2010-12-03

Principals' Opinions on the Impact of High-Stakes Testing on Teaching and Learning in the Public Elementary Schools in the State of Utah

Raylene Jo Hadley
Brigham Young University - Provo

Principals' Opinions on the Impact of High-Stakes Testing on Teaching and Learning in the Public Elementary Schools in the State Of Utah

Raylene J. Hadley

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree

Doctor of Philosophy

A. LeGrand Richards, Chair
Scott Ellis Ferrin
Sterling Hilton
Vance Randall
Joe Matthews

Department of Educational Leadership and Foundation

Brigham Young University

December 2010

ABSTRACT

Principals' Opinions on the Impact of High-Stakes Testing

on Teaching and Learning in the Public Elementary

Schools in the State Of Utah

Raylene Hadley

Department of Educational Leadership and Foundations

Doctor of Philosophy

The No Child Left Behind Act of 2001 (NCLB) brought high-stakes testing to the forefront of American public education. With its call for teachers and schools to be accountable for academic performance, NCLB has focused the spotlight on yearly progress, as measured by students' test scores. Issues associated with this charge include the questionable reliability of tests, the variation evident in state standards, and the consequences an emphasis on high-stakes testing may have on teaching and learning in the classroom.

The purpose of this study was to investigate the consequences of high-stakes testing on teaching and learning in public elementary schools in Utah from the vantage point of school principals. Although policymakers assume a direct correlation between increased test scores and academic achievement, this study went beyond test scores. Analysis of semi-structured interviews with 12 principals, selected through purposive sampling from both Title 1 and non-Title 1 schools, revealed both positive and negative themes. Principals appreciated the focus and collaboration that NCLB testing encourages among teachers, but they disliked the impact of poor test scores on faculty morale. Unlike respondents in previous studies, principals did not feel that NCLB diminished creativity in the classroom; they did worry, however, about the validity of scores as a measure of student learning, particularly in the case of a one-time, year-end test.

Keywords: No Child Left Behind, achievement tests, high-stakes testing, Title 1 funding, elementary education, criterion-referenced tests

ACKNOWLEDGMENTS

It may not be possible to acknowledge all of the individuals who have played a part in helping me reach this long-desired educational goal, so I apologize if I neglect to mention anyone by name. Of course, as with any accomplishment, I must start by acknowledging the members of my family, who have been my examples, my teachers, my cheerleaders, my friends. I especially need to recognize the encouragement and patience of my husband, Dale. I cannot count the number of times I turned down excursions to the cabin, or weekend trips to Lake Powell, or Saturday afternoon dates because of "the paper." I am so grateful for his unconditional support and understanding, and for his accompanying me on this arduous journey. I have greatly benefitted from the example set by our two oldest children, Jason and Alex, who showed me the meaning of perseverance and hardwork as they each fulfilled their desire for a juris doctorate degree. Our youngest son, Justin, has shown me what it is like to remain courageous in spite of personal challenges—which he never allowed to thwart his academic progress. Megan, our youngest child, provided me with a much-needed (although not always appreciated at the time) distraction that helped me keep things in perspective. She has been a buddy and confidante who, likewise, encouraged me by her example and love. My deepest affection and appreciation go to my beautiful daughters-in-law, Debbie and Holly, who have provided continual support and encouragement.

I owe a debt of gratitude to my wonderful parents, Robert and Roberta Cowley, who instilled in me, at a young age, a belief that I could accomplish anything I desired. Their support and confidence in me was invaluable during this academic journey. To this

day, their belief and participation in life-long learning and service inspire me. I am grateful to my in-laws, Charlotte and Earl Hadley, who have been a constant source of support and love, and have been with me every step of the way.

To my "New England" parents, Don and Dorothy Colley, I express my sincere gratitude. It was through their example that I learned to pursue learning for learning's sake. Although Don did not live long enough to see me complete this degree, I have felt his spirit and encouragement along the way. I express my great admiration and appreciation to Faith Colley Wadleigh, who has always been generous in sharing her thoughts, ideas, experiences, and insights with me. Her unwavering loyalty and love have been a source of comfort and joy in my life.

I have received encouragement and support from so many friends, particularly Kathy Crawford, who helped to keep me grounded during my doctoral studies. She always made herself available whenever I needed a break from writing. I owe a debt of gratitude to my sister, Gayle Murray, one of the smartest people I know. She was my first mentor and teacher. It was under her tutelage that my grammar, spelling, writing, and typing skills became such that this paper could eventually be written.

I am indebted to the wonderful members of my doctoral cohort who are like family to me: Patras Bukhari (Pakistan), Susan Huff, Cameron Martin, Christopher Mugimu (Uganda), Frances Naluwemba (Uganda), Hilda Sabbah (Israel), Janet Tanner, Annie Rumsey Wairepo, Sarah Westerberg (England), and Xiucheng Yang (China). Their friendship and association was well-worth the price of the journey.

The faculty in the Educational Leadership and Foundations department at Brigham Young University became our teachers, our mentors, our friends. Our successes

became their successes. Special gratitude goes to my committee chair, Dr. LeGrand (Buddy) Richards, whose enthusiasm for his own personal research is motivating and energizing. Sincerest appreciation goes to Dr. Steve Hite for the hours he spent helping me understand the difference between qualitative and quantitative research and who taught best by supplying the questions that needed to be asked; to Dr. Sterling Hilton, who consented to join my committee at the eleventh hour and yet spent countless hours assisting with the refining process; to Dr. Scott Ferrin, a long-time colleague and friend, whose understanding and passion for law enlightened us all; to Dr. Vance Randall, whose quiet assurance provided the much-needed support and motivation to continue improving; and to Joe Matthews, who graciously consented to remain on my committee after securing a new position at Utah State University in Logan. Thank you, Joe.

Appreciation and admiration go to Bonnie Bennett, the life-blood of the EDLF department. Without her assistance in helping us navigate the protocol, procedures, and processes of the system, no one would ever graduate. A sincere debt of gratitude goes to my editor, Hilary Hendricks, without whose help this paper would never have reached completion.

Last, I'd like to thank the principals who consented to be interviewed for this study and the many educators world-wide who share a love for learning and a commitment to a future where no child is left behind.

TABLE OF CONTENTS

LIST OF TABLES

Chapter One: Introduction

> *The purposes of schools and other social agencies are not "discovered" as a prospector strikes a gold-mine. They evolve.*
>
> —Educational Policies Commission, 1938, pp. 1

Although education has often been seen as the means in which society's best thinking is handed down from generation to generation, there is a question as to what this really means. One scholar wrote, "Americans love education, believe in education and pay big money for education, but few agree on what it is, what it should accomplish, or what methods should be used to achieve it" (DeMille, 2000, p. 19). Americans do seem to agree about one aspect of education, however: the American public education system is in need of repair (U.S. Department of Education, National Commission on Excellence in Education, 1984; U.S. Department of Education, 2002; Ravitch, 2000).

This is not a new idea. A cry for school reform has reverberated in every decade since the 1920s (Ravitch, 2000). Long before the landmark report "A Nation at Risk" (U.S. Department of Education, National Commission on Excellence in Education, 1984) and years beyond the No Child Left Behind (2001) legislation, widespread public concern regarding student outcomes and the performance of public education exists. Yet progress has been slow. In 2002 the U.S. Department of Education reported, "Despite hundreds of programs and hundreds of billions of dollars invested during the last generation, American students still lag behind many of their fellow foreign students and the academic achievement gaps in this country between rich and poor, white and minority students, remain wide" (p. 9).

At the beginning of the 21st century, in response to these achievement gaps, a focus on "leaving no child behind" reinvigorated the debate. Wood (2004) articulated the appeal of

legislative initiatives aimed at educating all children to higher competencies:

> Who could object to a law that promises no child left behind when it comes to our schools? After all, isn't this the great promise of our public school system—that all children, regardless of race, socioeconomic status, gender, creed, color, or disability will have equal access to an education that allows them to enjoy the freedoms and exercise the responsibilities of citizenship in our democracy? (Wood, p. vii)

Yet there is a lack of consensus as to what changes would actually enhance educational opportunities for all children and, at the same time, help them acquire the necessary competencies and skills to become highly functioning, productive, and successful adults.

One trend preferred by lawmakers is to seek accountability by emphasizing test scores. This tactic, of course, reflects the assumption that higher test scores accurately indicate increased learning, and that paper-and-pencil assessment is a good measure of the learning that has taken place in the classroom. Opposing this assumption, many researchers warn that test scores may not accurately reflect competency (Chapman, 2004; Harvey, 2003; Jones, 2004; Neill, Guisbond, Schaeffer, Madden, & Legeros, 2004; Popham, 2003; Rose & Gallup, 2003; Sizer, 2004; Wood, 2004). Still, legislators have been quick to rely on testing as both a measure of competency and an incentive to build that competency—as exemplified by No Child Left Behind.

No Child Left Behind

Hailed by the U.S. Department of Education as "a landmark in education reform," the 2001 No Child Left Behind Act was "designed to improve student achievement and change the culture of America's schools" (U.S. Department of Education, 2002, p. 9). In essence, No Child Left Behind (NCLB) is a reauthorization of the 1965 Elementary and Secondary Education Act (ESEA) that encompasses four main principles: greater accountability for results; increased

flexibility in the use of federal funds by schools, districts, and states; more parental choice; and an enhanced emphasis on educational practices "that work" (U.S. Department of Education, 2002). The basic components of the legislation are outlined below.

Title I. Title I is central to NCLB's mission "to ensure that all children have a fair, equal, and significant opportunity to obtain a high-quality education and reach, at a minimum, proficiency on challenging state academic achievement standards and state academic assessments" (U.S. Department of Education [n.d., a], Sec. 1001). In distributing more resources to schools with the highest percentage of disadvantaged (i.e., low-income) students, the expectation is that the quality of teacher preparation, academic assessments, curriculum and instructional materials will improve, thus increasing the quality of education for the students in greatest need. But the extra funding comes with hefty strings attached. By accepting Title I funding, states, school districts, and individual schools become "accountable for improving the academic achievement of all students and turning around low-performing schools" (U.S. Department of Education, 2002, p. 13).

Accountability. Making schools accountable for the increased academic achievement of students is a major component of No Child Left Behind. Former Secretary of Education Rod Page explained accountability in terms of investment returns:

> In our dedication to improving achievement for all students, we will put into practice a culture of accountability that extends to all education levels: federal, state, and local. To that end, states and districts that receive federal funds are expected to report improved student achievement as a return on the investment. (U.S. Department of Education, 2004, p. 33)

The clearly stated expectation of NCLB is that all children should have the same academic

opportunities and should meet the proficiency standards established by their state in language arts and mathematics (U.S. Department of Education, 2004).

Supporters of NCLB believe that progress is being made. After a national tour where she visited schools, educators, and legislators from 22 different states, former Secretary of Education Margaret Spelling was optimistic: "The discussions we had supported the idea that our nation is on the move and that our schools and students are making progress" (Spelling, 2008, Advancing Accountability 2008 national tour). Spelling also recognized the need for continued dialogue: "The one thing all these visits had in common was that we continued the national conversation about how to strengthen and improve NCLB."

On the other hand, many educators believe that it will take a lot more than continued discussion to correct the problem areas in public education—problems only compounded by NCLB. Linda Darling-Hammond (2004) summarized that NCLB provides for a "shortsighted, one-way accountability system that holds children and educators to test-based standards they are not enabled to meet" (p. 6).

Regardless of the disparity in opinions, NCLB has made a significant contribution in capturing the nation's attention and renewing interest in classroom equity and student success. At the same time, lawmakers' requirement that all states create definite proficiency standards has moved statewide responsibility to the top of the charts.

Adequate yearly progress (AYP). According to No Child Left Behind, all states must implement a statewide accountability system that clearly defines "adequate yearly progress." They must also create measurable yearly objectives that will help to determine whether all groups of students have reached "proficiency" within the expected 12-year timeframe. All districts and schools are expected to meet adequate yearly progress (AYP) as defined by their

individual state, or risk specific sanctions (U.S. Department of Education, 2002).

States must establish separate AYP goals for up to nine different categories of students: Asian, African American, American Indian, Caucasian, Hispanic, Pacific Islander, economically disadvantaged, English language learners, and students with disabilities. At least 95% of the students in each group must be tested in order to meet AYP (U.S. Department of Education, 2002). Schools that fail to meet their AYP goals for two consecutive years will face sanctions, and sanctions will continue to increase for each subsequent year AYP is not met.

High-Stakes Testing

Although accountability is only one piece of NCLB, it is a significant one. The legislation requires all states to give assessments that are aligned with their own established achievement standards—in particular, states are to "measure what children know and can do in reading and math in grades 3 through 8" (U.S. Department of Education, 2002, p. 9). Although connecting content standards, statewide exams, and accountability might seem a logical strategy, these state assessments have garnered their share of attention during the many deliberations over NCLB.

Part of the controversy stems from the high-stakes nature of the tests. When a student's grade level or ability to graduate are on the line, or when AYP for the school or district is determined by the test results, the test is referred to as "high-stakes" (Heubert & Hauser, 1999).

Although public schools have a long-standing tradition of testing, only since the 1960s has high-stakes testing became associated with achievement testing. In other words, students' performance on large-scale achievement tests now determines such "high-stakes" consequences as placement in special classes or programs, grade promotion, and graduation from high school (Heubert & Hauser, 1999).

Impact on teaching. There seems to be little question that an increased emphasis on testing necessitates a comparable focus on instruction in those areas most likely to be tested. The discussion taking place is whether this has had a positive or a negative effect in the classroom.

Traditionally, elementary schools around the country have provided students with a variety of experiences and opportunities that extended beyond the regular core subjects of reading and mathematics to include areas such as art, music, dance, history, and the sciences. In the years since NCLB, however, scholars have noted changes in the academic landscape at many elementary schools. Instruction in non-tested subjects, as well as experiences such as science labs, field trips, and assemblies, are being replaced by more classroom time spent on test preparation and subject-tested material (Jones, Jones, & Hargrove, 2003; Madaus, Russell, & Higgins, 2009; Pedulla et al., 2003; Wood, 2004).

While some educators feel the curriculum is suffering due to teachers' preoccupation with test preparation (Karp, 2004; Madaus, Russell, & Higgins, 2009; Pedulla et al., 2003), others have a different opinion. Winters, Greene, and Trivitt (2008) reported a direct correlation between gains in student learning in high-stakes subjects and an increase in student proficiency in other subject areas. They also suggested a spill-over effect from better teaching in core subjects:

> We have discovered evidence suggesting that student proficiency in science has increased under the high-stakes sanctions primarily because the improvements that students have made in math and reading have enhanced their ability to learn science material as well. (pp. 2-3)

Clearly, additional research needs to explore the impact high-stakes testing is having on teaching in the classroom.

Impact on learning. Another important area of study is the impact of high-stakes testing on the students themselves. Clarke et al. (2003) interviewed 360 educators from three states about the perceived effect of state-mandated testing on teaching and learning. Teachers in the state whose tests carried the lowest stakes (Kansas) were less likely to mention test-related stress for students, while educators from the state with the highest testing stakes (Massachusetts) frequently mentioned students' stress. On the other hand, the educators from Massachusetts also observed an increase in student motivation as well as an overall improvement in the quality of education as a result of high-stakes testing.

Addressing this paradox, Madaus, Russell, and Higgins (2009) indicated that motivation is an important determinant of academic achievement. However, they cautioned that expecting high-stakes testing to automatically motivate all unmotivated students is an oversimplification. They specified that "students must believe that the rewards attached to test performance are important and realistically within reach" (p. 159). This may help explain why some students seem more motivated than others by high-stakes testing.

Literature on high-stakes testing is replete with examples of its influence on students and learning—an influence that is interpreted as both good and bad (Clarke et al., 2003; Jones, Jones, & Hargrove, 2003; Madaus, Russell, & Higgins, 2009; Pedualla et al., 2003). The extent of this impact, and whether it benefits or hinders the students' academic accomplishments overall, however, is still being studied.

Statement of the Problem

Much has been written about both the expected benefits and the unanticipated negative consequences of the No Child Left Behind Act of 2001. As with other legislation of this magnitude, the outcomes were not certain before implementation. Still, one of the expectations

implicit with the passage of NCLB was increased attention to the high-stakes tests that would be required for accountability, grade promotion, and high school graduation.

The two central functions of state-mandated high-stakes testing are (a) to assess individual student achievement (learning) and (b) to determine the effectiveness of the school (teaching) (Pedulla et al., 2003). The discussion continues as to whether high-stakes testing is indeed promoting enhanced learning and teaching or actually fostering unfortunate and unintended consequences.

Purpose of the Study

The purpose of this dissertation is to examine the consequences of high-stakes testing on teaching and learning in the public elementary school setting in the state of Utah, as seen through the eyes of the elementary school principal. Although teachers are often the targets of such studies (Abrams, 2004; Abrams, Pedulla, & Madaus, 2003; Bussert-Webb, 1999; Cimbricz, 2002; Pedulla et al., 2003), public school principals, who usually bear the responsibility for poor test results, are frequently overlooked. Because most public school principals have the advantage of coming from classroom settings prior to becoming administrators, they can be a rich source of information. Especially important to this study, principals have had a front row seat in the evolution of pedagogical practices and education legislation, and they are ideally situated to recognize the influences such changes have had on teaching and learning.

Justification for the Study

If history repeats itself, the "school reform" pendulum will continue to swing. Will the use of tests for grade promotion be as prevalent ten years from now as it is today? Amrein and Berliner (2002) called for regular and repeated research to track reforms: "Because testing programs and their effects change all the time, reinterpretation of the research that bears on this

issue will be needed every few years" (p. 11). Other researchers also argued that, because of the complexity of the issues, meaningful studies on the impact of high-stakes testing should be ongoing (Cimbricz, 2002; Dorgan, 2004; Vogler, 2002).

In her 2004 study of one elementary school's attempt to meet a higher academic standard, Dorgan ran into as many questions as she answered: "The project showed the effects of the state testing program on classroom practices, both positive and negative, and it raised questions for further study" (Abstract section, para. 1). For example, although research such as Dorgan's has demonstrated that high-stakes testing does influence the objectives, activities, and materials used in the classroom, the significance of this influence is in debate (Cimbricz, 2002; Vogler, 2002).

Current research presents a complicated picture with passionate, knowledgeable, and well-meaning educators on both sides of the issues. Inasmuch as the current trend is to use some form of standardized assessments to determine everything from grade promotion and graduation to how well our schools are doing, additional studies seem imperative.

Further study regarding the relationship that exists between high-stakes testing and teaching and learning in actual school settings will help determine the direction of future decisions that impact our public school system.

Research Questions

To determine the impact high-stakes testing has had on teaching and learning in the state of Utah, from the perspective of the elementary school principal, the following research questions have been developed:

1. According to the opinions of Utah public elementary school principals, how have teaching strategies been influenced, if at all, by high-stakes testing?
2. According to the opinions of Utah public elementary public school principals, how are

students being prepared for high-stakes testing, and have the course curriculum and related school activities been influenced, if at all, by the time spent in preparing students for taking these tests?

3. According to the opinions of Utah public elementary school principals, how has student achievement been impacted, if at all, by high-stakes testing?
4. Are principal perceptions regarding high-stakes testing significantly affected by school demographics or whether the school demonstrated adequate yearly progress (AYP)?

Definition of Terms

Adequate yearly progress (AYP): Under NCLB, all states are to establish annual academic objectives in reading and mathematics that must be met by each school and school district. Progress is to be measured annually to ensure that all students reach the established proficiency level by the year 2014—including students who are from low-income homes, students with disabilities, minority students, and students with limited English proficiency. At least 95% of students in each of the sub-groups must be tested in order for a school or district to pass AYP (U.S. Department of Education, 2002).

Criterion-referenced test (CRT): On a CRT, a student's test score is interpreted by comparing it to a pre-specified standard of performance. The test questions are clearly written. Score interpretation is not tied to the performance of other students (Ediger, 2003).

Formative assessments: A formative assessment is an ongoing evaluation or activity usually used in the classroom to provide information that is used as immediate feedback for the teacher to adapt the teaching to meet the needs of the students (Black & William, 1998; Bransford, Brown, & Cocking, 2000).

High-stakes tests: A high-stakes test is a standardized assessment that includes substantial

sanctions for the individual student, such as grade retention or failure to graduate, for unsatisfactory test scores (Heubert & Hauser, 1999). For the district or school, a high-stakes test is an assessment linked with financial or other consequences. Based on NCLB, a school or district that fails to "make sufficient yearly progress toward state proficiency goals for their students [i.e., failure of the students to pass high-stakes tests] first will be targeted for assistance and then be subject to corrective action and ultimately restructuring" (U.S. Department of Education, 2002, p. 10).

Learning: Simply stated, for the purposes of this study, learning refers to the act, process or experience of gaining knowledge or a skill.

Low-stakes tests: Tests with low stakes are those with "no significant, tangible, or direct consequences . . . attached to the assessment results" (Heubert & Hauser, 1999, p. 35).

Norm-referenced tests (NRT): Norm-referenced tests are tests in which an individual's test score is interpreted by comparing it with scores received by a "norming" group that has the following factors the same: (a) time limits, (b) subject matter tested, and (c) directions for test taking (Ediger, 2003).

Performance standard: A performance standard is "a written description of the knowledge or skills that students must demonstrate to show that they meet a specified level of performance in a subject-matter area" (Brandon, 2002, p. 167).

Standardized tests: As set forth in NCLB, and for the purposes of this study, standardized tests refer to statewide tests that are administered to all the state's students in the same way. It does not necessarily refer to particular format or content.

Summative assessment: Summative assessments are administered toward the end of the school year to assess overall academic performance of course goals. They are intended to

measure the learning that has taken place and are specifically used to determine students' understanding of state and nationally developed standards (Bransford et al., 2000).

Teaching: Teaching refers to not only the act of instructing, but also the activities and pedagogical practices involved in educating or instructing. Not only how, but *what* is being taught is also closely associated to teaching, for the purposes of this study.

Teaching strategy: The style or method of teaching or the manner in which a teacher instructs.

Utah Basic Skills Competency Test (UBSCT): The UBSCT is a test that is administered to Utah students beginning in the 10th grade that includes a minimum competency in English, language arts, reading and mathematics (Utah State Office of Education, 2006b).

Utah Performance Assessment System for Students (U-PASS): U-PASS is the Utah state assessment program that includes (a) the norm-referenced achievement testing of students in grades 3, 5, 8, and 11 by means of tests decided by the Utah State Board of Education; (b) criterion-referenced achievement testing of students in all grade levels in basic skills; (c) direct writing assessments in grades 6 and 9; (d) a 10^{th}-grade basic skills competency test and (e) the use of student behavior indicators in assessing student performance (Utah State Office of Education, 2006c).

Chapter Two: Literature Review

Historical Elements Leading to Testing in America

The desire that "all children have the opportunity to obtain a high quality education" (U.S. Department of Education, 2002, p. 13) was not born with the No Child Left Behind Act of 2001. Opinions expressed at the end of the nineteenth century carried the same sentiment. For example, superintendent William A. Mowry of Rhode Island urged, "Let the doors of the school house be open to all the children; and the child once started on the career of learning, let him not find those doors ever closed against him." Mowry insisted that educating all children equally would benefit communities, and he envisioned that "your bootblack to-day may be your lawyer to-morrow" (cited in Ravich, 2000, p. 19-20).

The Prussian influence. Nor was America the first country to recognize the importance of centralized, compulsory, public education. In 1843, educator Horace Mann changed the landscape of American education by promoting a model of education he observed in Prussia (Gatto, 2003; Tyack, 1974). The Prussian model of education that was so attractive to Mann can be traced back to 1806, when the professional soldiers of the Prussian army suffered a major defeat at the hands of Napoleon's amateur army in the battle of Jena (Alden, 1998; Gatto, n.d.). In an effort to thwart any future attacks to their pride and power, the Prussians looked to their schools.

Prussian leaders believed that a major restructuring of the school system could produce the well-disciplined, obedient individuals essential to their military organization. In his famous "Address to the German Nation," philosopher Johann Gottlieb Fichte promoted this idea of a new education system that would create a nation of obedient subjects through mandatory schooling (Bowman, 2006; Gatto, n.d.). Gatto observed that following the implementation of the

school system, "Prussian industry boomed. . . . She was successful in warfare and her reputation in international affairs was very high" (n.d., para. 11). It was this Prussia that Horace Mann visited in 1843.

Horace Mann and the common school. Upon his return to the U.S., Horace Mann introduced Prussian education as a highly structured and tightly supervised system that included graded classes with a standardized curriculum (Tyack, 1974). He campaigned for a school that would "be available and equal for all, part of the birth-right of every American child" (Cremin, 1957, p. 8). Mann believed that all children, regardless of their religious, ethnic, social, or economic background, should be educated in a common curriculum. Bringing children of differing backgrounds together and educating them in the same way, he argued, would decrease tensions among social groups (Spring, 1990, p. 74). Traced to these early beginnings, the *common school* came to mean "a school that was attended in common by all children, in which a common political and social ideology was taught" (Spring, 1990, p. 74). In the words of Horace Mann, this universal education would become "the great equalizer," with the promise of eliminating poverty and crime along the way.

One of those who subscribed to Mann's vision and continued his drive to standardize public schools was the superintendent of schools in St. Louis, William T. Harris. Harris taught that "the first requisite of the school is *Order*: each pupil must be taught first and foremost to conform his behavior to a general standard" (quoted in Tyack, 1974, p. 43). Indeed, punctuality and precision, similar to what one might find in the military, became of utmost importance in a modern industrial society and Harris believed that it fell to the schools to instruct students accordingly: "The pupil must have his lessons ready at the appointed time, must rise at the tap of the bell, move to the line, return; in short, go through all the evolutions with equal precision"

(Tyack, 1974, p. 43). The implementation of a uniform course of study and standardized exams did not fall far behind.

As early as 1874, exams were considered a valuable instrument, not only for measuring what students had learned, but also as a means of "motivating" them to do better. Superintendent Samuel King wrote that "these examinations necessitate industry, foster promptness, and encourage pupils to do the right thing *at the right time*" (cited in Tyack, 1974, p. 47; emphasis in original).

The common school movement had become entrenched in this country's education system by the end of the nineteenth century. Most children ages five to thirteen were attending school for at least part of the year. Schools were locally controlled although they taught a fairly common curriculum (Ravitch, 2000). As Ravitch described, "The common schools emphasized reading, writing, speaking, spelling, penmanship, grammar, arithmetic, patriotism, a clear moral code and strict discipline, enforced when necessary by corporal punishment. The values they sought to instill were honesty, industry, patriotism, responsibility, respect for adults, and courtesy" (p. 21).

The common school did not stretch beyond the eighth grade, a level at which access to schooling was not always available or much of a priority. Only professions such as law and medicine required much beyond an eighth grade education (Ravitch, 2000).

This, however, was about to change, as the turn of the century ushered in a new era of industrialization and urbanization, along with a tremendous increase in immigration. Cities were growing, and with that growth came a host of challenges and issues that many sought to correct by focusing attention on education (Ravitch, 2000). Although there was consensus that changes

to the system of education were needed, there was little agreement on what those changes should be.

The Committee of Ten. The National Education Association (NEA) responded to this debate by creating a committee to review the issues and make recommendations (Resnick & Resnick, 1985). This committee, the Committee of Ten, published its report at the close of the nineteenth century. The report focused on disagreements: What subjects should be taught, and to what ages of students? What methods should be used? What should be the significance of a high school diploma? (Resnick & Resnick, 1985; see also Ravich, 2000, p. 41). However, rather than resolve the issues, this report only seemed to raise more questions than it answered.

The Committee of Ten opposed differentiating academic programs according to the possible career or vocational track a student might wish to pursue. They believed that all children could learn and would benefit from an academic education, although this could vary in length according to the individual interests, desires, and aptitudes of the student. Their recommendations went so far as to suggest that colleges admit students who may not have studied such widely accepted standards as Latin or Greek. At the time, to suggest that science, history and modern languages were equal to the classics was considered blasphemous by some (Resnick & Resnick, 1985).

Although educators agreed that the new century should bring with it a new system of education, they were not, for the most part, prepared to adopt the recommendations put forth by the Committee of Ten. The next few decades bear witness to the intensity of the debate, as the opinions were plenty and the agreements were few.

John Dewey. During the beginning of the twentieth century, public schools were made to fit the industrial efficiency of the time. Marshak (2003) wrote that "these 'industrialized

schools' were structured to maximize competition between students and to minimize the depth of relationships between students and their teachers" (p. 229). Opposing this rigidity, the psychologist and philosopher John Dewey became a prominent figure in education circles towards the end of the nineteenth century, with his influence increasing throughout the next thirty years. Dewey wished to replace precision and subject-centered curriculum with more "progressive" and "democratic" education (Pulliam & Van Patten, 1995; Ravitch, 2000; Tyack, 1974). He embraced the ideas of Francis W. Parker, a reformer of the preceding generation, who had "rejected discipline, authority, regimentation, and traditional pedagogical techniques and emphasized warmth, spontaneity, and the joy of learning" (Loss & Loss, 2002, p. 1935).

Dewey soon became the leading advocate for the "progressive education movement," envisioning a student-centered education system that would foster individual growth and differences. He believed that individuals' experiences greatly influenced their education, and his own ideas continued to evolve as a result of his own life's experiences (Pulliam & Van Patten, 1995). Ultimately, he desired to replace the concept of a common academic curriculum with a schooling model that emphasized problems, processes, and experience.

Bureaucracy and measurement. Although Dewey's name had long been associated with progressive reform in education, the movement took on a variety of meanings at the beginning of the twentieth century, many of which greatly diverged from Dewey's views. Along with the common desire to replace traditional schooling, some "progressive educators" also wanted to create stronger central controls for schools, through what Loss and Loss (2002) have called "vertically integrated bureaucracies, curricular differentiation, and mass testing" (p. 1935).

With attention focused on bureaucratic efficiency and a desire to meet the needs of an increasingly diversified student population, the progressive movement took a detour that

included the creation of both an academic and a vocational curriculum. The schools were shifting from simply providing intellectual and moral training to specifically preparing the American workforce (Loss & Loss, 2002). The creation of a vocational track was also recognized as helping assimilate a growing number of immigrants into the American way of life.

Advancements in intelligence testing around this time reinforced this new model of bureaucratic efficiency in the schools. The studies and research of Edward L. Thorndike, a leading educational psychologist and a contemporary of Dewey's, provided strong support in favor of intelligence testing and the categorizing of students. He was not alone. G. Stanley Hall (the first president of the American Psychological Association) and Charles Hubbard Judd (a prominent educational psychologist), among others, insisted that "the secrets of education would be unlocked by the persistent and meticulous efforts to measure every aspect of it" (Richards, 2008, p. 2).

By the beginning of World War I, measurement was at the core of educational research and was vigorously promoted by the newly formed American Educational Research Association (AERA). Although Dewey's works may have received greater attention in the academic community, it was Thorndike's belief about measurement and the education system that really took hold.

Testing in America's Public Schools

Although it intensified within the realm of No Child Left Behind (NCLB), America's infatuation with testing actually dates to 1709, when the oral exam was introduced in the United States (Madaus & O'Dwyer, 1999). Later, as Secretary of the Massachusetts State Board of Education, Horace Mann replaced the oral exam with a written essay exam in 1845 (Madaus & O'Dwyer, 1999). At the time, Mann was responding to resistance from some of Boston's

headmasters and thought "that school-by-school test results would give [reformers] political leverage" (Madaus & O'Dwyer, 1999, Premodern Section, para. 23). Mann intended to publish the results of the written exams in the local newspaper, hoping to hold individual teachers and administrators accountable to the public for poor test results. He felt strongly that written test results would illustrate the disparities among schools and prove that "children could learn, if teachers had taught" (Madaus & O'Dwyer, 1999, Premodern Section, para. 23).

Achievement tests. The written exam coincided with a significant increase in the population, accompanied by a disproportionate increase in more students entering school, and staying longer. Resnick and Resnick (1985) explained that the years 1890 to 1918 saw school enrollment grow ten times faster than population growth (p. 6). Heubert and Hauser (1999) wrote that "enrollments more than doubled as waves of immigrants created a newly diverse student population" (p. 31). To accommodate this growth, a uniform system of education became widely popular—accompanied by the need to ensure that appropriate content was being taught and the desired competencies were being learned. It wasn't long before the written exam entered the academic scenery as the acceptable means of assessing student achievement.

The uniformity and standardization inherent in this new model of education seemed a viable and popular alternative to a seriously lacking educational system. Relying on tests was seen as both fair and efficient (Heubert & Hauser, 1999, p. 31). Tyack (1974) compared education in this period to manufacturing: "Like the manager of a cotton mill, the superintendent of schools could supervise employees, keep the enterprise technically up to date, and monitor the uniformity and quality of the product" (p. 41). Soon the practice of using a written exam not only to measure the effectiveness of teachers and schools, but also as a means of determining

appropriate grade promotion of students, gained momentum throughout the country (Jones, Jones, & Hargrove, 2003; Madaus & O'Dwyer, 1999; Tyack, 1974).

Although some school superintendents liked Mann's idea of publishing test results in the local newspaper as a way to encourage accountability and, thus, improve the quality of what was being taught, the practice was short-lived (Madaus & O'Dwyer, 1999). A spirit of contention and competition among teachers seemed to be the unintended consequence. Reportedly, teachers went "to great lengths to protect their reputations, urging children to withdraw from school shortly before the examination and even advising the superintendent to suspend slow students for trivial offenses, so that they would not drag down the percentage of promotions" (Tyack, 1974, p. 48). Nevertheless, the written exam itself remained unscathed in its perceived value as a tool for accessing how much the student had learned—although the results of such tests were no longer shared with the public.

The pioneering efforts of Edward L. Thorndike were largely responsible for popularizing the achievement testing movement in the beginning years of the twentieth century. In developing the subject-matter tests, wrote Corbett and Wilson (1991), Thorndike and his colleagues influenced "what was taught and how it was taught." They also provided a "mechanism to classify students and further standardize the schooling process" (Corbett & Wilson, 1991, p. 18).

At the time, other leading public school crusaders agreed that the traditional heterogeneous grouping of students should be replaced with a system that categorized students by age and ability (Jones et al., 2003). They planned to accomplish proper classification of students through "a uniform course of study and standard examinations" (Tyack, 1974, p. 45).

Additional significant contributions to the testing movement occurred during the first few decades of the twentieth century as a direct result of the need for better efficiency and ease with which to test larger groups of students (Jones et al., 2003). E. A. Kirkpatrick suggested changes that would be more time efficient; Arthur Otis, in developing an exam that would test large numbers of recruits for the Army, created one that could be easily administered and scored; and Frederick Kelly suggested the concept of norm-referenced scoring as a means of easily categorizing students. Essay exams were gradually being replaced by short-answer test questions, followed closely by the multiple-choice test items that were first introduced by Kelly shortly after World War I (Madaus & O'Dwyer, 1999).

For administrators, it seems, switching to the new testing was a no-brainer. Ravitch (2000) explained that "most school districts adopted the new standardized tests because they were more objective than teacher-made tests, faster to administer and grade than essay examinations, and produced comparable results for different students and classes" (p. 131). With the swell in the immigrant population and the increasing access to and demand for public education, standardized testing became a way of determining whether all children were receiving a comparable education (Jones et al., 2003).

Following World War II, the student population continued to increase while additional attention focused on curriculum offerings. This was particularly true during the 1960s, when, in the words of Jones et al. (2003), "social, cultural, and racial disruption . . . opened the doors for new curricula, including student choice for courses and programs of study" (p. 14). A short time later, public education came under severe scrutiny as the curriculum, attempting to reach an increasingly broad number of students with varying degrees of ability and interests, was viewed as being "watered down" (Jones et al., 2003). Test scores also began to decline during the 1960s

and 70s, causing "many people to question the direction in which American education was headed" (West & Peterson, 2003, p. 4).

The late 1960s saw the creation of a new federally funded test, the National Assessment of Educational Progress (NAEP), whose purpose was to "provide a comprehensive picture of overall student performance" (West & Peterson, 2003, p. 4). Although the test was intentionally designed so that conclusions could not be drawn about the performance of any one particular school, the results of the test did confirm what many already suspected: the academic performance of students was declining (West & Peterson, 2003).

Intelligence testing. Walking hand-in-hand with the development of academic achievement tests was the intelligence test. Whereas achievement tests measured what students had learned, the intelligence test was designed to measure what students were capable of learning.

Alfred Binet created the first credible intelligence test on behalf of the French government in 1905 (Corbett & Wilson, 1991; Ravitch, 2000). Although Binet believed that a person's score on the intelligence test could change over time, many who followed after him had a differing opinion. Many psychologists and educators, including Lewis Terman, a Stanford University psychologist responsible for coining the term "intelligence quotient" (IQ), believed that a person's IQ was largely inherited and remained constant throughout life, and that these tests could calculate a person's natural intelligence (Ravitch, 2000).

Henry Goddard, the research director at a school for the "feebleminded" in New Jersey, also believed these tests were an effective means of determining a person's fixed intelligence (Corbett & Wilson, 1991; Ravitch, 2000). Further, like many others who supported the use of intelligence testing, Goddard was a staunch believer in eugenics, the idea that the human race

could be “scientifically improved . . . by careful selection of the best specimens and elimination of the weakest” (Ravitch, 2000, p. 134). Although the eugenics movement had a substantial following of prominent educators during the beginning of the twentieth century, it also attracted extremists and racists, who championed policies that would control propagation of those considered to be less intelligent. The movement eventually lost credibility and support largely due to the rise of Germany’s Nazism in 1930 (Gatto, 2003).

The intelligence test did enjoy a short stint in the Army, as it became a widely accepted tool for classifying new recruits while the country prepared for World War I. However, for a number of reasons, its utility as a military instrument was short-lived.

Nevertheless, even as intelligence testing was being severely criticized as an appropriate tool for the classification of army recruits, the popularity of intelligence testing in academia was gaining momentum. “By the mid-1920s,” wrote Ravitch, “psychologists had created more than 75 different tests of mental ability for students of all ages” (2000, p. 137).

Intelligence testing, of course, nicely complemented the popular movement of the time that encouraged the placement of students into different groupings according to their abilities. The single-track curriculum was quickly being viewed as obsolete, while the use of intelligence testing went initially unchallenged.

The College Entrance Examination Board (CEEB). The College Entrance Examination Board was established in 1900 and, unlike intelligence testing, initially focused its attention on determining what students had learned rather than what they were supposedly capable of learning. Even at that time, the most vocal critics of the exams offered by the College Board were concerned that the exams were shaping the secondary school curriculum (Ravitch, 2000; Resnick & Resnick, 1985).

Eventually, however, the College Board was influenced by the intelligence testing movement and, in 1926, introduced the first Scholastic Aptitude Test (SAT), an instrument for measuring a student's aptitude rather than achievement (Resnick & Resnick, 1985). By 1941, all other college entrance exams were discontinued and the SAT stood alone as the sole instrument for measuring a student's aptitude for college (Ravitch, 2000).

Educational Testing Services (ETS). Towards the end of 1947, the New York State Board of Regents granted a charter, incorporating five individuals and their associates under the name of Educational Testing Service (ETS). Three corporations were represented: the Carnegie Foundation, the American Council on Education, and the College Entrance Examination Board (Nairn, 1980, p. 2). The response to ETS was substantial. "IBM (International Business Machines), Pepsi Cola Corporation, the Association of American Medical Colleges, Harvard University, the U.S. Department of State, the U.S. Atomic Energy Commission, and more than fifty leading universities, foundations, government agencies and corporations greeted the new testing firm with advisory services and contracts" (Nairn, 1980, p. 2).

Henry Chauncey, the first president of ETS, defined the mission of ETS as serving "American education by providing tests and related services . . . which will aid in the guidance of students and in their self-understanding, and which will lead to proper selection and placement of students, not only within but also at the end of the educational process" (cited in Nairn, 1980, p. 3). Even though Chauncey believed that the military and private industry would also benefit from the ability of ETS to provide testing and related services, it was with education that he felt ETS would have the greatest impact. Chauncey further predicted that "we will become accustomed to [testing] and will find ourselves better off for it" (cited in Nairn, 1980, p. 4).

Elementary and Secondary School Reform

The Soviet Union's launch of Sputnik in 1957 signaled a message to Americans that resounded for years: the belief that the academic achievement of students in the United States was falling behind that of students in other countries (Amrein & Berliner, 2002; U.S. Department of Education, National Commission on Excellence in Education, 1984). The country viewed Sputnik as a major humiliation and, maybe for lack of a better target, focused the majority of the blame on a universal lack of rigor and standards in the American schools.

The Elementary and Secondary Education Act of 1965 (ESEA). Although the Eisenhower administration pumped funding into the defense and education budgets, the next major commitment toward improving America's schools did not occur until 1965, with the creation of the Elementary and Secondary Education Act. This act came about as a result of Lyndon Johnson's "War on Poverty" and provided special funding (Title I) to low-income children, schools, and communities. Section 201 of the Elementary and Secondary School Act explained the rationale:

> In recognition of the special educational needs of low-income families and the impact that concentrations of low-income families have on the ability of local educational agencies to support educational programs, the Congress hereby declares it to be the policy of the United States to provide financial assistance . . . to local educational agencies serving areas with concentrations of children from low-income families to expand and improve their educational programs . . . which contribute to meeting the special educational needs of educationally deprived children. (cited in Schugurensky, 2002, para. 2)

A Nation at Risk. When then-Secretary of Education, T. H. Bell, created the National

Commission on Excellence in Education in 1981, the frenzy created by the launch of Sputnik almost 25 earlier had been re-ignited. The country's perception was that "what was unimaginable a generation ago has begun to occur—others are matching and surpassing our educational attainment" (U.S. Department of Education, National Commission on Excellence in Education, 1984, p. 5). Any gains that had been made toward improving America's education system since Sputnik were seen as minimal. The Commission concurred that "history is not kind to idlers" (U.S. Department of Education, National Commission on Excellence in Education, 1984, p. 6). Their report was passionate in its plea:

> Our Nation is at risk. Our once unchallenged preeminence in commerce, industry, science, and technological innovation is being overtaken by competitors throughout the world. This report is concerned with only one of the many causes and dimensions of the problem, but it is the one that undergirds American prosperity, security, and civility. We report to the American people that while we can take justifiable pride in what our schools and colleges have historically accomplished and contributed to the United States and the well-being of its people, the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people. (U.S. Department of Education, National Commission on Excellence in Education, 1984, p. 5)

It took the commission 18 months to prepare their report. The concerns expressed were sufficient to re-focus attention on the public schools, with a demand for accountability. Included with their recommendations was a call for higher, measurable standards: "We recommend that schools, colleges, and universities adopt more rigorous and measurable standards, and higher expectations, for academic performance and student conduct" (U.S. Department of Education,

National Commission on Excellence in Education, 1984, p.73). This report may have actually been one of the first to focus attention on high-stakes testing for grade promotion. Recommendation B, #3 suggests that the administration of standardized testing take place between one level of schooling to another in order to determine a student's academic proficiency (U.S. Department of Education, National Commission on Excellence in Education, 1984).

Since that time, there has been a surge in the use and acceptance of high-stakes testing as a means of determining student competencies (Abrams & Madaus, 2003; Boardman & Woodruff, 2004; Goertz & Duffy, 2003; Gulek, 2003; Heubert & Hauser, 1999; Hoffman, Assaf & Paris, 2001; Lin, 2002; Plake, 2002). As explained by Heubert and Hauser (1999), "the logic seems clear: Unless we test students' knowledge, how will we know if they have met the standard?" (p. 13).

Goals 2000: Educate America Act. "A Nation at Risk" (1984) was followed a decade later by "a stimulus to improving learning and teaching in public schools through the establishment of national education goals" (Thorn & Mulvenon, 2002, p. 195). Goals 2000: Educate America Act of 1994 outlined eight educational goals that were to take place prior to the year 2000. Included in the goals was an emphasis on academic competency. Students in grades 4, 8, and 12 were to demonstrate minimum competency in a number of subject areas including English, mathematics, government and science before moving on to the next grade level. This, it was felt, would ensure that all students would be better prepared to face the future challenges of responsible citizenship, employment, and additional education (North Central Regional Educational Laboratory, 1994).

The Improving America's Schools Act of 1994 (IASA). Following closely on the heels of Goals 2000 was the Improving America's Schools Act of 1994 (IASA; see Rudalevige, 2003).

This was a reauthorization of the Elementary and Secondary Education Act of 1965 (ESEA), whose focus had been to change "the way we deliver education, encouraging comprehensive systemic school reform, upgrading instructional and professional development to align with high standards, strengthening accountability, and promoting the coordination of resources to improve education for ALL children" (U.S. Department of Education, 1965, Introduction section, para. 1).

ESEA, Goals 2000, and IASA were all intended to provide state and local school districts with the support necessary to assist all students in achieving high academic standards. Imbedded in each of these attempts at improving the nation's schools lay four key principles, which all state and local school districts had been encouraged to remember for decades: "1) high standards for all students; 2) teachers better trained to teach to high standards; 3) flexibility to stimulate local initiative coupled with responsibility for results; and 4) promoting partnerships among families, communities and schools" (U.S. Department of Education, 1965, Introduction section, para. 3).

IASA relied heavily on the states to decide what subjects should be included in testing and at what proficiency levels to determine the academic progress. While IASA provided states with a large amount of flexibility, progress was slow (Rudalevige, 2003). The experience of introducing new testing in Virginia demonstrates some of the challenges with the newly mandated assessment.

Virginia's Standards of Learning (SOL's). In 1998, the state of Virginia began implementing changes to their curriculum in preparation for high-stakes testing. Poor performance at the elementary school level would result in either retention or required attendance at summer school. Beginning with the graduating class of 2004, seniors were required to pass six "Standard of Learning" (SOL) exams in order to receive a diploma (Dorgan,

2004).

Seeking to determine the impact of SOL's on teaching and learning, Dorgan (2004) used interviews, discussions, and observations as she studied the efforts of a small elementary school in Virginia over the 1999-2000 school year. One of the first steps, initiated by the school principal, was to form a curriculum committee whose assignment was to align the school's curriculum with what was being tested. Content previously taught in the classroom was removed from the curriculum if it was not included on the tests, and "pacing guides" were established for teachers to follow so that everyone was introducing new content in the same sequence and spending the same amount of time on each area of study.

Dorgan (2004) noted that the greatest frustration for teachers was the challenge "to prepare all students to take the same test at the same time in May" (p.7). All of the teachers interviewed in this study believed that children differ, both in terms of how they learn and their attitude toward learning. One teacher expressed it this way:

> Some of them might be (more confident if they slowed down). Some of them just take a longer time to get a particular skill. We're expecting them all to get it done at the exact same time. They all can't do it that way! (cited in Dorgan, 2004, p.8)

As the time for taking the SOL's approached, everyone at the school was engaged in the process:

> Teachers reviewed and provided practice tests. They drilled students on skills. Parents received letters notifying them about the upcoming tests, and teachers lectured students about the importance of doing their best.
>
> The school was uncommonly quiet during the test period, which stretched over most of a week. As a visitor to the building, I was asked not to venture into hallways

> where students were testing. Custodians adjusted their daily routines so as not to have the sound of their work interfere with students' concentration. Resource teachers cancelled or rescheduled their art and physical education classes. All were attuned to the task at hand. (Dorgan, 2004, p.11)

When the test results were in, scores had improved in all but Grade 5 English, which showed only a slight decline. Although the teachers seemed pleased that their efforts had paid off, they wondered at what cost (Dorgan, 2004).

Massachusetts Comprehensive Assessment System (MCAS). A statewide look at testing, this time in Massachusetts, illustrates the cost of testing in another way. The state of Massachusetts also underwent a number of education reforms in the 1990s, with the current mandated assessment being the Massachusetts Comprehensive Assessment System (MCAS), a set of achievement tests administered every spring (Bolon, 2001).

Initially, the Massachusetts Board of Education felt that adverse long-lasting consequences would result from placing too much weight on test scores for the purposes of accountability (Bolon, 2001). Nevertheless, the board began developing a testing program that would eventually become MCAS. Beginning in 1998, all students in grades 4, 8, and 10 were expected to take the MCAS (Vogler, 2002) and, starting with the graduating class of 2003, expected to obtain minimum scores on the English and Mathematics sections in order to graduate from high school (Bolon, 2001).

After conducting a study on the MCAS, Bolon (2001) was especially critical. He felt that if Massachusetts was really serious about determining the competencies of the typical student, it would implement a number of different assessment instruments rather than rely solely on test questions focused mainly on assessing the skills and knowledge of the exceptional student.

Portfolios, projects, and other examples of work product would provide for better understanding of the general student's academic abilities. Bolon suggested that the $25 million price tag on developing MCAS may, in part, explain why "members of the Massachusetts Board of Education remain rigid, programmatic and hostile to facts that do not support their policies" (p. 42).

Haney (2002) conducted another study of the MCAS, in which he analyzed school average scores on the MCAS Grade 4 mathematics tests for the years 1998-2001. His findings did not support the use of the MCAS as a good indicator of school quality. He determined that "schools that show relatively large gains in score averages from one year to the next tend to show losses the following year" (p. 18). This he attributed to the fact that the makeup of students differs every year. This is especially true in smaller schools, where "having a few especially test savvy, or not so savvy, students may skew results from one year to the next" (p. 18).

Haney (2002) further claimed that the unreliability of the MCAS is a direct result of its questionable technical merit: "I was surprised to find many poorly worded questions" (p. 18). He also identified "some questions as having wrong answers, some as having more than one correct answer and some as misaligned with the Massachusetts curriculum frameworks" (p. 18).

While the MCAS was purported to be a criterion-referenced test, Haney's (2002) study reveals that test items were selected "in terms of item difficulty and discrimination [which] is standard practice for norm-referenced tests of aptitude, ability and achievement" (p. 22). He explained that lower test results are always the case when questions that are answered correctly 70% of the time in the pilot study are removed from the test. Haney concluded, "results from state tests, like that in Massachusetts, ought not to be used in isolation to make high-stakes decisions about students or schools" (p. 1).

Similar to Haney's study in Massachusetts was Kane and Staiger's 2002 study on math and reading test scores for nearly 300,000 students, grades 3 through 5, in North Carolina. From their findings, Kane and Staiger determined, like Haney, that "a school's mean test score will vary from year to year, simply because the particular sample of students in a given grade differs" (p. 240). They attributed such variation on two things: "the variance in test scores in the population of students from which a school is drawing and the number of students in a particular grade" (p. 240).

Utah's Performance Plus. In 2003, the Utah State Legislature passed Senate Bill 154, which charged the Utah State Board of Education with the task of implementing competency standards for graduation by focusing on core curriculum, increasing requirements for graduation, and implementing more challenging courses in high school curriculum (Utah State Board of Education, 2004). This directive resulted in the creation of a statewide competency-based student achievement plan titled Performance Plus (see Appendix A), and the establishment of a statewide accountability system, the Utah Performance Assessment System for Students (U-PASS; see Appendix B). Ongoing and frequent assessments are significant features of Performance Plus and U-PASS.

Included in the duties of the State Board of Education as stated in Utah State Code 53A-1-603, Subsection (2) (b), is the administration of criterion-referenced tests (CRTs; see Appendix C) in all districts for grades 2 through 12 to measure basic skills in core curriculum courses (Utah State Office of Education, 2006a). It should be noted that revisions have recently been made to accommodate the English language learner: "Beginning in 2008, the UALPA [Utah Academic Language Proficiency Assessment] will replace all CRT proficiencies and progress for Levels Beginner (B), Pre-Emergent (P), and Emergent (E) if the student has been in the United

States for less than three years" (Utah State Office of Education, 2006c, p. 3).

Although grade retention does not occur very often and varies from district to district, frequently with parent input (D. Johnson, personal communication, May 30, 2007), the opportunity for retention as a result of these criterion-referenced tests is clearly inherent in the state code: "Scores on the tests and assessments required under Subsection (2) (b) shall be considered in determining a student's academic grade for the appropriate course and whether a student shall advance to the next grade level" (Utah State Office of Education, 2006a, sec. 4, p. 86).

Further, Utah State Code 53A-1-611 states that a basic skills competency test (see Appendix D) will be administered in the 10th grade, upon which the awarding of a high school diploma will be based: "A student must pass the basic skills competency test, in addition to the established requirements of the state and local boards of education of the district in which the student attends school, in order to receive a basic high school diploma of graduation" (Utah State Legislature, 2002, Title 53A). However, according to a report by the Center on Education Policy, "Utah policymakers decided not to withhold diplomas as planned in 2006, opting instead to note on student's diplomas whether they passed or did not pass the state competency exam" (Kober et al., 2006, p. 1). It was recently confirmed with the Utah State Office of Education that, rather than being denied graduation, high school students now have one of the following designations on their diploma: "Passed the UBSCT" or "Did not pass the UBSCT" (D. Johnson, personal communication, June 13, 2007; Utah State Office of Education, 2007).

Texas Assessment of Academic Skills (TAAS). Out of all the statewide programs created to increase academic proficiency, probably none has garnered more attention than the programs in Texas. Even before the commission reported on "A Nation At Risk," Texas was

busy implementing their own programs. In 1980, the Texas Education Agency (TEA) adopted the Texas Assessment of Basic Skills (TABS), a series of criterion-referenced assessments designed to test basic skills competencies in mathematics, reading, and writing. These assessments were to be administered during the third, fifth, and ninth grades, and were based around statewide learning objectives created by the TEA (Cruse & Twing, 2000). By 1983, the Texas State Legislature amended the Texas Education Code to require all ninth graders who did not pass TABS to retake the exam every year thereafter until they passed. "In addition, for the first time in Texas, the results of the TABS testing for each campus (school) and district were released to the public" (Cruse & Twing, 2000, p. 328), forcing the schools to be accountable to the public in a very tangible way.

In 1984, the Texas Educational Assessment of Minimum Skills (TEAMS) took the place of TABS when "minimum" basic skills replaced "basic skills competencies" and the administration of the exam was expanded to include grades 1, 3, 5, 7, 9, and 11. The exit-level exam was now being given in the 11th grade, with remediation and retesting being required of all students who failed the test. "For the first time in Texas, individual students were denied diplomas based, in part, on their performance on the TEAMS test" (Cruse & Twing, 2000, p. 329).

The assessment program in Texas was to experience another overhaul in 1990 when additional significant changes were implemented. TEAMS became the Texas Assessment of Academic Skills (TAAS) with "an expansion of the content being measured, with more content being directly linked to the core curriculum's EE [Essential Elements], and greater emphasis on the assessment of problem-solving skills" (Cruse & Twing, 2000, p. 329). Since its beginning, TAAS has experienced minor modifications including the expansion of testing in reading and

mathematics to include grades 3 through 8, and grades 4 and 8 in writing. The exit-level exam was also moved from Grade 11 to Grade 10 to allow more time for remediation. Hoffman et al. (2001) summarized the significant decisions made as a result of this increase in testing:

> As the amount of testing has increased, so have the consequences. . . . For students, high school graduation is dependent on successful performance on TAAS. For schools and districts, accreditation is dependent in large part on TAAS performance. For principals and teachers, performance ratings and merit raises are influenced by TAAS performance of their students. (p. 483)

Evidence seemed to show that the high-stakes consequences of TAAS might have, indeed, been responsible for a significant improvement in students' academic performance (McNeil & Valenzuela, 2001). Although offering cautions about interpreting the data, Haney (2000) enumerated the gains:

> Between 1994 and 1998, the percentage of students passing the three grade 10 TAAS tests had grown from 52% to more than 70%. . . . Statistics from the Texas Education Agency showed that over the same interval dropout rates declined steadily. Finally, in 1997, release of results from the National Assessment of Educational Progress (NAEP) showed Texas 4th graders to have made more progress on NAEP math tests between 1992 and 1996 than those in many other states participating in state NAEP testing. (Recapping the Myth, section 8.2, para. 2)

Such significant gains over a relatively short period of time did not go unnoticed. Many referred to this as the "Texas Miracle" and suggested that other states would do well to implement the Texas model.

No Child Left Behind

In January 2002, President George W. Bush signed into law the No Child Left Behind Act (NCLB). In part, this was a response to the belief that "despite hundreds of programs and hundreds of billions of dollars invested during the last generation, American students still lag behind many of their fellow foreign students and the academic achievement gap in this country between rich and poor, white and minority students, remains wide" (U.S. Department of Education, 2002, p. 9). If Texas was able to make such major strides in improving student's academic achievement, it must have seemed logical to use the TAAS as a model for designing legislation that would include the rest of the country. With both Former President George W. Bush and then-Secretary of Education, Rod Page, hailing from Texas, many people surmise that this is exactly what happened (Darling-Hammond, 2004; Harvey, 2003; Neill et al., 2004). NCLB became a reauthorization of the ESEA, which had been written almost forty years earlier.

The basic premise of NCLB is that all children can learn and should be given the same opportunity for learning, regardless of race, color, or socioeconomic level. As such, the law is based on four guiding principles: (a) accountability by schools and teachers, (b) more flexibility and greater local control, (c) more choices for parents, and (d) a focus on pedagogical methods that work (U.S. Department of Education, 2002). NCLB intends to improve the academic competencies of all students by concentrating on raising test scores, allowing parents more educational options for their students, and requiring teachers to be better qualified (Darling-Hammond, 2004).

At first glance, it is hard to take issue with an educational policy whose intent is that no child be left behind. Linn (2003) wrote that the policy "is also praiseworthy for the emphasis on all children and particular attention it gives to promoting the learning of groups of students that

have lagged behind in the past" (p. 4). The promise that all children in this country have access to a free system of public education has a history dating back to Thomas Jefferson. And almost all educational policies since Jefferson's time have intended to enhance an education system that, in some way, seemed to be falling short. Wood (2004) analyzed that, like many before them, the "proponents of NCLB undoubtedly felt they were doing the right thing for our children and our public schools." In this case, "the right thing" seemed to be "increasing funding for schools that serve the poor; ensuring that every child would be taught by highly qualified teachers; and holding schools that take federal funds accountable for raising achievement of every student by 'disaggregating' their achievement data" (p. ix).

In May 2005, then-Secretary of Education, Margaret Spellings, publicly recommitted to the tenets found in NCLB: "I stand here today to re-affirm that *all* children deserve a high-quality education. . . . We must give them all a chance. It is the imperative of the 21st century" (p. 2).

Most opponents of NCLB are not opposed to school reform. As Theodore Sizer (2004) puts it:

> We agree on most of the ends. Where we disagree is on many of the means, including those embedded in NCLB. Some of us believe that these not only dodge today's major problems of educational excellence and democratic fairness, but, perversely, make them worse. (p. xxii)

Annual statewide testing. There is little question that the NCLB spotlight shines squarely on the accountability provision and its charge to all schools to increase academic proficiency by the year 2014. Rudalevige (2003) described this major component of NCLB as "a 'tripod' made up of standards, tests that measure whether those standards have been reached, and

penalties or rewards linked to performance on the tests" (p. 25).

In this provision, NCLB requires every state to develop the same "rigorous" academic standards for all students and to administer statewide tests annually in math and reading in grades 3 through 8 to make sure these standards are being met. Test scores are to be reported in the aggregate and should be categorized by race, ethnicity, socioeconomic status, and disability. At least 95% of the students from each group must take the test (U.S. Department of Education, 2002).

Any state that accepts Title I funding must agree to another battery of tests as well: the National Assessment of Educational Progress (NAEP) tests in reading and mathematics at grades 4 and 8. According to Chapman (2004), "NAEP tests help to identify states where standards and tests in these subjects are not sufficiently rigorous" (Some Key Terms in NCLB section, para. 6).

Adequate yearly progress (AYP). An underlying assumption of No Child Left Behind is that "for too long, America's education system has not been accountable for results and too many children have been locked in underachieving schools and left behind" (U.S. Department of Education, n.d., b, The Facts About . . . Measuring Progress section, para. 1). To address this concern, one of the major provisions of the NCLB legislation is that all states are required to report on whether their schools have made "adequate yearly progress" (U.S. Department of Education, 2002). Adequate yearly progress (AYP) is based, of course, on whether a certain percentage of students from a particular school pass "the test."

In designing tests, states are required to create objectives that are consistent with the conditions outlined in NCLB and that will assist them in reaching and maintaining AYP. Linn, Baker, and Betebenner (2002) outlined the rules:

1. States must develop AYP statewide measurable objectives for improved achievement by all students and for specific groups: economically disadvantaged students, students from major racial and ethnic groups, students with disabilities, and students with limited English proficiency.
2. The objectives must be set with the goal of having all students at the "proficient" level or above within 12 years (i.e., by the end of the 2013-2014 school year).
3. AYP must be based primarily on state assessments, but must also include one additional academic indicator.
4. The AYP objectives must be assessed at the school level. At the end of 2 years, schools that have failed to meet their AYP objectives for 2 consecutive years will be identified for improvements.
5. School AYP results must be reported separately for each group of students identified above so that it can be determined whether each student group met the AYP objective.
6. At least 95% of each group must participate in state assessments.
7. States may aggregate up to 3 years of data in making AYP determinations. (p. 3)

In summary, statewide tests are to be aligned to the state's curricular standards and are meant to determine a proficiency level for each grade and subject area. All schools must make AYP, as defined by their individual state. The problem is that there is still much dispute about what AYP is actually telling us.

One challenge to AYP is in determining the "proficient" level (Hess, 2003; Linn et al., 2002; Ryan, 2004). With the definition of "proficiency" being left up to each state, there is some divergence on where each state has established the proficiency level. There are those who

believe that some states have set the standard too high and others too low (Darling-Hammond, 2004; Hess, 2003; Ryan, 2004). The bottom line seems to be that, regardless of the approach taken, the variation among states in addressing what is proficient has certainly caused many to question its validity (Darling-Hammond, 2004; Linn, 2002; Pedulla et al., 2003; Ravitch, 2010; Ryan, 2004).

Sanders (2003) explained that "the AYP is not a value-added measure of progress for individual students, but rather it requires cross-cohort comparisons of the percent of students meeting the proficiency standards this year compared to last year's percent proficiency" (p. 1). Schools and districts fail to make AYP if their percentage of proficient students is not significantly greater than the previous year's and may face sanctions if they fail to make AYP for more than two consecutive years. As noted, this can be frustrating for schools with high student turnover and for small schools, where year-to-year variations in student ability can skew AYP numbers.

Controversy aside, the message inherent in requiring all schools to meet AYP is that previous progress was not being made fast enough and that educators can and should do better. Similar to the way many teachers in the public school system reward good academic performance among individuals in the classroom (i.e., with gold stars, smiley faces, prizes, praise, etc.), teachers, schools and school districts are encouraged by NCLB to perform well. If not, their name goes on "the chalkboard," or in this case, in a published report for the whole world to see. As Linn and colleagues (2002) articulated, "the notion is that, given enough pressure from the accountability system and some additional resources, the schools will improve and the goals will be met" (p. 16).

According to the NCLB website, improvements are being made. In 2006 the U.S. Department of Education wrote that "the long-term Nation's Report Card (NAEP) results, released in July 2005, showed elementary school student achievement in reading and math at all-time highs" (para. 3; see Appendix E). The Performance and Accountability report issued by the U.S. Department of Education in 2007 mentioned other noteworthy accomplishments since the passage of NCLB: accountability and assessment plans are in place in all 50 states for third through eighth grades, academic achievement is improving, and there are indications that the achievement gap is decreasing.

For their own report on student achievement, the Center on Education Policy (CEP; 2008) studied academic achievement trends in all 50 states since 2002. Their findings also show that, in most states, the percentage of students performing at the proficient level in reading and math has increased (see Appendix F). Nevertheless, they are quick to point out that it is impossible to determine whether these trends are related to the passage of NCLB because so many "interconnected policies and programs have been undertaken to raise achievement" (p. 2).

Although the proficiency trends, as reported by CEP, seem to be increasing, there are still a substantial number of schools that do not appear to be performing up to standards (see Appendix G). The CEP report, and others like it, only add to the discussion on whether or not NCLB is truly working.

Assessment

For years testing has been accepted as a fundamental component of teaching and learning. Black and William (1998) defended testing: "We start from the self-evident proposition that teaching and learning must be interactive. Teachers need to know about their pupils' progress and difficulties with learning so that they can adapt their own work to meet

pupils' needs—needs that are often unpredictable and that vary from one pupil to another" (The Argument section, para. 1). Ideally, teachers use the data from tests to develop or revise effective pedagogical methods to accommodate their students' needs (Ferrandino & Tirozzi, 2002). Ohanian (1999) confirmed the importance of the feedback to teachers provided through testing: Most educators understand that "just because you taught it you can't assume students learned it" (p.45). Sacks (1999) similarly pointed out that testing students on a regular basis can help their teachers understand what has been learned.

Of course, there are various ways for a teacher to determine a student's progress in the classroom, including observation, discussion, participation, and homework assignments, in addition to testing (Neill, 1997). But testing may have the advantage of serving multiple purposes at once. Beyond teachers' classroom use, Heubert and Hauser (1999) identified seven ways in which student assessments are used, "with the same test often serving multiple functions":

> 1) for identifying individual competencies or academic proficiency of students, often for the purpose of grouping, class advancement, or discovering special needs; 2) to help determine academic performance over time and among different subgroups; 3) to motivate teachers, schools and students to perform better; 4) to determine program effectiveness; 5) to hold teachers, schools, and districts accountable for student achievement; 6) to influence change in classroom instruction; and 7) to certify a minimum level of student proficiency for the purpose of grade promotion or high school graduation. (p. 33)

Similarly, Haladyna, Nolen, and Haas (1991) identified 29 ways in which standardized test scores are used (see Appendix H), ranging from "policy analysis at the national level to parental

review of their child's achievement" (p. 2).

After explaining the possible usefulness of testing, particularly testing with high stakes, Jones, Jones, and Hargrove (2003) offered a caution: "While on the surface testing appears to be a simple means to attain these outcomes, its practical implementation has proven much more difficult" (pp. 10-11). One problem is the "do-or-die" nature of high-stakes assessments. Some educators believe that, in order to be effective, assessments must be integrated into the teaching process, allowing students more than one attempt to demonstrate their proficiency. Guskey (2003) insisted that "to become an integral part of the instructional process, assessments cannot be a one-time, do-or-die experience for students. Instead, assessments must be part of an ongoing effort to help students learn" (p. 4). In other words, students deserve multiple chances to "demonstrate their new level of competence and understanding"—and evaluating their performance after a "second chance" can provide important clues about "the effectiveness of the corrective instruction" they received (Guskey, 2003, p. 4). Guskey noted that professional training in most fields offers multiple, low-risk ways for trainees to practice and evaluate their skills. By contrast, "only in schools do student(s) face the prospect of one-shot, do-or-die assessments, with no chance to demonstrate what they learned from previous mistakes" (p. 4).

Providing feedback. Ideally, academic assessments should provide information to both the student and the teacher (Bransford, Brown, & Cocking, 2000; Galley, 2001; Guskey, 2003; Krumboltz & Yeh, 1996; Sharkey & Murnane, 2003). Teachers should be able to use meaningful feedback from academic assessments to help guide the students' thinking. Too often, by the time a grade has been received from an assignment or exam, the student has already moved on to another topic or project (Bransford, Brown, & Cocking, 2000). Thus, the feedback becomes useless as a tool to help the student improve their understanding. Sharkey and Murnane

(2003) complained about this lag, warning that "if schools do not make constructive use of their test results, the net effect of the No Child Left Behind legislation will likely be a reduction in student learning" (Sharkey & Murnane, 2003, p. 80).

The process of assessment should also provide the teacher with information that would be useful in planning the next steps in learning (Neill, 1997). "Good teachers continually assess student progress using a variety of techniques. . . . This ongoing flow of information enables teachers to respond to students immediately with corrective feedback" (Rettig, McCullough, Santos, & Watson, 2003, p.73).

The assessment instrument. In terms of utility, many educators agree that testing is especially useful when it is used in identifying students who have special strengths or weaknesses, rather than as an instrument to rank and categorize students (Powell, Farrar, & Cohen, 1985). Bersin (2005) asks a question that serves as a reminder that testing should not overshadow the fact that it's really the teaching and learning that should matter most: "Of every program in place, and every reform proposed, we must ask: Does it improve student achievement?" (p. 30). Rettig and colleagues (2003) stated it this way: "Those who complain about accountability miss the main point. Our mission is student success" (p. 71).

In designing an assessment instrument that would accurately measure school achievement, Chudowsky and Pellegrino (2003) suggest asking the following questions:

1. What are the most essential domains of knowledge and skill that students should master in school to be productive members of society?
2. What kinds of performance differentiate beginning, competent, and expert learners in each domain?

3. What are the central conceptual structures within each domain that students must grasp to successfully move on to higher levels of understanding?
4. What are typical difficulties or misconceptions that learners have in each domain that, if identified early, could be remediated with instruction? (p. 76)

Popham (2003) identified five characteristics of an "institutionally useful test" (p. 48). First, he suggested that the curriculum being measured should be *significant.* It should be worthwhile information as opposed to "trifling knowledge as esoteric scientific terms or dates associated with obscure historical events" (p. 48). Secondly, the test should measure something that is *teachable*; in other words, something that the teacher can reasonably teach to the majority of students, rather than something that requires special skills or innate intelligence. A useful test, Popham continued, should also be one that "provides or is directly based on sufficiently clear descriptions of the skills and knowledge it measures so that teachers can design properly focused instructional activities" (p. 49). Fourth, the results of a test should provide specific enough information that the effectiveness of the instruction is evident. Popham's fifth and last characteristic of an instructionally useful test instrument is that it not be so time consuming that it excessively supplants instructional activities.

Standards and Standardized Tests

Since 1965 and the creation of the Elementary and Secondary Education Act (ESEA), testing has played a significant role in the federal support for education, especially for low-income students. Beginning in the 1970s, the minimum competency movement made its appearance, giving standardized achievement tests a visible and viable place in the annals of American education (Heubert & Hauser, 1999).

Standardized testing. Since that time, the most popular and widespread use of the term "standardized test" refers to a norm-referenced, multiple-choice achievement test (Koretz, 1996). Nevertheless, there is some misunderstanding over exactly what is meant by "standardized." According to Koretz (1996), "standardized" does not necessarily refer to test format and content, but rather "to the fact that the test items, administrative conditions, and test scoring are uniform to all students" (p. 174). The requirements as set forth in NCLB reflect that definition: "The tests called for in the new law are to be *standardized* statewide tests—that is, the same tests are to be administered to all the state's students in the same way" (Popham, 2004, p. 17).

National standards. Although there is much discussion on what the standardized testing should be, public support for the widespread use of academic "standards" does not appear to be in question. Phelps (2005) reported that " decades of surveys reveal that a very large majority of parents want standards to matter and want them to be enforced uniformly" (p. 21). Even before NCLB, there was much discussion on the benefit of having national standards for improving the American education system (O'Neill, 1993).

In 1992, the National Council on Education Standards and Testing attempted to clarify what a national system of education standards would look like:

- National standards should be developed that include *content standards* (what students should know and be able to do), *student performance standards* (the level(s) of student competence in the content), and *system performance standards* (to assess the success of schools, districts, states, and the nation as a whole in helping all students attain high performance standards). In addition, the Council said that states should develop *school delivery standards* to judge whether schools are providing students with the opportunity to attain these high standards.

- A national assessment system linked to these standards should be developed. The assessment system would feature two components: large-scale sampling assessments, which would be provided by the National Assessment of Educational Progress, and other assessments capable of producing results for individual students. In other words, no *single* national test would be created, but multiple assessments would be linked to common national standards to provide a complete portrait.
- The National Education Goals Panel would appoint members to a new group, called the National Education Standards and Assessments Council. The Council would establish guidelines and . . . would certify content and student performance standards and criteria for assessments linked to those standards. (O'Neill, 1993, p. 5)

Reeves (2001) insisted that standards—and standards based assessment—"can be overwhelmingly positive for the thinking, reasoning, and communication skills of students" (p. 5). He identified four key advantages of using academic standards in the classroom:

> First, student performance is compared to a standard rather than to a norm, an average, or to other students' performances. Second, students are required to demonstrate their proficiency. This is an improvement over traditional assessment because guessing a response has little or no value for the teacher or student. Third, standards-based assessments are not veiled in secrecy; what is expected of students is clear to teachers, parents, and students. The gamesmanship typically surrounding test preparation is replaced with a focused curriculum that relates to academic standards. Fourth, the focus of standards-based assessments is the improvement of student learning, not merely the rendering of an evaluation and the announcement of a score. (p. 6)

A common fear among standardized testing advocates is that, without standards, the quality of educational opportunities will continue to fall along socioeconomic lines (Nash, 2000). Other educators believe that standards are a good idea and that they can work, but only if certain changes are made to our current education system. For example, Chase (2000) gave the following conditions:

> First of all, the standards must reflect the wisdom of parents and classroom teachers. Second, the curricula we teach must be aligned with the new standards. Third, teachers must be provided the professional development they need to incorporate the new standards into their teaching practice. Fourth, we all must insist that no single high-stakes test can measure the academic progress of any student—that multiple indicators must be employed. (p. 41)

Common Core State Standards Initiative. The first step towards creating national academic standards has already taken place. On June 2, 2010, after a year-long process, the National Governor's Association Center for Best Practices (NGA Center) and the Council of Chief State School Officers (CCSSO) announced the creation of state-led education standards in English language arts and mathematics. Although these standards are not part of a federal mandate, it appears that states are quickly jumping onboard. In August 2010, the Utah State Board of Education joined more than 30 other states in approving the common core state standards in language arts and mathematics for grades K-12 (Schencker, 2010). It is anticipated that other states will follow.

The standards, developed in collaboration with teachers, school administrators, and other educational experts, solicited initial feedback from a wide range of invested parties, including civil rights groups, post-secondary educators, students with disabilities, and English language

learners. A revised draft of the standards was then made available to the public for additional input.

Key points for reading, writing, speaking and listening, language, and media and technology as well as for mathematics are listed on the Common Core State Standards website. In essence, these standards:

- Are aligned with college and work experience;
- Are clear, understandable and consistent;
- Include rigorous content and application of knowledge through high-order skills;
- Build upon strengths and lessons of current state standards;
- Are informed by other top performing countries, so that all students are prepared to succeed in our global economy and society; and
- Are evidence-based. (Common Core State Standards Initiative, 2010)

The NGA Center and CCSSO acknowledge the importance of implementation and they plan to work together to assist states and districts in this process. Eventually, state education leaders are also anticipating the replacement of current tests with new statewide tests that will be more aligned with the new standards (Schencker, 2010)

Impact of High-Stakes Testing

Although high-stakes have been associated with achievement testing since the 1960s, only with the passage of NCLB has high-stakes testing been placed front and center on the school reform agenda throughout the country (Robbins, 2006). "High-stakes testing" simply refers to "tests that have serious consequences for students, teachers, schools, and/or school systems" (Jones et al., 2003, p. 2). By contrast low-stakes tests have "no significant, tangible, or

direct consequences attached to the results, with information alone assumed to be a sufficient incentive for people to act" (Heubert & Hauser, 1999, p. 35).

There are many who believe that high-stakes testing is necessary in order to motivate both teachers and students to "work harder and learn more" (Amrein & Berliner, 2002, Arguments in Support of High-Stakes Testing section, para. 1). In addition, some believe that high-stakes testing will bring attention to those who need assistance in reaching their goals of academic achievement (Kornhaber & Orfield, 2001). Advocates further believe that high-stakes tests "are good measures of the curricula; . . . provide . . . an equal opportunity for all students to demonstrate their knowledge; and . . . are good measures of an individual's performance, little affected by differences in students' motivation, emotionality, language, and social status" (Amrein & Berliner 2002, Arguments in Support of High-Stakes Testing section, para. 2).

The discussion is still continuing on whether high-stakes testing has a negative or a positive impact on student performance, the curriculum and instruction, and the public education system in general. High-stakes testing advocates believe that such exams are an accurate measurement of student performance and a good safeguard that prevents unprepared students from advancing inappropriately. On the other side of the aisle, critics feel that high-stakes testing has a foundation mired in false assumptions and, further, that such testing practices may increase the number of high school dropouts (Natriello & Pallas, 1998).

Reporting on their study of the effects of high-stakes testing on student learning, Amrein and Berliner (2002) connected high-stakes testing to the Heisenberg Uncertainty Principle: "The more important that any quantitative social indicator becomes in social decision-making, the more likely it will be to distort and corrupt the social process it is intended to monitor." In other words, "attaching serious personal and educational consequences to performance on tests for

schools, administrators, teachers, and students, may have distorting and corrupting effects" (The Heisenberg Uncertainty Principle Applied to the Social Sciences section, para. 1). Kohn (1999) concurred, "The more a test is made to 'count' . . . the more that anxiety is likely to rise and the less valid the scores become" (p. 76).

An additional concern identified in the research is whether high-stakes test scores are a true indication of what the student has learned (Amrein & Berliner, 2002; McNeil & Valenzuela, 2001). The problem, according to Amrein and Berliner (2002), is threefold: (a) test items are, at best, only a sampling and rarely represent the entire arena of content to be learned; (b) testing time is limited and restricts the opportunity to really assess what has been learned; and (c) test data may be polluted as a result of the teacher narrowing the curriculum to only that which is being tested.

Teaching to the test. In his 1988 study, Madaus identified one of the unintended consequences of high-stakes testing as its "major influence on what takes place in the classrooms, often resulting in an emphasis on test preparation that can compromise the credibility or accuracy of test scores as a measure of student achievement" (quoted in Abrams & Madaus, 2003). It seems to be a popular notion that when high stakes are imposed upon test scores, the ante increases and teachers are more apt to narrow the curriculum in order to "teach to the test" (Abrams & Madaus, 2003). According to Firestone and Schorr (2004), "'teaching to the test'. . . is a term with many meanings, but the implication is . . . that teachers are doing something special to help students do well on a test, often without helping them to better understand the underlying subject matter" (p. 2). This undue attention given to individual test questions rather than to the knowledge they are supposed to represent challenges the validity of test results.

There are many who would agree that there are unanticipated and unfortunate consequences resulting from high-stakes testing (Amrein & Berliner, 2002; Domenech, 2000; Dorgan, 2004; Gulek, 2003; Jones et al., 2003; Karp, 2004; Krumboltz & Yeh, 1996; Lin, 2002, McNeil, 2000; Pedulla, 2003; Popham, 2004; Rotberg, 2001; Smith, 1991; Williams, 2003). Karp (2004) asserted that "when schools become obsessed with test scores, they narrow the focus of what teachers do in classrooms and limit their ability to serve the broader needs of children and their communities" (p. 57). Linn (2003) pointed out that whenever high stakes are attached to testing, additional attention seems to be given to teaching the test material to the exclusion of any subject not covered on the test:

> It is no surprise that attaching high stakes to test results in an accountability system leads to a narrowing of the instructional focus of teachers and principals. There is considerable evidence that teachers place greater emphasis on material that is covered on a high-stakes test than they do on other material. . . . Furthermore, the concentration on tested content areas often comes at the expense of content domains that are not tested, such as science, history, geography, and the arts. (p.4)

Similarly, Rotberg (2001) explained that "the tests themselves become the curriculum " (p. 170).

In 2002, Amrien and Berliner reported on their study in which 18 states (Alabama, Florida, Georgia, Indiana, Louisiana, Maryland, Minnesota, Mississippi, Nevada, New Jersey, New Mexico, New York, North Carolina, Ohio, South Carolina, Tennessee, Texas, Virginia) were examined to see how student learning was impacted by high-stakes testing. They reported that "in all but one analysis, student learning is indeterminate, remains at the same level it was before the policy was implemented, or actually goes down when high-stakes testing policies are instituted" (p. 2).

Not all researchers agree, however. In their rebuttal of Amrein and Berliner's 2002 study, Raymond and Hanushek (2003) stated, "The controversial nature of high-stakes testing has led to the hurried release and dissemination of research that lacks scientific rigor" (p. 50). They found the Amrein and Berliner study to be seriously flawed due to "inappropriate methods" (p. 53) and concluded that, when properly administered, the research paints a much brighter picture for accountability: "Rigorous analysis reveals that accountability policies have had a positive impact on test scores during the past decade" (p. 50). Either way, high-stakes testing influences not just the students who take the tests but also the teaching they receive in preparation.

Impact on the quality of teaching. Years before NCLB, Mehrens (1998) asserted, "if stakes are high enough and if content is deemed appropriate enough by teachers, there is likely to be a shift in the curriculum and instruction to the content sampled by the test" (p. 11). This was reinforced with the 2003 national survey of teachers, when the National Board on Educational Testing and Public Policy concluded that the higher the consequences, the more impact testing has on the curriculum. As a result of the survey, it was determined that greater test preparation, both in time and focus, is common among teachers in states that have implemented some form of high-stakes testing (Pedulla et al., 2003, p. 5).

Nevertheless, the opinions are split as to whether this impact on the curriculum is good or bad. On one side are those who believe that teaching to the test, and the resulting narrowing of the curriculum, negatively impact the quality of teaching that goes on in the classroom. Popham (2004), for example, argued that the high cost of the current emphasis on test preparation is a much lower quality academic experience for students.

Between 1972 and 1980, a time when accountability testing was being instituted in the public schools, the National Center for Educational Statistics (NCES) collected data showing "a decline in the use of student-centered discussions, extended writing, and project or laboratory work" (Darling-Hammond, 1997, p. 56). As opposed to many foreign countries who rely heavily on essay and oral exams and examples of work product, schools in the United States consistently use basic skills tests that are of the commercially prepared, norm-referenced, multiple-choice variety (Darling-Hammond, 1997). As a result, some educators believe that we have sacrificed in-depth, high-level teaching and learning in order to cover the material that will be included on the test (Darling-Hammond, 1997; Dorgan, 2004; Kohn, 1999; Pedulla, 2003; Ravitch, 2010; Thompson, 2001).

Strategies used in teaching mathematics, for example, seem focused on helping students memorize facts as opposed to helping them understand number relationships. The shortage of time on extended problem-solving further erodes "the development of understanding" (Darling-Hammond, 1997, p. 86). Kohn (1999) agreed: "Once teachers and students are compelled to focus only on what lends itself to quantification, such as the number of grammatical errors in a composition or the number of state capitals memorized, the process of thinking has been severely compromised" (p.76).

Even if the process of thinking and understanding were not negatively impacted, some feel that high-stakes testing does not present an accurate picture of what students know and says very little about the quality of the education they are receiving (Meyers & Rust, 2000; Ravitch, 2010; Rotberg, 2001; Smith, 1991). They feel that tests are naturally limited since they can only assess a sampling of what the student has learned. Score inflation results when curriculum is narrowed to only cover that which is to be tested. Neill (2003) wrote that score inflation "largely

explains the difference between results on state-mandated tests and those on such neutral measures as the NAEP" (p. 44).

Fans of high-stakes testing offer a much different perspective. They suggest that teaching to the test is exactly what should be happening—a renewed focus by teachers on the established and desired curriculum. In other words, "teaching to the test" is actually "teaching to the curriculum" (Posner, 2004, p. 749).

Another frequently expressed concern is what impact the focus on high-stakes subjects has on student proficiency in low-stakes subjects, such as science and social studies. If most class time is spent in preparing students to do well in reading and mathematics, what happens to student achievement in other subjects? This is the question posed by Winters and colleagues in their 2008 study of Florida's high-stakes testing system. Although there are many who feel that, as a result of high-stakes testing, students have been short-changed when it comes to subjects outside of reading and mathematics, Winters was quick to point out that "there is currently very little empirical evidence of the impact of high-stakes testing policies on measured student proficiency in subjects that are not part of the accountability system" (Winters, Greene, & Trivitt, 2008, p. 2). What their study of Florida's system did suggest, however, is that the emphasis and incentives associated with high-stakes testing actually resulted in greater achievement in low-stakes subjects. This they attributed to the fact that (a) reforms were adopted as a result of high-stakes testing that improved the overall quality of instruction, and (b) acquiring a mastery of high-stakes subjects sharpened basic skills that resulted in a mastery of other subjects as well.

Impact on teachers. Studies from around the country relay teachers' feelings about high-stakes testing. A teacher in Texas admitted, "I was so afraid of getting a bad evaluation

from my former principal that I compromised what I believed to be good teaching" (Bussert-Webb, 1999). Another teacher stated, "I was really angry because so many of the things I had taught those children about math were not on the test. . . . And, indeed, the following year they did extremely well in the third grade . . . a great many of my children were in the advanced classes doing better than some of the children who had scored higher than they had on the ITBS" (cited in Smith, 1991, p. 8). Still another teacher said, "With the testing programs we have in this school, there isn't much leeway for me to be creative or innovative and create excellent lessons. The test is the total goal" (cited in Jones et al., 2004, p. 37).

There is evidence from the research that cheating is prevalent, and not just among students. Sacks (1999) wrote, "test scores have become so politically charged that some teachers, in addition to spending huge amounts of time teaching to tests, have resorted to cheating to make their numbers look good" (p. 14-15).

Teachers are also prone to direct student-learning experiences more closely when high-stakes consequences are involved. Amrein and Berliner (2003) explained, "high-stakes tests cause teachers to take greater control of the learning experiences of their students, denying their students opportunities to direct their own learning. When the stakes get high, teachers no longer encourage students to explore the concepts and subjects that interest them" (p. 32).

On the flip side of the argument is the suggestion that, for too long, teachers have not been accountable for what has taken place in their classrooms. In addition to accountability, NCLB mandates that "all teachers teaching in core academic subjects are highly qualified" (U.S. Department of Education, 2002, p. 13). This does not seem like an unreasonable request. But how is a "highly qualified" teacher defined? Credentials alone do not necessarily make a good teacher. Under NCLB, testing has become a way of measuring not only the student learner, but

the instructor as well. Regardless of how teachers feel about being evaluated in this way, most would agree that the recent surge in high-stakes testing has helped to focus the nation's attention and has been "a driving force behind fundamental change within schools" (Abrams, 2004, p. 1).

Impact on learning. One of the major assumptions attached to testing is the belief that the higher the stakes, the more motivated a student is to work harder to do well on an exam (Kohn, 1993). In their 2002 study of states that have implemented high-stakes testing, Amrein and Berliner (2003) discovered just the opposite:

> Unfortunately, the evidence shows that such tests actually decrease student motivation and increase the proportion of students who leave school early. Further, student achievement in the 18 high-stakes testing states has not improved on a range of measures, such as the National Assessment of Educational Progress, despite higher scores on the states' own assessments. (p. 32)

Stiggins (1999) explained that the pressure to do well on a test can have negative consequences for students: "Increasing pressure to score high on tests, combined with a lack of focused opportunities to learn, can lead to a sense of futility—a feeling of hopelessness—that can cause [students] to stop caring and stop trying. . . . Those who stop believing that they are capable of learning will stop trying" (Unwanted consequences section, para. 4). Another educator stated it this way: "Our public schools are here to teach, support and encourage our students to maximize their potential, not to penalize and destroy their future at the age of 17" (Domenech, 2000, Ignoring Counsel section, para. 6).

Test anxiety is a very real issue for many students, and can greatly impact the results of high-stakes tests (Gulek, 2003; Haladyna, et al., 1991; Supon, 2004). Haladyna, Nolen, and Haas (1991) refer to this as "test score pollution." In addition to the way in which high stakes may

impact test results, they believe that many testing practices affect the integrity of exam results and score interpretation. They conclude, "Until there is serious reform in the way schools prepare students to take standardized achievement tests, test results will continue to misrepresent American public education and its accomplishments" (p. 6).

Clearly, the issues surrounding accountability systems, test scores, teachers, curriculum, and students is a complicated one. The fact that there is still so much debate on the consequences of high stakes testing, both good and bad, may indicate the need for more research.

Summary

The volume of literature on educational issues and school reform speaks to America's interest in education. To quote, once more, from DeMille (2000): "Americans love education, believe in education and pay big money for education" (p. 19). Commissions have been created, research has been conducted, new programs and institutions have been established, all with the intent of supporting school reform that will result in better schools and a more robust learning environment, available to all children.

Unfortunately, school reform encompasses such a huge arena that some of the country's most respected and experienced educators cannot agree on the best way to proceed. The dialogue has continued for decades without consensus on the best way to ensure equal access to a high quality education. Although discussion may be healthy, it will not solve the problem. Indeed, educators can offer thoughtful opinions and well-researched findings, but legislators make the decisions regarding the American school system—not the professionals who are in the classroom every day.

The increase in charter schools, private schools and homeschooling across the country also speaks to the disenchantment many parents and communities have with the current

education system and the public schools to which their children are assigned. Although these efforts are commendable, creating more alternatives for only a select few is not the answer, nor will it result in long-term benefits for America's children. Public schools, accessible to all youth, should be the focus of attention.

In theory, accountability makes sense. Districts, schools and teachers should bear responsibility for what takes place in the classroom, and testing seems a reasonable method of determining teacher effectiveness. Yet the way policymakers and administrators currently use high-stakes testing seems more punitive than informative. Placing so much weight on a one-time, year-end, single assessment instrument is not in the best interest of teachers, students, or education in general. If the intent is really is make sure that all children have access to the same high quality education, our thinking needs to change.

Chapter Three: Research Design and Methodology

Introduction

Performance Plus and UPASS are Utah's attempt at helping students reach the academic destination NCLB has charted for public education. To this end, they are intended to do more than just provide data on the academic performance status of Utah's youth. Significant benefits, particularly with regard to increased academic proficiencies, are expected as a result of the attention given to test results. It is hypothesized that high-stakes testing does have an impact, whether good or bad, on teaching and learning in the school. Exploring how elementary school principals view the impact of high-stakes testing in their schools is the intent of this study.

If the aim of high-stakes testing is simply to raise test scores, the outcomes would be fairly easy to obtain and evaluate. Comparing test scores from one year to the next is not hard to do. What is not certain is whether test scores tell the whole story. Beyond the question of whether higher test scores are a true measurement of increased knowledge acquisition, the literature suggests that the focus on test results is bearing unintended consequences.

It does seem evident from the research that educators, parents, and even policymakers really do want the same thing: a "good" education for *all* students. They just don't always agree on which path to take to get there. What does good mean, and what will indicate that good has been attained? Since high-stakes testing has permeated academic culture, particularly through NCLB, more study is needed regarding the influence such testing has on teaching and learning in the public school sector.

Purpose of the Study

The primary purpose of this study is to examine the impact of high-stakes testing on teaching and learning in the public elementary school setting in the state of Utah. Specifically,

this study will explore elementary school principals' beliefs about the influence high-stakes testing has had on (a) teaching strategies (i.e., the style or method of teaching or the manner in which a teacher instructs), (b) the curriculum (e.g., course or subject offerings, time spent on certain subjects, elimination or reduction of other subjects, etc.), and (c) student achievement.

As mentioned above, it would be easy enough to compare this year's test results with those of previous years. This, in fact, is the focus of the National Assessment of Educational Progress (NAEP), an appendage to the National Center for Educational Statistics (NCES). NAEP assessments are conducted in all 50 states and in a number of subject areas. The results of NAEP are also presented by gender, race, and ethnicity, and socioeconomic level (U.S. Department of Education, 2008). A product of their research and assessments, the Nation's Report Card provides information on "subject-matter achievement, instructional experiences, and school environment for populations of students (e.g., all fourth graders) and groups within those populations (e.g., female students, Hispanic students)" for both elementary and secondary students throughout the country (U.S. Department of Education, 2007).

However, national assessments and test results alone are not sufficient to paint the entire picture. As indicated in the literature review, test scores have increased in many schools. Most often, however, a follow-up question would be: "What does this mean?"

Research Objectives

To explore the impact high-stakes testing has on teaching and learning (i.e., teaching strategies, curriculum and student achievement) in the state of Utah, from the perspective of public elementary school principals, the following research questions have been developed:

1. According to the opinions of Utah public elementary school principals, how have teaching strategies been influenced, if at all, by high-stakes testing?

2. According to the opinions of Utah public elementary public school principals, how are students being prepared for high-stakes testing, and have the course curriculum and related school activities been influenced, if at all, by the time spent in preparing students for taking these tests?
3. According to the opinions of Utah public elementary school principals, how has student achievement been impacted, if at all, by high-stakes testing?
4. Are principal perceptions regarding high-stakes testing significantly affected by school demographics or whether the school demonstrated adequate yearly progress (AYP)?

These questions provide the thematic foundation upon which further questions have been developed. As is the nature of a qualitative study, additional questions were identified and posed during the interview process as the need arose.

Research Design

There are many approaches to research inquiry and different options that could have been taken in the type of study suggested here. This study will use a qualitative approach to develop "insights and generalizations out of the data collected," as suggested by Neuman (2000, p. 122).

Participants. Teacher's opinions are often solicited with regard to the impact high-stakes testing has had on their teaching methods, the curriculum, and the students they teach (Abrams, 2004; Abrams, Pedulla & Madaus, 2003; Bussert-Webb, 1999; Charles, 2008; Cimbricz, 2002; Jones & Egley, 2004; Kaplan & Owings, 2001; Pedulla et al., 2003; Reese, Gordon, & Price, 2004; Shepard & Dougherty, 1991; Taylor, Shepard, Kinner, & Rosenthal, 2001). Another group that shares in teacher accountability are the school principals, who carry major responsibility for improvement in the classroom. Most, if not all, principals have already experienced the classroom setting and have acquired a good understanding of the variety of

challenges teachers face each day. In addition, through their administrative responsibilities, they have not only an opportunity, but also the responsibility, of looking beyond the individual classroom or teacher to consider a more holistic point of view.

As the educational leaders for their institutions, principals are responsible for promoting an "instructional program conducive to student learning" (National Policy Board for Educational Administration, 2008, p. 1). To do so, principals must help create a strong, coherent curricular program and learning environment for their students, along with developing and monitoring an "assessment and accountability system to monitor student progress" (National Policy Board for Educational Administration, 2008, p. 1). At a time when schools are being held accountable to improve student proficiency, everything under the principal's stewardship, from curriculum planning to professional development to the use of resources and the implementation of new educational programs or materials, could impact student academic achievement.

The subjects for this study are 12 elementary school principals from public schools in the state of Utah. According to the Utah State Office of Education website, there were 514 elementary schools in the state of Utah for the 2008-2009 school year. This does not include either charter or private schools in the state (Utah State Office of Education, 2009, p. 5).

Socioeconomic levels. Implicit in the NCLB legislation is the idea that, when it comes to education, low-income students are at a disadvantage compared to their middle- and upper-income classmates. According to the research, the percentage of students reaching proficiency levels is lower in high-poverty areas (U.S. Department of Education, 2002). Although a number of factors contribute to a student's or school's academic performance, NCLB seeks to help ameliorate the disparity through Title I funding, which was first introduced with the Elementary and Secondary Education Act of 1965.

Title I is specifically targeted to assist schools that have a significant number of economically disadvantaged students, with the intent that the funding be used to balance out the scales by providing "additional instructional staff, professional development, extended-time programs, and other strategies for raising student achievement" (U.S. Department of Education, 2002, p. 13). These poverty levels are determined annually by the number of students eligible for free or reduced-price lunch (A. C. Frost, memorandum to Chief State School Officers, March 18, 1996).

All children enrolled in an elementary or secondary school are eligible to apply for free- and reduced-price benefits. Eligibility is determined by the federal poverty guidelines. Students whose family income is above 130% but below 185% of these guidelines are entitled to reduced-price meals. To be eligible for free meals, a student's family income must be at or below 130%.

Although the percentage of students eligible for free and reduced lunch is used to determine the distribution of Title I funds, not all schools who meet the criteria are recipients of Title I funding. School districts have been given discretion in the disbursement of these funds. Kreig Kelley, Utah Title I Program Specialist, described how disbursement works:

> The District determines which schools receive Title I funds. However, a school must have a poverty level equal to or higher than the district poverty level to qualify for Title I funds. The district must provide Title I funds to schools in poverty level rank order, not skipping schools. The law allows for an exception—districts may choose to focus their Title I funds to certain grade-level groupings such as elementary schools or middle schools or high schools as long as all schools with 75% poverty are served. The district can decide how many schools receive Title I funds. Districts with high poverty schools may provide Title I funds to fewer schools that qualify for Title I . . . In districts with

higher poverty, the law allows districts to serve additional schools (below district average poverty) down to 35% poverty.

Additionally, the law also allows school districts with unusually small schools that receive additional state and federal funding to not be included in the rank order of served Title I schools if that school receives as much or more funding (per poverty student) when compared to Title I schools. (personal correspondence, 2009)

Clearly, administering Title I funds is a complicated endeavor. Because the authors of NCLB believed that the income level of families has some effect on students' access to quality education, this study will explore whether principals' experiences were different at schools with low family income levels. Included in the sampling are principals from Title I schools, principals who come from non-Title I schools but who have a high percentage of students eligible for free and reduced-price lunch, and principals from schools in wealthier areas. Whether the principals in this study believe that the funding provided to Title I schools has contributed to their students' achieving proficiency levels will be explored.

Adequate yearly progress (AYP). All public schools in the state of Utah are required to participate in the Utah Performance Assessment System for Students (U-PASS), Utah's AYP Plan. High quality, standards-based, criterion-referenced tests (CRTs) that are course and grade specific are included in U-PASS and are used to report academic achievement (Utah State Office of Education, 2005). The Utah State Board of Education has defined four proficiency levels for the Core CRTs (see Table 3.1).

A combination of participation rates, academic achievement, and attendance or graduation is used to determine whether individual subgroups, schools and districts have made AYP (Utah State Office of Education, 2005). All student subgroups, schools and districts in the

Table 3.1

Utah Student Achievement Levels

Levels	Descriptor	Federal levels
Level 1: Minimal	Not yet proficient on measured standards and objectives of the core curriculum in this subject; minimal understanding and application of key concepts	Basic
Level 2: Partial	Not yet proficient on measured standards and objectives; partial understanding and application of key concepts	Basic
Level 3: Sufficient	Proficient on measured standards and objectives; sufficient understanding and application of key concepts	Proficient
Level 4: Substantial	Proficient on measured standards and objectives; substantial understanding and application of key concepts	Advanced

Note. Adapted from "State of Utah Consolidated State Application Accountability Workbook," by Utah State Office of Education, 2005, p. 6. Retrieved from http://www.usoe.k12.ut.us/nclb/documents/pdf/AYPPlan.pdf

state are expected to meet AYP in mathematics and reading/language arts by 2014.

Methodology and Data Collection

Qualitative research lends itself to this study inasmuch as it is intended to "provide an in-depth description of a specific program, practice, or setting" (Mertens, 1998, p. 159)—in this case, by interviewing principals in their schools. Fundamental to qualitative research methods is the assumption that "people, institutions and interactions are involved in producing the realities in which they live or occur" (Flick, 2007, p.12). My task has been to acknowledge the reality of multiple perspectives and then to interpret how principals give meaning to their different experiences and events (Neuman, 2000). With this understanding, I attempted to collect meaningful information that would clarify the consequences, if any, associated with high-stakes testing.

Phenomenological research. Four major types of qualitative research are phenomenology, ethnography, grounded theory, and case study. Briefly, phenomenology describes one or more individuals' experiences of a "phenomenon," with data collection usually entailing 10 to 15 in-depth interviews with individuals. Ethnography is used to understand cultural characteristics, such as behavior, beliefs and knowledge, of a group from participant observation over a period of time. Case study research is used to describe one or more cases (i.e., individuals or groups) in detail and can use multiple methods to do so (e.g., observations, interviews, documents). Grounded theory is used to describe research that generates a theory that is "grounded" in the data, with interviews as well as observations being common means of data collection (Mertens, 1998).

Because the focus of this study is principals' understanding of their own experiences related to high-stakes testing in their schools, this study can be described as phenomenological

research. Mertens (1998) wrote that "typically, phenomenological research asks, what is the participant's experience like? The intent is to understand and describe an event from the point of view of the participant" (p. 169).

Sampling. I made every attempt to have the sampling in this study represent "core cases for the experience, knowledge, practice, etc." (Flick, 2007, p. 29). I used purposive sampling, with principals being selected from eight school districts (Alpine, Granite, Jordan, Murray, Nebo, Provo, Salt Lake, and Weber) in Salt Lake, Utah, and Weber counties. Time constraints and resources prohibited a greater representation of the state's school districts for the purposes of this study. I determined, however, that the principals from these school districts provided a good sampling, inasmuch as these districts include 55% of the elementary-school-aged students in the state of Utah.

In addition to representing the eight school districts mentioned above, the elementary school principals came from public schools that are receiving Title I funding and from schools that are not. A conscious decision was also made to include principals from schools that have both passed and failed AYP. Half of the principals selected represent schools that have passed AYP at least three out of the past five years, while the other half represent schools that have failed AYP at least three of the past five years (see Table 3.2.)

The decision to use three out of the past five years for determining whether schools passed or failed AYP was personal. Two out of three years did not seem sufficient, while five out of seven years didn't seem necessary. Although four of the principals in the sampling represent schools who never failed AYP, all of the schools represented passed AYP at least once since 2002. There are principals from schools who may have passed AYP five years ago, but failed in 2009, and some who failed in 2005 but passed most recently. There are two Title I schools in the

Table 3.2

Sampling of Utah Elementary School Principals

	Title I	Non-Title I
Passed AYP	Principal A1	Principal C1
(3 out of last 5 years)	Principal A2	Principal C2
	Principal A3	Principal C3
Failed AYP	Principal B1	Principal D1
(3 out of last 5 years)	Principal B2	Principal D2
	Principal B3	Principal D3*

* School D3 became a Title I school for the 2009-2010 school year. Because the data used for this study only went to the 2008-2009 school year, the decision was made to categorize D3 as a non-Title I school for the purposes of this study.

sampling that are currently in need of "Program Improvement," meaning that they failed AYP for at least two consecutive years. The intent for including principals from schools that have both passed and failed AYP was to explore whether the pressure to pass AYP may have created undue stress or influenced in any way the principal's opinion of the impact high-stakes testing has had at his or her school.

Although school districts exercise some discretion in determining which of their schools should be the recipient of Title I funding, all of the principals from Title I schools in this study have at least 40% of students eligible for free and reduced lunch. However, two of the six non-Title I schools represented also have over 40% of their students on free and reduced lunch. For the purposes of this study, eligibility for Title I funding is being used to identify principals from schools with a significant population of low-income students. My intent in looking at eligibility for funds in addition to actual receipt of funds is to examine whether principals believe that low-income students are being disadvantaged academically (i.e., are at a disadvantage when it comes to test taking), as well as whether they feel that additional funding from Title I is being used for resources that may influence test results.

Because there were multiple principals from schools that could have been selected in each of the specific categories (e.g., there were nine Title I schools in the Alpine district alone that passed AYP), a conscious decision was made to include a principal from at least one school from each of the eight school districts selected. Although both male and female principals are equally represented, this was not intentional, as gender differences were not a part of this study. Included in the sampling are also experienced principals, with 10 or more years of administrative experience, and those who are fairly new to their position, with three or fewer years as administrators, although, again, I did not intentionally plan for this attribute to be part of the stratification.

Interviews. This study follows the reasoning of Kvale (2007), who asked, "If you want to know how people understand their world and their lives, why not talk with them?" (p. 1). Although researchers can gather much information from a variety of sources, such as surveys and questionnaires, a face-to-face interview can add richness, insight and clarification to any exchange of information not readily available through a simple survey (Lofland & Lofland, 1995). Personal conversations highlight others' experiences, preferences, observations, and histories. In addition, interviewers have an opportunity to ask additional questions to clarify and broaden understanding. They get a chance to dig a little deeper.

Each of the individuals in this study was asked to participate in one 40- to 60-minute interview session. I began each interview by explaining the nature of the research and what the participant could expect during the course of the interview. I reviewed the interviewee's rights as a participant in the study and provided a copy of the Consent to be a Research Subject Form (Appendix I). I also invited the interviewees to ask questions before and during the interview process.

In order to ensure data quality, I followed a standard procedure throughout data collection. I used an interview guide to direct the line of questioning (see Appendix J), attempted to establish a rapport with each participant, sought to avoid language that could be misunderstood or was ambiguous, and made sure that responses were recorded accurately (Frey & Oishi, 1995). In some instances, I used probing questions to solicit additional information or to clarify.

The interview guide became the script for what took place during the interview and contributed both thematically and dynamically to the research topic. Kvale (2007) taught that "a good interview question should contribute thematically to knowledge production and dynamically to promoting good interview interaction" (p. 57). The "what," or thematic questions, normally come before the "how," or dynamic questions, in an interview. In all cases, the questions were brief, simple, and easy to understand, not only for the benefit of the interviewee, but also for the sake of the researcher, who was involved in the transcription and coding process.

During the course of the interview, 37 questions were asked (see Appendix J), including four that were demographic in nature. The questions were divided into three main sections, each related to the impact of high-stakes testing on (a) teachers and instruction; (b) students and learning; and (c) the responsibilities and decisions of the administrator.

Each interview session was audio-recorded with a digital recorder. Emerging themes aided in the formation of additional, probing questions in an attempt to discover additional insight and topics of interest. A second round of interviews was not found to be necessary, although most interviewees invited additional contact for clarification purposes, if necessary.

NVivo. NVivo 2008 computer software was used to help manage, organize and analyze

the data. In addition, I kept a journal to provide an audit trail of the development of insights and ideas. After each interview I captured observations, impressions, meaningful thoughts or key points arising from the data. I then created a log of each interview, along with any related journal entries, demographics or other attributes pertinent to the case. Storing attribute information was particularly useful in making comparisons across the data. Using headings helped me to organize and quickly identify interview questions, major ideas, or themes.

Transcribing. Transcribing took place following each interview session. Kvale (2007) noted that "transcribing from tape to text involves a series of technical and interpretational issues . . . for which there are few standard rules" (p. 95). In an effort to represent the conversation as accurately as possible, I followed the guidelines offered by Bazeley (2007), which I have adapted, below:

- A full transcript will include all ums, mmms, repetitions and the like. Repetitions communicate something about the thinking or emotion of the interviewee.
- In the same vein, don't correct incomplete sentences or poor grammar; it is important to capture the form and style of the participant's expression.
- Note events which create interruptions to the flow of the interview, for example: (tape off), (telephone rings). Note also other things that happen which may influence interpretation of the text.
- Record nonverbal and emotional elements of the conversation, such as (pause), (long pause), (laughter), (very emotional at this point).
- Digressions from the topic of the interview are a controversial issue. The decision about whether or not to include that text centers on whether there is any meaning in the digression.

- Some researchers like to comment on (annotate) the text as they are transcribing it. If you do this, then enclose your comment in unique markers to set it off from the transcribed text (e.g., <<comment>>). (pp. 45-46)

Coding. Following the transcribing of the interviews, NVivo 8 (2008) was used to code the transcriptions. I began the process of organizing and interpreting data by reviewing the transcriptions of the interviews with the intent of discovering common themes, strategies, observations and experiences that could be grouped into major and meaningful categories. This initial procedure for analyzing data is referred to as open coding. Flick (2006) called open coding "the operation by which data are broken down, conceptualized, and put back together in new ways. It is the central process by which theories are built from data" (p. 296). As Gibbs (2007) advised, I examined the text of interviews "by making comparisons and asking questions" (p. 50). Opening up the text in this way helps in sorting the data into major categories.

As suggested by Strauss and Corbin (1998), I reviewed the text "line by line" and "sentence by sentence" as I looked for significant ideas and themes to code. Questions such as who, what, when, where, and how were repeatedly asked to aid in uncovering emerging themes (Gibbs, 2002).

The next stage in the process is axial coding, by which "categories are enriched by their fit with as many passages as possible" (Flick, 2006, p. 301). In this stage, I looked for relationships among categories. The basic tasks involved in axial coding include the following:

- Laying out the properties of a category and their dimensions, a task that begins during open coding;
- Identifying the variety of conditions, actions/interactions, and consequences associated with a phenomenon;

- Relating a category to its subcategories through statements denoting how they are related to each other; and
- Looking for cues in the data that denote how major categories might relate to each other. (Strauss & Corbin, 1998, p. 126)

Following axial coding, which developed and related categories to subcategories, the process of integration and refinement took place during selective coding. In this initial step in integration, I looked for central categories or main themes to emerge, with all of the other major categories being organized around these principal themes.

In NVivo, coding is stored at a node, which is the way in which an idea or concept is attached to its corresponding passage of text (Gibbs, 2002). Nodes need to be clearly defined and can also be linked to other nodes or memos. Relationships between the coded texts at different nodes was explored, along with concepts, ideas, and interpretations of the text. Originally, 91 nodes were identified, which, after removing redundancies and combining common concepts, were synthesized down to 67. After another review of the nodes to search for major themes and patterns, these 67 nodes were eventually organized into 12 main categories or themes.

Informed consent. Each participant was advised of the procedure, risk, and benefits of participating in this study, as outlined by the IRB, and was given the opportunity to ask questions prior to and during the interview.

Limitations. In order to obtain information that was felt to be of most benefit to the study, it was necessary to be selective in the sampling. Rather than using random sampling to select participants, purposive sampling was used.

Participants for this study included principals from 12 public elementary schools located

in eight school districts along the Wasatch front in the state of Utah. Represented in the sample were both Title I and non-Title I schools, as well as schools that have passed and schools that have failed AYP for three of the past five years.

Delimitations. The decision was made to only include principals from Utah public schools for this study. Representation from charter or private schools was not included.

Summary

For the purposes of this study, interviews were conducted with 12 elementary school principals selected from the following eight school districts in Utah: Alpine, Granite, Jordan, Murray, Nebo, Provo, Salt Lake, and Weber. Six of the principals represented Title I schools and six principals came from non-Title I schools. The purposive sampling was also arranged so that schools that have been successful at making adequate yearly progress and those who were less successful were equally represented.

The intent of the study was to determine the extent to which teaching strategies, curriculum, and student achievement have been impacted by high-stakes testing, namely the criterion-reference test (CRT), according to the opinion of elementary school principals. Also investigated was whether the principals' opinions and experiences may have been influenced by Title I funding or the students' success in passing AYP.

A 40- to 60-minute interview was conducted with each principal, after which NVivo 2008 computer software was used to organize and analyze the data. Similarities and common themes were identified and particular attention was given to the variety of experiences and differences of opinions.

Chapter Four: Research Findings

Principal and School Profiles

Again, the principals in this study represented public elementary schools from the following school districts in the state of Utah: Alpine, Granite, Jordan, Murray, Nebo, Provo, Salt Lake, and Weber. Four school districts, Alpine, Nebo, Provo, and Salt Lake, had representation from two schools. Table 4.1 provides information on the basic characteristics of these schools and the identifiers used in this chapter when referencing a particular principal.

Principals. The impact high-stakes testing and the No Child Left Behind (NCLB) legislation is having on Utah's public elementary schools seems to be a topic of great interest to elementary school principals in the state. Each principal interviewed for this study seemed very willing to discuss personal experiences and observations with regards to mandated annual testing and its impact on their school, teachers, and students. They were all gracious and accommodating in answering questions, sharing observations, and voicing opinions on the way their schools experience the impact of high-stakes testing, particularly the criterion-referenced Test (CRT), adequate yearly progress (AYP), and the NCLB legislation.

In an attempt to be sensitive to the demands and time constraints placed on elementary school principals, I invited each principal to determine when and where the interview was to take place. One interview was scheduled after school hours and one was scheduled during spring break, both at the request of the principal, with the intent of avoiding unnecessary interruptions. In another case, the principal requested meeting in the lobby of a local hotel where he was attending a workshop. This interview took place prior to the beginning of his workshop in a hallway corner off the main lobby. All other interviews took place during the school day in the principal's office.

Table 4.1

Elementary School Characteristics

School	Pass	Fail	Title I	Enrollment	Free/reduced lunch (%)
A1	✓		✓	449	77.53
A2	✓		✓	545	48.52
A3	✓		✓	621	76.17
B1		✓	✓	522	54.17
B2		✓	✓	412	77.26
B3		✓	✓	589	92.80
C1	✓			624	24.72
C2	✓			810	20.27
C3	✓			549	16.90
D1		✓		438	37.07
D2		✓		614	61.07
D3		✓	*	762	48.68

Note. Information on enrollment numbers was obtained from the Utah State Office of Education, School Finance and Statistics Report, 2009, retrieved July 2, 2010, from http://www.schools.utah.gov/finance/other/ AnnualReport/ar2009.htm. Information on the percentage of students on free/reduced lunch was obtained from the Utah State Office of Education, Child Nutrition Programs, retrieved July 2, 2010, from http://www.schools.utah.gov/cnp/Files/cnp_data_statistics.html

* School D3 became a Title I school for the 2009-2010 school year. Because the data used for this study only went to the 2008-2009 school year, the decision was made to categorize D3 as a non-Title I school for the purposes of this study.

The initial contact with each principal occurred through e-mail. In four instances, the principals responded back quickly that they would be very willing to meet with me and asked that I call to set up an appointment. Shortly after the email, I made personal phone calls to each principal to set up the time and place of the interview. In some instances, it took multiple attempts to connect with the principal and establish an actual time to meet. In only one instance did the principal refuse my request for an interview. This principal was experiencing a recent reassignment, mid-year, to a school that had a history of failing AYP, and, understandably, she indicated that she was unable to reserve time to be interviewed. To find a principal from a school that met similar demographics and characteristics, I went outside of Utah and Salt Lake counties and included a school from the Weber School District. That interview proved to be a rich and rewarding one.

Although many of the principals who were interviewed shared similar experiences and observations, there were also a variety of differences. Some of these differences may be better understood by learning more about the variety of backgrounds from which these principals come.

With only two exceptions, the principals in the sampling have spent at least 15 years in public education, although six of the principals interviewed have been in administration six or fewer years. Three principals actually have only two years of administrative experience. Nevertheless, I found, without exception, that each principal interviewed shared a commitment to excellence and a strong loyalty and concern for teachers, staff, and students. They were professional, articulate and knowledgeable about the issues at hand and, unless I had asked, I would never have been able to differentiate between the more seasoned administrator and the less experienced one.

All of the principals had a background in elementary education teaching before becoming

principal, while two of them were specifically trained in special education. Three of the principals also had experience working in a middle or junior high Sschool prior to arriving at the elementary school level. When asked, all of them responded that they love it where they are. One principal stated, "I love elementary. It's a great place to be. Junior high is tough (laughs) . . . very, very tough" (C1).

In the course of their academic careers, many of the principals worked in at least one other school outside of their current district, but only one had worked outside the state of Utah. Most of the principals had both a pre- and a post-NCLB experience, either as an administrator or as a teacher, and half of them had the opportunity to work in both a Title I and a non-Title I school. The insight of these principals was particularly helpful when addressing issues related to the relationship between Title I funding and the passing of AYP. Before becoming principals, three of the participants also worked in the district office.

Schools. I worked hard to include principals from schools that represented different demographics and characteristics. As stated earlier, all of the schools represented have, at one time or another, passed AYP since its beginning in 2002-2003, while four of the schools in the sampling have never failed AYP. Two of the Title I schools are currently in the category of "Needs Improvement," with one school in its first year and one school in its second year. One of the schools in "Needs Improvement" was actually categorized as a "passing" school for the purposes of this study since it passed AYP for three of the past five years. However, since the two years it failed AYP were back to back, it came up as "Needs Improvement."

In addition to representing the experience of both passing and failing AYP, principals in this study also represent both Title I and non-Title I schools. This is important because the implications for "failing" AYP are different for Title I schools than for their non-Title I

counterparts. According to NCLB, Title I schools and districts "that do not make sufficient yearly progress toward state proficiency goals for their students first will be targeted for assistance and then be subject to corrective action and ultimately restructuring" (U.S. Department of Education, 2002, p. 10).

Although there are technically no rewards for schools that pass AYP, one of the non-Title I principals interviewed represented a "Blue Ribbon" school. Consequently, this school received national recognition or their academic excellence at an awards ceremony in Washington, D.C.

Consequences of NCLB

Early in this study, it became quite apparent that the question "What impact is high-stakes testing having on teaching and learning in the public elementary schools in Utah?" was not an easy one to answer. Behind every seemingly simple yes/no question was a variety of explanations and clarifications that could usually be illuminated by additional questioning. Such probing was minimal, however, because, with maybe one or two exceptions, the principals often volunteered more than was asked.

In some instances, I obtained clarification by simply allowing the individuals to take their time in answering. For example, when asked whether it was fair to use test results to evaluate teachers, one principal quickly stated, "Most definitely!" and then proceeded to explain why testing was not a good tool, in itself, for determining a teacher's ability. Such apparent contradictions actually turned out to be a rather good illustration of the fact that many of the questions had more than one answer. Allowing time for the principals to think through the question, provide examples, and offer additional explanation and information proved invaluable in truly understanding their perspective and experience.

Although it appears evident from the literature that the No Child Left Behind Act of

2001, and its subsequent mandates, were not very popular with those who work within the system of public education, there are also benefits from the legislation that were mentioned in this study. From my viewpoint as a researcher, it was difficult to ascertain just how much the opinion of these principals may have been influenced by having to live with NCLB for most of their administration. Have they been working with NCLB guidelines for so long that they are simply resigned to its mandates, or do they really subscribe to its principles and policies?

For whatever reason, all of the principals in the sampling gave thoughtful and thorough consideration to both the positive and the negative impact of high-stakes testing and AYP from their current vantage point.

Impact on classrooms. There was no disagreement that the pressure to pass AYP influenced both the principals' responsibilities and what took place in the classroom. From the tendency to want to "teach to the test," to the variety of creative strategies and activities, to the amount of time set aside for test preparation, the desire, by teachers and principals, to make adequately yearly progress touched almost every corner of the education experience.

Impact on principals. Principals in this study agreed that the 2002 legislation has had a definite influence on how they organize and plan their day. Even those principals who only became administrators since the implementation of NCLB and have no basis for comparison recognize the impact high-stakes testing and making AYP has had on their daily responsibilities and assignments. One newer principal observed, "[AYP] has to be what we base a lot of our judgments on and what we're trying to do" (B2). Another principal concurred,

> We rely on them [CRT results] a lot, and maybe more than we should. Even though it's not the only piece of data, I would say that it's the strongest piece . . . most of our improvement plan goals are based around the CRT. (C3)

Principals who've experienced both a pre- and post-NCLB landscape spoke specifically to the changes they have felt since 2002. One principal estimated, "my responsibilities as a principal are probably twice what they were" (B3). Some principals spoke of the increased pressure and consequences associated with high-stakes testing and AYP, in terms of teacher turnover. Principal A1 shared, "In the NCLB era, huge expectations with consequences are present, where a lot of people have left. People want out of the consequences, want out of the mess that it's created." Another principal stated that AYP "influences what is the emphasis for the school; influences professional development; it influences what we talk about in our team meetings." The principal continued, "As much as I might not fully agree with the premise of AYP . . . I don't have the power or wherewithal to say, 'I'm just going to ignore it'" (A2).

Even with the feelings of increased pressure coupled with the negative observations of the consequences of NCLB, many principals did not want to ignore what they view as being some of the positive outcomes of the legislation. Principal A2 said, "I hope I'm a better principal. . . . I do things differently . . . than I did before." "Part of that," this principal explained, "is paying more attention to student achievement. . . . I am paying more attention to the teaching and instruction going on in the classroom." This increased awareness and focus on student achievement and instruction were cited numerous times as a positive benefit of high-stakes testing and AYP. Further, AYP has had a noticeable influence on the School Improvement Plan goals at a number of schools.

Many of the principals gave credit to AYP and CRT results for an increase in dialogue and genuine conversations they are now having with their teachers, regarding ways in which academic proficiency could be improved: "How can we make this stronger? How can we do a better job? What else could we do?" are questions Principal A1 often asks. An additional benefit

seemed to come from the district office. Several principals reported an increase in positive interaction with district administrators and a noticeable feeling of trust, related to the shared goal of passing AYP.

Principals reported that data from CRTs influences their decisions on funding, helping them to focus on gaps in their programs or areas that might be in need of additional attention and resources. "Where can we apply a little more money that might make a difference in how our kids do?" Principal C2 asked. One principal realigned funds to enhance her school's science program when she noticed that scores in science weren't as high as in other areas of the curriculum.

Renewed focus on the classroom. Regardless of whether or not the principal represented a Title I or a non-Title I school, or whether they had passed or failed AYP, every principal, without hesitation, stated similar benefits of NCLB and AYP: No Child Left Behind and AYP have been beneficial in turning their attention to student achievement. For the most part, principals felt that they have become more purposeful and strategic in planning their day, in working with the teaching staff, and in structuring their support systems within the school, with the intent of improving academic proficiency levels. As one principal stated, "No Child Left Behind has had a lot of negative things, but it's caused us to really look at what we're doing in the classroom" (B3).

Principals mentioned being very aware of what is taking place in their classrooms. One principal discussed the time she spent in the classroom "making sure that the instruction's being taught; that they're following the core [and] getting out old materials that no longer apply to the core" (D2). With high visibility, principals may be motivating teachers to keep more focused on core curriculum guidelines. Staying abreast of what's happening in the classroom on a daily, or,

at least, weekly basis may also provide principals with an opportunity to identify any concerns or weaknesses in teaching strategies or classroom protocol before they become problematic. Principal A3 specified that in visiting classes, "I want to see objectives written on the board. I like to see them using strategies that we've been studying."

In most instances, teachers are encouraged to review the CRT data with the intent of improving their instruction for the upcoming year. One Title I principal spoke of the importance of a quick turnaround time in reporting the CRT results:

> As soon as I get the test scores, I create a chart for [teachers] to see . . . what percentage of their kids passed each major area for both math . . . and language arts . . . so they can actually look and say, "How did I do compared to my grade level?" (A3)

Another principal, this time from a non-Title I school, mentioned that some of her grade-level teams will review the data with the intent to know where to begin. She hopes that they are saying, "Here's where the kids are right now; here's where our starting place is" (D3). Another principal agreed that reviewing CRT results has made a difference in improving instruction:

> It's done a lot for improving teacher quality . . . teachers are more in tune to what they know and do not know. . . . I think that that's improved their teaching and, in turn, a benefit for the students. (C2)

The increase in collaboration and working in grade-level teams was another stated benefit of NCLB and AYP. One principal mentioned the emphasis placed on Professional Learning Communities and collaboration in his district: "Our district is very big into Professional Learning Communities and collaboration, and we spend each Monday going over data . . . . That helps us unpack the core and . . . focus students' learning" (C1). Another principal talked about the value of collaboration among numerous staff members and how they use data to improve their

strategies and programs:

> Our staff developer, and literacy coach, and . . . there's a whole team of us that get together and go through the data as well . . . and then we meet with the teachers. . . . We spend a great deal of time going through that [the data] and seeing where the loopholes are and then try to mend those holes up. (B1)

What apparently starts in the fall with the review of CRT data continues throughout the school year. In anticipation of the following year's CRTs and AYP, collaboration meetings take place in many schools each month or, in some cases, weekly, with teachers and staff reviewing formative assessments to determine how their students are progressing and, particularly, to determine which students might benefit from additional attention or remediation, and what strategies to put in place that will help them. Again, it was not uncommon for teachers to check out each other's data and to help each other. Principal C2 praised such collaboration: "The fact that we've started to collaborate and teachers have started looking at areas where different teachers do well and . . . gathering from other teachers, their strengths, and sharing those ideas through collaboration has really made an improvement."

A number of principals mentioned how intimidating it can be for some teachers to share test results with their coworkers. Principals said that teachers are afraid of being looked at as inferior, afraid of not measuring up. In principals' estimation, the teachers who seem most open to sharing information and learning from their colleagues are the ones who seem to be the most confident and successful in the classroom. One principal made this observation:

> The teachers that are feeling comfortable and realize that, "Hey, the CRT is not my grade, it is not my salary," they're feeling comfortable enough to share and to get ideas. Those that are feeling . . . "Don't bug me" . . . haven't quite caught that vision that "If I need

help, I need to go to teachers." (D1)

Open dialogue and discussion among teachers, especially in grade-level teams, is particularly significant for Title I schools who, most likely, have the greatest incentive to direct their attention and resources to the areas with the greatest need. One principal stated, "All of our interventions are a result of looking at the test results" (A3).

The principals agreed that reviewing the CRT results each year has helped teachers to make data-informed decisions about their instruction. Principal C1 spoke of better instruction for better learning:

> I don't believe we're teaching to the test as much as we are trying to look at the core . . . and say, "How can we make sure students are learning . . . not just that they've been taught it, but actually learning it?"

In addition to grade-level teams and encouraging collaboration among their teaching staff, Title I principals, particularly, meet with each other frequently to discuss best practices and what each school is doing to make AYP. They openly share ideas, what is working and what is not. As one principal stated, "We're all in this together" (A1). One example is an emphasis on "double-dosing," or giving students additional opportunities to learn something. In Principal A1's school, double-dosing occurs both inside and outside of the regular school day, and has resulted in very positive outcomes. Other Title I principals in the school district have met together to share ideas on how double-dosing works and how they might adapt it to their particular situation. They discuss strategies and what kind of investment needs to be made in terms of resources, including technical and staff support and time commitments.

Although most teachers see the value of collaboration, some teachers prefer to be left alone to do their own thing, while others are resistant to change. Even though most of the

principals in the sampling, regardless of whether or not they were passing AYP, indicated a strong desire to continue working on improvement, principals from schools who had failed AYP can use CRT data to support needed changes. As one principal explained to his staff, "We're not perfect yet. Until we're perfect, we're going to keep changing, we're going to be finding all the new strategies, and we're going to be talking about how to improve" (D1). He further stated that he wanted his teachers to always be searching for new ideas, to find out what the current research is saying, recognizing that today's students need to be prepared for a world that is different than what their parents experienced at their age. He tells his faculty, "If you retire teaching the same thing you did thirty years ago, we've got a problem" (D1). Evidently, data from CRTs and pressure from AYP requirements are tools principals in this study use for leverage with their staff.

Most of the principals in this study acknowledged that the faculty, for the most part, will do their best to try and cover all of the core information that most likely will be tested, because scores are important to teachers. But these efforts come at a cost: according to these principals, teachers end up trading "depth" for "breadth" because there is simply not enough time to cover all of the information in any detail. This sacrificing of in-depth teaching and higher-level learning in order to cover the material included on the test was also mentioned frequently in the literature (Darling-Hammond, 1997; Dorgan, 2004; Pedulla, 2003; Thompson, 2001).

In summary, principals should know what is going on in their classrooms. Although they cannot be everywhere at once, they should have a general understanding of who is teaching what and how in each of their grade levels. If principals are more actively engaged and aware of what's going on in the classroom as a result of NCLB, that is a positive consequence to the legislation.

Harder to determine is the positive or negative nature of the classroom focus created by AYP. When principals talk about high-stakes testing and the pressure to make AYP creating a renewed focus on the classroom, including better collaboration among grade-level teams and the impetus for change, their expectation seems to be that this should ultimately strengthen and support teaching and learning. However, critics of NCLB could view this increase in time and attention as no more than a continued focus on the ever-present goal of preparing students to do well on the CRTs—the ever-lurking "teaching to the test." And the principals, themselves, mentioned the trade-offs in depth of learning resulting from high-stakes tests. Is this renewed attention on the classroom really encouraging better teaching practices and, ultimately, an increase in learning, or is it simply reinforcing whatever practices might work to meet adequate yearly progress and having all students "proficient" by 2014?

Pressure to pass AYP. Although all of the principals in this study expressed the commitment to work towards making AYP each year, none seemed to feel it as intensely as the principals from Title I schools. Even principals from non-Title I schools recognized the difference in the pressure that is felt by their Title I colleagues. One Title I principal mentioned that his teachers were conducting weekly assessments to make sure expectation levels were being met and to prepare students for the "heavier" test at the end of the year (B1). This kind of regular evaluation places the stress up front and center every week of the school year. This principal also mentioned that the pressure to achieve a certain standard was so intense at times that he had to remind the faculty that "this test is not the most important thing. The kids are the most important thing" (B1).

The awareness of making AYP seemed always constant even for high performing schools. One principal from a very successful non-Title I school explained,

> If we have 94% of our kids making AYP, we've got to keep that 94% and . . . continue and continue to improve so that we can maintain what we have with new groups of students . . . moving in. Those things we have no control over. (C2)

On the other hand, some non-Title I principals, even though they value strong academic performance, know they have the luxury of not being as concerned as their Title I colleagues. As one such principal stated,

> Personally, I don't care what the CRTs are if my kids are understanding. If they know what they're talking about when they leave here, and they're prepared to face the world, I don't care if we have a zero, and our results are published as such . . . a test doesn't mean anything to me. What matters is "Are we understanding?" (D1)

For the Title I school, the consequences of not making AYP comes with the threat of being placed on "Needs Improvement" with the possibility of further sanctions being imposed if academic progress is not forthcoming. However, in the case of non-Title I schools where Title I funding is not an issue, principals' aim in meeting AYP standards often stemmed from a desire to meet the expectations of parents and public.

For both Title I and non-Title I principals, the pressure to increase test scores for the purpose of passing AYP seems to be a consequence of the NCLB legislation. Although fans of NCLB view an increase in test scores as proof that the legislation is working (refer to Appendix E), there are others, as previously stated in the literature review, who do not feel that higher test scores are necessarily a true indication of either the teaching or the learning that has taken place (Kohn, 1999; Meyers& Rust, 2000; Ravitch, 2010; Rotberg, 2001; Smith, 1991). There are many ways to improve test scores. Ravitch (2010) states, "The pressure to increase test scores is likely to produce higher scores, whether by coaching or cheating or manipulating the pool of test

takers" (p. 161). But if the goal is simply to increase test scores, schools may very well be going in the right direction.

Impact on teachers. According to the principals in this study, test scores are important to teachers, whether the teacher is from a Title I school or a non-Title I school, and whether the school has a history of failing or passing AYP. In fact, test scores are more meaningful to the teacher than to the student, particularly at the elemenary school level. From these principals' perspective, some teachers have a tendency to view a student's test score as an indication of how well they are performing as a teacher and, therefore, teachers are concerned about having certain groups of students in their classes because of the impact it could have on test scores.

More than one principal mentioned that many teachers assume that their principal, parents, and coworkers will view test scores as being a true indication of their teaching abilities. Whether this is an accurate assumption by teachers or simply perceived, anxiety greatly increases for the teacher who feels that test scores are being viewed as a barometer of their teaching ability. Principal D1 worried about teacher morale and misinterpretations by the public:

> Some teachers see that AYP as just being so threatening and, as being a concern to them that they get disheartened. . . . We're getting to the point where our society views things printed as being a little too truthful and they don't look into and they're not involved enough in the schools to know.

Another concern is interpretation among the staff. With test scores becoming more transparent through the increase in collaboration, teachers are very much aware of how they measure up to their grade-level colleagues. One principal stated, "So, it becomes obvious to me that they're concerned about how that [low test scores] reflects on who they are as a teacher, and their ability" (C3). If not tangible sanctions, low test results certainly have an impact on

teachers' feelings about their success, as evident from this observation from a principal: "I showed her the data, and she . . . is now whipping herself in the closet" (D1).

Such experiences do not appear to be isolated. Both Title I and non-Title I principals mentioned how discouraging it was for their teachers when their school failed to make AYP, or when teachers noted that their scores were lower than those of their grade-level counterparts. It was not just the fear of having sanctions imposed on them that was disheartening. It had a lot to do with the fact that they were working hard to prepare their classrooms for the CRT and were discouraged when there was little or no improvement in their students' academic proficiency. Principal D3 empathized, "The teachers who really are trying to do a great job, I think they get a little discouraged because they're working very, very hard and maybe not always seeing the results they hope for."

Although some teachers may feel that test scores are used to evaluate their performance, all of the principals in the sampling did not feel that the results of the CRT, by themselves, were a fair way to access the quality of teaching in the classroom. Principals frequently stated that test results may be one indicator of what is going on in the classroom, but every principal expressed resistance in allowing them to be the only measure of a teacher's success.

One of the concerns in using test results to evaluate teacher performance is the differences in the makeup of the classes and the challenge of controlling for these variables. As one principal explained,

> It [CRT data] tells us something, but not all classes are created equal. I mean, in a school like ours we do everything to balance classes. When we make those class lists, they are balanced by gender; they're balanced for ESL; they're balanced for behavior; they're balanced for resource; they're balanced for everything. And then, during the summer, a

> bunch of them move out, a bunch of new kids move in, and you just point them to the lowest class when they come because you don't have any assessment data on them when they arrive; you just put them in a class and, by the end of the year, those classes are not created equal. (A1)

This is true for most schools where the transient population is high. Teachers whose classrooms see an unusual influx of English language learners, for example, could potentially be at a disadvantage.

On the other hand, the reverse can also be true: groups of students with unusually high aptitude can raise CRT scores for an entire class or grade level. One principal mentioned how phenomenal this year's sixth grade class has been, regardless of who their teacher was. Another principal talked about his incredibly bright group of kindergarteners this year: "on paper it may look good, but I cannot attribute it solely to good teaching" (A2).

Another problem with linking test scores to teacher performance is the influence of other educators, such as literacy specialists and paraeducators. The Title I principals in this study observed that schools often have multiple players invested in and working towards student success. In addition to the regular classroom teacher, many Title I schools have classroom support such as a reading specialist and a special education teacher in both the lower and upper grades. In these cases it would be difficult to determine who had the most influence on a student's academic performance.

Given all of these limitations, teachers who stress about the implications of test scores may benefit by hearing from their principal, repeatedly, that test scores alone are not a good indication of teaching success.

With a few exceptions, principals reported that it was their more seasoned teachers who

were most discouraged over failure to pass AYP. Principal D3 said, "I think the seasoned teachers are a little more discouraged. . . . They tend to be a little more stuck, generally." Another principal mentioned,

> I have teachers who have been in education for thirty or more years and they seem to be the ones that their morale is impacted the most . . . I just feel like sometimes the message they feel, even though that's not what we intend to send . . . is that what they've done for all these years isn't good enough. (C2)

For seasoned teachers, negative feelings about CRTs may be compounded because they have experienced the pre-NCLB era and remember what teaching was like before the pressure of passing high-stakes tests.

The principals were quick to point out that not all experienced teachers are negatively impacted by the pressure to pass AYP. In most instances, the principal only identified one or two teachers who seemed particularly affected. And the teachers mentioned were not always more experienced. In some instances, a less experienced teacher seemed overly preoccupied with preparing students to do well on the CRT.

Whether new to teaching or well established, all teachers are affected somewhat by failure to make AYP, according to principals in this study. One principal remembered, "To not achieve AYP was very devastating to the teachers" (A2). When schools failed to pass, it most often fell to the principal to lend additional support and encouragement so the disappointment did not linger too long. One principal divulged her efforts to boost morale among her staff: "Lots of food; lots of praise; lots of music; lots of visiting. . . . [This was a] big, big time for morale building . . . because the teachers work very, very hard, very long hours" (D2).

The principals in this study placed a lot of trust in their teachers and looked to them to

make a positive difference in the academic life of their students. Most of these principals not only encouraged, but also expected, that their teachers would thoroughly and thoughtfully review the CRT data, as well as data from other formative assessments, with the intent of identifying ways to improve their instruction to assist students in increasing their academic proficiency. Principal A2 emphasized,

> A good teacher . . . [is one who] *will* effectively and regularly monitor progress, who adjusts teaching to meet the needs given that progress . . . who works with others to identify and find the best practices to meet the needs of our students.

So, a good teacher, from the principal's vantage point, isn't the one who necessarily has the highest test scores but, rather, is the one who reviews the data and is open and amenable to searching for better ways to help the students, even if that means asking for help from his or her colleagues. Good teachers are willing to change their teaching practices to the benefit of the student.

Principals whose teachers bring last week's assessment data to their weekly collaboration meetings are encouraged by the discussions that are taking place throughout the year. These teachers not only identify students who are having trouble in one particular area, but they also help each other in determining how and what kind of remediation should take place. Teachers whose students have the highest scores may be depended on to help re-teach the group of students who are struggling. As mentioned earlier, this kind of collaboration has been a major benefit of AYP.

One Title I school principal mentioned that, for him, the first indication of good teachers is that they love the kids. He explained that if teachers really love their students, they will care about them to the point of knowing what's going on in that child's life and, particularly, of

sensing whether or not something is amiss. It is difficult for learning to take place when the child is dealing with other, sometimes personal, unresolved issues. The teacher who can identify these problems and help to remove potential roadblocks to learning will have better success in teaching and connecting with the student. "It's not just about teaching; it's about kid's learning" (B1).

A number of principals concurred that a good teacher should be able to build a solid connection with the students and develop, in the words of one principal, "very strong relationship-building skills with students that excite them about the learning, that excite them about the content, and then they learn . . . because of the teacher" (A1). This was further clarified by the principal, who believes a good teacher to be one who sets clear expectations and then helps the child accomplish them; a good teacher does not allow students to settle for anything less than their best (A1).

In addition to collaborating with each other and connecting with their students, teachers should be adept at delivering instruction and, when necessary, differentiating instruction for the variety of student needs in their classroom. In other words, good teachers consider what's best for the individual child and are "constantly checking for understanding in their classes . . . [to know] where their students are at" (C1). One principal offered the following measure of success:

> You walk into their classroom and they're up there having . . . a good time; the kids are attentive and actively involved . . . the activity that they've designed . . . is teaching the child to understand versus to do worksheets . . . (the staple of the 80s and 90s [that] needs to be thrown out). (D1)

The value of teaching for understanding was mentioned a number of times during the interviews. One principal insisted that the time for preparing students for an assembly-line type

job has passed. With the availability of the Internet, encouraging students to memorize facts and figures is no longer necessary. Teachers should be inspiring kids to want to learn, leading them to have the discussions that help them understand. He was not sure that the CRT encouraged that kind of learning.

Rewards and sanctions. There were no rewards or incentives, to speak of, for teachers whose classrooms perform consistently well on the CRT. A few of the principals mentioned having a "celebration" to recognize the school's efforts in making AYP, although caution was taken not to single out specific teachers whose class may have scored particularly well that year. As one principal stated, "I'm a little afraid of that practice. I'm trying to maintain a collaborative atmosphere " (A3). In addition, most of the principals hesitated in crediting particular teachers with their class's academic progress because a number of factors could be responsible for their success. As mentioned earlier, a classroom may have benefited from additional resources, such as a literacy or reading specialist, or one-on-one mentoring from a paraeducator. The principals mentioned that, in such instances, it was difficult to give credit or cite one particular individual or reason for the academic improvement. In addition, a teacher may have simply been blessed with an exceptional group of students who would have excelled regardless of who was teaching.

Similarly, when the idea of merit pay was addressed in these interviews, it did not receive a very strong vote of support, possibly for the reasons already mentioned. One pilot program in Utah does offer teachers at five different schools (three elementary and two charter schools) rewards of "up to $2000 more a year for improving student performance, parent satisfaction and giving quality instruction" ("5 Schools," 2010)—but this program does not include any of the schools from this sampling.

The rationale for not singling out teachers for CRT-based rewards was brought up again

when discussing when and where to place sanctions for low academic performance. The principals felt that there were too many variables that could influence academic outcomes to zero in on any one particular cause or teacher. Thus, currently, no sanctions are being imposed on individual teachers in any of the districts represented in this sampling. However, principals did mention that the fear of sanctions hangs heavy over their heads, particularly for the Title I schools. Some principals noted scenarios across the country where teachers were fired en masse because of low test scores (e.g., McKinley Tech High School in Washington, D.C., and Central Falls High School in Rhode Island).

Although teachers are not currently singled out if their students do not do well on the CRT, Title I schools may receive sanctions if they or their district fail to make AYP for two consecutive years. This can weigh heavily on both the principal and the teacher. One principal noted, "We don't have sanctions, but right now, the district's in sanctions . . . . So, because the district's in sanctions, we're all in sanctions" (D2).

I agree with the principals in this study who believe that the idea of imposing sanctions on low-performing schools in an effort to motivate them to "do better" is based on faulty thinking. Such an approach tends to presume that students aren't performing well because teachers and the schools aren't working hard enough. The feeling is that, if the sanctions are tough enough, educators will do whatever they can to make sure students reach academic proficiency. Implicit in this reasoning is the idea that, no matter the set of challenges students bring to school, adequate teaching will result in adequate yearly progress. For teachers, principals, and students alike, this is not a fair assumption.

In training elementary teachers, Collins and Tamarkin (1982) urged, "When a child gets something wrong, don't simply red-mark the paper; take the student aside and help him get it

right. Remember, if he knew how to do it correctly, he would have done so in the first place" (p. 222). I think the same could be said of teachers and schools. If they knew what to do to produce not just high test results, but legitimate proficiency levels, they would do it. Thus, professional development opportunities—not sanctions—are the most promising recourse to improve teaching strategies and enhance a teacher's management skills. Ironically, professional development is often one of the first things to go as a result of budget constraints.

Not only is the rationale behind sanctioning low-performing schools based on faulty thinking, but the sanctions are based on a one-time, year-end test that may not have scientific validity. Ravitch (2010) observed,

> The problem with using tests to make important decisions about people's lives is that standardized tests are not precise instruments. . . . Tests vary in their quality, and even the best tests may sometimes be error-prone, because of human mistakes or technical foul-ups. . . . Sometimes questions are poorly worded. Sometimes the answers are wrongly scored. Sometimes the supposedly "right" answer to a question is wrong or ambiguous. Sometimes two of four answers on a multiple-choice question is equally correct. (p. 152)

The principals in this study are wise in their efforts to not attach too much weight to test results—but they are seeking to do so in an educational climate that works against these efforts.

Another form of sanction proposed by education reformers consists of encouraging parents to be aware of test results and then choose their child's school (U.S. Department of Education, 2002). While this may be helpful for an individual child or family, policies that allow parents to transfer their child out of a low-performing school may have negative consequences overall. Some students come from families who may not care, or do not have the means, time or understanding to take advantage of the school choice option. More important, moving students

to other schools does nothing to solve the performance problem. Speaking of sanctions, VanderArk (2004) stated, "The overarching goal must be to improve student outcomes, which means a focus on turning around (rather than simply punishing) low-performing schools" (p. 41).

Impact on curriculum. Overall, it appears that most of the schools represented in this study teach to both the state core curriculum and the CRTs. Although both the state core and the CRTs include reading/language arts and mathematics, the core curriculum established by the state of Utah, particularly for grades 3-6, also includes science, social studies, arts, visual arts, music, dance, theatre, health education, physical education, educational technology and library media (Utah Administrative Code, 2010). Principals noticed that particular subjects were not given as much attention after the increased emphasis on language arts and mathematics. The amount of time spent in each of the core subjects seemed to vary from school to school and from teacher to teacher.

For Title I schools, the lack of time for non-CRT subjects was especially evident. All of the principals acknowledged the additional pressure placed on Title I schools to perform sufficiently on the CRTs and mentioned that Title I schools spend even less time than non-Title I schools on subjects outside of language arts and mathematics. Principal A1 said, "In a Title I school, specifically, we tend to do more time in reading and math than we might have the luxury of in a non-Title I school. We'd have the luxury of more . . . time for music and art." Another Title I principal concurred that "the greatest attention is given to language arts and math, and then, after that, science and social studies . . . everything else comes in as they can" (A2). Even though non-Title I teachers may have more freedom in deciding their curriculum, the CRTs are still a constant reminder of the need for academic proficiency in the language arts and math. The principal of a non-Title school insisted, "We look at how well the students did on the CRT to

guide our instruction for the next year . . . So it does affect what we teach to" (C3).

Although the emphasis on language arts and mathematics was common among all of the schools represented in this study, there was not a consensus on what subject areas were neglected as a result of attending to the CRTs. Some principals noted that social studies was being neglected, while others felt it was health and P.E. Even though the fine arts, including music and visual arts, were mentioned more than a few times, there were a number of schools who discussed the importance their school or school district placed on the arts. In each case, these were non-Title I schools. Principal C2 mentioned school board support to continue arts education:

> Our school board is very passionate about the arts, and so . . . our school improvement plan has arts written into it, throughout the district. . . . So, it may not have funding and it may not have as much emphasis but they're saying, "Find ways to integrate."

Another non-Title I teacher mentioned that they even have a paid art teacher while science ("somewhat") and social studies "get left behind" (C3).

There was some discussion on the value of integration and how some teachers were able to integrate some of the less important topics into the core. Principal B1 mentioned this strategy: "I think the arts are kind of neglected, but they don't have to be. So, we're working at integrating those into the other subject matter." Another principal discussed how his teachers came together to create a grade-level collaborative where music, P.E., library, and computer literacy were taught for 30 minutes to an hour each week. Incorporating other subject areas, such as science, while teaching reading concepts was also mentioned as a way to integrate different topics into the core curriculum.

A commonly expressed frustration was over the volume of information that needed to be

covered during the year. As stated by one principal, "There's so much information to be taught throughout the school year. . . . The breadth is there, but if you want to really get deep into a lot of core information, there's not enough time" (B2). Considering these constraints, many principals praised their teachers for being able to meet all of their core objectives in spite of the volume of material they need to get through.

The principals realized that this volume of information increases teacher responsibility. Teachers must decide what subjects are most important while, at the same time, deal with the ever-present threat of making AYP. Some principals felt that the established core and the CRTs actually benefitted instruction by focusing teachers' attention on essential learning outcomes and diminishing the amount of time spent in areas of less value.

Teaching to the test. The principals in this study agreed that the fixation on test scores seemed to promote an inclination to "teach to the test." They confessed to being influenced by wanting to produce good test results and, therefore, focusing more attention on the subjects to be tested. One principal stated, "My teachers' worry that 'Well, if we don't teach to the test, then how will they ever do well?'" (A1). And principals differentiated between test prep that does little to educate students (sometimes they used the words "teaching to the test") and test prep that is actually focused, in-depth teaching of concepts.

A number of principals mentioned the desire to have their students really understand what is being taught, and not just memorize information in order to do well on the test. "We want to provide the depth so that . . . the kids, no matter what is on the test . . . they're learning more," Principal C1 explained. "By doing depth, a student understands a whole lot more than just a specific concept."

However, principals acknowledged that it takes time for some teachers to understand that

they can still get good test results even if they do not "teach to the test." Principal D1 spoke of this shift in perspective

> The teachers that are still wrapped up in "I've got to get this taught" and teaching the how, teaching what the answer is versus understanding, they're teaching to the test still. . . . We're getting there. It does take time to change teachers' perspectives, but the more I'm getting them out and viewing other teachers and realizing that, "You know, there are other ways of approaching things," they're starting to get it.

Some of the principals were not as concerned with "teaching to the test" as they were with the test itself, as represented by this principal's observations

> I think the critical thing here is what is on the test really testing what we want them to know? And, I suspect that that is not happening because you can't do a multiple-choice test and get more complex thinking. . . . Ideally, I would say we need to adjust the test, too, so they are testing what we really want the students to know. I'm not sure we're doing that. (D3)

Other principals had similar feelings and wished that they had an opportunity to provide more input on what was on the year-end test and how it should be structured.

As a researcher, I believe that the real danger in "teaching to the test" is not in the actual instruction but in the implied importance this kind of teaching places on the test itself. Teaching solely to the test reinforces the misconception that higher test scores are the goal. Additionally, if the test is flawed to begin with, then teaching to the test is even more problematic. On the other hand, if the test is used as an information and diagnostic tool, it can be valuable in improving teaching instruction and, ultimately, improving student achievement.

Creativity in the classroom. Seven of the principals interviewed felt teachers did not have to compromise creativity to teach the core. In fact, when asked about the impact the focus on making AYP had on teachers' creativity in the classroom, these principals stated that teachers have to be more creative in order to incorporate the core across content areas. Rather than giving up creativity, some teachers were actually quite clever in integrating the core so that every content standard and objective could be met. They did not let the pressure of AYP or teaching the core curriculum negatively influence how and what they taught in their classrooms. "I think my teachers . . . have a great level of creativity when it comes to teaching to the core," Principal C3 said. Principal B1 concurred "They have to use a great deal of creativity to hit all the core areas." Another principal stated that teachers are "going to insert their creativity whether we like it or not" (D1) and, for the most part, that's a good thing.

The observations and experiences of these principals differ from those found in much of the literature review. By comparison, they seem much less daunted or cramped by CRT requirements. Ravich complained that "one of the unintended consequences of NCLB was the shrinkage of time available to teach anything other than reading and math" (2010, p. 107). Jones, Jones, and Hargrove (2003) reported one teacher as saying "The tests have affected my instruction because they have taken away the flexibility from my teaching. . . . My teaching has changed; I used to teach ideas rather than testing" (p. 1). Another teacher said,

> With the testing programs we have in this school, there isn't much leeway for me to be creative or innovative . . . . The test is the total goal. We spend time every day doing rote exercises. Forget ever doing hands-on activities, science or math games, or creative writing experiences. We do one hour of sit and drill in each of the subjects of math, reading, and writing. . . . You can't improvise, add, or take away. . . . This is how testing

has impacted my school and my teaching. As a first year teacher, I feel like I don't have a choice to deviate from this awful test preparation. (quoted in Jones et al., 2003, p. 37)

The more positive outlook expressed by principals in this study regarding creativity and flexibility may reflect administrator bias. On the other hand, it may reflect an academic climate in which teachers accept that preparing students for assessment is paramount, and they use their creativity to make that happen in positive ways.

Preparing students to take the CRTs. All principals, including those whose schools had a history of success, agreed that not making AYP was a disheartening experience. They were hopeful, however, that their teaching staff would use the data to reflect on their teaching practices and determine what changes they could make in the classroom in order to improve instruction. "If they don't know how to look at data, change their teaching instruction to meet the needs of kids . . . [they] cannot change and become a teacher that meets the needs of students at their level," Principal B2 said.

According to principals in the sampling, teachers are using a number of different strategies to prepare students for taking tests, although, again, most try to resist the tendency to "teach to the test." There are a number of programs and materials available that are commonly used by teachers as a means of preparing their students to do well on the CRT or other assessments. Attention is given to teaching both test-taking strategies and the core content.

The principals mentioned a number of programs and products that assist the teacher with these two goals, including UTIPS (Utah Test Item Pool Server) and DIBELS (Dynamic Indicators of Basic Early Literacy Skills). UTIPs is an online testing program created by the Utah State Board of Education. It includes both an assessment engine and a pool of potential test items that are aligned to the state curriculum. DIBELS is used to monitor the development of

early literacy and reading skills. Other programs mentioned that assist teachers in preparing their students to develop their test-taking skills include Acuity (math test), DRA (Direct Reading Assessment), PM Benchmark Assessment (reading assessment program), RAP (Reading Achievement Program), Success Maker (math and language skills computer program),Test Ready (test preparation materials offered in a variety of subjects), Treasures (a reading/language arts program), UALPA (Utah Academic Language Proficiency Assessment), Waterford (early reading program) and YPP (Yearly Progress Program for math and reading).

A review of the data from CRTs and other tests seems to influence what and how much test preparation is incorporated into the daily class routine, although most of these programs are, at some time or another, used during the school year. Frequent practice with these programs, explain the principals, assists students in becoming familiar with the type of test questions that might be asked, as well as becoming comfortable with the test format.

Regular test practice throughout the year also provides additional, ongoing data for the teachers. They use this data to measure the progress of their students and remediate, if possible, any weaknesses before CRT test time at the end of the year. All of the principals mentioned the value of formative assessment as a way to know what's going on in the classroom on a fairly regular basis. There was a consensus that waiting until the end of the year to find out how a student was doing was too late. These principals felt that CRT results were more valuable in capturing an accurate picture of a student's academic progress when reviewed and considered alongside data from formative assessments. "To make a judgment about the student's progress throughout the year has to come from a culmination of data," Principal B2 said. CRT data was also helpful in identifying patterns, especially with regards to a particular teaching strategy or a particular teacher's instruction.

In addition to the online and hard copy test preparation materials already mentioned, a few of the principals noted that teachers often post learning objectives and curriculum maps in their classrooms as a reminder of the essential learning areas. Even with the incorporation of such strategies and tools throughout the year, some principals admitted to an extra concentration on the core subjects two or three weeks prior to test time.

With all of the expense and time devoted to test preparation, it is clear that at least some focus has been diverted from evaluating student learning to simply doing well on the test. Ravitch (2010) discussed what she calls "gaming the testing system; that is, tricks and shortcuts to achieve the desired results, without improving education" (p.154). She views test preparation as one of the most common gaming methods

> Most districts . . . relentlessly engage in test-prep activities. Some preparation for test-taking is valuable; reading and studying, learning new vocabulary; and solving math problems are good ways to get ready for the tests. But school districts have invested hundreds of millions of dollars in programs and training materials that teach students the specific types of questions that will appear on the state tests . . .
>
> They master test-taking, but not the subject itself. In the new world of accountability, students' acquisition of the skills and knowledge they need for further education and for the workplace is secondary. What matters most is for the school, the district, and the state to be able to say that more students have reached "proficiency." (p. 159)

This kind of focus seems more prevalent among Title I schools, which are required to meet AYP, as opposed to their non-Title I colleagues. The three principals in this study who represented non-Title I schools and had experienced less success in making AYP expressed a

preference for knowledge acquisition and understanding over simply receiving a good test score.

Impact on student learning. Even though Title I principals felt more pressure to produce good test scores, all of the principals in this study named student learning as the primary goal. One concern expressed was that student stress levels about testing may harm the learning process. Although some principals did not feel that students, as a whole, knew much about the significance of the CRTs and making AYP, and were, therefore, not bothered by it, others observed increased anxiety among some of their students as test time approached. Principal A2 mentioned that different students respond differently to testing "Generally speaking, there are some kids I wished cared more about how they're doing in school. There are some kids who care too much . . . who get really stressed out."

A few principals expressed frustration over children who don't try their best. Principal D1 related,

> Some of our smart kids realize that CRTs have no impact on them, so . . . they want to read. "Guess what the activity is as soon as I'm done with my test? I can read." So, they'll just zoom through and even though they may be able to figure out the answer, they go through and do the test quickly and they get to read. I heard one teacher in here yesterday, just frustrated. "That kid took the test in under 20 minutes! I couldn't even . . . read through the thing before he was answering questions. So I know he wasn't even reading it."

Principals recognized the influence teachers have on their students, particularly when it comes to testing "I've got probably 10% of my faculty who are so stressed out about it; they stress their kids out about it," one principal said (D1). If teachers are relaxed and do a good job in setting up the context for the test, most often, they can do a lot to assuage any concerns or

fears their students may have. Some principals mentioned attempts to create a positive test-taking environment by providing students with a small snack either before or after the test. The result: "A lot of kids love testing now" (D1). It appears that, in such instances, a fairly simple gesture reaped positive benefits.

Disaggregation of data. Principals mentioned one facet of CRT that shows great promise in terms of student learning the ability to separate, or disaggregate, the data for different groups of students. The disaggregating of data is particularly beneficial in pinpointing specific areas of concern. One Title I principal, for example, learned that the boys at her school were not performing as well as the girls in math, science, and reading by studying the trend data provided by the end-of-the-year AYP report. Although this information arrived too late to remediate individual students, it did help steer the school's focus and attention for the following year. Male discussion groups were created in reading while more competitive games and activities that catered specifically to boys were developed in math. "We've tried to bring in more games that are more competitive . . . for practice and reinforcement. [It] hasn't changed our primary instruction, but has had to change our practice piece because that gets them more engaged," this principal explained (A1).

All principals mentioned that the NCLB legislation has helped to focus their attention on individual academic progress for all students and that the disaggregation of data provided a clear vision for areas that needed their attention. One principal noted that, previous to high-stakes testing, "We were pretty satisfied if the majority of our kids got it. . . . We don't do that any longer. [We now] focus on each individual child and their learning" (B1). Another principal mentioned, "When you disaggregate the data, it's real important to look at all of your groups. . . . iI it wasn't for AYP or No Child Left Behind, we'd still have huge gaps all over the place" (B3).

Other principals reaffirmed this increased awareness of and progress being made in closing the achievement gap as a benefit of NCLB and AYP. Regardless of how many minority students are enrolled, the feeling from many of the principals was that it was no longer sufficient to simply meet a good overall academic proficiency level. They now have to look at each subgroup and determine who needs help and what can be done to help them.

These previously neglected smaller subgroups of students include Asian, African American, American Indian, Hispanic, and Pacific Islander groups, economically disadvantaged students, students with limited English proficiency, and students with disabilities. Including these special populations in the AYP report is particularly significant for the Title I schools, which experience a more transient population that often includes special subgroups. Although demographic makeup was never used by the Title I school principals as an excuse for failing to make AYP, it was noted that this mix of students does come with a unique set of challenges seldom shared with their non-Title I counterparts.

In most cases, all students, regardless of their nationality, cultural background, disability or English language proficiency, are placed in a regular classroom. For the non-native English speaker, this is a particular challenge. "Our students who are learning another language as they're learning the content . . . are not going to perform as well as our Caucasian kids," Principal A1 said. Although English language learners (ELLs) can be found in both Title I and non-Title I schools, they are often quite transient, which poses its own set of problems. One Title I principal noted the challenge this presents in trying to raise academic proficiency each year:

> We get [students who are] less and less and less proficient in English coming in to us. . . . So, we're getting lower and lower and lower kids and expect them to perform higher and higher and higher. So, the basic assumption that you can compare last year's data to this

> year's data is flawed. . . . As long as you always have new kids moving in who haven't been taught and you have the kids who are here exiting as soon as they get to proficiency, you're never going to have the data show what you think it's showing. (A1)

It was recognized that even some native English speaking students, particularly those from an economically disadvantaged background, come to school with a language deficit because they come from home situations where reading and conversations are minimal if existent at all. These same groups of students have a tendency to come with additional "baggage" as well. "I have a whole lot of kids who come to school with a whole lot of issues that prevent them from learning to the best of their ability," said Principal A2. Principals discussed a number of these factors that could potentially influence a student's ability to perform in the classroom, including coming from a single parent home where the parent is often working more than one job and is seldom, if at all, available to help with homework or offer other kinds of support. Some students come from homes where English is not spoken, or where emotional or physical abuse might occur. Some students come to school hungry.

All of the principals recognized how detrimental such circumstances can be in the life of a child. The home situation is more difficult for principals and teachers to address than language deficiency. Although they realize their limitations, principals established, wherever appropriate, special accommodations to try to compensate for situations that place students at a disadvantage in the classroom. Special attention and instruction, for example, are often provided to English language learners to help them catch up to their classmates in reading, writing and comprehension. This is particularly challenging when considering that over half of the student body in some of these schools are ELL students. One principal mentioned that the school's ELL population is 86%, with 33 different languages being spoken!

Formative assessments play a significant role in helping teachers regularly monitor the progress of these special populations. Reading specialists, special education teachers, and paraeducators, who are often bilingual, supplement the regular teaching staff and provide necessary support for the ELL student. This, of course, cannot be realized without the necessary resources, most often provided through federal funding in the case of a Title I school.

Cut scores and proficiency level. All of the principals expressed concern with the manner in which cut-scores and proficiency standards are determined. The suggestion that standards should be the same across the country came with some hesitancy inasmuch as no one seemed to want to encourage more national mandates. However, these principals also recognized that the way in which states are currently allowed to set their own standards leads many educators to question AYP's validity.

With no national guidelines or standards, states set proficiency levels wherever they deem appropriate. Again, with the pressure to make AYP, it is not surprising that some states have chosen to set a lower proficiency level than others. A study conducted by the Thomas B. Fordham Institute in 2007 reported that "'proficiency' varies wildly from state to state, with 'passing scores' ranging from the 6^{th} percentile to the 77^{th}" (p. 3). This range of proficiency levels calls into question what the results are really saying and how much weight they should receive.

In the forward to the Fordham study, Finn and Petrilli (2007) stated that "America is awash in achievement 'data,' yet the truth about our educational performance is far from transparent and trustworthy. . . . Gains may be illusory. Comparisons may be misleading. Apparent problems may be nonexistent or, at least, misstated" (p. 3). They further emphasized the danger of placing so much importance on a testing infrastructure that "is unreliable—at best"

(p. 3). Comparing one year's data to the next can very well be misleading as previously mentioned. With the influx of students each year, particularly in Title I schools, which are also receiving the most scrutiny, it is difficult to know what CRT scores are really saying.

Three principals specifically mentioned that they felt questions on the test were posed in an attempt to "trick" students rather than to measure their academic proficiency. As one principal passionately stated, "Make it fair. Say . . . 'Here's where we want our kids to be.' And make it specific. Don't say, 'We want all of our kids to be on grade level in reading by Grade 3.' What does that mean?" (D1).

The principals in this study mentioned their preference that individual student progress be emphasized more than meeting a certain standard. These principals had a strong desire to celebrate the accomplishments of the individual, especially when considering English language learners, students with disabilities, students from low socioeconomic backgrounds, and other disadvantaged students, rather than the group as a whole. They expressed frustration that many students will never meet academic proficiency standards, especially when cut-scores continue to rise every few years. Principal A3 asserted, "I have a problem with math testing when 42% cannot pass. . . . To me that says that there's something wrong with the test." On the other hand, there are other students who should be held to a higher standard than the current cut scores.

Every principal, including those whose schools have never failed AYP, felt that the mandate to have every single student on grade level by 2014 is unrealistic. Representative of their comments is the principal who stated, "It makes me crazy when we are working very hard to help our students learn and grow and they may not be on grade-level, but if I look at the progress they've made, it's wonderful" (A2). Another principal concurred, "It is unrealistic to assume that everybody's going to be on [grade] level at any given time" (D3). Ravitch (2010)

argued that having all students proficient in reading and mathematics by 2014 is "the most toxic flaw of NCLB" (p. 102) and that it "was out of reach, unless states deliberately dumbed down the meaning of proficiency" (p. 109).

Principals expressed concern over the amount of trust that has been placed in the cut scores. In the words of one principal, "You could set them anywhere and they're going to be too high for somebody and too low for somebody" (A1). Two principals referred to them as being "ridiculous." This seemed to be particularly true of their feelings when discussing special education or ELL students. One principal offered this example:

> I don't give up on my special ed. kids, but they're special ed. for a reason. . . . They're fifth graders but they're reading on a second-grade level and so we address their issues on a second-grade level, and we push them to a third-grade level, but we test them on a fifth-grade level. I really have problems with that. (A3)

End-of-the-year, one-time test. Each principal in this study also took issue with using a single test score, at the end of the year, to determine academic proficiency. A few referred to this as "autopsy data" because, as one explained, "they're after the fact. They're probably the least helpful as far as informing practice, but they're mandated by law and we just live with them" (A1). Another principal continued the analogy:

> Just like a doctor . . . if you want to know how a person died, great! You can go back through and do the autopsy. But, if you want to know how to change and be better, you've got to do it all the way along. (B2)

Another challenge is in determining the best way to evaluate a student's academic proficiency, when considering the following: "There are kids that are just poor test-takers, and there are kids that are just slow, and there are kids that are always going to guess" (D2). Other

principals concurred that, particularly a one-time test, is not the best measurement of a child's progress since different students may connect differently with the test. Again, they were quick to suggest that the CRT should be used as only one indicator of achievement; that it is a one-time snapshot of what's going on with that child at that particular moment in time.

Most principals placed a lot of value on tests conducted throughout the school year, mentioning that their teachers conduct a lot of formative assessments in the classroom. These tests not only provide more timely assessments, but also help students understand the core and give them experience in taking tests in preparation for the year-end exams.

Demographics. As a researcher, I noted differences in principal responses based on the demographics of their schools and the types of funding they receive. Some of those differences are described below.

Parents and public perception. Parent involvement and awareness varied from school to school, although most principals were still concerned with the perception the public has of their school and how that might be influenced by the publishing of CRT results and whether or not they made AYP. Most of the principals, however, admitted that the majority of their parents knew very little about AYP or what it meant to pass or fail.

There was a noticeable difference in parent involvement taking place between the Title I and non-Title I schools. While non-Title I schools experienced the benefits of strong parental involvement, most Title I schools did not. It was not surprising that, in those instances where parents were involved, students as well as the schools were the beneficiaries. One non-Title I principal stated,

> We have a very involved clientele. We have pretty affluent parents here. . . . I think that kids get more and so they naturally have a little bit of an edge . . . our students come to

> school more ready. They get more support at home so their homework is done and they have the support system so . . . we don't have to do so much at school to make sure that learning happens. (C2)

Even if parents are not fully aware of what making AYP really means, it does appear, at least in some instances, that high test results are still looked upon favorably by the community. Another non-Title I principal observed,

> A lot of people want to come to our school because of the results we get on the CRT and because those are published and on the website. . . . I have 46 people who want their child to come to our school . . . who are not in our school boundaries. (C3)

On the other hand, a non-Title I principal whose school did not make AYP did not feel that the community was very concerned about the test results. This principal explained that the positive personal experiences and observations of some of the parents overrode any negative impressions created by low test scores: "The public perception and most parents who are involved in this school love this school. They know what's going on; they see the teaching that's going on. . . . Overall, they like what's going on" (D1).

Another non-Title I principal in a similar situation explained that student, parent, and teacher satisfaction is monitored by reviewing the results of the Indicators of School Quality survey (ISQ). Created by the Center for the School of the Future, at Utah State University, the ISQ is a survey system that provides school administrators with an instrument in which they can evaluate and monitor perceptions and feelings about their school. The ISQ organizes school functions into the following domains Parent Support, Teacher Excellence, Student Commitment, School Leadership, Instructional Quality, Resource Management and Schools Safety. In their last ISQ review, this principal explained that they received high ratings from both the students

and their parents—therefore, the principal concluded, the CRT results did not seem to make a difference in the way they were perceived.

In terms of parent opinion and parent support, Title I principals are faced with a different set of circumstances than non-Title I principals. They were quick to point out that, even though they may not have the same parental support that their neighboring non-Title I schools might enjoy, it wasn't because their parents cared less. The parents were simply dealing with a different set of circumstances. As one Title I principal explained, many of his parents are in "survivor mode . . . they're just trying to feed their kids" (B1).

Title I principals also mentioned that many of their parents either did not finish school or did not have good experiences when they were students, so they are somewhat cautious of participating or volunteering in their children's school. "They aren't the type of parents who come in and help out in the classroom because they just don't feel comfortable there," Principal B3 shared. From their comments, these principals seem to be as concerned with trying to build up the parents' trust level as they are about how their test results are going to be viewed.

Since Title I schools are more likely to have a larger number of ELL students than non-Title I schools, Title I principals have the dual challenge of working with students who come with limited English proficiency and with parents who are not native English speakers. Limited English-speaking ability may also account for lack of engagement on the part of the parent. One Title I principal made a concerted effort to get all of his parents to their Student-Parent Conferences and, impressively, was successful with all but three parents. He explained, "Two of them are Burundi and it's been hard getting translators" (B3).

So, in addition to working with their students, who may or may not be receiving much support in the home, these principals were also spending a significant amount of time trying to

keep their parents informed and engaged. They wanted to help parents understand what "needed to happen at home to help in their [child's] education" (B1). There's no question that the demographics of the school greatly influence the kind of parental involvement as well as the teaching and learning that take place within the school walls.

Title I funding. When the No Child Left Behind legislation was implemented in 2002, it was not without teeth. Although its intentions seemed honorable and few could disagree with the desire that "no child be left behind" in the opportunity to obtain a good, quality education, it seemed to be a major assumption, inherent in NCLB, that this would not happen unless certain guidelines were enforced. One of the carrots that hung over the heads of public education and the one that seemed to carry the most weight most certainly had to do with funding; Title I funding to be exact.

As mentioned in Chapter Three, Title I funding is intended to provide monetary assistance to schools that have a large number of economically disadvantaged students, with the expectation that such funding be dedicated to professional development, additional staff, after-hours programs or any other strategies that might result in increasing students' academic proficiency. Most school principals are given the discretion in how to use this funding to meet the individual and collective needs of their students, and they stretch that funding as far as they can. One principal from a very successful Title I school re-evaluates every year the way in which funding is used in her school and applauds the autonomy she is given by the district office to use funding how she deems most appropriate. She says, "You can't make a difference if you're not given the ability to make a difference" (A1).

The most common uses of funding among the Title I schools in this study were (a) after-school programs, (b) additional teaching staff (e.g., paraeducators, literacy coaches, etc.), and (c)

smaller class sizes. The after-school programs appear to serve more than one purpose. Not only do they provide a chance for remediation for students who are lagging behind their classmates in reading and math, but they often provide a safe setting for some of these students whose parents work long hours. Rather than go home after school to an empty house and unsupervised activity, the students are given a productive and secure environment in which learning and mentoring can take place.

Many principals also talked about the value of hiring additional teaching staff, such as paraeducators, literacy coaches, reading or math coordinators, and bilingual aids. Power Hour is used in many classrooms as a way in which these additional staff members can work with children, one on one, to target specific needs and deficiencies. Since many Title I schools have a significant number of ELL students, some classroom teachers have found it particularly useful to hire bilingual assistants, usually Spanish speakers. This has benefited the classroom teachers, who can proceed with the scheduled curriculum instead of taking class time to address the needs of a handful of students.

Perhaps because they come from a teaching background, most principals do see the difference in class environment that something as simple as class size can make. Many use Title I funding to ensure their class sizes do not reach an unreasonable level. This is particularly significant when considering the complex socioeconomic and cultural mix that can be found in some Title I classrooms. Smaller class size allows the teacher more one-on-one time with students as well as more flexibility in the use of teaching strategies and learning activities.

Principals also mentioned using Title I funding to maintain or enhance teacher development and training programs in an attempt to bring some of their teachers up-to-par with their more successful colleagues. In some instances, teacher training programs have been

diminished or eliminated as a result of budget constraints at the district level. One non-Title I teacher explained,

> Two years ago it was easier to provide them with time to do that [get together in grade-level teams] because we had collaboration time provided by the district. This year we lost it . . . so we had to be very creative in how we're able to get time for the teachers. (D2)

At least in this instance, funding for Title I schools places them at an advantage over non-Title I schools who have lost their teacher development programs. Often, funding is simply used to supplement the salary of a teacher who might be asked to work after-school hours, either for instructing students or for professional development opportunities which are no longer offered during the course of a regular school day.

Principals who had used federal funding to supplement their resources recognized that, without such funding, they would not be as successful. "Would I be able to do the things I do to impact kids without funding? Never!" Principal A1 insisted. As mentioned earlier, because the dispersement of Title I funding is often discretionary, there may be some non-Title I schools, as there were in this study, who qualify for Title I funding but because of decisions made at the district level are excluded from receiving funds. These schools face the same challenges as Title I schools, but without the same funding and support. Since Title I funding seems to make a significant difference in a school's ability to provide the resources that enhance the chances of passing AYP, failure to receive additional funding is largely detrimental to schools whose profile is so similar to that of a Title I school. One principal made this observation:

> If you're a Title I school, great! You get those resources, but the sister school down the street might be just a few percentage points lower and not receive any money, and they're

> still expected to make the same progress with less money. . . . Whether that's fair or not, the expectation's the same. (B2)

One principal said that Title I funding may even tip the scales in a school's effort to pass AYP. She mentioned that every Title I school in her district meets AYP "They have aids in every grade-level . . . they have money to pay teachers for after-school programs; they have all kinds of resources and support that we do not have" (D2). With recent budget cuts, this principal lost funding to pay her teachers for after-school programs, so she had to rely on volunteers in order to continue their after-hours activities.

All of the principles agreed that additional funding can make a significant difference in the programs and resources they are able to provide their students. Title I and non-Title I principals alike would love to see resources used to lower class size. Some of the Title I schools are using their funding to do just that, but schools that do not receive additional funding don't have that luxury. One non-Title I principal stated,

> I think class size makes the biggest difference in our success. I would probably have smaller class sizes so that teachers could give more individual attention to students because . . . human interaction is really what makes the difference in how students do. (C2)

Besides lowering class size, additional funding could also be earmarked for much-needed supplies and equipment in the classroom. Up-to-date computers and computer programs, for example, would prove useful in supplementing the core curriculum and learning activities. Extending the school day was also at the top of the list of ways in which funding could help accommodate the needs of children who lag behind. Whether used for resources, supplies, additional personnel, or an increase in salaries for teachers who participate in after-school

activities, funding is a big part of the equation.

Summary

Twelve elementary school principals were interviewed for this study. Asked to describe and share their experiences and observations with regards to the impact of high-stakes testing on teaching and learning in their schools, they were candid and forthright. Some were more passionate about the issues; some were less talkative. All have been affected by the NCLB legislation and the mandate for high-stakes testing as a means of encouraging accountability.

I expected the principals from AYP-passing schools to differ greatly in their experience and opinions from principals whose schools have had less success with showing AYP. I did not find that to be the case. Although there may have been a difference in the degree to which principals expressed their objections or concerns, the fact that their school either passed or failed AYP did not seem to alter their opinion significantly.

Even though all of the principals shared common experiences, they also expressed various viewpoints based on their personal philosophies and unique backgrounds. Out of all the characteristics that separated them from one another, none was more influential in determining their experiences than school demographics. Every principal in this study agreed that principals from Title I schools have a unique set of challenges. Non-title I schools are certainly not exempt from enrolling students from low socioeconomic brackets, ELLs, or students with special needs, but students from these special populations tend to be more prevalent in Title I schools. The non-Title I principals could easily envision the challenges that large populations of these students might present in the school setting.

Regardless of the observations expressed in this study, all principals interviewed seemed to appreciate the opportunity of being asked their opinion. They tried to share both positive and

negative experiences, wanting it known that, with all of their concerns, they are very much aware of the good things that come out of NCLB and high-stakes testing. They believed that because of this legislation, they, as administrators, were doing a better job. They seemed committed to continue doing the best job they could in working toward making AYP—hoping that this, indeed, is a positive step in providing a good quality education for all of their students.

Chapter Five: Discussion of Findings

Summary

One reoccurring theme to come out of the literature review and the research findings in this study is that there is no simple answer to educational reform. Although deficiencies in the educational system have long been acknowledged, there has rarely been consensus on how to proceed. Decade after decade, legislators and educators involved in school reform seemed to believe that they had the long sought-after solution that would eliminate the academic achievement gap and result in major improvements in schools. However, with no indication that any one solution has been successful in accomplishing its intended purpose, each new school reform movement or legislation has been followed by another.

When the No Child Left Behind Act was launched in 2001, there were high expectations that finally the nation would accomplish the long-term goal of providing all children an equal opportunity to obtain a high-quality education (U.S. Department of Education, 2002). The assumption that seemed to accompany NCLB was that the problem of underachievement could be corrected by making students, teachers, schools, and principals more accountable for what went on behind classroom doors. In an honest effort to get America's schools back on track and raise the bar for academic proficiency for all students, NCLB, along with its mandates and sanctions, was born.

Like every educational reform that has come before, NCLB has had its share of critics and supporters. The proponents of NCLB point to increasing test scores as evidence of its success. The critics are not as quick to use test results as the measuring stick by which success can be determined. As in other reform legislation, an outside observer may find in NCLB both things to celebrate and things to deplore.

One of the most debated provisions of NCLB was the demand for accountability, monitored by requiring all schools to make adequate yearly progress (AYP) by meeting certain academic requirements for all subgroups of students in their schools. States were given the charge to establish proficiency levels and administer a year-end test that would assess students' academic performance. The pressure to meet AYP was meant to motivate teachers and schools in their efforts to assist students in performing well on the required, year-end test. Failure to make adequate progress each year could result in possible sanctions for the school and school district.

The intent of this study was to explore the impact high-stakes testing has had on teaching and learning (i.e., teaching strategies, curriculum, and student achievement) in the state of Utah, from the standpoint of the public elementary school principal. Questions were developed with this goal in mind and a concerted effort was made to obtain all points of view, both positive and negative.

The impact of high-stakes testing. With any new school reform, impact evaluation takes years. The consequences of NCLB—both intended and unintended—are surfacing over time. The mandate that all states administer testing every year in reading and mathematics in grades 3 through 8 was designed to impact teaching in the classroom through enhanced accountability. But reformers didn't know for sure how teaching would be impacted.

Teaching. The principals in this study agreed that No Child Left Behind certainly got their attention and helped direct their focus to what was going on in the classroom. Overall, the principals reported feeling more organized and more purposeful in their daily activities. They felt that NCLB encouraged dialogue among their teaching staff and between teachers and administrators. They were grateful for the collaboration and team-building that was taking place,

and thought it beneficial to their teachers and, ultimately, to the students.

The mandate to get all students to a state-approved proficiency level by 2014, coupled with the stigma attached to not making adequate yearly progress, determines the principals' priorities. I expect that, even without NCLB, most principals were aware of what their teachers were teaching at any given time. What they may not have had prior to NCLB is a common goal. Although there were no specific standards attached to it, NCLB created a nation-wide objective. Principals were very concerned that teachers were teaching a curriculum that would result in improved test scores, and they encouraged teachers to use the results of the CRTs to guide their instruction so improvements could be made for the following year—improvements, of course, that they hoped would result in better test scores. Higher test results became the goal.

By not using the year-end test to inform the student of their academic progress, the test has become more of an instrument in determining what the teacher is doing than what the student is learning. It is ultimately the teacher, school, and district that are being held accountable for test results, not the student. It's no wonder that teacher morale is impacted by their students' test scores.

Although accountability is not necessarily a bad thing, placing sanctions on schools, districts, or states for insufficient test results is not an effective means to encourage change. The threat of sanctions diverts attention from actual student achievement to increasing test scores, which is not necessarily synonymous. Since educators and policymakers can't even agree that the one-time, end-of-the-year test scores are a good indication of learning, it seems inappropriate to give them so much weight as to result in schools' being restructured or closed.

Also problematic in NCLB is the provision allowing states to set their own proficiency standards. This provides a great opportunity for inconsistencies from year to year, and from

state to state, possibly resulting in arbitrary standards implemented solely at the discretion of the state superintendent or school board. The pressure to meet AYP encourages state-level administrators to lower achievement standards. Of course, creating a national standard presupposes a larger role for the federal government which expressly violates one of the objectives of NCLB in allowing states more flexibility and freedom.

Curriculum. With the pressure to make AYP each year, it is not surprising that teachers sometimes "teach to the test" and, thus, narrow the curriculum. The objective imbedded in AYP of reaching proficiency in language arts and mathematics is already pretty narrow. The implication of "teaching to the test" is that all other subjects and activities not included on the year-end CRTs will be ignored or minimized. There is no doubt that reading and problem-solving are foundational to a good education, but there's a concern that the way in which they are being taught excludes other opportunities for in-depth, higher-level thinking (Darling-Hammond, 1997; Dorgan, 2004; Kohn, 1999; Pedulla, 2003; Ravitch, 2010; Thompson, 2001).

The principals in this study were aware that "teaching to the test" ran the risk of diminishing opportunities for higher-level learning and understanding. The non-Title I principals whose schools were challenged in making AYP were the most vocal in expressing their desire that their teachers were teaching for "understanding" as opposed to teaching solely for the purposes of improving test scores. Without the fear of sanctions, they seemed more accepting of their AYP results. One of the principals insisted, "I don't care if we have a zero [on the test], and our results are published as such . . . a test doesn't mean anything to me. What matters is, 'Are they understanding?'" (D1).

Some of these same principals, however, were not concerned with the amount of attention given to language arts and mathematics, as they felt some teachers needed to be more

focused. Giving faculty clear parameters on which to place their attention was viewed as a good thing.

Even though the schools represented in this study also used the state core curriculum (science, social studies, art, visual arts, music, dance, theatre, health education, physical education, educational technology and library media) as a guideline, the amount of time and focus on the core curriculum varied from school to school and from teacher to teacher. In order for the core curriculum to be of value, I concur with Ravitch (2010) who suggests that "these subjects should not be discretionary or left to chance. Every state should have a curriculum that is rich in knowledge, issues, and ideas, while leaving teachers free to use their own methods" (p. 236). I further believe that instruction in every subject, whether it is English, science, art, or technology, should include reading, writing, and problem-solving. The subject matter should be the means by which these foundational skills are taught and reinforced.

The principals in this study did not observe a significant lack of teacher creativity in the classroom as a result of NCLB and AYP. Although teachers did not have as much flexibility as in the past, they still used some discretion on what took place in the classroom. The principals saw teachers as being creative in finding ways to incorporate core subjects across the curriculum. However, these observations do not seem to concur with the literature review and surveys which have been conducted with teachers. It may be that teachers, as opposed to principals, did not consider what they were doing with the integration of curriculum as necessarily being very creative. What they seem to have lost, as a result of the NCLB legislation, is the encouragement and motivation to get creative in the classroom in ways that take time. Test subjects and test-prep activities restrict the amount of time they have to spend on a particular subject and may dictate the kind of teaching strategies used, resulting in fewer activities, outings, or events that

may be creative, and might benefit students, but are too time-consuming to fit into the new approach.

Learning. The principals in this study agreed that the disaggregation of data was very beneficial in making them aware of how certain populations of their students were performing academically. Particularly with Title I principals, the disaggregation of data was extremely significant as it helped them determine how to direct funds to where they would do the most good in moving students closer to the proficiency level. All of the principals in this study represented schools which dealt with special student populations, such as English Language Learners, multi-cultural students, economically disadvantaged students, and students with disabilities. Although students from many of these subgroups experience academic challenges, only the schools with a high percentage of economically disadvantaged students receive the additional funding. Non-Title I schools with a significant population of ELL students, for example, often face the same challenges as schools with a high percentage of lower income students, but lack the funding to address their academic deficiencies. The successful Title I principals agreed that the additional funding they received made a significant contribution to what they were able to provide for their academically challenged students.

Additional funding is most often earmarked towards programs and materials intended to help the students perform better on the test at the end of the year. So, once again, test results become the target of attention and funding. Although teachers and administrators are spending so much time and effort on improving test scores, students, according to the principals in this study, do not seem to be as concerned with test results. Students never know how they do on the test. It is not graded. They still get promoted to the next year in school. There are no personal ramifications whether they excel or perform poorly on the test.

I suspect that high school students view high-stakes testing a bit differently than their elementary school counterparts. Unlike the young students discussed in this study, high school students from states that require an exit exam for graduation may be directly impacted by high-stakes testing results. After 12 years of formal education, not being able to receive a high school diploma due to a one-time, year-end test could be devastating.

Even without the same consequences faced by high school seniors, the principals in this study agreed that placing so much weight on a one-time, year-end test is inappropriate. I agree. I am particularly concerned that the assessment instrument may not be a valid and reliable measurement of academic achievement. Unless CRTs are accompanied by other formative assessments, teacher observations, homework, and participation in other classroom activities, it is hard to accurately evaluate a student's academic proficiency level. To base school and teacher performance on how well the student has performed on a one-time test administered at the end of the school year is inappropriate and damaging.

Even if test results may be able to provide some information on knowledge acquisition at a particular moment in time, they are only a snapshot of what a student may actually know. I do agree that some kind of assessment or evaluation needs to be done to determine a student's competencies, but unless it is used to correct or strengthen their academic deficiencies, it seems of little value. In order to declare that all students have reached a certain proficiency level, legislators would need to require that the same students be measured year after year and that their individual academic performance be compared to previous years, taking into account that their expected proficiency level should continue to increase from grade to grade. In schools with high turnover, the current system of scoring whole classes and grade levels cannot be an accurate measure of progress.

It may be helpful to also have a means by which teachers are regularly evaluated, possibly through classroom observation and personal interviews. To use students' test scores as a way to measure a teacher's success does not seem to be a fair or accurate indication of competencies and teaching style.

The impact of demographics on principals' observations. It was the consensus that almost all Title I schools have a different and more challenging demographic than do non-Title I schools. The only exception here is for schools that qualify for Title I funding but have been excluded due to decisions made in the district office. (Refer to Table 4.1, School D2, for example.) These schools have the same challenges as Title I schools, but do not receive compensation by way of additional funding. Non-Title I principals, although just as committed as their Title I colleagues, did display some relief in not having to deal with the requirements for schools that are receiving Title I funding.

In addition to having a fairly low number of economically disadvantaged students, the highest performing non-Title I schools also have the lowest number of students in the other subgroup categories. This is illustrated by the information in Table 5.1 that compares the demographics of two schools in the sampling from the same school district.

Table 5.1

Demographic Comparison of a Title I and non-Title I School in the Same School District, by Number of Students in Each Subgroup

School	Title I	Pass/ Fail AYP	Asian	African Am.	Am. Indian	Pacific Islander	Hisp.	Econ. Disadv.	ELL	Students w/disabil.	Total
C3	No	Pass	6	9	3	43	8	52	12	40	173
B3	Yes	Fail	14	16	3	40	158	238	178	31	678

Note. School C3 has a total enrollment of 534; School B3 has a total enrollment of 553.

Even with funding, Title I principals have a number of distractions not always common to the non-Title I school. In addition to focusing on the academic needs of their students and deciding where to direct additional funding to supplement the regular school day, Title I principals also acknowledged special situations that are often connected to their particular demographic. For example, the challenges of dealing with a transient population are numerous, with kids coming and going, some with incomplete student records or academic histories.

Conclusion

The question of whether high-stakes testing is a good thing is very complicated and cannot be answered by a simple yes or no. High-stakes testing, just like NCLB, has its supporters and its critics. However, there is little dispute that whenever high stakes are attached to testing, school policies, procedures and methodology will be impacted. Whether this has been beneficial or harmful to actual learning depends on the individual asked. What researchers do know is that there is no simple answer to improving or measuring student achievement.

As a result of the literature review and this study, I have added my own observations and opinions on where we should go from here. I believe that fundamental to changing the inequalities evident from state to state and school to school is the way in which education is funded. Besides comparing the differences between Title I and non-Title I schools, school funding was not a part of this study, so I do not address this issue specifically. However, I mention it here only briefly in context of the larger question of what should be addressed if future school reform is to be more successful than what has happened in the past.

The differences between schools in low socioeconomic neighborhoods and schools that serve high to middle-class populations point to inequalities based mainly on finances. Although Title I funding provides some assistance, it is not the solution. When first implemented in 1965,

Title I funding was intended to help repair such income-based inequalities (see Kozol, 1991). It seems that 45 years is long enough to determine whether something is working—and so far, Title I funding does not seem to be enough. Something more needs to be done to balance out the inequities in current school funding policies if positive changes are to occur in schools.

In addition to funding, educators and policymakers need to have a clear vision of what a good education looks like. Ravitch (2010) stated that

> The goal of education is not to produce higher scores, but to educate children to become responsible people with well-developed minds and good character. . . . The unrelenting focus on data that has become commonplace in recent years is distorting the nature and quality of education. (p.228)

I would like to see more attempts at bringing well-respected and experienced educators to the table with legislators to create a clear definition of what education should look like in this country. Until the nation can agree on something so fundamental, policymakers are apt to repeat mistakes from the past.

As part of this dialogue, stakeholders must arrive at a clear understanding of what should be included in a high-quality curriculum. Although language arts and mathematics are fundamental to any curriculum, they should not be taught to the exclusion of other subjects such as science, civics, literature, the arts, geography, history, foreign language, or technology. Although there is some discussion on creating a national curriculum, I'm not ready to support that recommendation. It would be well for states to create a good, robust curriculum for their students based on standard recommendations, but I'm concerned with too much involvement at the federal level.

To help them teach this curriculum, teachers should be provided with regular professional development opportunities that support good teaching strategies and classroom practices. There are so many ways in which content material can be taught and reinforced. Teachers should be exposed to as many of these teaching strategies as possible and should be familiar with the multiple ways in which students learn.

In conjunction with quality curriculum is a system of high-quality evaluation. Accountability should only be measured through accurate and reliable evaluations that align with the curriculum and are appropriate to both the student and the school. There should be multiple ways in which students can demonstrate their proficiency and understanding of subject matter, including formative assessments, homework, observation, classroom activities and discussions, projects and presentations, both written and oral. The intent of evaluations should be to remediate and improve, not to discipline or punish. Removing students from a "failing" school or increasing the number of charter and private schools does not solve the problem.

Teachers are expected to be highly qualified in the subject matter, pedagogy, and human development, and they should be compensated as professionals. I don't support the practice of merit pay, as I believe merit is too difficult to determine. Further, if schools pay teachers appropriately to begin with, schools should not need to compensate them for doing what is expected in the first place. Through regular evaluation, interviews and observations, teachers, just like students, should be encouraged to do their best and should be provided with the resources to continually improve.

A good, quality education for all students will not be achieved until policymakers (a) change how schools are funded; (b) agree on what education should be, the best measurements

and tools for assessing knowledge attainment, and how to use test results to actually improve proficiency levels; and (c) treat teachers like the professionals they expect them to be.

Future Research

The Obama administration has announced their intent to revert back to the title used in 1965, when the Elementary and Secondary Education Act was first created, for referring to their educational reform legislation (U.S. Department of Education, 2010). Although No Child Left Behind, as a common frame of reference, may be replaced with ESEA, its impact will be felt for decades to come. The literature continues to increase with articles, opinions and research on the consequences of high-stakes testing and the pressure to make adequate yearly progress. Unless there are significant changes to the reauthorization of ESEA with the current administration, I expect that policymakers and educators will continue to dwell on studies targeted at the provisions and mandates first introduced by NCLB.

There were some differences between what the principals in this study observed and experienced and the experiences mentioned in the literature review. As an example, the literature clearly articulated the diminished opportunities for teachers to be creative in the classroom, whereas the principals in this study did not feel that was a major concern. The renewed focus on the classroom and the value of disaggregating the data was noted by the principals in this study as being a benefit of the NCLB legislation, but did not receive as much attention in the literature. I don't know that additional studies with other populations, possibly principals in other states or even at other schools within the state of Utah, would uncover any new observations or experiences.

There is a consensus, however, apparent in both the literature and this study, that high-stakes testing does have an impact on teaching and learning in the classroom. Rather than

continue to address the significance of this impact, I would like to see additional research directed toward the value of high-stakes testing in determining academic proficiency. Is it really measuring what policymakers think it is? Is it an accurate indication of a teacher's or school's performance? If not, then why is it given so much credence? Is there a better way to measure a student's performance? Is there a better way to encourage accountability in the classroom?

Do we really believe that requiring schools to make AYP and have all of their students reach academic proficiency by a certain year will produce desired outcomes? What will legislators do in 2014 when students cannot meet the required proficiency level? What then?

America's public education system is in need of repair. Although previous school reform measures have attempted to address the issues, they have not been successful. And reforms may never succeed until educators and legislators come together to resolve the fundamental questions that will lead to a public education system that will benefit all children.

References

Abrams, L. M. (2004). Teachers' views on high-stakes testing: Implications for the classroom. *Educational Policy Studies Laboratory*. Retrieved November 29, 2008, from http://epicpolicy.org/files/EPSL-0401-104-EPRU.pdf

Abrams, L. M., & Madaus, G. F. (2003). The lessons of high stakes testing: Research shows that high-stakes tests can affect teaching and learning in predictable and often undesirable ways. *Educational Leadership, 61*(32), 31-55.

Abrams, L. M., Pedulla, J. J., & Madaus, G. F. (2003). Views from the classroom: Teachers' opinions of statewide testing programs. *Theory into Practice. 42* (1), 18-29.

Alden, D. (1998). Are we teaching American citizens or training Prussian serfs? *Nevada Journal, 6*(5). Retrieved February 17, 2007 from http://nj.npri.org/nj98/05/prussian.htm

Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10*(18). Retrieved May 27, 2005, from http://epaa.asu.edu/epaa/v10n18/

Amrein, A. L., & Berliner, D. C. (2003). The effects of high stakes testing on student motivation and learning. *Educational Leadership, 60*(5), 32-38.

Bazeley, P. (2008). *Qualitative data analysis with NVivo* (2nd ed.). Thousand Oaks, CA: Sage.

Bersin, A. D. (2005). Making schools productive: The point of accountability and the key to renewal [Electronic version]. *Education Week, 24*(32), 30, 40.

Black, P. & William, D. (1998). Inside the black box: Raising standards through classroom assessment [Electronic version]. *Phi Delta Kappan*, *80*(2).

Boardman, A. G., & Woodruff, A. L. (2004). Teacher change and "high-stakes" assessment: What happens to professional development? *Teaching and Teacher Education*, *20*(6), 545-557.

Bolon, C. (2001, October 16). Significance of test-biased ratings for metropolitan Boston schools. *Educational Policy Analysis Archives,* 9(42). Retrieved August 22, 2005, from http://epaa.asu.edu/epaa/v9n42

Bowman, C. (2006). Johann Gottlieb Fichte (1762-1814). In *The Internet encyclopedia of philosophy.* Retrieved from http://www.iep.utm.edu/f/fichtejg.htm

Brandon, P. R. (2002). Two versions of the contrasting-groups standard-setting method: A review. *Measurement and Evaluation in Counseling and Development, 35*, 167-181.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Bussert-Webb, K. (1999). To test or to teach: Reflections from a holistic teacher-researcher in south Texas. *Journal of Adolescent and Adult Literacy, 42*(7), 582-585.

Chapman, L. H. (2004). No child left behind in art? [Electronic version]. *Arts Education Policy Review, 106*(2), 3-17.

Charles, D. (2008). *Teachers' perspectives on the unintended consequences of high-stakes testing* (Doctoral dissertation, Louisiana State University).

Chase, B. (2000). Making a difference. In D. Meier (Ed.), *Will standards save public education?* (pp. 40-43). Boston, MA: Beacon Press.

Chudowsky, N., & Pellegrino, J.W. (2003). Large scale assessments that support learning: What will it take? [Electronic version]. *Theory into Practice, 42*(1), 75-83.

Cimbricz, S. (2002). State-mandated testing and teachers' beliefs and practice. *Education Policy Analysis Archives, 10*(2). Retreived from http://epaa.asu.edu/epaa/v10n2.html

Clarke, M., Shore, A., Rhoades, K., Abrams, L., Miao, J., & Li, J. (2003). Perceived effects of state-mandated testing programs on teaching and learning: Findings from interviews with educators in low-, medium-, and high-stakes states. Chestnut Hill, MA: National Board on Educational Testing and Public Policy, Boston College. Retrieved September 22, 2009, from http://www.bc.edu/research/nbetpp

Collins, M., & Tamarkin, C. (1982). *Marva Collins' way: Returning to excellence in education.* New York, NY: Penguin Putnam.

Common Core State Standards Initiative. (2010). About the standards. Retrieved from http://www.corestandards.org/about-the-standards

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

Cremin, L. (Ed.). (1957). *The republic and the school: Horace Mann on the education of free man.* New York, NY: Teachers College Press.

Cruse, K. L., & Twing, J. S. (2000). The history of statewide achievement testing in Texas. *Applied Measurement in Education, 13*(4), 327-331.

Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work.* San Francisco, CA: Jossey-Bass.

Darling-Hammond, L. (2004). From "separate but equal" to "No Child Left Behind": The

collision of new standards and old inequalities. In D. Meier & G. Wood (Eds.), *Many children left behind: How the No Child Left Behind Act is damaging our children and our schools* (pp. 3-32). Boston, MA: Beacon Press.

DeMille, O. V. (2000). *Teaching a generation of leaders: A Thomas Jefferson education for the twenty-first century.* Cedar City, UT: George Wythe College Press.

Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education.* New York, NY: The Free Press.

Domenech, D. (2000, December 1). My stakes well done [Electronic version]. *School Administrator*.

Dorgan, K. (2004). A year in the life of an elementary school: One school's experience in meeting new mathematical standards. [Electronic version]. *Teachers College Record, 106*(6), 1203-1228.

Educational Policies Commission. (1938). *The purposes of education in American democracy.* Washington, DC: National Education Association of the United States and the American Association of School Administrators.

Ediger, M. (2003). Philosophy and measurement of school achievement. *Journal of Instructional Psychology, 30*(3), 231-236.

Ferrandino, V. L. & Tirozzi, G. N. (2002). Beyond standards [Electronic version]. *Principal, 81*(3), 56.

Finn, C. E., Jr., & Petrilli, M. J. (2007). Forward: The proficiency illusion. In J. Cronin et al. (Eds.), *The proficiency illusion* (3rd ed., pp. 2-5).Washington, DC: Thomas B. Fordham Institution and Northwest Evaluation Association.

Firestone, W. A. & Schorr, R. Y. (2004). Introduction. In W. A. Firestone, R. Y. Schorr, & L. F. Monfils (Eds.), *The ambiguity of teaching to the test: Standards, assessments, and educational reform* (pp. 1-17), Mahwah, NJ: Lawrence Erlbaum Associates.

5 schools to test teacher performance pay program. (2010, June 24). *Daily Herald.* Retrieved from http://www.heraldextra.com/news/state-and-regional/article_d5ad30a7-feea-5dc4-8571-db3bd8af5eb6.html

Flick, U. (2006). *An introduction to qualitative research* (3rd ed.). Thousand Oaks, CA: Sage.

Flick, U. (2007). *Designing qualitative research*. Thousand Oaks, CA: Sage.

Frey, J. H., & Oishi, S. M. (1995). *How to conduct interviews by telephone and in person.* Thousand Oaks, CA: Sage

Galley, M. (2001). State tests don't support good instruction [Electronic version]. *Education Week, 21*(9), 5.

Gatto, J. (n.d.). History: The birth of formal schooling in the USA. Retrieved February 17, 2007, from the CT Homeschool Network website: http://www.cthomeschoolnetwork.org/EducationalFreedom-History.htm

Gatto, J. (2003). *The underground history of American education: An intimate investigation into the prison of modern schooling.* New York, NY: The Oxford Village Press.

Gibbs, G. (2007). *Analyzing qualitative data.* Thousand Oaks, CA: Sage.

Gibbs, G. R. (2002). *Qualitative data analysis: Explorations with NVivo.* Philadelphia, PA: Open University Press.

Goertz, M., & Duffy, M. (2003). Mapping the landscape of high-stakes testing and accountability practices [Electronic version]. *Theory into Practice, 42*(1), 4-12.

Gulek, C. (2003). Preparing for high-stakes testing [Electronic version]. *Theory into Practice, 42*(1), 42-51.

Guskey, T. R. (2003). How classroom assessments improve learning [Electronic version]. *Educational Leadership, 60*(5), 6-11.

Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher, 20*(5), 2-7.

Haney, W. (2000, August 19). The myth of the Texas miracle in education. *Education Policy Analysis Archives, 8*(41). Retrieved June 16, 2005, from: http://epaa.asu. edu/epaa/v8n41

Haney, W. (2002, May 6). Lake Woebeguaranteed: Misuse of test scores in Massachusetts, part 1. *Education Policy Analysis Archives, 10*(24). Retrieved August 17, 2005, from http://epaa.asu.edu/epaa/v10n24

Harvey, J. (2003). The matrix reloaded: The internal contradictions of No Child Left Behind are creating a nightmare. *Educational Leadership, 61*(3), 18-21.

Hess, F. (2003). Refining or retreating? High-stakes accountability in the states. In P. E. Peterson & M. R.West (Eds.), *No child left behind? The politics and practice of school accountability* (pp. 55-79). Washington, DC: The Brookings Institute.

Heubert, J. P., &. Hauser, R. M. (Eds.). (1999). *High stakes: Testing for tracking, promotion, and graduation.* Washington, DC: National Academy Press.

Hoffman, J. V., Assaf, L. C., & Paris, S. G. (2001). High-stakes testing in reading: Today in Texas, tomorrow? *The Reading Teacher, 54*(5), 482-492.

Jones, B. D., & Egley, R. J. (2004). Voices from the frontlines: Teachers' perceptions of high-stakes testing. *Educational Policy Analysis Archives, 12*(39), 1-34.

Jones, G., Jones, B. D., and Hargrove, T. Y. (2003). *The unintended consequences of high-stakes testing.* Lanham, MD: Rowman & Littlefield.

Jones, K. (2004). A balanced school accountability model: An alternative to high-stakes testing. *Phi Delta Kappan*, *85*(8), 584-590.

Kane, T. J., & Staiger, D. O. (2002). Volatility in school test scores: Implications for test-based accountability systems. In D. Ravitch (Ed.), *Brookings papers on educational policy* (pp. 235- 283). Washington, DC: Brookings Institution.

Kaplan, L. S., & Owings, W. A. (2001). How principals can help teachers with high-stakes testing: One survey's findings with national implications. *NAASP Bulletin, 85*(15), 15-23.

Karp, S. (2004). NCLB's selective vision of equality: Some gaps count more than others. In D. Meier & G. Wood (Eds.), *Many children left behind: How the No Child Left Behind Act is damaging our children and our schools* (pp. 3-32). Boston, MA: Beacon Press.

Kober, N., Zabala, D., Chudowsky, N., Chudowsky, V., Gayler, K., & McMurrer, J. (2006). *State high school exit exams: A challenging year.* Retrieved from the Center on Education Policy website: http://www.cep-dc.org

Kohn, A. (1993). *Punished by rewards: The trouble with gold stars, incentive plans, A's, praise and other bribes.* Boston, MA: Houghton Mifflin.

Kohn, A. (1999). *The schools our children deserve: Moving beyond traditional classrooms and "tougher standards."* New York, NY: Houghton Mifflin.

Koretz, D. (1996). Using student assessments for educational accountability. In E.A. Hanushek & D.W. Jorgenson (Eds.), *Improving America's schools: The role of incentives* (pp. 171-195). Washington, DC: National Academy Press.

Kornhaber, M. L., & Orfield, G. (2001). High-stakes testing policies: Examining their assumptions and consequences. In G. Orfield & M. L. Kornhaber (Eds.), *Raising standards or raising barriers? Inequality and high-stakes testing in public education* (pp. 1-18). New York, NY: The Century Foundation Press.

Kozol, J. (1991). *Savage inequalities: Children in America's schools.* New York, NY: HarperCollins.

Krumboltz, J. D., & Yeh, C. J. (1996). Competitive grading sabotages good teaching [Electronic version]. *Phi Delta Kappan*, *78*(4), 324-326.

Kvale, S. (2007). *Doing interviews.* Thousand Oaks, CA: Sage.

Lin, Q. (2002). Beyond standardization: Testing and assessment in standards-based reform. *The Education Digest, 23*(4), 43-49.

Linn, R.L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher*, *32*(7), 3-13.

Linn, R. L., Baker, E. L., & Betebenner, D. W. (2002). *Accountability systems: Inplications of requirements of the No Child Left Behind Act of 2001* (CSE Technical Report 567). Los Angeles: University of California, Los Angeles.

Lofland, J. & Lofland, L. H. (1995). *Analyzing social settings: A guide to qualitative observation analysis.* Belmont, CA: Wadsworth Publishing.

Loss, C.G. & Loss, C.P. (2002). Progressive education. In J. W. Guthrie (Ed.), *Encyclopedia of Education* (2nd ed., Vol. 5, pp. 1933-1938). New York, NY: Macmillan.

Madaus, G., Russell, M., & Higgins, J. (2009). *The paradoxes of high stakes testing: How they affect students, parents, teachers, principals, schools and society.* Charlotte, NC: Information Age.

Madaus, G. F. & O'Dwyer, L. M. (1999). A short history of performance assessment [Electronic version]. *Phi Delta Kappan, 80*(9), 688-696.

Marshak, D. (2003). No Child Left Behind: A foolish race into the past [Electronic version]. *Phi Delta Kappan, 85*(3), 229-231.

McNeil, L., & Valenzuela, A. (2001). The harmful impact of the TAAS system of testing in Texas: Beneath the accountability rhetoric. . In G. Orfield & M. L. Kornhaber (Eds.), *Raising standards or raising barriers? Inequality and high-stakes testing in public education* (pp. 127-150). New York, NY: The Century Foundation Press.

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

Mehrens, W. A. (1998). Consequences of assessment: What is the evidence? *Education Policy Analysis Archives*, *6*(13). Retrieved May 24, 2005, from http://epaa.asu. edu/epaa/v6n13.html

Mertens, D. M. (1998). *Research methods in education and psychology: Integrating diversity with quantitative and qualitative approaches.* Thousand Oaks, CA: Sage.

Meyers, E., & Rust, F. O. (2000). The test doesn't tell all [Electronic version]. *Education Week, 19*(38), 34, 37.

Nairn, A. (1980). The reign of ETS: The corporation that makes up minds. *The Ralph Nader report on the Educational Testing Service*.

Nash, G.B. (2000). Expert opinion. In D. Meier (Ed.), *Will standards save public education?* (pp. 44-49). Boston, MA: Beacon Press.

National Policy Board for Educational Adminstration. (2008). *Educational Leadership Policy Standards: ISLLC 2008.* Retrieved from www.npbea.org/projects.php

Natriello, G., & Pallas, A. (1998). *The development and impact of high stakes testing.* Retrieved December 31, 2005, from Columbia University website: http://www.columbia.edu/~gjn6/histake.html

Neill, M. (1997). Transforming student assessment. *Phi Delta Kappan,* 79, 34-40.

Neill, M. (2003, February). The dangers of testing. *Association for Supervision and Curriculum Development,* 43-46.

Neill, M., Guisbond, L., Schaeffer, B., Madden, J., & Legeros, L. (2004). Failing our children. Retrieved from the FairTest website: http://www.fairtest.org/node/1778

Neuman, W. L. (2000). *Social research methods: Qualitative and quantitative approaches* (4th ed.). Boston, MA: Allyn & Bacon.

North Central Regional Educational Laboratory. (1994). *Summary of Goals 2000: Educate America Act*. Retrieved from http://www.ncrel.org/sdrs/areas/issues/envrnmnt /stw/sw0goals.htm

NVivo 8 [Computer software]. (2008). Cambridge, MA: QSR International.

Ohanian, S. (1999). *One size fits few: The folly of educational standards.* Portsmouth, NH: Heinemann.

O'Neill, J. (1993). Can national standards make a difference? [Electronic version]. *Educational Leadership, 50*(5), 4-8.

Pedulla, J. J. (2003). State-mandated testing: What do teachers think? *Educational Leadership, 61*(3), 42-46.

Pedulla, J. J., Abrams, L. M., Madaus, G. F., Russell, M. K., Ramos, M. A., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers.* National Board on Educational Testing and Public Policy. Boston, MA: Boston College.

Phelps, R. P. (2005). Persistently positive: Forty years of public opinion on standardized testing. In R. P. Phelps (Ed.), *Defending standardized testing* (pp. 1-22). Mahwah, NJ: Lawrence

Erlbaum Associates.

Plake, B. S. (2002).Evaluating the technical quality of educational tests used for high-stakes decisions. *Measurement and Evaluation in Counseling and Development, 35*, 144-152.

Popham, W. J. (2003). The seductive allure of data. *Educational Leadership, 60*(5), 48-51.

Popham, W.J. (2004). *America's "failing" schools: How parents and teachers can cope with No Child Left Behind.* New York: RoutledgeFalmer.

Posner, D. (2004). What's wrong with teaching to the test? *Phi Delta Kappan, 85*(10), 749-751.

Powell, A. G., Farrar, E., & Cohen, D. K. (1985). *The shopping mall high school: Winners and losers in the educational market place.* Boston, MA: Hougton Mifflin.

Pulliam, J. D. & VanPatten, P. J. (1995). *History of education in America.* Englewood Cliffs, NJ: Prentice-Hall.

Ravitch, D. (2000). *Left back: A century of battles over school reform.* New York, NY: Touchstone.

Ravitch, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education.* New York, NY: Basic Books.

Raymond, M. E., & Hanushek, E. A. (2003). High-stakes research. *Education Next, 3*, 48-55.

Reese, M., Gordon, S. P., & Price, L. R. (2004). *Teachers' perceptions of high-stakes testing*. Texas State University-San Marcos. Retrieved on June 15, 2009, from http://pdfs.scarecroweducation.com/SC/TJS/SCTJSLReeseManuscriptSep2004.pdf

Reeves, D. B. (2001). Standards make a difference: The influence of standards on classroom assessment [Electronic version]. *NASSP Bulletin, 85*(621), 5-12.

Resnick, D. P., & Resnick., L. B. (1985). Standards, curriculum, and performance: A historical and comparative perspective. *Educational Researcher, 14*(4), 5-20.

Rettig, M. D., McCullough, L. L., Santos, K., & Watson, C. (2003). A blueprint for increasing student achievement. *Educational Leadership, 61*(3), 71-76.

Richards, A. L. (2008, October). The new orgy of tabulation: Still trying to gather figs from thistles. Presented at the Far Western Philosophy of Education Society, Orem, UT.

Robbins, A. (2006). *The overachievers: The secret lives of driven kids.* New York, NY: Hyperion.

Rose, L. C. & Gallup, A. M. (2003). The 35th annual Phi Delta Kappa/Gallup poll of the public's

attitudes toward public schools [Electronic version]. *Phi Delta Kappan, 85*(1), 41-52.

Rosenshine, B. (2003). High-stakes testing: Another analysis. *Educational Policy and Analysis Archives, 11*(24). Retrieved May 27, 2005, from http://epaa.asu.edu/epaa/v11n24

Rotberg, I. C. (2001). A self-fulfilling prophecy [Electronic version]. *Phi Delta Kappan, 83*(2), 170-171.

Rudalevige, A. (2003). No child left behind: Forging a congressional compromise. In P. E. Peterson & M. R.West (Eds.), *No child left behind? The politics and practice of school accountability* (pp. 23-54). Washington, DC: The Brookings Institute.

Ryan, J. E. (2004). The perverse incentives of the No Child Left Behind Act. *New York University Law Review*, *79*, 932-989.

Sacks, P. (1999). *Standardized minds: The high price of America's testing culture and what we can do to change it.* New York, NY: Da Capo Press.

Sanders, W. L. (2003). *Beyond "No Child Left Behind."* Paper presented at the 2003 annual meeting of the American Educational Research Association, Chicago, Illinois. (ERIC Document Reproduction Service No. TM034898)

Schencker, L. (2010, August 6). Utah education leaders adopt new academic standards. *The Salt Lake Tribune.*

Schugurensky, D. (2002). *Selected moments of the 20th century: Elementary and Secondary School Act, the "War on Poverty," and Title I.* Toronto, Canada: The Ontario Institute for Studies in Education, the University of Toronto (OISE/UT).

Sharkey, N. S. & Murnane, R. J. (2003). Learning from student assessment results: How can schools use assessment data to improve student achievement? *Educational Leadership, 25*(3), 299-318.

Shepard, L. A. & Dougherty, K. C. (1991, April). Effects of high-stakes testing on instruction. Paper presented at the annual meetings of the American Educational Research Association, Chicago, IL.

Sizer, T. R. (2004). Preamble: A reminder for Americans. In D. Meier & G. Wood (Eds.), *Many children left behind: How the No Child Left Behind Act is damaging our children and our schools* (pp. 3-32). Boston, MA: Beacon Press.

Smith, M. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher, 20*(5), 8-11.

Spellings, M. (2005, May 6). Is America really serious about educating every child? Prepared remarks for Secretary Spellings at the Education Writers Association 58th National

Seminar, St. Petersburg, FL. Retrieved from http://www.ed/gov/news/speeches/2005/05/05062005.html

Spring, J. (1990). *The American school 1642-1990.* New York, NY: Longman.

Stiggins, R. J. (1999). Assessment, student confidence, and school success [Electronic version]. *Phi Delta Kappan, 81*(3), 191-198.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks, CA: Sage.

Supon, V. (2004). Implementing strategies to assist test-anxious students. *Journal of Instructional Psychology, 31*(4), 292-296.

Taylor, G., Shepard, L., Kinner, F., & Rosenthal, J. (2001). *A survey of teachers' perspectives on high-stakes testing in Colorado: What gets taught, what gets lost.* National Center for Research on Evaluation, Standards, and Student Testing. Los Angeles: UCLA.

Thompson, S. (2001). The authentic standards movement and its evil twin [Electronic version]. *Phi Delta Kappan, 82*(5), 358-362.

Thorn, A. R. & Mulvenon, S.W. (2002). High-stakes testing: An examination of elementary counselor's views and their academic preparation to meet this challenge. *Measurement and Evaluation in Counseling and Development, 35*, 195-206.

Tyack, D. B. (1974). *The one best system: A history of American urban education.* Cambridge, MA: Harvard University Press.

U.S. Department of Agriculture, Food and Nutrition Service. (2009). National school lunch, special milk, and school breakfast programs, national average payments/maximum reimbursement rates. 74 Fed. Reg. 134. Retrieved August 3, 2009, from http://www.fns.usda.gov/cnd/Governance/notices/iegs/IEGs09-10.pdf

U.S. Department of Education. (n.d., a). *Elementary & secondary education: Title I—Improving the academic achievement of the disadvantaged.* Retrieved November 18, 2009, from http://www.ed.gov/policy/elsec/leg/esea02/pg1.html

U.S. Department of Education. (n.d., b). *The facts about...measuring progress.* Retrieved February 24, 2005, from http://www.ed.gov/print/nclb/accountability/ ayp/testing.html

U.S. Department of Education. (1965). *Elementary and Secondary Education Act of 1965.* Retrieved from http://cwx.prenhall.com/bookbind/pubbooks/burns3/medialib/docs/esea1965.htm

U.S. Department of Education. (2002). *No Child Left Behind: A desktop reference.* Washington, DC: Office of Elementary and Secondary Education.

U.S. Department of Education. (2004). *U.S. Department of Education FY 2004:Performance and accountability report.* Washington, DC.

U.S. Department of Education. (2006). *No Child Left Behind Act is working.* Retrieved November 11, 2006, from http://www.ed.gov/nclb/overview/importance/nclbworking.html

U.S. Department of Education. (2007). *2007 Performance and accountability report highlights.* Retrieved February 1, 2008, from http://www.ed.gov/about/reports/annual/2007report/report-highlights.pdf

U.S. Department of Education. (2008). *An introduction to NAEP: National Assessment of Educational Progress.* Retrieved April 30, 2009, from http://nces.ed.gov/nationsreportcard/about/

U.S. Department of Education. (2010). *A blueprint for reform: The reauthorization of the Elementary and Secondary Education Act*. Retrieved July 2, 2010, from http://www2.ed.gov/policy/elsec/leg/blueprint/blueprint.pdf.

U.S. Department of Education, National Commission on Excellence in Education. (1984). *A nation at risk: The imperative for educational reform.* Washington, DC.

Utah Administrative Code. (2010). Rule 277-700-4. *The elementary and secondary school core curriculum.* Retrieved from http://www.rules.utah.gov/ publicat/code/r277/r277-700.htm#T4

Utah State Board of Education. (2004). *Performance Plus: A competency-measured achievement plan for all Utah students with progress-based assessment and accountability.* Retrieved May 28, 2007, from http://www.schools.utah.gov/BOARD/summary/Performance+0804.pdf

Utah State Legislature. (2002). *Standards and assessment processes to measure student performance – Basic skills competency test.* 53A-1-611. Retrieved May 29, 2007, from http://le.utah.gov/~code/code.htm

Utah State Office of Education. (2005). *State of Utah consolidated application accountability workbook.* Retrieved August 8, 2009, from http://www.usoe.k12.ut.us/nclb/documents/pdf/AYPPlan.pdf

Utah State Office of Education. (2006a). *Comprehensive assessment system:Utah State Office of Education* (Section 4, pp. 84-89). Retrieved May 28, 2007, from http://www.schools.utah.gov/assessment/Default.aspx

Utah State Office of Education. (2006b). UBSCT clarification of practice. Retrieved from http://www.schools.utah.gov/eval/documents/UBSCT_Clarification_of_Practice.pdf

Utah State Office of Education. (2006c). *U-Pass accountability system*. Retrieved from http://www.schools.utah.gov/assessment/documents/UPASS_Accountability_System.pdf

Utah State Office of Education. (2007). *Secondary school completion and diplomas.* Retrieved May 29, 2007, from http://www.rules.utah.gov/publicat/code/ r277/r277-705.htm

Utah State Office of Education. (2009). *Fingertip facts 2009*. Retrieved from http://www.schools.utah.gov/main/DATA-STATISTICS/Fingertip-Facts/DOCS/2009_FingertipFacts.aspx

VanderArk, T. (2004). Getting high school accountability right [Electronic version]. *Education Week*, *23*(41), 41, 52.

Vogler, K. E. (2002). The Impact of high-stakes, state-mandated student performance assessment on teachers' instructional practices. *Education*, *123*(1), 39-55.

West, M. R., & Peterson, P. E. (2003). The politics and practice of accountability. In P. E. Peterson & M. R.West (Eds.), *No child left behind? The politics and practice of school accountability* (pp. 1-20). Washington, DC: The Brookings Institute.

Williams, N. M. (2003). Thinking outside the bubble. *Educational Leadership, 61*(3), 82-83.

Winters, M. A., Greene, J. P., & Trivitt, J. R. (2008, July). Building on the basics: The impact of high-stakes testing on student proficiency in low-stakes subjects. *Civic Report*, No.54. Manhattan Institute for Policy Research.

Wood, G. (2004). Introduction. In D. Meier & G. Wood (Eds.), *Many children left behind: How the No Child Left Behind Act is damaging our children and our schools* (pp. vii – xv). Boston, MA: Beacon Press.

Appendices

Appendix A: Performance Plus Executive Summary

Performance Plus is a progress-based plan for education that focuses on the growth of each student toward full competency.

The Utah State Board of Education, along with state and local leaders in education, have called for increased competency of all high school graduates and a public education accountability plan that measures results. Utah's plan requires:

- Proficiency in reading, writing, and math, measured every year at every grade so as to track continual progress and provide early warnings of student failure
- Performance standards in each subject each year
- Instruction in explicit, systematic, and intensive phonics
- Instructional methods and materials that are research based and have proven track records of increasing student achievement
- Increased graduation competencies for all students
- Advanced coursework and learning experiences beyond basic proficiencies
- Measurable performance goals for students, schools, and districts
- Determination of progress and gain scores in reading, writing, and math for students, classes, grades, schools, districts, and the state of Utah
- Full implementation of Utah's Performance Assessment System for Schools (U-PASS), including annual reports of progress and results on the Consolidated Utah Student Achievement Plan
- K-16 articulation to create a seamless system

In addition, the plan recognized that education goes well beyond the basics and includes a bevy of elective choices such as fine arts, applied technology, and advanced, applied, or accelerated studies. In addition, leadership, extracurricular, and academic service learning opportunities are acknowledged as essential elements of a good education.

To ensure each child in Utah succeeds, competency-measured classrooms will be characterized by:

1. Clear academic content standards embedded in rigorous instruction
2. Targeted professional development focused on student achievement results
3. High performance standards for all children
4. Differentiated instruction to provide individual growth in learning
5. Ongoing and frequent assessment to measure learning growth
6. Immediate interventions for students who struggle
7. Accountability for results for all stakeholders and for Utah's public education system

Retrieved from http://www.schools.utah.gov/BOARD/summary/Performance+0804.pdf

Appendix B: Utah Performance Assessment System for Students (U-PASS)

Purpose (retrieved from http://u-pass.schools.utah.gov/u-passweb/)

The purpose of the Utah Performance Assessment System for Students is to provide information about how students in Utah schools are doing relative to the Utah State Core Curriculum and State requirements for school performance. Information about Utah schools is provided to help in the interpretation of the achievement information. Schools vary significantly in size, characteristics of students, faculty, staff, and facilities, and all of these factors influence how the students do. In addition, student performance is influenced by many factors outside the school environment. Nevertheless, it is the goal of the State education system that all students learn the Core Curriculum each year so that they have the fundamentals necessary for a solid academic future.

U-PASS Legislation (retrieved from: http://www.schools.utah.gov/assessment/ info_upass.aspx)

The U-PASS legislation was enacted in 2000. This law requires an annual report of assessments and behavior indicators and includes a state accountability plan.
Achieved State Level of Performance (retrieved from: http://www.schools.utah.gov/ assessment/info_upass.aspx)
Schools are identified as Achieving the State Level of Performance if:

- The percentage of students participating on each assessment is 80% (or greater) **AND**
- The **Total School** overall status is 80% (or greater) or the **Total School** overall progress score is 190 (or greater) **AND**
- The **Subgroup** overall status is 80% (or greater) or the **Subgroup** overall progress score is 190 (or greater)

Overall Status (retrieved from: http://www.schools.utah.gov/assessment/ info_upass.aspx)

The overall status of a school is acceptable or unacceptable based on the total score calculated as follows:

Grades 3-8: 35% language arts, 35% math, 20% science, 10% attendance
High School: 30% language arts, 25% math, 25% science, 10% attendance, 10% graduation rate

Overall Progress (retrieved from: http://www.schools.utah.gov/assessment/Default.aspx)

The overall progress of a school and/or subgroup is a longitudinal measure defined as low, medium or high by comparing the achievement levels of the same student from one year to the next year.

The progress of a school and/or subgroup is determined as low, medium or high based on the total score calculated as follows:
Grades 3-8: 35% language arts, 35% math, 20% science, 10% attendance
High School: 30% language arts, 25% math, 25% science, 10% attendance, 10% graduation rate

Appendix C: Criterion-Referenced Test (CRT) Policy

English Language Arts CRT

This CRT is a grade specific test. Students should take the test that corresponds to the grade in which they are enrolled. Example: 9th grade student takes the 9th grade CRT.

Math CRT

For elementary students, this CRT is grade specific. Students should take the test that corresponds to the grade in which they are enrolled. Example: 3rd grade student takes the 3rd grade math CRT.

For secondary students, this CRT is course specific. Students should take the CRT at the completion of the course. Example: Students enrolled in pre-algebra should take the pre-algebra CRT.

Science CRT

For 4th-8th grade students, this CRT is grade specific. Students should take the test that corresponds to the grade in which they are enrolled. Example: 5th grade student takes the 5th grade science CRT.

For secondary students, this CRT is course specific. Students should take the CRT at the completion of the course. Example: Students enrolled in biology should take the biology CRT.

There are a variety of science courses that are based on the same core curriculum. Example, biology, human biology, AP biology, and agricultural biology are all responsible to the same core curriculum. Therefore, all students enrolled in these courses should take the biology CRT.

CRT Retakes

Students are not required to take the same CRT more than once. If a student takes two courses, two separate years (Chemistry, AP Chemistry), it is a local decision if the student is to retake the chemistry CRT.

For the purposes of U-PASS and NCLB, the first time a student takes a course for which there is an associated CRT, the student is required to take that associated CRT. The score/proficiency determination from CRTs will only be used from the first time the CRT was administered for a given course/grade. Students may take the same CRT (course/grade) again and that test would be scored by USOE; however, this score would be excluded from all U-PASS and CRT calculations.

This policy applies to all grades/courses assessed by CRTs.

USOE/Assessment & Accountability/1/19/06

Retrieved from http://www.schools.utah.gov/eval/Documents/CRT_Policy.pdf

Appendix D: The Utah Basic Skills Competency Test (UBSCT)

1. UBSCT shall be given twice annually.
   - Tuesday, Wednesday and Thursday of the **1st week of February**
   - Tuesday, Wednesday and Thursday of the **3rd week of October**
2. Each subtest shall be given on a separate day.
   - The same subtest shall be given to all students on the same day.
   - All sections of a subtest shall be completed in a single day.
3. The test is not timed.
   - Each subtest is designed to take approximately 90 minutes to complete. However, students shall be allowed enough time to complete each subtest, as long as it is completed within the designated testing day for each subtest.
4. School districts shall provide a makeup window not to exceed five school days immediately following the last day of each administration of the UBSCT.
5. All students should be given the opportunity to participate in all five administrations of the UBSCT, scheduled according to Board rule, during the time they are enrolled in Utah schools between the spring administration of their 10th grade year and the spring administration of their 12th grade year.
   - If a student takes the UBSCT one or more times in a district and then moves to another district, that student should be given the opportunity to participate in all remaining UBSCT administrations that are given through the end of the student's 12th grade year.
6. If a student has passed a high school exit exam in another state, a district may substitute that exam for the UBSCT.
   - Documentation should be the student's/parent's responsibility.
   - If there is any question that the student has not passed an adequate high school exit exam in a reciprocal state, the student should be required/urged to take every possible UBSCT administration offered while the student is enrolled in Utah schools.
   - The reciprocal state's exit exam must test the same basic areas covered in the UBSCT.
   - The student must take the exit exam within the grade levels required by the UBSCT.
   - The student must pass the exit exam at the reciprocal state's required level.
     - The student must pass the reciprocal state's exit exam in its entirety to qualify to substitute the reciprocal state's exam for the UBSCT; otherwise the student must take all three subtests of the UBSCT. (**It's an "all or nothing" policy; no credit for individual subtests will be allowed**.)
7. Students enrolling in a Utah school after February of their 10th grade year can attempt the UBSCT for every remaining administration after their enrollment. A student enrolling when less than three attempts are possible, if that student is not able to pass all three sections, but that student attempts every administration possible, will qualify for the Diploma—Did not pass UBSCT.
8. Diploma—Passed UBSCT
   Student passed all three sections of UBSCT
   Student met all graduation requirements

Districts design Diploma—Must say "Passed UBSCT" on Diploma

9. Diploma—Did **not** pass UBSCT
   Student did not pass all three sections of UBSCT
   Student attempted all three sections of the UBSCT three times
   Student meets all graduation requirements
   Districts design Diploma—Must say "Did not pass UBSCT" on Diploma
10. Certificate of Completion
    Student completed 12 years of education
    Student did not pass all three sections of UBSCT
    Student did not attempt all three sections of the UBSCT three times
    Student did or did not meet all graduation requirements
    Districts design certificate—Must say "Certificate of Completion"
11. Diploma for adult education
    Adult education students are not "required" to take the UBSCT
    There is no provision for K-12 students to "upgrade" a diploma through an adult education program

    For this program year (2006-2007), "graduation 2006" K-12 students who attend adult education programs may upgrade their awarded diploma to a "regular" high school-specific diploma by successfully passing the UBSCT and all other graduation requirements. Adult education programs are responsible to facilitate the process by working with the district assessment coordinators to see that "graduating 2006" students are afforded this opportunity. After this program year, upgrading a K-12 diploma will not be an option.

    School districts shall:
    Establish policies allowing or disallowing adult education students from attempting the UBSCT
    Provide for wording on ADULT EDUCATION SECONDARY DIPLOMAS including allowing or disallowing adult education diplomas to state that the student did or did not pass the UBSCT.

Retrieved from
http://www.schools.utah.gov/eval/documents/UBSCT_Clarification_of_Practice.pdf

Appendix E: The No Child Left Behind Act of 2001 Is Working

December 2006

Because of *No Child Left Behind's* accountability provisions, schools and parents are getting the information and help they need to focus attention and resources on the children who need it most—and it's working.

NCLB Benefits Children, Empowers Parents, Supports Teachers and Strengthens Schools.

- All children are counted under NCLB, and schools are responsible for making sure every child is learning.
- Parents are given unprecedented information and new options for their children, which may include free tutoring.
- Teachers utilize assessment data and scientifically based teaching methods to improve classroom instruction.
- Schools identified as in need of improvement receive extra help and resources to raise student achievement.

Multiple studies and reports show that student achievement is rising across America:

- The long-term Nation's Report Card (NAEP) results, released in July 2005, showed elementary school student achievement in reading and math at all-time highs and the achievement gap closing.
    - For America's nine-year-olds in reading, more progress was made in five years than in the previous 28 combined.
    - America's nine-year-olds posted the best scores in reading (since 1971) and math (since 1973) in the history of the report. America's 13-year-olds earned the highest math scores the test ever recorded.
    - Reading and math scores for African American and Hispanic nine-year-olds reached an all-time high.
    - Math scores for African American and Hispanic 13-year-olds reached an all-time high.
    - Achievement gaps in reading and math between white and African American nine-year-olds and between white and Hispanic nine-year-olds are at an all-time low.
- The state-by-state Nation's Report Card results, released in October 2005, showed improved achievement in the earlier grades in which NCLB is focused. In the last two years, the number of fourth-graders who learned their fundamental math skills increased by 235,000—enough to fill 500 elementary schools!
    - Across-the-board improvements were made in mathematics and in fourth-grade reading.
    - African American and Hispanic students posted all-time highs in a number of categories.
    - Forty-three states and the District of Columbia either improved academically or held steady in all categories (fourth- and eighth-grade reading and fourth- and eighth-grade math).
- The Nation's Report Card Trial Urban District Assessments for Reading and Math, released in Dec. 2005, showed students in select urban school districts improving faster than their peers over the last two years.

    - Fourth-graders in 8 of 10 urban districts made larger gains in math than the national average.
    - Fourth-graders in 7 of 10 urban districts made larger gains in reading than the national average.
    - Eighth-graders in 7 of 10 urban districts made more progress in basic math skills than the national average.
- The Nation's Report Card Science 2005 Report found significant academic gains by fourth-graders.
    - Overall, fourth-graders improved four points in science achievement over 1996 and 2000 levels, with the lowest-performing students making the largest gains.
    - African American and Hispanic fourth-graders made significant gains as well, narrowing the achievement gap.
- And the Nation's Report Card Trial Urban District Assessment for Science, released in Nov. 2006, showed narrower achievement gaps for low-income students than for the entire student body, between nearly all of the participating school districts and the nation.

Retrieved January 27, 2007, from
http://www.ed.gov/nclb/overview/importance/nclbworking.html

Appendix F: State-by-State Summary of Overall Achievement Trends

This table is not intended to support comparisons between states. Each state has its own standards, tests, proficiency definitions, and cut scores. Because of differences in testing systems, some states may be more likely to show gains than others.

Legend: PP= Percentage Proficient
↑ = Moderate-to-large gain
↓ = Moderate-to-large decline
↗ = Slight gain
↘ = Slight decline
△ = No change
◊ = Not enough years of data (only 1-2 years) to determine trend
O = Data not available

States and years analyzed	Reading			Mathematics		
	Elem PP	Middle PP	High PP	Elem PP	Middle PP	High PP
Alabama (2004-7)	↑	↑	↘	↑	↑	↑
Alaska (2005-7)	↑	↑	◊	↑	↑	◊
Arizona (2005-7)	↗	↘	↘	↑	△	↘
Arkansas (2002-7)	↑	↑	↑	↑	↑	↑
California (2003-7)	↑	↑	↘	↑	↘	↑
Colorado (2002-7)	↗	↗	↗	↗	↑	△
Connecticut (2002-7)	◊	◊	↗	◊	◊	↗
Delaware (2002-7)	◊	◊	◊	◊	◊	◊
Florida (2002-7)	↑	↗	↘	↑	↑	↑
Georgia (2002-7)	◊	◊	◊	↑	↑	↗
Hawaii (2002-6)	↗	↘	↗	↗	↑	↘
Idaho (2003-6)	↑	↑	↗	↑	↑	↗
Illinois (2006-7)	◊	◊	◊	◊	◊	◊

States and years analyzed	Reading			Mathematics		
	Elem PP	Middle PP	High PP	Elem PP	Middle PP	High PP
Indiana (2002-7)	↗	↑	△	↗	↑	↗
Iowa (2004-7)	↑	↑	↘	↑	↑	△
Kansas (2006-7)	◊	◊	◊	◊	◊	◊
Kentucky (2002-6)	↑	↑	↑	↑	↑	↑
Louisiana (2002-7)	↑	↑	↗	↑	↑	↑
Maine (2003-7)	◊	◊	↓	◊	◊	↑
Maryland (2003-7)	↑	↑	↑	↑	↑	◊
Massachusetts (2002-7)	↗	↑	↑	↑	↑	↑
Michigan (2006-7)	◊	◊	◊	◊	◊	◊
Minnesota (2006-7)	◊	◊	◊	◊	◊	◊
Mississippi (2002-7)	↑	↗	↓	↑	↑	↑
Missouri (2006-7)	◊	◊	◊	◊	◊	◊
Montana (2004-7)	↑	↑	↑	↑	↓	↓
Nebraska (2002-7)	↑	↑	↑	↑	↑	↑
Nevada (2004-7)	↑	↑	↓	↑	↑	↓
New Hampshire (2002-7)	◊	◊	↑	◊	◊	↑
New Jersey (2002-7)	↗	↗	↗	↑	↑	↑
New Mexico (2005-7)	↑	↑	↓	↑	↑	↗
New York (2006-7)	◊	◊	O	◊	◊	O

States and years analyzed	Reading			Mathematics		
	Elem PP	Middle PP	High PP	Elem PP	Middle PP	High PP
North Carolina (2003-7)	↗	↗	O	◊	◊	O
North Dakota (2005-7)	↑	↑	↑	↗	↗	↑
Ohio (2004-7)	↑	↗	↑	↑	↑	↑
Oklahoma (2002-7)	↑	↗	↑	↑	↑	↑
Oregon (2002-7)	↑	↗	↗	↑	↑	△
Pennsylvania (2002-7)	↗	↑	↑	↑	↑	↗
Rhode Island (2004-6)	◊	◊	↗	◊	◊	↘
South Carolina (2002-7)	↑	↘	↗	↑	↗	↗
South Dakota (2005-7)	↗	↓	↓	◊	◊	◊
Tennessee (2004-7)	↑	↑	↑	↑	↑	△
Texas (2005-7)	↑	↑	↑	↑	↑	↑
Utah (2004-7)	↗	↗	↗	↗	↑	↑
Vermont (2006-7)	◊	◊	◊	◊	◊	◊
Virginia (2006-7)	◊	◊	◊	◊	◊	◊
Washington (2002-7)	↑	↑	↑	↑	↑	↑
West Virginia (2004-7)	↑	△	↘	↑	↑	↑
Wisconsin (2006-7)	◊	◊	◊	◊	◊	◊
Wyoming (2006-7)	◊	◊	◊	◊	◊	◊

Appendix G: Adequate Yearly Progress and School Ratings

State Assignment of School Ratings, Percent of Schools not Making Adequate Yearly Progress, and Percent of Schools Identified as In Need of Improvement, by State, 2006

State	State assigns ratings to all schools: Based on adequate yearly progress	State assigns ratings to all schools: Based on additional state-developed criteria [1]	Percent of schools that did not make AYP based on data from 2004-2005 [2]	Percent of schools identified as in need of improvement based on data from 2004-2005 [2,3]
United States	51 [4]	28 [4]	26 [4]	14 [5]
Alabama	Yes	No	47	34
Alaska	Yes	No	41	38
Arizona	Yes	Yes	14	9
Arkansas	Yes	Yes	--	24
California	Yes	Yes	38	19
Colorado	Yes	Yes	25	6
Connecticut	Yes	No	20	16
Delaware	Yes	Yes	26	21
District of Columbia	Yes	No	55	35
Florida	Yes	Yes	64	32
Georgia	Yes	Yes	18	17
Hawaii	Yes	No	66	48
Idaho	Yes	No	43	15
Illinois	Yes	No	27	17
Indiana	Yes	Yes	40	5
Iowa	Yes	No	9	6
Kansas	Yes	Yes	9	1
Kentucky	Yes	Yes [6]	26	11
Louisiana	Yes	Yes	16	13
Maine	Yes	No	23	8
Maryland	Yes	No	25	18
Massachusetts	Yes	Yes [6]	43	24
Michigan	Yes	Yes	12	13
Minnesota	Yes	Yes	13	4
Mississippi	Yes	Yes	11	9
Missouri	Yes	No	35	7
Montana	Yes	No	7	9
Nebraska	Yes	Yes	--	--
Nevada	Yes	No	53	29
New Hampshire	Yes	No	47 [7]	42 [7]
New Jersey	Yes	No	39	25
New Mexico	Yes	Yes	53	30
New York	Yes	Yes	20	18

State	State assigns ratings to all schools: Based on adequate yearly progress	State assigns ratings to all schools: Based on additional state-developed criteria [1]	Percent of schools that did not make AYP based on data from 2004-2005 [2]	Percent of schools identified as in need of improvement based on data from 2004-2005 [2,3]
North Carolina	Yes	Yes	42	9
North Dakota	Yes	No	9	4
Ohio	Yes	Yes	24	13
Oklahoma	Yes	Yes	3	7
Oregon	Yes	Yes	32	26
Pennsylvania	Yes	No	19	13
Rhode Island	Yes	Yes	56 [7]	18 [7]
South Carolina	Yes	Yes	53	15
South Dakota	Yes	No	16	15
Tennessee	Yes	Yes	7	9
Texas	Yes	Yes	13	3
Utah	Yes	No [8]	13	1
Vermont	Yes	No	14 [7]	21 [7]
Virginia	Yes	Yes	19	6
Washington	Yes	No	20	9
West Virginia	Yes	Yes	17	5
Wisconsin	Yes	No	2	2
Wyoming	Yes	No	20	4

-- Not available

1 Column refers to states that use additional information to rate schools beyond that required by federal law for AYP, or that apply a separate rating to schools statewide in addition to AYP ratings. For example, a state may have a method of assigning letter grades to schools based on the schools' performance on statewide tests, or may require that schools meet criteria in addition to meeting AYP in order to receive high ratings.

2 Some percentages are based on preliminary data, pending the outcome of appeals.

3 All states must include Title I schools in their designation of schools "in need of improvement." Federal law allows states to choose whether non-Title I schools are assigned a school improvement status. As a result, some schools rated for AYP may not receive a school improvement designation. Percentage was calculated by dividing the total number of schools identified as in need of improvement by the total number of schools rated for AYP.

4 National total reflects the number of "Yes" responses for the column.

5 National total is based on states for which data were available.

6 Ratings based on state-developed criteria are assigned biennially.

7 New Hampshire, Rhode Island, and Vermont results for the percent of schools not making AYP and identified as "in need of improvement" include high school only.

8 Utah plans to assign ratings to all schools based on additional state-developed criteria in the 2006-2007 school year.

Note. ES=Elementary School; MS=Middle School; HS=High School. Adapted from "Standards and Accountability Table," from Editorial Projects in Education Research Center, *Quality Counts at 10: A Decade of Standards-Based Education*, Washington, DC: Education Week, 2006, retrieved from http://nces.ed.gov/programs/statereform/saa_tab6asp

Appendix H: Consumers and Uses of Standardized Test Information

Consumer	Unit of analysis
National level	
Allocation of resources to programs and priorities	Nation, state
Federal program evaluation	State, program
State legislature/state department of education	
Evaluate state's status and progress relevant to standards	State
State program evaluation	State, program
Allocation of resources	District, school
Public(lay persons, press, school board members, parents)	
Evaluate state's status and progress relevant to standards	District
Diagnose achievement deficits	Individual, school
Develop expectations for future success in school	Individual
School districts – central administrators	
Evaluate districts	District
Evaluate schools	Schools
Evaluate teachers	Classroom
Evaluate curriculum	District
Evaluate instructional programs	Program
Determine areas for revision of curriculum and instruction	District
School districts – building administrators	
Evaluate school	School
Evaluate teacher	Classroom
Group students for instruction	Individual
Placement into special programs	Individual
School districts – teachers	
Group students for instruction	Individual
Evaluate and plan the curriculum	Classroom
Evaluate and plan instruction	Classroom
Evaluate teaching	Classroom
Diagnose achievement deficits	Classroom, individual
Promotion and graduation	Individual
Placement into special programs (e.g., gifted, handicapped)	Individual
Educational laboratories, centers, universities	
Policy analysis	All units
Evaluation studies	All units
Other applied research	All units
Basic research	All units

Note. Adapted from "Raising Standardized Achievement Test Scores and the Origins of Test Score Pollution," by T. M. Haladyna, S. B. Nolen, and N. S. Hass, 1991, *Educational Researcher, 20*(5), pp. 2-7.

Appendix I: Consent to Be a Research Subject

Introduction
This research study is being conducted by Raylene Hadley, a PhD candidate in the Department of Educational Leadership and Foundations at Brigham Young University, in partial fulfillment of a doctoral program. The purpose of this study is to learn of the consequences high-stakes testing is having on teaching and learning in the public schools in Utah from the vantage point of the elementary school principal.

Procedures
As an elementary school principal, you will be asked a series of questions regarding your opinion on the impact high-stakes testing is having on teaching and learning in your school. This will take approximately 50-60 minutes and will be tape-recorded and, later, transcribed.

Risk/Discomforts
There are no anticipated risks associated with participation in this study. It is remotely possible, however, that you might feel some slight discomfort answering a particular question or line of questioning. In such instances, you will have the option of discontinuing the interview at any time.

Benefits
There are no direct benefits to subjects. However, it may be helpful to understand the experiences and viewpoints of other principals, both those that share similar demographics and those who do not. In addition, it is hoped that such information will prove to be of value to the Utah State Board of Education as they continue to plan for and implement policies and programs in alignment with the stipulations as outlined in NCLB.

Confidentiality
Although demographic information will be collected, all raw data linked to individual participants will be kept strictly confidential. This will be accomplished by providing you with a unique identifier. Demographic and descriptive data will be managed carefully to protect your identity. All raw data will be kept in a secure location and will be destroyed at the conclusion of the study.

Participation
Participation in this study is voluntary. You have the right to stop the interview at any time without any negative consequences.

Questions about the Research
If you have any questions regarding this study, you may contact Raylene Hadley at 801-422-2246 or raylene_hadley@byu.edu.

Questions about your Rights as Research Participants
If you have questions regarding your rights as a research participant, you may contact the BYU IRB Administrator at (801) 422-1461, A-285 ASB, Brigham Young University, Provo, UT 84602, irb@byu.edu.

I have read, understood, and received a copy of the above consent and desire of my own free will to participate in this study.

Name:_____ Date:_____

Signature:_____

Appendix J: Interview Guide

Used to measure learning outcomes, influence classroom instruction, and make teachers and schools accountable for what is being taught, testing has become an integral part of pedagogical practices in the public education system in this country. With the passage of No Child Left Behind and the attention placed on Adequate Yearly Progress (AYP), testing and test results have become even more significant.The purpose of this study is to gather information about high-stakes testing and its impact on teaching and learning in the public elementary schools in Utah from the perspective of the principal. Specifically, with regards to this study, high-stakes testing in the elementary school grades refers to the Core Criterion Reference Tests or CRTs, since the CRTs are used to report academic achievement for the purposes of AYP.

I'm glad you've agreed to be interviewed. I want to explain how this will work. I will conduct about a 60-minute interview that will be tape-recorded, transcribed, and edited.

I will ask you a number of questions having to do with the impact testing is having on your school—your teachers, your students, and your administration—from your perspective. I want to understand what you see as significant when considering the impact of high-stakes testing at your school, both positive and negative. I am interested in your candid beliefs, observations, and practices about important issues related to these tests. Your individual responses will be kept strictly confidential and will not be provided to any other person or group. I will not use your profile in any way that you don't personally approve.

Introduction

I would like to ask you a few general questions to start:

1. In general, how do you view CRT's?
2. What are your impressions of AYP?

Teaching

Next, I will ask you a series of questions that are specificially related to teachers and instruction in your school.

In your opinion:

3. What influence, if any, has the CRT had on *teaching strategies* in your school? (Note: Teaching strategies refer to the style or method of teaching or the manner in which the teacher instructs.)
4. What effect, if any, has the emphasis on CRT results had on *teachers' creativity* in the classroom?
5. Are there any sanctions imposed on teachers as a result of poor test scores? If so, what are they?
6. Are there positive rewards or incentives for teachers whose students do well on the CRTs?
7. What are accurate indications of a good teacher?
8. Has teacher morale been affected due to the emphasis on test results?
9. In what ways are your teachers preparing students for the CRT?
10. Are there any areas of the curriculum that have been neglected as a result of the

emphasis on teaching the core subjects that are going to be tested?
11. Do you see a difference between "teaching to the test" and "teaching for the test?"

Learning

The next series of questions has to do with students and the relationship of testing to learning in the classroom.

12. In your opinion, what do improved CRT scores mean?
13. How effective have the results of the CRT been in evaluating student achievement?
14. In what ways do you rely on the results of the CRT in making decisions concerning students?
15. How does testing fit within your ideal vision of student learning?
16. Has testing impacted other programs in your school, such as the resource classes or the gifted classes?
17. Do you have a significant population of students where English is not their native language?
18. Has student morale been affected as a result of the emphasis placed on CRTs?

Administration

The last series of questions are related to the affect of testing and test results, particularly AYP, on the decisions you make as an administrator.

19. Have there been any program and policy changes, either at the district or school level, as a result of the emphasis on making AYP? If so, what are they?
20. What have been the benefits and/or drawbacks of the emphasis on making AYP for you, as principal?
21. Do the CRTs influence the decisions that you make as principal? If so, how?
22. Is it fair to use test results to evaluate teachers?
23. Do you think cut scores are set at an appropriate level?
24. How do you feel about the amount of emphasis placed on AYP?
25. How has the emphasis on test results affected your daily responsibilities as principal?
26. Do you have a lot of parent support?
27. What changes have you made in an effort to pass AYP? (Failing schools only)
28. Do you see any correlation between Title I funding and passing AYP?
29. How have you used Title I resources to better prepare your students for passing AYP? (Title I schools only)

Other

I have just a few more general questions:

30. If you could make any changes to your school to improve student achievement, without regard to AYP or NCLB, what would you do?
31. If you could make any recommendations to change NCLB, as it comes up for reauthorization, what would you recommend?

32. Do you think your experiences are similar to or different from principals at other elementary schools in the district?

Background Information

33. How long have you worked in the district?
34. How long have you worked as an administrator?
35. How long have you worked in public education?
36. How big is your school? How many students do you have?
37. Anything else you'd like to add?