve of the largest districts showed five districts had made reductions in the achievement gap between high poverty and low poverty students and seven had not. (Mitchell & Hubbard, 2005). The findings on the relationship between student achievement over time and poverty, school size and charter designation were not conclusive in the literature. Hertert et al. (2003) explained that, "the findings from much of this work are mixed and provide little guidance in designing discrete solutions with reliable and predictable results. To the extent conclusions have been drawn, they indicated that no silver bullet exists" (p.4).

The researcher found that in the model studied there was not a relationship between poverty, school size, or charter designation, and student achievement over time. While all middle schools in Colorado improved their OAPI scores slightly from 2001-2004, the school scores were highly correlated with each other from 2001 to 2004.

Recommendations for Practice

On the basis of findings from this study, the researcher discusses the recommendations for improving the Colorado Student Assessment Program with the

findings on poverty, school size, charter designation, and change in student achievement. By examining the results of the research and implications of this study, the researcher intends to add to the body of knowledge on the extremely complex issue of improving student achievement with state-wide high-stakes testing.

CSAP and poverty. State policy makers have responded to the need for greater accountability in the public education system from the landmark federal legislation of "No child left behind" in 2001. NCLB mandates all states to implement statewide high-stakes testing systems to hold schools accountable for student learning and take corrective action when students do not learn. The Colorado School Accountability Program was designed as a measuring stick to hold schools accountable for all students' learning. By using the same stick for measuring all public schools, the results have been fairly one-dimensional. From a research perspective, the Colorado model is making a false assumption that each of the 176 school districts in the state serves a similar student population. Over the past four years, the average scores for middle schools have improved slightly. What we don't know from the accountability model is whether that improvement is because of increased student learning, or accountability measure design issues.

On the surface, the political rhetoric of both NCLB and CSAP sound good as accountability models to ensure that no children are being left behind. Below the surface, this study helped confirm the literature findings that a school's poverty level is a significant predictor of how well the school will perform on state-wide student achievement tests. The greater than typical finding, was the extremely high correlation of

low test scores with high poverty in both 2001 and 2004. This finding was not surprising based on previous studies in the literature.

The use of high-stakes tests in Colorado to create the school accountability reports, have confirmed that the higher the average poverty level in a school, the lower the student achievement. While these accountability measures were intended to leave no children behind, the accountability reports may have simply identified schools with the most disadvantaged students. Currently, one Denver Public Middle School is being converted to a charter school for failing to improve, and one high school is being closed for failing to improve on the CSAP. Before state policy makers consider making school changes mandated by the CSAP program, this researcher would recommend that using one measuring stick for all schools may not the best indicator of student achievement.

Because of the researcher's findings on poverty's impact on student achievement, Colorado's school accountability program and the way schools are "rated" could be improved with the following recommendations:

1. Since school achievement on the CSAP test was highly predicted by the percent of students on free and reduced meals, the accountability model would be improved by controlling for the socio-economic variable to provide a clearer picture of student achievement. The accountability system should measure schools on meeting expectations based on how well they are doing considering the population they are serving.
2. The accountability model should also incorporate a "value added" component that examines actual student achievement longitudinally. The current model uses the average proficiency levels on an uneven scale equivalent to -.5 for unsatisfactory, .5

for partially proficient, 1 for proficient, and 1.5 for advanced. The accountability system should measure schools on improving "value added" student achievement. This approach would be one way to control for socio-economic differences in student populations.

3. The School Accountability Report should discontinue comparing school ratings with other schools ratings within a 75-mile radius. The School Accountability Report should rate the school on meeting expected achievement performance based upon demographics and prior performance.
4. The Colorado Student Assessment Program and use of the OAPI to determine a school's academic achievement rating focuses only on improving school averages. The formula used to calculate the OAPI does not take into consideration the variance within a school or the performance of disaggregated groups, for example those students on free and reduced meals. Reliance on this high-stakes formula discourages the use of "value added" analytical methods.
5. Recently, Colorado added the Academic Growth of Students Measurement to the Colorado School Accountability Report. This longitudinal use of individual student CSAP scores is more "value added" than the academic performance rating assigned to the schools. However, this rating is not part of the high-stakes accountability model. The new measure uses a comparison of students' scale scores on the CSAP from one year to the next. The results are totaled and averaged by the number of students who gained, lost or were stable in their individual scale score from year to year. The model however doesn't treat each rating category equally; students in the advanced category both years are exempt from any decline in their individual scale score. The model is

heavily impacted by the use of student percentages in each category. Since this was just implemented in 2005 it was outside of the scope of this study. However, it does indicate that Colorado is moving in a more longitudinal and "value added" direction on the SAR's.

CSAP and school size. The researcher also found that small school size only in combination with low poverty contributed to higher than average CSAP scores. This finding had a very small, almost negligible effect size when compared to the huge correlation of poverty and poor student achievement. The researcher was surprised to find that the schools that improved the most and declined the most were much smaller than the average size school. The analysis concluded that smaller schools had much greater volatility in their scores from year to year. The model used averages for calculating scores and those averages were greatly affected by the number of students included in the average. The findings call into question whether the volatility in small schools scores was because of improved learning, declined learning, variability in class demographics, regression to the mean or chance.

Because the research model found no individual correlation of school size and CSAP performance and a small but significant correlation when combined with poverty, the recommendations for practice are:

1. Increase the chances for academic improvement in schools with high levels of poverty by reducing the overall size of the school.
2. Change the OAPI calculation model to control for the greater volatility in small school scores from year to year and account for that volatility in the calculation of the Overall Academic Performance Index.

3. In order to be used for accountability purposes, the model should change the minimum number of students needed in a grade to be greater than 30. In schools where most grades have less than 30 students per grade the accountability model could look at the district as a whole. Currently small schools that have below 30 students in a grade are averaged and included in the accountability model. Making the aforesaid change may be a more valid measure for statistical accountability purposes, especially from year to year in a small school.
4. Change the way the overall academic performance rating is calculated so that the scale score is equal between all performance levels. This would benefit the schools with a small number student samples. Currently, a small school's score can be skewed dramatically by unsatisfactory scores of -.5, while partially proficient is .5, proficient is 1, and advanced is 1.5.
5. Since the researcher found greater volatility in the average proficiency scores of smaller schools from year to year, there is a question whether student achievement gains and losses were a direct result of changes in instructional approaches. This researcher would suggest that the results were influenced by how the model compared a small group to a different small group year to year. This type of comparison is often called comparing apple to oranges. The use of individual student longitudinal testing data or "value added" accountability measures could provide a better picture of student achievement and achievement gains within a school.

From the literature, student achievement seemed to benefit more when small schools served high populations of poorer students. The larger, more comprehensive schools seemed to benefit students of more affluence. The challenge for district

administrators and policy makers is to create schools that are the optimal size for the type of student population that the school serves to improve the chances of student achievement.

CSAP and charter designation. The researcher found that charter school designation individually had a small to medium correlation with higher student achievement. When combined with poverty and school size, the combination was negligible but slightly negative in 2004. This may be due to both overlapping variance, since most charter schools also tend to be small, and also because of the overpowering influence of poverty on the model. While, individually, charter schools on average scored higher on the CSAP in 2001 and 2004, it is difficult to conclude from the model studied that charter schools are outperforming non-charter schools. Charter schools serve a wide variety of purposes and student demographics in Colorado. In Colorado, charter schools continue to grow in popularity as part of the school choice reform movement (Fitzgerald, 2002). Charter designation did not correlate with school improvement on the CSAP in the model studied.

The current accountability model mandates that failing schools must be converted to charter schools after three years of failing scores. The current research does not provide scientifically-based support for this intervention. While only one non-charter school has been forced to convert to a charter school in Colorado, the new charter school has yet to open. Before the Colorado Department of Education converts more failing schools into charter schools, the researcher makes the following recommendations:

1. Policy makers should change the current accountability rules that specify converting failing schools to become charter schools. Since charter schooling as an approach to

school choice is just emerging in the literature, scientifically-based research is also just emerging. While this study found charter schools performed better on average than non-charter schools, there was no relationship to school improvement and charter designation. By forcing a school to become a charter school, it creates a completely different type of school than the charters in this study. All of the charters in this study chose to be charter schools, and the students chose to attend.

2. Policy makers should implement measures of student achievement that more accurately identify schools as failing or succeeding based on the type of student community they serve. For example, controlling for the socio-economic variable of poverty and including value added measures. The current model may be identifying schools as failing and converting them to charter schools or closing them down with out a clear picture of how that would change student achievement.
3. The combination of school choice options in Colorado's charter law and a one-dimensional student achievement measuring system like the SAR may be causing an unintended consequence of state-wide high-stakes testing; socio-economic segregation. According to Hadderman (2002) North Carolina experienced socio-economic segregation with the phase out of busing programs to end racial segregation, the implementation of high-stakes testing programs and school choice options. This researcher would recommend that some urban Colorado school districts consider adopting "socio-economic integration" as a model to overcome the unintended consequence of the charter school movement and high-stakes testing programs. For example school attendance boundaries and attendance choices would be re-configured based upon a balance of students from varying socio-economic

backgrounds. According to Hadderman (2002) this approach of reconfiguring school choice plans would give all students access to schools that have a core number of middle-class families which has been shown as a reliable predictor of school quality.

CSAP and student achievement over time. While this research study found no correlation between school change scores, school poverty, school size and designation as a charter school, this may be due to the high correlation of OAPI scores from 2001 to 2004. Overall, most schools improved their scores slightly from 2001 to 2004 on the CSAP assessment. However, it is difficult to conclude that the implementation of the CSAP has improved student achievement in Colorado based on the current OAPI calculation model. Because of the higher weight given to raising students from unsatisfactory at -.05 to partially proficient at .5, the OAPI measure is unequally weighted to give more credit to those schools that started with a higher percentage of students in the unsatisfactory rating.

It was clearly the intent of this study to be able to determine if school size and charter designation could help overcome the strong predictive relationship between high poverty and low student achievement over time. Because the change scores didn't change that much from 2001 to 2004 and the poverty variable was equally as strong of a predictor in both years, it is questionable to conclude that CSAP as a statewide high-stakes test has improved student achievement based on the model studied.

The research model studied found that poverty was a strong predictor of student achievement in both 2001 and 2004. While poverty levels accounted for a greater than expected correlation with student achievement it did not explain all of the variance in the model studied. Therefore, the theoretical question about how much impact teacher and

administrator actions can have on student achievement over time will be discussed in the Recommendations for Further Research section of this chapter.

In summary, the Colorado Student Assessment Program was designed to hold educators accountable for improved student achievement over time. The literature and research findings clearly showed a strong correlation between low achievement and high poverty. The literature and research showed some relationships of size and charter designation with student achievement. The current School Accountability Report (SAR) used to communicate student achievement does not factor any of these attributes into the model. The researcher suggests the use of a "value added" accountability model that would report school achievement based on longitudinal measures of individual student achievement. The researcher also recommends controlling for poverty when reporting and comparing student achievement data between schools. The benefits would be a clearer picture of student achievement in Colorado. By identifying the strong impact poverty has on student achievement policy makers could focus on intervention strategies that help schools overcome the impact of poverty and student achievement.

Limitations and Delimitations

There are several inherent limitations and delimitations associated with this study. A major delimitation is that the data only includes Colorado students, and the Colorado Student Assessment Program is only administered in Colorado. This limits the ability to generalize and replicate the study since the CSAP is unique to Colorado.

A statistical limitation of this study was the skewed rating scale that the Colorado Department of Education used in calculating the Overall Academic Performance Index score. The scale was weighted to show greater gains in moving students from the

unsatisfactory range to the partially proficient range. This limits the use of the OAPI as an overall indicator of student achievement since the gains or losses were not statistically calculated equally in each proficiency category. For example, student achievement gains are potentially greater if a school has more students in the "Unsatisfactory" proficiency level initially because the chance to move them to partially proficient creates a change score of one, while all other score improvements create a score change of .5.

Another limitation is the reliance on district reported data to the Colorado Department of Education both in 2001 and 2004. In this ex-post-facto design, the researcher was unable to use 39 schools' data because the data was erroneous or missing from the Colorado Department of Education databases.

Another limitation in this research design is that the tests the Colorado Department of Education used to calculate the OAPI in 2001 for middle schools were slightly different than the tests used in 2004. Although the Colorado Department of Education contends that the OAPI was statistically adjusted for the inclusion of new assessments, the author recognizes the inherent differences between a math test and a reading test and the research implications.

The author also recognizes that another limitation in this research design is that all middle schools in Colorado do not serve the same grade configurations. Fifty-eight percent of schools were 6-8, thirty eight percent were 7-8, three percent were 7-9 and two percent were 8-9. Again, the Colorado Department of Education contends that school configuration was adjusted statistically in the OAPI index scores; however, the researcher recognizes the variability in middle school grade configurations and research implications.

Another limitation was that charter school law for the creation and implementation and accountability program is very different in Colorado. Through the literature review the researcher acknowledges that in other states charter school criteria is vastly different than in Colorado's, which limits the ability to generalize the findings of charter school achievement beyond Colorado.

A final limitation was the researchers use of the change in OAPI scores from 2001 to 2004 as an indicator of student achievement or decline over time was limited both statistically and theoretically. Since the population of students was completely different in 2001 and 2004, the change in scores could be attributed to different students taking the tests.

Recommendations for Research

The researcher investigated the relationship of percent of students on free and reduced meals, school size, and charter designation to student achievement on the CSAP in both 2001 and 2004. Replication of, and expansion of this study with additional years of data and the use of another state's high-stakes testing program would add to the findings.

Since this study found no variables studied predicted CSAP improvement or decline from 2001 to 2004, it is recommended that using a "value added" model for measuring student achievement over time should be studied. This would lead to comparing student achievement in an "apples to apples" comparison instead of an "apples to oranges" comparison.

Because of the greater than typical relationship found between high poverty and low student achievement in this study, the researcher recommends studying schools with

high poverty that performed better than expected and worse than expected and investigate why. While this research study found a high correlation between student achievement and poverty it did not account for all of the variance. Therefore, additional research examining what additional factors beyond poverty impact student achievement is essential for educational reform initiatives.

Additional research should be conducted to determine what relationship exists between school size and student achievement and whether school size can mitigate the strong influence of poverty on student achievement. Additional research should be conducted on the relationship between charter schools and student achievement and whether charter schools can mitigate the strong influence poverty has on student achievement. The researcher recommends using multiple measures of student achievement since the model studied found no correlation with student achievement over time.

Further research should investigate more closely the variables initially planned for this study (average teacher salary, years of teaching experience, student teacher ratios and school location) and their correlation with student achievement. While these variables were omitted because of multi-collinearity with percent of students on free and reduced lunch and school size, the fact that they have multi-collinerarity is another research study in itself. The study might ask why were the lowest paid teachers with the least amount of experience mostly teaching in the largest schools serving the highest concentration of students of poverty?

Finally, the relationship between statewide high-stakes testing programs and their impact on student achievement needs to be investigated over time and with a broad

measure of student achievement. The benefits would be a clearer picture of how the statewide high stakes testing programs are meeting the federal expectations of "Leaving No Child Behind" and state expectations of increased accountability as well as improved student achievement.

Conclusions

This study examined the influence of poverty, school size, and charter designation on student achievement as measured by the Colorado Student Assessment Program (CSAP) in 2001 and 2004. The results of this study found a greater than typical effect size relationship between high poverty and low achievement. Charter school designation alone indicated a small to medium effect, and school size only in combination with poverty showed a small effect. This research study found no correlation between school change scores and school poverty, school size or designation as a charter school. The Colorado Department of Education currently uses the OAPI index score as the only measure of student achievement in the state. Based on the findings of this research, school poverty, school size, and charter designation, to varying degrees seem to influence student achievement. This researcher suggests using a "value added" measure as part of the OAPI index score. By including this measure, Colorado would have a clearer picture of student achievement. When policy makers and educators have "value added" information on student achievement, the use of high-stakes statewide assessments can become less political, and more about implementing programs, resources, and interventions to truly "leave no child behind".

REFERENCES

Abbot, M. L., Joireman, J., & Stroh, H. R. (2002). The influence of district size, school size and socioeconomic status on student achievement in Washington: A replication study using higher archical linear modeling. Technical Report (Washington School Research Center) Seattle Pacific University. Lynwood, WA.

Ainsworth-Darnell, J. W., & Roscigo, V. J. (1999). Race, culture and educational resources: Persistent inequalities and achievement returns. *Sociology of Education, 72,* 158-178.

Amrein, A. L., & Berliner, D. C. (2002). High Stakes testing, uncertainty and student learning. *Education Policy Analysis Archives, 10* (18). Retrieved January 25, 2003, from http://epaa.asu.edu/epaa/v10n18/

Amrein, A. L., & Berliner, D. C. (2002). T*he impact of high stakes tests on student academic performance: An analysis of NAEP results in states with high- stakes tests and ACT, SAT and AP test results in states with high school graduation exams.* Tempe, Arizona: Arizona State University, Education Policy Research Unit.

Bracey, G. W. (2001). The 11th Bracey Report on the condition of public education. *Phi Delta Kappan, 83*, 157-169.

Bracey, G.W. (2003). The 13th Bracey Report on the condition of public education. *Phi Delta Kappan, 85*, 148-164.

Broadfoot, P. (2000). Comparative education for the 21st century: Retrospective and prospect. *Comparative Education, 36,* 357-371.

Bulkley, K., & Fisler, J. (2002). A review of the research on charter schools. Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education.

Cibulka, J. G., & Derlin, R. L. (1998). Accountability policy adoption to policy sustainability: reforms and systematic initiatives in Colorado and Maryland. *Education & Urban Society, 30*, 502-515.

Cizek, G. J. (2000). Pockets of resistance in the assessment revolution. *Educational Measurement: Issues and Practice, 19,* 16-23.

Cizek, G. J. (2001). *Setting performance standards.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Clark, D. C., & Clark, S. N. (2000). Appropriate assessment strategies for young adolescents in an era of standards-based reform. *The Clearing House, 73,* 201-204.

Cohen, D. K. (1996). Standards-based school reform: Policy, practice, and performance. In Helen Ladd (Ed.), *Holding schools accountable.* Washington, D.C.: Brookings Institution.

Coleman, J. W. (1966). *Equality of educational opportunity.* Washington, DC: U.S. Department of Health, Education and Welfare.

Cooley, W. W. (1993). *The difficulty of the educational task.* Pittsburgh, University of Pennsylvania, Learning Research and Development Center. (ERIC Document Reproduction Service No. ED358527)

Colorado Department of Education (2005). *School accountability reports: A five-year review of progress 2000-2005.* Retrieved January 8, 2005, from http://reportcard.cde.state.co.us/reportcard/pdf/sar5yearreport.pdf

Colorado Department of Education (2003). School accountability report (SAR) information. Retrieved July 15, 2003 from http://www.cde.state.co.us/cdeassess/SAR/index.htm

Darling-Hammond, L. (2000). *Transforming urban public schools: The new role of standards and accountability.* East Lansing, MI: National Center for Research on Teacher Learning. (ERIC Document Reproduction Service No. ED459290)

DeVito, P. J., & Koening, J. A. (2001*). NAEP reporting practices: Investigating district-level and market-basket reporting.* Washington D.C.: National Holiday Press.

Downey, L. W. (1988). *Policy analysis in education.* Calgary, Canada: Detselig Enterprises Ltd.

Eisner, E. W. (2001). What does it mean to say a school is doing well? *Phi Delta Kappan, 82,* 367-372.

Elmore, R. F., & Fuhrman, S. H. (2001). Holding schools accountable: Is it working? *Phi Delta Kappan, 83,* 67-72.

Evans, R. (1999). The great accountability fallacy. *Education Week, 18,* 52.

Firestone, W. A., Mayrowetz, D., & Fairman, J. (1998).Performance-based assessments and instructional change: the effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis, 20,* 95-113.

Fitzgerald, J. (2003). *The state of charter schools in Colorado: 2001-2002.* Denver: Colorado Department of Education.

Gill, B.P., Timpane, P.M., Ross, K.E. & Brewer, D.J. (2001). Rhetoric versus reality: *What we know and what we kneed to know about vouchers and charter schools.* Santa Monica, CA: Rand Corporation.

Goldhaber, D. D. (1999) School choice: An examination of the empirical evidence on achievement, parental decision making and equity. *Educational Researcher, 29,* 16-25.

Gliner, J. A., & Morgan, G. A. (2000). *Research methods in applied settings. An integrated approach to design and analysis.* Mahwah, NJ: Lawrence Erlbaum Associates.

Gratz, D. B. (2000). High standards for whom? *Phi Delta Kappan, 81,* 681-687.

Grissmer, D., Flanagan, A., Kawata, J., & Williamson, S. (2000*). Improving student achievement, what state NAEP tests scores tell us.* Santa Monica, CA: RAND.

Guskey, T. R. (1997, March). *The relationship between socioeconomic characteristics and school- level performance assessment results.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Hadderman, M, (2002). *Trends and issues school choice* (Report No.EA032330). Eugene, OR: ERIC Clearinghouse on Educational Management. (ERIC Document Reproduction Service No. ED472995)

Haycock, K. (2001). Closing the achievement gap. *Educational Leadership, 58,* 1-9.

Hertert, L., & Teague, J. (2003) Narrowing the achievement gap: A review of research, policies, and issues. EdSource. http://www.edsource.org

Howley, C. B. (1994). *The academic effectiveness of small-scale schooling (an update).* (Report No. EDO-RC-94-06-00) Charleston, WV: ERIC Clearinghouse on Rural Education and Small Schools.(ERIC Document Reproduction Service No. ED372897).

Howley, C. B. (1996) *Ongoing dilemmas of school size: a short story.* (Report No. EDO-RC-96-6). Charleston, WV:ERIC Clearinghouse on Rural Education and Small Schools. (ERIC Document Reproduction Service No. ED401089)

Howley, C. B., & Bickel, R. (2000). *When it comes to schooling...small works: School size, poverty, and student achievement.* (Rural School and Community Trust). Randolph, VT.

Hirsch, E. D., Jr. (1996). *The schools we need.* New York, NY: Doubleday.

Irmsher, K. (1997). *School size.* (Report No. EDO-EA-97-5). Eugene, OR: ERIC Clearinghouse on Educational Management. (ERIC Document Reproduction Service No. ED414615)

Johnson, J. D., Howley, C. B., & Howley, A. A. (2002). *Size, excellence and equity: A report on Arkansas schools and districts.* East Lansing, MI: National Center for Research on Teacher Learning. (ERIC Document Reproduction Service No. ED459987)

Johnson, D. D., & Johnson, B. (2002). *High stakes: Children, testing, and failure in American schools. A year in the life of one rural schools and its children.* Blueridge Summit, PA: Rowman and Littlefield.

Jones, K. (1997). Kentucky's conflicting reform principles. *Phi Delta Kappan, 79,* 276-281.

Jones, G. M., Jones, B. D., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impact of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan, 81,* 199-203.

Khattri, N., Reeve, A. L., & Kane, M. B. (1998). *Principles and practices of performance assessment.* Mahway, NJ: Lawrence Erlbaum Associates, Inc.

Kohn, A. (2000). Burnt at the high stakes. *Journal of Teacher Education, 51,* 315-327.

Kornhaber, M. L., & Orfield, G. (2001). *Raising standards or raising barriers? Inequality and high-stakes testing in public education.* New York: The Century

Linn, R. L. (2000). Standards: The policy environment [Electronic version]. *ERIC Digest, 138,* 1-7.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 29,* 4-16.

Linn, R. L., Baker, E.L., & Betenbenner, D. W. (2002). Accountability systems: Implications of requirements of No Child Left Behind Act of 2001. *Educational Researcher, 31,* 3-16.

Lissitz, R. W., & Schafer, W. D. (2002). *Assessment in educational reform: Both means and ends.* Boston: Allyn & Bacon.

Lockwood, A. T. (1998). *Standards. From Policy to Practice*. Thousand Oaks, CA: Corwin Press, Inc.

Loveless, T. (2002). *Charter school achievement and accountability*. Paper presented at the Program on Education Policy and Governance, John F. Kennedy School of Government, Harvard University Conference on Taking Account of Accountability: Assessing Politics and Policies. Boston, MA.

Madaus, G., & Clarke, M. (2001). *The adverse impact of high-stakes testing on minority students: evidence from one hundred years of test data.* In Orfield, G., Kornhaber, M. (Eds). (2001). *Raising standards or raising barriers? Inequality and high-stakes testing in public education*. New York: The Century Foundation Press

Massachusetts Department of Education. (2005). *Assessment and Accountability Division*. http://www.doe.mass.edu/Assess/

Maryland State Department of Education (2004). Testing Division, http://www.marylandpublicschools.org/MSDE/testing/msa/

McMillan, J. H. (2001). *Essential assessment concepts for teachers and administrators.* Thousand Oaks, CA: Corwin Press, Inc.

McNeil, L. M. (2000). *Contradictions of reform. The educational costs of standardized testing.* New York, NY: Routledge.

McNeil, L. M. (2000). Creating new inequalities. Contradictions of reform. *Phi Delta Kappan, 81,* 729-734.

Mehrens, W. A. (1998). Consequences of assessment: What is the evidence? *Education Policy Analysis Archives*, 6, 1-30.

Meier, D. W. (1996) The big benefits of smallness. *Educational Leadership, 54,* 12-15.

Meier, D. W. (2000). *Will standards save public education?* Boston, MA: Beacon Press.

Miller-Whitehead, M. (2001). Educational equity in Alabama: What we learned from report card 2000. *Research Brief.* (ERIC Document Reproduction Service No. ED459198)

Miron, G., & Nelson, C. (2001). *Student academic achievement in charter schools: What we know and why we know so little.* (Tech. Rep. No. 41). New York: Columbia University, National Center for the Privatization in Education.

Mitchell, N., & Hubbard, B. (2005, July 16). Test gaps linger. *The Rocky Mountain News*, pp. 2A, 11A.

Muller, C., & Schillar, K. S. (2000). Leveling the playing field: Educational attainment and states performance testing. *Sociology of Education, 73,* 196-218.

National Research Council. (2001). *NAEP reporting practices.* Washington, D.C.: National Academy Press.

Nelson, B., Berman, P., Ericson, J., Kamprath, N., Perry, R., Silverman, D., & Soloman, D. (2000). *The state of charter schools, 2000, The national study of charter schools. Fourth-year report.* (OERI report No. SAI 2000-3000) Washington, DC: US Government Printing Office.

North Carolinas Department of Public Instruction (2005). *ABC's of education background packet.* http://www.dpi.state.nc.us/docs/newsroom/news/2004-05/05backgroundpacket.pdf

Ohanian, S. (2001). News from the test resistance trail. *Phi Delta Kappan, 82,* 363-366.

O'Neill, P. T. (2001). Special education and high stakes testing for high school graduation: An analysis of current law and policy. *Journal of Law & Education, 30,* 185-222.

Pearson, P. D., Vyas, S., Sensale, L. M., and Kim, Y. (2001). Making our way through the assessment and accountability maze. Where do we go now? *The Clearing House, 74,* 175-182.

Pipho, C. (2000). Politics, elections, and education. *Phi Delta Kappan, 81,* 725-726.

Pipho, C. (2000). The sting of high-stakes testing and accountability. *Phi Delta Kappan, 81,* 645-646.

Popham, J. W. (1998). Farewell, curriculum: Confessions of an assessment convert. *Phi Delta Kappan, 79,* 380-384.

Ramirez, A. (1999). Assessment-driven reform. The emperor still has no clothes. *Phi Delta Kappan, 81,* 204-208.

Rose, L. C., & Gallup, A. M. (2001). The 33rd annual Phi Delta Kappa/Gallup Poll of the public's attitudes toward public schools. *Phi Delta Kappan, 83,* 41-58.

Salowe, A., & Lessinger, L. (2001). *Healing public schools.* Lanham, MD: Scarecrow Press, Inc.

Sanders, W. L. (2003). *Beyond "No Child Left Behind".* Paper presented at the annual meeting of the American Educational Research Association. Chicago, IL.

Schiller, K. S., & Muller, C. (2000). External examinations and accountability, educational expectations, and high school graduation. *American Journal of Education, 108,* 73-99.

Schmoker, M. (1999). *Results: The key to continuous school improvement*, Alexandria, Va.: Association for Supervision and Curriculum Development.

Sirotnik, K. A., & Kimball, K. (1999). Standards for standards-based accountability systems. *Phi Delta Kappan, 81,* 209-214.

Slovacek, S. P., Kunnan, A. J., & Kim, H. (2002). *California charter schools serving low-SES students: An analysis of the academic performance index.* California State University, Los Angeles, College of Education.

Smith, S. S., & Mickelson, R. A. (2000). All that glitters is not gold: School reform in Charlotte-Mecklenburg. *Educational Evaluation and Policy Analysis, 22*, 101-127.

Stake, R. (1999). The goods on American education. *Phi Delta Kappan, 80,* 668-672.

Stotsky, S. (2000). *What's at stake in the K-12 standards wars?* New York, NY: Peter Lang Publishing, Inc.

Texas Education Agency (2005). Student assessment division. http://www.tea.state.tx.us/student.assessment/

Thomas, D.M. & Bainbridge, W.L. (2001). 'All children can learn.' Facts and fallacies. *Phi Delta Kappan, 82,* 660-662.

Thompson, S. (2001). The authentic standards movement and its evil twin. *Phi Delta Kappan, 82,* 358-362.

Timble, S. (2002). Common elements of high performing, high poverty middle schools. *Middle School Journal, 3*, 7-16.

Wassermann, S. (2001). Quantum theory, the uncertainty principle, and the alchemy of standardized testing. *Phi Delta Kappan, 83,* 28-40.

Zeller, L., & Jinkins, D. (2001, December). *Consequences of high stakes testing on the literacy programs of high-performing learning communities*. Paper presented at the National Reading Conference, San Antonio, TX.

Zigo, D. (2001). Constructing firebreaks against high-stakes testing. *English Education, 33,* 214-232.