THE INFLUENCE OF HIGH STAKES TESTING ON ELEMENTARY CLASSROOM INSTRUCTION

A Dissertation by

James Finkeldei

Master of Education, Wichita State University, 2002

Master of Education, Wichita State University, 2000

Bachelor of Arts in Education, Wichita State University, 1995

Submitted to the Department of Counseling, Educational Leadership,
Educational and School Psychology
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Doctor of Education

May 2016

THE INFLUENCE OF HIGH STAKES TESTING ON ELEMENTARY CLASSROOM INSTRUCTION

The following faculty members have examined the full copy of this dissertation for form and content, and recommend that it be accepted in partial fulfillment of the requirement for the degree of Doctor of Education with a major in Educational Leadership.

_____
Jean Patterson, Committee Chair

_____
Eric Freeman, Committee Member

_____
Susan Bray, Committee Member

_____
Kristin Sherwood, Committee Member

_____
Rhonda Lewis, Committee Member

Accepted for the College of Education

_____
Shirley Lefever-Davis, Dean

Accepted for the Graduate School

_____
Dennis Livesay, Dean

Dedication

I dedicate this dissertation to my wife Nina and our children,

Elleana, Emma and Elias for their patience and support throughout this journey.

Acknowledgements

First, I would like to thank Mr. Bob Voboril and the Diocese of Wichita for supporting me the last three years. Without this support, I never would have been able to finish this dissertation. Second, I would like to thank the members of cohort 22, Neil, Cathy, Charla, Martyna, Brad, and John for all their support during our journey. I will always remember our time together. Each of you changed me and made me a better educator and person. For that, I will be eternally grateful.

I would also like to thank my committee members, Dr. Freeman, Dr. Sherwood, and Dr. Lewis. Each of you challenged my thinking and guided me along my path. I appreciate all the guidance and expertise you shared with me.

Finally, I would like to thank Dr. Jean Patterson for being my Committee Chair and friend. You were with me every step of the way, inspiring and pushing me to succeed. I want to thank you for all the time spent reading my drafts and all the insightful comments that shaped the final version.

Abstract

Policymakers from both political parties and the general public see standardized tests as easily quantifiable measures of school quality and student learning. They assume these high stakes tests will motivate teachers and students to try harder and that the results will be used systematically to benefit all students. Since schools operate with a finite amount of resources, any time or money they dedicate to passing the test is time and money they cannot spend on educating all students.

The purpose of this study was to investigate how the new high stakes testing environment and the new state assessments have influenced elementary teachers' approach to preparing students for standardized tests. Campbell's law was used to examine whether the utilization of educational and instructional triage was an inevitable result of high stakes testing. Study participants included nineteen experienced third through fifth grade teachers in a suburban district in a small, Midwestern town.

The research indicated that education and instructional triage was occurring, although there was more instructional triage then educational. Teachers are resigned to high stakes testing, therefore assessments drive their instruction, and they define student success by their score on the test. High stakes testing has led to increased accountability and data driven instruction, however low stakes testing also creates these benefits without the undue pressure. Future research is needed in light of the passage of Every Student Succeeds Act (ESSA).

TABLE OF CONTENTS

Chapter **Page**

TABLE OF CONTENTS (continued)

Chapter **Page**

LIST OF TABLES

CHAPTER 1

If you are over the age of 30 and grew up anywhere in the Midwest, you probably remember sitting down with a Number 2 pencil to fill in oval bubbles while taking the Iowa Test of Basic Skills (ITBS). The ITBS' *Interpretive Guide for School Administrators* explains that test results were a supplement to, not a replacement for, other information about a student's performance. The guide further warned that a score from a single test should not be used to assign students to special education, hold students back, screen students for first-time enrollment, evaluate the effectiveness of entire educational systems, or identify the best teachers or schools (Hoover et al., 2003). Most state departments of education, educators, school psychologists and professional evaluation organizations would agree with this assertion that no one test can ascertain whether a student is learning or not and that multiple measures are essential to a well-rounded view of a student's learning (Lucido, 2010).

Even though most educators and ITBS' leaders still believe one test is not the best indicator of student learning, students today have a very different experience. A third grader sits down in front of a computer screen to take the new common core assessments and he or she may be nervous because there is significant pressure to do well (Watson, Johanson, Loder, & Dankiw, 2014). Outgoing Secretary of Education Arne Duncan said tests are becoming more important because, "We should be able to look every third grader in the eye and say, 'You're on track, you're going to be able to go to a good college, or you're not'" (Hernandez, 2009). This type of expectation is creating anxiety among students and parents.

Similarly, there is growing anxiety among school personnel. Administrators know their building's test scores are going to be ranked from highest to lowest performance, published in the newspapers, and broadcast by the media. Teachers know that their performance review, pay

raises, and possibly their jobs may depend on what that third grader does on that one day in front of that one computer (Kohn, 2000). However, many state legislators, congressional representatives, businesspersons, and journalists believe this is a healthy anxiety. They argue that teachers and students will try harder and do better if there is a consequence involved (Koretz, 2008). They believe having an easily quantifiable number allows for better accountability and results-based distribution of dwindling funds. Conservatives tend to view testing as creating accountability over a failing system, while liberals are prone to seeing high stakes testing as a way to make school districts accountable for the success of desegregation and minority students (Mehta, 2008). As a result, high-stakes testing has become entrenched in the fabric of our culture. Children and adults alike are tested so often and for so many reasons, testing is ingrained and citizens do not question its legitimacy (Moses & Nanna, 2007).

While the debate over whether one test can accurately predict student success and teacher effectiveness has historical roots that go back to the beginning of the 20th century, the current iteration began with the reauthorization of the Elementary and Secondary Education Act in 2002, commonly referred to as No Child Left Behind (NCLB). The centerpiece of this legislation was the requirement of yearly testing in grades 3 through 8 and once during high school (U.S. Department of Education, 2002). Accordingly, schools and districts had to meet Adequate Yearly Progress (AYP) goals that determined whether schools must take steps to raise achievement and, if low scores persisted, whether the school would face sanctions (No Child Left Behind, 2002). This brought about a massive change in state testing systems and vastly increased the consequences of high stakes testing, up to and including the possibility that teachers and administrators would be replaced (Chudowsky, Chudowsky, & Kober, 2007).

Publicizing assessment results and comparing schools based on them has increased public scrutiny and intensified the pressure for principals and teachers to perform well.

Since the passage of NCLB, the pressure put on the assessment results has only intensified. In 2010, when it became obvious that the nation's schools would never be able to reach NCLB's goal of 100% proficiency by the 2014 deadline, President Obama did not do away with the high stakes tests as many hoped. Rather, he awarded Race to the Top (RTTT) grants to states that met certain mandates, and NCLB waivers to those who did not. Both RTTT and the NCLB waivers required states to adopt more rigorous college and career standards and to tie teacher evaluations to assessment results (Guisbond, Neill, & Schaeffer, 2012). The new college and career standards have increased anxiety for students because they will now be assessed utilizing tests that measure higher-level thinking and problem solving that make them more difficult than the current tests (Forgione, 2012). The new teacher evaluations have increased anxiety for teachers because the scores their students earn on the assessments will now determine a significant percentage of their performance rating. This, in turn, is leading schools and teachers to require mandatory remediation, summer school, or even retention for students who do not pass, which is likely to further increase the pressure on students to perform to the standard. All these factors are increasing the anxiety surrounding high stakes assessments. In this study, I examined teachers' perceptions about how their classroom instruction has been altered by the new high stakes testing environment.

Research Problem

Policymakers and the public in general see standardized tests as easily quantifiable measures of school quality and student learning. They hope that these high stakes tests will motivate teachers and students to try harder and that the results will be used systematically to

benefit all students (Kearns, 2011). Understanding that high stakes testing is part of a larger shift towards neoliberal economic and social policies clarifies why, regardless of whether Republicans to Democrats have the majority in Washington DC, the high stakes assessment environment continues unabated (Hursh, 2007).

Neoliberals assert that free market competition and privatization is the best way to promote economic growth (Apple, 2011). It capitalizes on the American belief in rugged individualism and promotes that every person's success and failure is a result of individual choice. Society, and therefore government, cannot create or solve problems. The free market, guided by quantitative information and shaped by competition, will yield the best results (Hursh, 2007). The prevalence of this thinking has led most policymakers to act as if all children come to school equally ready to learn and that a student's test score is a result of the hard work and effort of individual teachers and students.

However, districts and teachers are utilizing educational and instructional triage to increase test scores without necessarily increasing student learning (Jennings & Sohn, 2014). Instructional triage occurs when teachers emphasize test-taking skills, narrow the curriculum to maximize time spent on tested indicators, and focus on matching their teaching to the format and type of questions most likely to be found on the assessment (Ambrose, 2012). Educational triage is using assessment results to target interventions on "bubble students" at the expense of low and high achieving students (Nichols & Berliner, 2007). Bubble students are those who are close to attaining the next achievement level. Even though educational and instructional triage have been utilized in schools since testing began, the more high stakes assessments become the more these strategies will be utilized (Koretz, 2011). It is now possible to find advice in major education

publications that unabashedly offer teachers researched based strategies on how to best "teach to the test" (DeFauw, 2013).

Since schools operate with a finite amount of resources, any time or money they dedicate to passing the test is time and money they cannot spend on educating all students (Koretz, 2008). Those who believe in increased accountability will argue that high stakes assessments improve results for all students. However, one of the unintended consequences of these assessments is that schools focus on educational and instructional triage at the expense of high and low achieving students. High-stakes testing has forced schools and teachers to focus on students they deem most likely to respond to educational interventions with the least amount of time, effort, and expense (Booher-Jennings, 2005). They get more "bang for their buck" by focusing their resources on this relatively small number of students who will have maximum impact on their overall accountability rating (Jennings & Sohn, 2014). As a result, districts, schools, and teachers are not using assessment data to help all students learn; rather they are using it to fulfill their accountability requirements.

The ever-increasing high-stakes testing environment that districts are forced to work within contributes to the discrepancy between the intent of high stakes testing and the lack of subsequent results. The more emphasis put on only one measure to identify and reward so-called quality schools and teachers the more that indicator will corrupt the very institution that it is trying to measure (Campbell, 1975). In this study, I examined teachers' perceptions about how their classroom instruction has been altered by the new high stakes testing environment.

Theoretical Framework

The purpose of a theoretical framework in research includes not only setting the boundaries for a study (Merriam, 2009) but also guiding the development of research questions

that might explain or suggest a relationship between concepts or ideas (Maxwell, 2012). The theoretical framework I chose to guide this study is Campbell's Law (Campbell, 1975), which suggests that the more a quantitative social indicator is used for social decision making, the more apt it will be to distort and corrupt the very social processes it is intended to monitor. In recent years, Nichols and Berliner (2007) and Booher-Jennings (2005) have used Campbell's law to explain why high stakes testing is impacting classroom instruction.

Campbell's Law

In his book *Experimental and Quasi-Experimental Designs for Research (*Campbell, *Stanley, & Gage, 1963),* Donald Campbell, a social scientist, proposed a validity typology that is still a standard for evaluating public policy (Shadish & Luellen, 2004). In 1975, while serving as president of the American Psychological Association, Campbell summarized his views on federally funded program evaluation (Campbell, 1975). Among the topics he covered was the corrupting effect of quantitative indicators on social programs. He claimed, "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor." (p.34). His contemporaries began referring to this assertion as Campbell's Law. He gave an example of how census data, which at the time had little political significance, was considered trustworthy while voting statistics, which had great political and economic significance, were not. This is because there was much more incentive to corrupt the voting statistics because of the amount of money involved. The more voting statistics were used to make decisions, the less accurate they become. He expounded upon this assertion by citing research about President Nixon's war on crime and his harsh demands that police decrease the crime rate. Campbell asserted that police simply changed how they cleared and labeled cases

rather than actually decreasing crime rates. Similarly, he cited research from Blau (1963) about how the practice of evaluating U.S. government employees on how many public complaints they dealt with did not increase program effectiveness. Rather, it led to employees speeding through their cases and concentrating on the cases where citizens had the least number of needs while ignoring the cases where citizens had multiple needs. Campbell contended that one of underlying causes of the infamous 1968 My Lai massacre during the Vietnam War was the fact that field commanders were being rated based on one quantitative number - body count- rather than the more ambiguous, "meeting mission objectives."

Since it was first introduced almost 40 years ago, Campbell's law has been applied to many social indicators. Stiglitz, Sen, and Fitoussi (2010) used it to warn that improper measurement will lead to inappropriate economic policies. Koretz (2011) found that 83% of cardiologists reported that the publication of mortality rates stopped them from operating on patients who might benefit from heart surgery but had a low chance of survival. In business, Sears had to abolish a plan to pay mechanics based on the number of repairs they made rather than a salary when officials discovered mechanics were making unnecessary repairs (G. Baker, Gibbons, & Murphy, 1994).

Campbell (1975) concluded by saying the most clear-cut example comes from education. He cited Stake (1971) who discovered that educator contractors in Texas, whose pay was based upon student achievement on test scores, were teaching to the test and cheating. He concluded by saying,

> From my own point of view, achievement tests may well be valuable indicators of general school achievement under conditions of normal teaching aimed at general competence. But when test scores become the goal of the teaching process, they both

> lose their value as indicators of educational status and distort the educational process in undesirable ways. (p. 35)

In their book, Nichols and Berliner (2007) applied Campbell's law to the new era of high stakes testing and outlined how it had distorted the educational process. According to their findings, the pressures of high stakes testing have led to student cheating, teacher cheating, principal cheating, and whole scale district cheating. In a 2003 national survey, 1.5% of teachers admitted to changing a student's answer, 10% to providing hints about the answers during tests, and 15% to providing more time for the test than was allowed (p. 36). More recently, award-winning gains by Atlanta schools were found to be the result of teachers and principals cheating (Jonsson, 2011).

Less obvious is schools exclusionary practices, schools that allow low achieving students to drop out or transfer out just before testing. In addition, administrators often systematically exclude immigrant and special education students from even taking the test. As Apple (2006) put it, the emphasis has switched from what the school does for the student to what the student does for the school. If a student is not going to be helpful to the school's test scores, they will often be excluded from the testing.

Nichols and Berliner (2007) compiled evidence that as schools and teachers start to narrow the curriculum, teach to the test, and focus on bubble students, the tests lose validity. "No longer are we measuring real world math or reading skills. Instead, it becomes a test of how well students memorized the math content or how adept students are at filling in test booklet bubbles" (Nicolas and Berliner, 2007, p. 122). Campbell's law would predict that these actions are a logical result of the effect of high stakes testing. Booher-Jennings (2005) found that teachers justified engaging in these practices for two main reasons. First, since the school,

district, and public have narrowed the definition of good teaching to good test scores, teachers used that same institutional logic to justify the use of these educational triage practices. Second, teachers used the rationale of data-driven instruction to justify teaching to the test and focusing on bubble students.

Campbell's law theorizes that educational and instructional triage are inevitable and logical strategies for districts to utilize to meet the ever-increasing requirements of the high-stakes testing environment. The question is whether these changes are improving test scores and student learning, just improving test scores, or neither. As one Florida Superintendent put it, "When a low performing child walks into a classroom, instead of being seen as a challenge, or an opportunity, for the first time since I've been in education, teachers are seeing [that child] as a liability" (Nicolas and Berliner, 2007, p.58). Campbell's Law was suitable as a guiding theoretical framework for analyzing the impact of high stakes testing on classroom instruction.

Research Questions

Students are now accountable for passing increasingly difficult assessments. At the same time, schools are not only being judged on how many students pass, but how many student scores exceed standards or are college and career ready. In addition, up to 40% of a teacher's professional evaluation is now based on how well students do on the assessments (Kansas State Department of Education, 2015b). The new Common Core assessments, increasing media coverage on test scores, and the RTTT's requirement that teacher evaluations be at least partially based on test scores has created a new testing environment for schools and teachers. The purpose of this study was to investigate how the new high stakes testing environment has influenced teachers' classroom instruction.

This study sought to answer the following research questions:

1. What professional guidance do 3rd- 5th grade teachers receive for improving student outcomes on the new state assessments?
2. What techniques do 3rd-5th grade teachers employ in their classrooms to improve student outcomes on the new state assessments?
3. What do 3rd-5th grade teachers see as the effects of the new state assessments on classroom instruction and student learning?

CHAPTER 2

Literature Review

Any study requires a comprehensive understanding of the current research problem (Creswell, 2013). To understand the effects of high stakes testing on classroom instruction it is first necessary to trace the history of high stakes testing and how testing became ingrained into our educational system. Second, I discuss the reasons schools have moved from using assessment for instructional purposes to using assessment for accountability and how neoliberalism has influenced this transition. Third, I review the literature on educational and instructional triage to situate this study within the current research (Marshall & Rossman, 2010). Finally, I examine the research on how high stakes testing is impacting overall student achievement which situates this study within the broader context (Merriam, 2009).

History of High Stakes Testing

A standardized assessment is traditionally defined as high stakes when its results are used to influence policies or procedures that effect students, teachers, schools or districts (Madaus, 1988). In the United States, standardized tests were first introduced in the early part of the 20th century but were used sparingly to evaluate specific children so they were not considered high stakes (Leman, 1999). This began to change with the publication of *The Principles of Scientific Management* by Fredrick Taylor (1914), which encouraged all businesses to become more scientific and efficient. School districts adopted this model of efficiency and began to look for ways to measure their success by determining averages of their performance. In education, Joseph Mayer Rice, the publisher of educational reform magazine *The Forum*, published an influential article that proclaimed test scores should be used to evaluate teachers and rate principals and superintendents (Rice, 1913). Soon educators wanted a way to measure individual

performance, particularly among struggling students (A. M. Ryan & Stoskopf, 2008). Educators reasoned that if they could use standardized test to identify students learning difficulties they could more efficiently operate their schools.

French psychologist Alfred Binet developed the first individual intelligence test to determine if some students had what we refer to today as a learning disability. He thought the tests could rectify the learning problems experienced by a number of students. In 1916, Stanford University psychologist Lewis Terman used Binet's work to simplify the process even more and create a single test, which yielded what he called an intelligence quotient (IQ). This streamlined test was successfully used by the U.S. Army to sort recruits during World War I into different jobs based on their IQ (Cherry, 2014). The Army ignored the fact that all immigrants showed shockingly low intelligence compared to established families from the Northeast (Rothman, 1995). The fact that the test results favored soldiers from the Northeast over immigrants led Carl Brigham, the original author of the Scholastic Aptitude Test (SAT), to conclude, "The test scores very definitely are a composite including schooling, family background, familiarity with English, and everything else, relevant and irrelevant. The natural intelligence hypothesis is dead" (Leman, 1999, p. 34). Unfortunately, it was not. The use of one number to place recruits won out and by World War II IQ testing of recruits increased. Once again educators like Davis and Havighurst (1948) spoke out and noted that these tests had a strong cultural bias against poor, immigrant children and were being wrongly used to justify the superiority of the white upper-class. But once again opinions of educators were drowned out by those looking for easy answers to complex questions (Leman, 1999).

During the 1930s and 1940s schools followed the lead of the military and used the IQ test to sort students into programs that best met their academic needs but here again immigrants were

routinely placed in lower classes then students from the Northeast (Labaree, 2012). E.L. Thorndike pushed to apply the rigorous techniques of hard sciences to psychology and education by measuring the outcomes of learning. He was quoted as saying, "whatever exists at all, exists in some amount" (Garrison, 2009). This belief that learning was quantifiable by a number won out over reformers like John Dewey who believed that education, like the amount of love or humanity in a person, was more complex than a single number (Labaree, 2012).

By the time of the Korean War in the late 1940s, both the US military and schools were using standardized tests created by Henry Chauncey and the Educational Testing Service (ETS) to sort students and recruits based on academic ability. Chauncey was troubled by the fact that on his unbiased military test only 42% of Southerners showed proficiency compared to 73% of New Englanders but decided to withhold that information from the public (Leman, 1999). This was because the idea that tests like the SAT could accurately rank and sort students regardless of their background was already deeply ingrained in the public view. All the while, in New York, Stanley Kaplan was disproving this public opinion by tutoring immigrant children on the supposedly un-coachable SAT and dramatically increasing their scores (Leman, 1999). However, the SAT continued to dominate the field of educational testing as a means of determining which students were best suited for college.

While these early tests were high stakes for students trying to get into college and could have profound impact on an individual student's future success, these assessments had limited impact on teachers, schools, or districts. These assessments were used for placement, not accountability. The release of the Coleman Report in 1966 changed that. This long-term study on class size and student performance focused attention on using an individual student assessment as a measure of accountability (Coleman et al., 1966). In 1969 the federal

government created the National Assessment of Educational Progress (NAEP) assessment to do just that (Leistyna, 2007). This also grew into the creation of minimum competency tests in the early 1970s. Students were required to demonstrate a minimal level of knowledge but since the vast majority of students easily passed these low-level tests, neither teachers nor students considered them high stakes.

Most of what are considered high stakes tests today began in the 1980s with the advent of the standards movement (Gunzenhauser, 2003) which accompanied the philosophy that individualistic, market-driven system would produce the best results (Parkison, 2009). In the broadest terms, the standards movement had its roots in neoliberalism, which gained prominence with the election of Ronald Regan in 1980. Neoliberals believe that measurement and accountability will motivate people and business to do better. If a school's performance can be quantified then its level of success, or failure, can be measured. If its level of success or failure can be measured then success can be rewarded and failure can be addressed.

With the publication of *A Nation at Risk* (1983), neoliberalism was translated to education when the commission began to rank schools according to test data and make funding decision based on those results. Despite its many critics who pointed out the tests were meant to evaluate individual students and not schools (Bracey, 1991; Guthrie & Springer, 2004), *A Nation at Risk* popularized the notion that public schools were failing. This perception prompted politicians and educators to launch a multitude of state and federal reforms, demanding standardized tests to measure students, rank schools, and evaluate teachers so consumers could make the best decisions (L. M. McNeil, 2000).

Throughout the 1980s and 1990s neoliberalism continued to entrench itself in educational testing. Beginning with the presidential election of 1982, every presidential candidate and the

vast majority of gubernatorial candidates have supported high stakes testing as an accountability tool for what they considered to be failing public schools (Orfield & Kornhaber, 2001). Throughout the presidency of George H. Bush (1988-1992), the Business Roundtable, the US Chamber of Commerce, the American Legislative Exchange Council, and politicians pushed for mandatory testing as the primary means of improving education (Tienken, 2012). Neoliberalism in education thus became so entrenched that it superseded political party lines. In 1996, President Clinton called the National Education Summit to ensure every state had rigorous state standards (Leistyna, 2007). He also passed the Improving America's Schools Act and Goals 2000 Educate America Act, which both pushed for higher standards monitored by one assessment (Groen, 2012). By 2000 the bipartisan goal of testing had been achieved; every state except Iowa had some sort of mandatory test (Au, 2009).

Throughout this time, the state most active in high stakes testing was Texas. When its former Governor, George W. Bush, was elected President in 2000, one of the cornerstones of his first term was the No Child Left Behind Act (NCLB), which received enormous bipartisan support. The centerpiece of NCLB was that by 2006 all children would be assessed once in grades three through eight and once in high school in both reading and math and every school would make Adequate Yearly Progress (AYP) so that by 2014 all students would score proficient or above (U.S. Department of Education, 2002). In addition, states were required to assess students in science once in elementary, once in middle school and once in high school, although there were no AYP requirements.

States struggled to implement this testing and struggled even more in meeting AYP, especially because of the requirement that each subgroup must meet AYP as well. As Rose (2004) demonstrated in Indiana, if NCLB were enforced as written in the first year, 269 of the

state's 293 schools districts and 68% of the schools would have failed to achieve AYP. Over the next five years, the U.S. Department of Education continued to approve changes to make it easier for schools and districts to demonstrate AYP (Chudowsky & Chudowsky, 2007). Despite these changes states, districts, schools, teachers, and students entered a new era of high-stakes testing.

The pressure to meet AYP became so onerous that some schools began to reallocate their instructional time to focus on the tested content areas at the expense of the non-tested subjects (Rentner et al., 2006). Children who did not pass were filled with shame and further marginalized (Kearns, 2011). Teachers began to change their classroom instruction to focus on memorized skills most likely to appear on the test, rather than higher-level thinking skills (L. M. McNeil, 2000). Teachers became frustrated with all the testing requirements (Jones & Egley, 2004). But most disappointing was the fact that NCLB had not increased test scores as expected (Chudowsky et al., 2007).

Policymakers examined the negative aspects of high stakes testing and determined the problem was not with high stakes testing but with the fact the standards and assessments were not rigorous enough (Guisbond et al., 2012). As a result, in the late 1990s, the National Governors Association (NGA) and the Council of Chief State School Officers (CCSSO) began work developing the new Common Core State Standards that 44 states have since adopted. These standards expanded the content areas to be tested and focused on higher level thinking skills in an attempt to address some of the common complaints against increased testing.

In 2009, President Obama announced Race to the Top (RTTT) grants as part of the American Recovery and Reinvestment Act, which required assessment results to be a significant part of all teacher evaluations (American Recovery and Reinvestment Act of 2009, 2009). In 2010, the U.S. Department of Education awarded two multi-state consortia, Smarter Balanced

Assessment Consortium (SBAC) and Partnership for Assessment of Readiness for College and Careers (PARCC), with funding to create so-called next generation assessments aligned with the Common Core. These assessments were expected to address many of the shortcomings of the NCLB high-stake assessments, particularly in the area of higher-level thinking, and usher in a new era of more successful high stakes testing (Herman & Linn, 2013).

Despite these improvements, in 2013, the Kansas legislature voted to withdraw from the national consortiums and create its own common core assessments because it wanted more local control of the test content (Llopis-Jepsen, 2013a). However, the Center for Educational Testing and Evaluation (CETE) out of the University of Kansas assured the legislature that these tests would be similar to the national tests in that they would be rigorous, align with common core, and promote higher-level thinking (Llopis-Jepsen, 2013b). The CETE piloted individual questions in 2014 and piloted the complete test in 2015. The test given in 2016 will establish a baseline for subsequent administrations. So far, the legislature has been pleased with the rigor of the test and its ability to rank and sort students and schools.

Assessment for Instruction versus Assessment for Accountability

Most Americans presume education creates a meritocracy where ability is the determining factor and the best and brightest children will inevitably rise to the top and get the best grades (Garrison, 2009). As Labaree (2012) described in his book, there is an inherent contradiction in education between democratic and capitalistic ideals. According to the Democratic ideals, every child can learn and education is the great equalizer that allows every child the opportunity to join the meritocracy. Capitalistic ideals, however, allow for winners and losers, rich and poor, business owners and workers; therefore, not everyone can join the meritocracy. If democratic ideals prevail, then testing is used to demonstrate proficiency and

help students improve. If capitalistic ideals prevail, then testing is a way to sort students and begin to identify who will be a producer and who will be a consumer.

The debate over whether assessments should be used to measure proficiency or sort has raged throughout the history of standardized testing. Currently, the neoliberal view has become so entrenched that states and politicians argue over which standards are more rigorous and which tests are more accurate. However, they do not even analyze two underlying assumptions. First, all students have an equal chance to succeed on the test if they, and their teachers, just try hard enough (Kearns, 2011). Second, increasing the consequences for not passing will naturally result in improved outcomes (Tienken, 2012). Such notions notwithstanding, many educators challenge the paradigm that a single test score should drive all decisions (Parkison, 2009). It is my contention that unless more educators challenge these underlying assumptions instead of focusing on the merits and demerits of the new Common Core assessments vis-a-vis the previous State Assessments, they will never address, much less resolve, one of the major difficulties in education – the growth of high stakes testing as the sole determination of success.

NCLB and RTTT use a single test score to determine student proficiency, rank schools, and evaluate the effectiveness of teachers. Politicians and the media prefer to simplify academic success by reducing it to a single number even though statistical analysis has repeatedly demonstrated that a single test score is not a reliable method for measuring student academic performance (Mason, 2007). This is equally true if a single test score were to be used to evaluate the effectiveness of a teacher (Haertel, 2013). Since using single test scores is not a reliable way to measure effective teachers, it is possible these new high stakes assessments will drive out more effective teachers who choose to work with difficult learners while less effective teachers who work higher performing children will be rewarded (E. L. Baker et al., 2010). In fact, Nye,

Konstantopoulos, and Hedges (2004) noted that teachers account for only 13% of the variance in students' scores in math and 7% in reading. Therefore, removing low performing teachers will have little impact on scores overall. In general, if a test is designed to measure student performance then it cannot be reliably used to measure teacher performance and vice versa (Casey, 2013).

The rhetorical claim that all students are equally ready to learn if teachers and schools would just work harder is difficult to substantiate. Rothstein (2004) argued it is the effects of poverty, not effort, which determines academic success. This is because students from low-income households suffer from health issues such as poor vision, poor nutrition, more exposure to smoke, and less adequate health care. All of these have been found to impact academic achievement. They also tended to have fewer literary and cultural experiences and fewer positive role and peer models. This lack of life experiences inhibits the creation of background information, a fundamental component to learning. Finally, Rothstein asserted these students tended to be more mobile and have less consistency of instruction. Both of these have significant impact on learning and have nothing to do with effort.

If effort was the determining factor then one would expect test scores in low-income schools and high-income schools to be similar, unless it is assumed all low-income students are lazy. Koretz (2008) analyzed test scores from both low and high-income schools and found this was not the case. The income level of the students had a substantial impact on test scores. Even where two schools are equally funded, lower income students within those schools still tend to do less well academically (M. Baker & Johnston, 2010). Similarly, if one compares large schools to small schools or public schools to charter schools, low-income students still underperform (Chamberlin, 2007). Likewise, even in schools with high standards and increased

accountability for teachers, students' economic status was the number one predictor of success on standardized tests (Borg, Plumlee, & Stranahan, 2007). In fact, Kohn (2000) contends if any school was selected and all the students were asked, "How much money does your mom make?" the results would be almost identical to the school's standardized test results. No matter what variable is controlled for, income level always has a significant impact on the results and that would not consistently occur if effort were the determining factor.

Despite this evidence, the vast majority of politicians and American citizens believe in high stakes tests and are convinced that accountability will improve effort and results (Moses & Nanna, 2007). One of the reasons adduced for this is that these ideas developed slowly over time and never involved a watershed event or public debate, thus many people believe that failure is the result of lazy students, ineffective teachers, poor leadership, or a combination of the three (Leman, 1999). The public tends to believe if the standards were more rigorous or the tests more accurate, schools would perform better. They believe a singular number conveys objectivity and certitude and it feeds into the national obsession with ranking everything from top 100 cities to top 10 songs of the week (Lucido, 2010). The result is a testing culture that is self-fulfilling. With an occasional individual exception, testing continues to sort students into the same economic categories as their parents and reinforces an economic disparity comforting in its predictability, but one that falls short in achieving increased learning for all.

There are many ways standardized tests scores are artificially inflated and most are ignored by the public, politicians, and even the testing profession (Koretz, 2008). The following sections develop more the problems of test score manipulation, particularly instructional and educational triage, and their subsequent lack of impact on student achievement.

Instructional Triage

Instructional triage occurs when teachers emphasize test-taking skills, narrow the curriculum to maximize time spent on tested indicators, and focus on matching their teaching to the format and type of questions most likely to be found on the assessment (Ambrose, 2012). The current testing environment has led many teachers to implement some sort of test preparation: pep rallies to encourage effort, practice tests, or test taking strategies (Rothman, 1995). There are two main ways teachers might practice instructional triage (Booher-Jennings, 2006). First, they might narrow their curriculum to tested subject areas, namely math, reading, and science. Secondly, they narrow what they teach within those subject areas to the specific indicators most likely to appear on the test. While there is nothing intrinsically immoral about trying to get students to do their best on the test, it limits time for other pursuits and influences the results of the test. That is because standardized tests are incomplete measures; they are a representative sample of the body of knowledge being tested that is supposed to reflect the whole. If a teacher teaches primarily to the tested standards, no matter how innocently, it invalidates the test by making it no longer a true reflection of the whole (Koretz, 2008). Teaching to the test is like a doctor giving an aspirin to all their patients prior to taking their temperature. Their score might be lower, but they are no less sick (Kohn, 2000)?

Narrowing the curriculum. One consequence of instructional triage is narrowing the curriculum to focus on tested subjects. In a study of 500 school districts, Berliner (2011) found, as a result of NCLB, 80% of elementary schools increased reading instruction by 75 minutes a week and 63% of schools increased math instruction by 75 minutes a week. He found these minutes were often taken from social studies and science. Similarly, a survey of elementary teachers revealed that on average they spent 75% of their instructional time on reading and math,

leaving little time for other subjects (Cawelti, 2006). Even when districts and schools did not mandate a change in the number of minutes, 61% of middle school teachers felt that high stakes testing had narrowed the focus of the school to tested subjects (Musoleno & White, 2010).

These instructional choices do have an impact on overall student learning. Maltese and Hochbein (2012) found schools that demonstrated improvements in reading and math scores almost always demonstrated losses in science. Groen (2012) found that many schools showing improvement in reading and math scores had reduced the amount of time spent on art and music.

Narrowing the curriculum within subjects. In addition to narrowing the amount of time spend on each subject; high stakes testing has caused schools and teachers to narrow their focus within those subjects. In a national survey of schools who had failed to meet standard on their state assessment, 84% of elementary teachers identified highly assessed standards and focused a significant amount of their instructional time on that narrowly defined list of standards (Jennings & Sohn, 2014). In another survey, teachers reported feeling they had to narrow the curriculum to the testable indicators because they were under extreme pressure to have their students pass the test (Guilfoyle, 2006). Consequently, elementary teachers indicated their primary goal became to increase test scores, not learning (Hout, 2012). At the high school level, Journell (2010) found that civics teachers were hesitant to spend class time on the 2008 presidential election because it would not be on the mandatory civics test. Likewise, science teachers indicated they would eliminate important elements of their class if they were not on the state mandated test (Aydeniz & Southerland, 2012).

Finally, there is evidence that due to the pressures of high stakes tests teachers at all levels are spending an inordinate amount of time reviewing for the test (Mora, 2011). In a study of classes where more than 60% of students were minority, 75% of the teachers reported

teaching test taking skills and test prep almost exclusively for a month prior to the test (Rothman, 1995). In a similar study of high poverty schools, Reback, Rockoff, and Schwartz (2011) found that 100% of teachers took time to have their students practice questions with particular styles or formats similar to those on the assessment. All these instructional changes reduce higher level thinking skills that demonstrates real learning and replace it with test taking skills that help students, teachers, and schools avoid accountability sanctions.

Educational Triage

Educational triage occurs when schools and teachers allocate scarce resources mostly to students who are closest to advancing an accountability level at the expense of low achieving students who have no hope of passing and high achieving students who are guaranteed to score exemplary (Jennings & Sohn, 2014). For example, in the state of Kansas, new assessments use four proficiency ratings to categorize students: exemplary, exceed standards, meets standards, and below standards (Kansas State Department of Education, 2015a). Since each category has a range of results, for instance 60% to 80% denotes "meets standards," there can be a wide range of difference between two students who are rated as "meets standards." More troubling is the fact that teachers now have the incentive to focus more on the student who scores a 78% than one who scores a 65% (Koretz, 2011). The students who are on the verge of moving up a performance level are often referred to as "bubble students." While schools and teachers are focusing on bubble students, high achieving and very low achieving students receive less attention (Lucido, 2010). It is a phenomenon researchers in England have called "hugging the middle" and has been prevalent ever since high stakes testing was introduced by Margaret Thatcher in the 1980s (Cuban, 2007). Later in the United States, Gillborn and Youdell (1999)

coined the term “educational triage” to describe situations where teachers selectively chose to help the students who had the best chance of “surviving” the assessment.

In America, Booher-Jennings (2005) adopted the term education triage when she began documenting Texas teachers’ efforts to raise test scores in that state’s accountability milieu. As NCLB’s high stakes test spread from Texas, Booher-Jennings (2006) found this phenomenon being replicated in classrooms across the country. Likewise, White and Rosenbaum (2013) provided more examples of how teachers focused on borderline students in an effort to help their schools meet the assessments requirements during the transition to NCLB.

Despite this evidence, several studies have challenged the idea that NCLB negatively affects low and high achieving students. If educational triage was occurring, quantitative researchers expected to find students in the middle of the distribution scale achieving higher gains than the students at either end (Springer, 2008). Several quantitative researchers have demonstrated this is not occurring in a variety of school districts around the nations (Ladd & Lauen, 2010; Reback et al., 2011; Springer, 2008). Other quantitative researchers found the opposite. In several districts, students in the middle were achieving more significant gains than high and low achieving students (Krieg, 2008). Analyzing test scores from the early 2000s when the Chicago public schools switched to high stakes assessments because of NCLB, Neal and Schanzenbach (2010) found that students who had little chance of passing the new test demonstrated significantly fewer gains than the average and bubble students outgained expected growth. This indicated that teachers were focusing on the bubble students.

More recently, researchers have tried to determine why quantitative studies have not produced a more clear-cut answer. In their meta-analysis, Lauen and Gaddis (2012a) determined that inconsistency in the quantitative studies came from the rigor of the standards. In

quantitative studies conducted in high achieving districts or districts in states in low expectations, there was little indication of educational triage. However, in districts which were "on watch" because of low achievement or in states with hard to achieve proficiency standards, there was more quantitative evidence of educational triage. They concluded by theorizing that the higher the standards the more educational triage existed. Therefore, they believe the new rigorous college and career standards will increase educational triage.

Jennings and Sohn (2014) took a different approach and analyzed the contradictory quantitative research in terms of methodology. They determined that quantitative studies showed varying degrees of educational triage because they were examining entire districts, crunching huge amounts of data that can have various statistical results. In addition, they argued that quantitative studies are not the ideal way to identify educational triage because quantitative studies can only determine if teachers were successful in trying to aid bubble students. Qualitative studies can identify if teachers try to focus on bubble students, but fail. Finally, they pointed out that all qualitative studies conducted to date, which asks teachers in struggling districts how they attempt to improve test scores have determined educational triage to be one of their strategies.

In summary, despite the fact that some quantitative studies have demonstrated no significant statistical difference between bubble students and the average, many other quantitative, and most qualitative, studies have demonstrated that teachers make a concerted effort to effect change with bubble students. Only through further qualitative studies can the amount of educational and instructional triage be determined and to what extend that triage is impacting student achievement, because high-stakes testing, as well as, educational and instructional triage could be acceptable if they resulted in increased student achievement for all.

However, the next section outlines how high stakes testing has no significant impact on overall student achievement.

Effects High Stakes Testing on Student Achievement

The preponderance of qualitative and quantitative research since the passage of No Child Left Behind indicates that high stakes tests decrease higher-level thinking, impact teacher effectiveness, discourage students, and have no measurable effect on standardized test scores.

Decreased higher level thinking. Higher-level thinking is the ability of students to analyze and synthesize information rather than regurgitate facts. Au (2007) in a metasynthesis, of 49 qualitative studies found, overall, teachers were prone to fragment knowledge closely aligned to what was on the test rather than having students integrate and assimilate that knowledge. In addition, teachers tended to use low-level teaching strategies when their goal was to have their students pass a test that consisted primarily of low-level multiple-choice questions (Berliner, 2011). Abrams, Pedulla, and Madaus (2003) conducted a national survey of teachers from 28 states from all grades levels and found that 76% of teachers reported a decrease in sound educational practices because of preparing for the test. Similarly, Hargrove et al. (2000) surveyed 236 elementary teachers and learned the majority of teachers believed they had to limit higher level thinking activities to prepare their students for the test.

This phenomenon is not just limited to reading and math. Science teachers have increasingly begun to teach toward students' ability to answer multiple-choice questions and, as a result, have reduced the amount of higher level thinking activities (Aydeniz & Southerland, 2012). Likewise, social studies teachers were increasingly likely to encourage memorization of facts and not higher-level thinking (Au, 2009). As AYP pressures increased, Plank and Condliffe (2013) found the quality of teaching decreased as teachers felt more pressure to pass

the assessments and therefore taught only the skills they felt would be on the test. For instance, teachers in West Virginia admitted to dropping irregular polygons from their geometry curriculum because only regular polygons were on the state assessment, until they found out irregular polygons were on the NAEP assessment and had to add it back in (Koretz, 2011). In view of all these, I submit that high stakes tests are driving instructional decisions rather than what teachers know students need to learn.

High stakes testing is also narrowing the definition of what is "smart" and reducing any expression of creativity, especially among the poor (Ambrose, 2012). The recent rise in high stakes testing has demonstrated that students have increased their test taking skills, not that they have learned anything more (Goodwin, 2014). Again, as research indicates the impact of this phenomenon is more pronounced in low-income schools. The discrepancy is so stark that Groen (2012) argues that high stakes testing has created a two-tier educational system where the lower class is taught a narrow curriculum while the upper class, who already passed the test, receive a more well-rounded education.

Impact on teacher morale. Teachers are aware they are narrowing the curriculum and limiting higher level thinking in the name of standards and accountability and it greatly hurts their morale (Cawelti, 2006). Qualitative research has revealed that high stakes testing affects the morale of teachers in several ways. First, teachers feel like they have less time to form caring relationships with their students (Wellman, 2007). Yet, according to research on the impact of affective qualities on learning, students desperately need caring teachers to succeed (Jeffrey et al., 2013). Unfortunately, students who do best on the test have the most time to form caring relationships with their teachers while the students who do worst on the test have the least amount of time to form strong bonds with their teachers, which creates a self-fulfilling prophecy.

Secondly, teachers face a dilemma of either spending their time making sure everyone passes the test for accountability purposes or differentiating instruction to help the high and low achievers. After interviewing teachers, Willis (2011) reported teachers wanted to help all students learn but when they were forced to teach to the middle they knew they were not meeting the needs of all learners. Therefore, the teachers suffer through the irony that No Child Left Behind is forcing them to leave children behind.

The most significant impact on teacher morale is that teachers do not believe that high stakes tests are a good reflection of the teacher's ability to teach or how much the students have learned (Vernaza, 2012). Relying on one test to determine teacher effectiveness demonstrates a profound lack of trust in their profession (Feuerstein, 2011). Other professions, like psychologists, are not judged based on one arbitrary indicator. Rather, they are judged on their adherence to professional standards and effort. The fact that hard working teachers are demeaned and not treated as professionals has had a significant impact on teacher retention rates (Goldstein, 2011). The impact has been especially high in schools serving low-income schools (Crocco & Costigan, 2007).

Decreased student morale. In addition to teachers' morale suffering because of high stakes testing, students suffer as well. von der Embse and Hasson (2012) measured the anxiety levels of 369 10^{th} graders and found that test anxiety is a contributing factor to poor student performance on high-stakes tests. As early as 3^{rd} grade, Dutro and Selland (2012) found that students were aware there was a possibility of retention associated with state assessments and therefore approached the test with trepidation. Cornell, Krosnick, and Chang (2006) interviewed 911 students in Minnesota who were wrongly informed in 2000 that they did not pass the mandatory state math assessment. Of those, 80% reported depression or worry, 50% decreased

their involvement in other school activities to focus on schoolwork, just over 50% reported feeling stupid, and 4% dropped out of school before they learned of the mistake. In another study on the impact of high stake tests on students' perceptions, Pershey (2010) found across all grade levels students' self-perceptions of academic ability were tied primarily to their high stakes assessment score, even more than their grades. In addition, Huddleston (2014) found high stakes testing had greatly increased the numbers of student retentions nationwide despite evidence that benefits of retention are short-lived. He concluded this was because short-term gains were sufficient to benefit schools and districts on academic reports, regardless of the long-term impact on the student. Finally, Brown (2007) found the long-term impact of retention was an increased likelihood of a student dropping out.

Impact on student achievement. There are some discrepancies in the research as to whether making students take a high stakes test improves student achievement. While there are some instances where research has shown that high stakes testing improves tests scores (Yeh, 2005), most studies find little or no impact. Dee and Jacob (2010) found significant gains in math but not in reading. Marchant, Paulson, and Shunk (2006) used NAEP scores to determine the effectiveness of high stakes testing in several states and found although there were gains in a few subgroups, there was no achievement gains overall. Nichols, Glass, and Berliner (2012) found although there were slight gains in 4th grade scores, no other grade levels showed improvement. Martin (2012) studied inner city schools and determined that test scores should not be used for accountability because the data shows they are skewed against low-income schools. Berliner (2011), Sahlberg (2010), and Chudowsky et al. (2007) reviewed studies on the effectiveness of high stakes testing and student achievement and concluded that, overall, NCLB has not resulted in increased test scores.

Some proponents of high stakes testing will point to increased state assessment scores proof that high stakes testing improves learning. However, research has indicated when scores on state assessments increased there was often no corresponding increase on NAEP scores, raising suspicion that score inflation was occurring. For example, in 2005, Alabama reported 83% of its fourth graders proficient in reading, while the National Assessment of Educational Progress (NAEP) found only 22% to be proficient (Goldstein, 2011). In another instance, Koretz (2008) conducted longitudinal studies in Kentucky and Texas where it was observed that assessment scores would drop rapidly when a new test begun and rise steadily until a new test was introduced and suddenly the scores would drop a second time before beginning to climb again. He then went back and tested students using the original test and found their scores were the same as when they took the original test the first time. He concluded these results were a clear indication that teachers were effectively finding ways to improve students' scores on the specific test presented to them, without necessarily improving overall learning.

Likewise, Goldstein (2011) reviewed the NAEP scores for all states that had shown significant growth in state assessment scores and found there was no correlation between increased state assessment scores and increased NAEP scores. Goldstein concluded that such states simply confirmed the trends that NAEP scores have stagnated over the past 20 years. Therefore, while teachers are succeeding in improving the results on the state assessment, as they are required to do, there is no corresponding increase in student learning on an objective test they have not specifically prepared for. Lauen and Gaddis (2012b) found evidence for such an assertion in in North Carolina. A year after a school failed to make AYP; the scores on their minority subgroups showed an increase but the school admitted they could not determine whether that increase was a result of increased learning or just increased test prep and narrowing

the curriculum. There is simply no corroborating evidence that a once a year test reflects what a student knows or learned (Lucido, 2010).

Next generation assessments. Despite the criticisms of narrowing the curriculum, decreasing higher level thinking skills, lowering teacher morale, and discouraging students, high stakes tests do not seem to be going away any time soon. Eighty percent of Americans surveyed support exams for promotion to the next grade and 85% support a high school graduation exam, if they are well written (Howell, West, & Peterson, 2007). Politicians support high stakes testing and believe it an easy way to measure success (Orfield & Kornhaber, 2001). Therefore, much of the recent research has shifted away from the relative benefits of high stakes testing to trying to define the characteristics of a well-constructed high stakes test. In a metasynthesis of 49 studies, Au (2007) concluded that in the rare cases an assessment authentically tested a well written curriculum, high stakes testing led to expansion of the curriculum and a better integration of knowledge.

Other researchers have identified various aspects of a well-constructed high stakes assessment. First, the tests must include social studies and science so teachers will have no reason to focus solely on math and reading (Hout, 2012). Second, they must be rigorous enough that teachers need not narrow their curriculum (Hong & Youngs, 2008). Third, they must assess higher level thinking skills (Cawelti, 2006; Dee & Jacob, 2010). Fourth, they must be written in a way that teachers can reclaim their role and find ways to teach vital skills (Reich & Bally, 2010). Finally, they cannot be only one-time summative tests (Madaus & Russell, 2010). Instead, they must be formative and summative so they can be used to inform decisions about learning, instead of being simply an accountability tool (Supovitz, 2009). If the next generation of assessments have these characteristics combined with a solid curriculum, data driven

instruction, and the support of building and district leadership, high stakes assessments have a chance of impacting learning (Williams, Kirst, & Haertel, 2005). If they are successful at properly reflecting student learning then they might be fair (Porter & Chester, 2002).

Organizations such as CETE, SBAC, and PARCC have developed tests that exhibit these characteristics. Darling-Hammond (2010), after carrying out an extensive review of SBAC and PARCC test design for the Council of Chief State School Officers (CCSSO), found that by using technology enhanced questions it is possible to create assessments that properly reflect the curriculum, encourage higher level thinking, provide formative information for teachers, and provide summative information to parents and schools. Researchers have illustrated several improvements of next generation assessments. First, they have the strong collaboration between curriculum designers and assessment developers necessary to minimize some of the validity concerns of previous assessments that narrowed the curriculum (Lai, 2011). Second, an analysis of the SBAC content specifications found that more than 66% of the questions in reading and math asked higher level thinking questions, compared to 20% in reading and 2% in math on traditional state assessments (Darling-Hammond & Adamson, 2013). Third, the new assessments are formative, guiding the teachers and students to improve their learning rather than simply being a punitive onetime assessment (Heritage, 2010).

These next generation assessments have only been given over the last five years but some researchers have found these tests are improving learning. Pecheone, Kahl, Hamma, and Jaquith (2010) analyzed the performance assessment results from many states and determined that "designed and used well, development of the next generation of state accountability systems has the potential to strengthen instruction, curriculum, and assessment as well as serve as a catalyst to reform schools and districts" (p. 41). Herman and Linn (2014) analyzed the PARCC and

SBAC assessments and found they contained higher rigor and high level thinking then most state assessments. Despite early successes, there has been backlash against these tests in some states and many are opting to write their own assessments rather than rely on a federally funded one (M. McNeil, 2013). Kansas was one of those states, selecting CETE to write their new assessment.

To examine how high stakes testing environment and the next generation CETE assessment has influenced classroom instruction I conducted a qualitative study, to what extent educational and instructional triage is occurring public schools. In the next chapter, I explain the methodology of the study.

CHAPTER 3

Methodology

This chapter includes a description of the research design, research context, participant selection, data collection methods, data analysis methods, research quality and my positionality.

Research Design

A research design is the format around which the data are collected. In view of the research problem, theoretical framework, and relevant literature, the research design I selected for this study is a basic qualitative study (Merriam, 2009). This design allowed for a preliminary structure which gave cohesion to the study (Maxwell, 2012) while retaining flexibility in exploring emergent data (Patton, 2002). It allowed a specific setting to be the direct source of data and the researcher to be the primary instrument of data collection (Biklen & Bogdan, 2006). As a result, qualitative studies are predicated on the theory of social construction where meaning is derived from the perceptions of individuals (Creswell, 2013). The researcher is not discovering the one true meaning, rather the meaning is constantly being constructed by the participants and interpreted by the researcher (Merriam, 2009). This meaning and its context are richly described which provides for a thorough understanding of the phenomenon being studied (Stake, 1995).

Based on the literature review and theoretical framework used, I determined educational and instructional triage could best be observed through a qualitative research study, as this allowed teachers to explain their instructional decisions and expound on the reasoning behind their choices. Since this study focused on how teachers teach, a qualitative study was preferable to a quantitative one (Neal & Schanzenbach, 2010). From the teacher interviews, I garnered a clear picture of how the high stakes testing environment had influenced their instructional

practice. In addition, a qualitative case study allowed me to provide an in-depth description and analysis of these changes.

Research context: Appraisal Public Schools

Appraisal School District is a located in a small suburban town in the Midwest. It has a total student population of 3,749 students spread out over four elementary buildings, one 5/6 building, a 7/8 building, and a high school. Appraisal Public schools offered a good location to examine the influence of high stakes testing on classroom instruction because its size and composition made it a typical Midwestern town, which, for several reasons, enhance the transferability of the results. First, the fact that it is a small city is characteristic of the Midwest. A small city allows comparisons to other small cities, as well has making it more comparable to both urban and rural schools. Second, the scores at Appraisal Public schools have fluctuated over the last few years. When scores are consistently low, assessments tend to be overemphasized. When they are consistently high, assessments tend to be deemphasized. Therefore, examining teachers in a school with fluctuating scores enhanced the results. Third, the growing Hispanic population of the City of Appraisal is being mirrored across the country and teachers' reactions in Appraisal might be more indicative of what is happening in other locations. Finally, with five schools, the size of Appraisal was manageable enough to allow interviews with all its 3rd -5th grade teachers who have taught for more than three years, while being large enough to provide a variety of conditions at each school.

Participants

In qualitative studies, participants are purposefully selected based on their knowledge or relationship to the topic of the study (Marshall & Rossman, 2010). Because this study focused on how teachers' classroom instruction had been impacted by high stake assessments, I planned

to gathered data from approximately 24 third through fifth grade teachers in five schools. The choice of 24 participants is based on Merriam (2009) observation that interviews should be conducted until redundancy is attained.

Data Collection: Focus Groups. To get perspectives from a variety of teachers, focus groups consisting of four to six 3rd, 4th, and 5th teachers were held at each of the five selected buildings. The exact number at each school varied because I chose not to interview all the 3rd, 4th and 5th grade teachers at each school, rather I only selected teachers who were able to provide the most insight into my topic. This purposeful selection was based on teachers' ability to offer understanding into how high stakes testing has influenced their classroom instruction (Creswell, 2015). Therefore, the focus groups involved veteran teachers, defined in this study as any teacher who had taught a minimum of four years and who had a perspective about how classroom instruction might have changed with the advent of the new assessments. Within these five schools, I utilized focus groups, rather than interviews to capitalize on the complex interplay between a group of participants (Merriam, 2009).

Since my methodology called for 3rd-5th grade teachers with more than three years' experience, I was unable to conduct interviews at two of the schools. In the smallest K-4 school, there was only one section per grade and both teachers happened to have less than three years' experience. In a second, larger building, the principal reported that a combination of retirement, transfers, and turnover had resulted in all but one of the 3rd-4th grade teacher having less than three years' experience. Since my methodology relied on focus groups to enhance discussion and honesty, I chose not to do a personal interview with that one teacher but rather focus on the participants in the other three schools.

I conducted two focus groups in each of the three buildings for a total of six. The total number of participants was 19. Of those, there was one male teacher and 18 female teachers. Thirteen of them had taught their entire career in Appraisal and six had moved from a nearby district, in most cases that was from the nearest urban district. In addition, two teachers had taught for 3-5 years, eleven for 5-10 years, three 11-20 years, and three for more than 20 years.

Focus groups are used to collect the observations and interpretations of participants (Stake, 1995). The assumption with focus groups is the perceptions of others are meaningful both individually and as part of a collective narrative (Patton, 2002). Participants share their perceptions on the topic but are guided by a predetermined set of questions which are meant to guide the participant to address the research questions (Merriam, 2009). It gives participants the opportunity to compare their experiences and construct a common understanding while giving the researcher the opportunity to observe the extent and nature of the interviewees' agreement or disagreement (Morgan, 1996).

The focus groups were semi-structured interviews that began with a predetermined set of questions but allowed for exploration into unexpected areas of inquiry (Patton, 2002). I used eight open-ended questions with probes because they allowed participants the opportunity to expound on their answers with specific details or illustrations (Creswell, 2013). While conducting the focus groups I listened keenly, avoided interruptions, tolerated silence, avoided leading questions, followed up on relevant participant comments, and remained neutral (Gay, Mills, & Airasian, 2012). Each focus group was conducted on site, after school hours and lasted about 45 minutes. All focus groups interviews were audio recorded.

Focus groups transcriptions were be digitally recorded and secured in a password-protected program for the purpose of analysis and safety/privacy. Later the recordings were be

transcribed and unitized in an excel spreadsheet. Field notes were taken to document participant interaction (Morgan, 1996) and comments about specific responses (Patton, 2002).

Permission to be involved in the study was obtained from all participants using the Informed Consent documents provided by Wichita State University. The Informed Consent form described the objectives of the study and the protections provided to the participants. Each participant was given a copy of their consent and was reminded that they can opt out of the study at any time. I maintained confidentiality for all participants by ensuring that no names or other identifiable information appeared in the final report. All of these documents were submitted to the Wichita State Internal Review Board (IRB) for approval (See Appendix A).

Data Analysis

Data analysis is a process of building a holistic and in-depth description of a set of data (Patton, 2002). Qualitative data analysis occurs simultaneously with data collection (Merriam, 2009). At the conclusion of each focus group interview, I transcribed the audio-recorded data and began to identify initial themes and patterns (Maxwell, 2012). These initial themes were used to modify questions during subsequent focus groups and help determine if they were confirmed or denied. This iterative process allowed for data analysis to begin with the first interview and develop over the course of the research (Marshall & Rossman, 2010).

Once all the focus group interviews were completed and transcribed, I read the transcriptions to help determine if my initial themes held up under the scrutiny of all the data and allowed me to develop a first list of comprehensive themes (Patton, 2002). Themes can be developed by searching in the data for repetition, metaphors, analogies, similarities, differences, linguistic connectors, and missing data (G. W. Ryan & Bernard, 2003). Next, each transcript was unitized and entered into an Excel spreadsheet and categorized by my themes. With each

entry, I utilized a constant comparison method where patterns or reoccurring themes were identified and either accepted or rejected based on my initial list (Creswell, 2013).

Once I completed all five focus groups, I had a second, more final list of themes. I then began to scrutinize the data and the themes to draw some conclusions. This analysis was also an iterative process that relied on me constantly reexamining my assumptions, staying focused on what was actually found in the data, and comparing those findings to my theoretical framework and literature review (Lapan, Quartaroli, & Riemer, 2011). The confluence of this information provided me with some insights into the views of educators on how their instructional practices have been influenced by high stakes testing.

Research Quality

Research quality is essential in a qualitative research design because it is based on the theory of social construction. When truth is socially constructed terms like validity and reliability take on a new definition (Kvale, 1995) and a term like reality becomes multi-dimensional and ever-changing (Merriam, 1995). This is because qualitative data is naturally interpreted by the perceptions of the participants and by my analysis of that data (Freeman, Preissle, Roulston, & Pierre, 2007). Therefore, in qualitative research, Lincoln and Guba (1986) contended that quality research should have trustworthiness, which includes credibility, dependability, transferability, and conformability. This section will describe how I achieved and sustained research quality.

Credibility. Credibility is the validity of the inferences drawn from the data (Freeman et al., 2007). It is correctly choosing among competing interpretations and depends on continually checking, questioning, and comparing interpretations against the data, prior research, and

theoretical framework. I demonstrated credibility through quality interpretation and member checks.

First, as I developed my themes and drew my conclusions, I continually checked whether the theme was truly representative of data. I checked for outliers that might contradict the theme and looked for negative evidence to support my belief (Kvale, 1995). I also looked for triangulation of the data. Not in a pure quantitative sense, but in the sense that the information I found in each school corresponded to each other and in the sense that the data I collected triangulated with my literature review and theoretical framework (Lincoln & Guba, 1986). Secondly, I conducted member checks. I emailed the participants a summary of my findings to determine if my themes and conclusions rang true with them (Merriam, 1995). This check of validity limited the possibility of me misinterpreting the data (Lapan et al., 2011).

Dependability. Dependability or reliability is the degree to which the findings are consistent with the data (Merriam, 2009). I addressed dependability in a variety of ways. Rich, thick description and after interview clarification with participants helped ensure that the results of the study were consistent with the participants' reality (Merriam, 1995). Interview transcription and analysis were used to ensure that participant's subtle or unspoken subtext were revealed and considered (Patton, 2002). Finally comparing the variety of teacher perceptions across focus groups and buildings allowed the data sets to be compared to ensure consistency of emergent themes.

Transferability. Transferability is how applicable the study is to other situations (Lincoln & Guba, 1986). Since transferability is determined by the reader, a rich, thick description is the best way to allow the reader to make that determination themselves (Merriam, 2009). The more the description captures all the nuances of the setting and the complexities of the findings the

more readers can determine how applicable the study is to their situation (Peshkin, 2000). I described carefully and accurately my setting, participants, data collection method, and findings with enough detail and acumen to paint a clear picture for my readers.

Confirmability. The test of confirmability is whether the reader believes that the findings verify the assumptions, elaborate on existing concepts, provide some new insight, or clarify the complexity of the influence of high stakes testing on teacher instruction (Peshkin, 1993). This is based on whether the research has a worthy topic, rich rigor, credibility, resonance or transferability, significant contributions, ethics, meaningful coherence and sincerity (Tracy, 2010). I have already addressed credibility, transferability, and meaningful coherence. I will address the remainder of Tracy's suggestions below.

Worthy topics are relevant, timely, interesting, and significant. The advent of the common core standards and their associated assessments, coupled with the long-overdue reauthorization of NCLB in Congress, has created a political environment rife with controversy (Schaeffer, 2012). Politicians and citizens alike are debating the merits of national assessments and whether assessment scores should be tied to teacher evaluations. How high stakes testing influenced teacher instruction is a topic relevant at the district, state, and national level. The more significant research conducted in this area, the better-informed state and national policy can become.

Rich rigor requires an appropriate theoretical framework, well-crafted data collection and analysis methods, and rich, informative descriptions and conclusions. Campbell's Law is a theoretical framework that has been successfully applied to a wide range of social science topics and has been specifically utilized to examine high stakes testing by several of the leading experts in the field (Nichols & Berliner, 2007). My data collection and data analysis methods mirror the

efforts of many qualitative researchers who have established the effectiveness of these designs (Creswell, 2013). My findings and conclusions are elaborately described and well thought out utilizing the methods outlined above. This created a study with enough rigor to inspire confidence in its conclusions.

If a topic is worthy and rigorous so that it is researched in a way as to be credible, coherent, and offer resonance to the reader then it has the potential to make a significant contribution (Tracy, 2010). My study examined how high stakes testing has influenced teacher instruction in the classroom. This study could be significant to any social scientist who is utilizing Campbell's law to analyze the impact of a social indicator on human reactions. More specifically it will be of interest to building and district leaders who have the goal of maximizing student learning regardless of the political environment (Kohn, 2000). It should have interest from state and national politicians hoping to craft legislation that maximizes learning for all children regardless of their background (Koretz, 2011).

Positionality

To ensure the trustworthiness and dependability of a research study it is important to identify anything that could potentially influence the study design, decisions, inferences, or conclusions that guided the study. Because truth is socially constructed, it is impossible for me to be completely neutral when studying any phenomenon (Freeman et al., 2007). The reader's interpretation is not based on the data but on my rendering of the data (Noblit, 1999). Therefore, it is vital that I disclose to the reader, and admit to myself, any subjectivity I have about this topic and its possible findings (Peshkin, 1988).

I am the assistant superintendent in charge of curriculum and assessment in a Midwestern, Catholic Diocese that encompasses 38 schools and nearly 11,000 students. In this

role, I oversee the implementation of the state assessment protocol and am ultimately responsible for the success, or failure, of our schools and students on those tests. I have witnessed firsthand how our schools and teachers have tried a variety of techniques aimed at improving the Catholic Diocese's schools success on these assessments. I have personally led strategic planning sessions that explored ways to increase test scores. I have helped schools develop action plans to address score deficiencies that resulted in these schools being labeled "on watch" or "on warning." I have overseen the implementation of Multi-Tiered Systems of Support (MTSS) in several of our schools. MTSS is Kansas' version of Response to Intervention (RTI), which encourages the use of data to created tiered instruction time to maximize learning for all students. In addition, I have felt the pressure that accompanies the publication and ranking of schools in the local newspaper. Finally, as the spouse of a public school elementary teacher, and a parent of three school age children, I have witnessed firsthand the anxiety and stress that these assessments inflict upon students and teachers alike.

In my professional capacity as an Associate Superintendent, I have tracked student and school data that clearly demonstrate these same assessments have resulted in dramatic student gains across multiple subgroups. Because of the high stakes aspect of these assessments I have witnessed teachers implement change and advocate for students in ways they never did when the tests were low stakes. In addition, I have written and mandated grade level assessments over our Diocesan Religion, art, physical education, and music standards because our leadership realized that if not all content areas were assessed with the same vigor as the core subjects then they would not be taught with the same vigor. I have given speeches and spoken with parent groups about the benefits of state accreditation and necessity of state assessments.

Since I have witnessed both the benefits and limitations of high stakes assessments it was necessary for me to separate my personal beliefs from the collection and analysis of the data (Marshall & Rossman, 2010). I accomplished this first by consciously acknowledging my positionality and being self-reflective enough to realize if left unchecked it could bias my findings (Chiseri-Strater, 1996). The process by which I separated my assumptions and biases from the data during data collection and analysis is called reflexivity (Kleinsasser, 2000). I built in this reflexivity by creating dedicated times, spaces and contexts within which to be reflexive (Mauthner & Doucet, 2003). Specifically, after each focus group interview, I went back through the transcription and identified which parts I impulsively agreed and disagreed with. I thought about which teachers I sympathized with and which comments particularly resonated with me. Then I made a conscious effort to disregard those first impulses and analyze the data from a more neutral lens. By taking the time to purposefully identify what comments supported my preconceived opinions I was better able to factor those out of my thinking.

CHAPTER 4

Findings

Qualitative research is emergent. It begins with collecting the data but continues through transcription, organization, theme development, and writing (Merriam, 2009). This chapter presents the findings from the data collected from six focus groups comprised of 19 teachers. This chapter is organized around recurrent themes expressed by the participants on the influence of high stakes testing on classroom instruction. The research site is reintroduced through the eyes of its teachers. In some cases, quotes and details were modified to protect the anonymity of the participants and the district.

Appraisal School District

Originally founded in 1871 along a major transcontinental railroad route, the city of Appraisal is located 25 miles from a Midwestern city. It has an ethnically diverse population of approximately 19,000. Because of its size and proximity to a city, Appraisal is labeled a small suburban city (National Center for Ecuation Statistics, 2015). The railroad is still a significant economic driver in the community. Other major employers are the Regional Medical Center and Appraisal Public Schools.

Appraisal Public Schools employees almost 800 teachers and support staff who serve a total student population of 3,749. This district has four K-4 elementary schools, a 5/6 building, a 7/8 building, a high school, and a Career and Technical Educational Center. More than half of the teachers hold a Masters Degree and they have won a combined 10 Teacher of the Year awards in the last two years.

State Assessment Scores

Since the NCLB requirements began in 2002, students in the Appraisal Public Schools have scored at or slightly above the state average on the State Assessments. The district has been put on watch in a few grade levels on several separate occasions but they have never been put "on improvement" under the NCLB requirements. In recent years, the city of Appraisal and the Appraisal Public Schools have undergone a demographic shift as an increasing number of economically challenged Hispanic families moved into the community. This influx of new, young families is demonstrated by the fact that while 16.3% of the overall community is Hispanic, the school district is 27.2% Hispanic. In addition, 55.5% of the students in Appraisal public school are considered economically disadvantaged. These changes and the increasing difficulty of the new assessments impacted their scores. In 2013, the last year of the old state assessments, 3rd and 4th grade reading scores were above the state averages but their 5th grade reading and their 3rd through 5th grade math scores were all slightly below the state average. They were more comparable when taking into account their high percentage of economically disadvantaged students. Table 1 below provides a breakdown of demographic information and state assessment results by school.

Table 1

Appraisal School Demographics and State Assessment Results

School	Grade	% Ethnic Minority	% Low SES	% Passing Math	% Passing Reading
Elementary 1	3rd	22.2	59.2	74	77
Elementary 1	4th	22.2	59.2	94	86
Elementary 2	3rd	43.8	75.2	85	90
Elementary 2	4th	43.8	75.2	78	79
Elementary 3	3rd	33.6	59.4	93	83
Elementary 3	4th	33.6	59.4	92	78
Elementary 4	3rd	29.2	66.1	82	74
Elementary 4	4th	29.2	66.1	90	85
Middle school 1	5th	28.2	58.4	85	81

Response to Intervention Process

Appraisal School District implements the Response to Intervention (RTI) program as suggested by the State. The state model is specialized variation of RTI called the Multi-tier system of support (MTSS). This model involves beginning the year with a benchmarking test. The results of this benchmark test are used to assign student to Tier I, Tier II, or Tier III intervention groups. The majority of students will be placed in Tier I. These are students who are on or above grade level and receive regular instruction. Tier II students are at risk students who are slightly below grade level on these skills. Tier II students receive regular instruction and then receive additional tier time, usually 30 minutes per day, which is done in small groups. Tier III students are severely at risk. They receive both Tier I and Tier II instruction and an additional 30 minutes a day of pull out, specialized instruction. Students in Tier II and Tier III

are progress monitored weekly to ensure they demonstrate growth. Students who demonstrate sustained growth are moved up a tier where services are reduced and students who are showing no growth on the weekly progress monitoring might be moved down a Tier. The placement of students in tiers varies by subject area and by content covered. For example, a student might be in Tier III during the geometry unit, but Tier I during computation. Benchmarking tests are given again in the winter and spring to verify the proper placement of students, while progress monitoring is intermittently used to move students between tiers.

Participants reported that the Appraisal School District began to implement the rigorous State RTI program in reading in 2012. In addition, they purchased a new reading series in 2013 that aligned with the new Common Core standards, and provided resources for progress monitoring and tiered instruction time. The teachers reported they began giving benchmarking tests three times a year, in September, December, and May. As one teacher explained,

> The test we gave in September determined which tier the students went into. They are progressed monitored in the tier II and tier III groups until they reach a certain level and then they graduate to the next tier. That can happen at any time during the semester based on enough positive scores. Then when we do benchmarking again in December we will do a different sort to see how the kids fit into the different tiers. So they can move at the benchmarking time or with enough progress monitors scores showing they are where they need to be. So it is kind of in flux a lot.

While this RTI process and new textbook series has worked well in reading, the District has chosen not to adopt a new textbook series in Math and is only beginning to implement RTI in math. The teachers were frustrated that they were expected to teach the new common core math standards but not given the resources to do it. One teacher said, “We were told that the State was

going to be rewriting the math standards in the next year or two anyway so why buy a new series?" As a result, the teachers were left augmenting the old math series with outside lessons as best they could. One teacher explained, "We are using our old curriculum and trying to pull from everyplace we can to meet the new standards and that is all on our own time." Likewise, the teachers were frustrated that the RTI system was not as far along in math. A teacher reported, "We do have a benchmark test for math but I don't think it is quite as clear where to put the kids in tiers…. We also have some questions about the validity that the benchmarking test matches the standards." One school said they were going to implement tier time in math next year. Another stated that they have time set aside for math tier time, "but we don't use it as well as we do for reading. It is just going to take time." The varied success of the schools within the district on the state assessments and recent implementation of the RTI process influenced the teachers' perception of high stakes testing and classroom instruction.

Teachers Perceptions of High Stakes Testing and Instruction

Throughout the interviews, the teachers' perceptions of the influence of high stakes testing on classroom instruction were fairly consistent. The first section describes the general belief that the new common core assessments are more rigorous. The second section focuses on instructional triage, specifically how this increased rigor, combined with limited time and resources, has led teachers to feel forced into making hard decisions on what skills and content they must teach based on the test rather than their professional judgment. The third section shifts to educational triage and a description of how the school has implemented policies and procedures built specifically around increasing test scores. Next, the teachers describe how high stakes testing has significantly increased the pressure on the school, the teachers, and the students and what influences that pressure is having on their classroom. The fifth section

highlights what the teachers see as the two positive effects of high stakes testing: accountability for all students succeeding and the advent of data driven instruction. The last section focuses on the teachers' perceptions of the future of high stakes testing and whether they see the positive outweighing the negative.

New Common Core State Assessments

The state's new common core aligned assessments were piloted in spring of 2015 and in late fall 2015 the teachers had only recently seen their students' scores from those assessments. Unanimously all 19 participants believed the new common core aligned assessments were significantly more rigorous than the old state assessments, which were discontinued in 2012. They cited two reasons for this increased rigor. Several pointed out the higher order thinking skills required, while others felt like "the entire curriculum has been moved down a grade," as one put it. Some saw this increased rigor as having positive benefits. One teacher contended, "They have to go back and find evidence to support their argument, which is better. I think that is beneficial all the way around." Another complimented the amount of writing expected, "especially on the science test where they have to think about how they can apply science and write about it." A third simply said, "I think it is good that they have to use more brain power."

Others thought the increased rigor was unrealistic, especially in Math. One teacher lamented the difficulty of the assessment, even wondering if the teachers could pass it,

> I thought the questions were written in such a way that it made the problems very difficult for most kids. In fact, there were several of us that when the kids were finished with the assessment, we looked at each other and said, "I don't really have any idea how to do some of those problems."

Another teacher focused on how the math assessment negatively affected students, "It is very intimidating to our children at first. It has been an adjustment to be able to no longer just click an answer. Now with those deeper levels of thinking…it has been difficult." A third teacher summed it up this way, "They definitely seemed to be higher level thinking questions, which don't always seem to match where students are in their life. It always seems like they try to trick them…so I thought it was kind of hard for them." The perceived difficulty of the state assessments led the teachers to acknowledge the fact that a conscious effort had to be made to help ensure that students would succeed.

Instructional Triage

Prior research has indicated that instructional triage is one of the most prevalent results of high stakes testing (Booher-Jennings, 2006). With instructional triage, much like medical triage in an emergency, teachers are assessing which skills students need to know versus what can be eliminated. With a limited amount of time and resources, teachers who are under extreme pressure to demonstrate positive results will often feel they have to make tough choices about what to teach. This will naturally lead teachers to put more emphasis on concepts and skills that are on the assessments. This enhances learning when all teachers use a common set of standards to vertically and horizontally align the curriculum. However, if teachers only teach the portion of the standards that are assessed rather than all of the standards, the advantage of standards based education is lost (Koretz, 2008). Likewise, if teachers believe they must spend an inordinate amount of time teaching test taking skills it might limit the amount of time they have to teach the remaining standards (Rothman, 2005). Finally, practice tests, motivational activities, and the tests themselves all take away precious instructional time that teachers need to teach all

the required skills. Overall, it seems that teachers believe that too often they are subjugating their professional expertise to teaching to the test.

Teaching to the test. During the interviews, teachers indicated they were resigned to the fact that they have to teach what is on the assessments. When asked, "How did the new state assessments change what you teach," One teacher simply replied, "Everything. It is a different test." And when asked, "What happens if the test changes again," another teacher answered, "You just go with what they tell you." In fact, teachers displayed frustration that they did not know enough about what was on the new assessments. One teacher stated,

> It is kind of like trying to hit a bull's-eye but you have no idea what you need to do to get there…. If you want me to teach something, just tell me. I would be glad to do everything I can…. Just give me an idea of what is going on.

Another teacher cited a specific example of changing what she taught for a formative math assessment, "Maximum and range are things that are not in the math standards at our grade level so I wasn't going to teach them, but they are on our screener assessment so I went ahead and taught them today."

The teachers also specified numerous ways they changed what they taught to align to the new assessments. One said, "We tweaked our schedule a bit this year to make sure we hit some of the math skills because there were some we taught before the test last year." Another reported,

> We chose to transition to a new curriculum because we had so many gaps and holes, we needed one that was already established…and we knew that this new curriculum was going to be closely aligned to the new assessment, which is what we needed.

This change occurred across grade levels as well. A teacher stated, "We spent a lot of time realigning grade level standards because of the new test." Admittedly, at least two teachers saw many of these changes as simply adjusting to the new standards, not teaching to the test. For example, one teacher stated,

> We have changed so much in our curriculum that there has been so much focus on what we need to teach, how we are going to teach it, and how we are going to fit it all in that honestly preparing for the assessments has been the last thing on my mind.

Another said, "I spend a lot of time looking at the standards trying to figure out what they want me to teach." These two teachers were resisting the temptation to utilize instructional triage because they were trying to teach to the standards rather than the test. However, they were absolving themselves from the problem by teaching what "they" want them to teach. Whenever teachers begin to subjugate their professional opinion in the classroom for the opinions of the "they," they are practicing another variation of instructional triage, but one that relies on "I was just doing what they told me to do."

Other teachers resisted the idea that teaching what someone else thought was important was really best for their students. One teacher wondered aloud, "On my practice tests they look like they are making progress but is it the right kind of practice? I don't know if it is the right kind of practice." She was conflicted that even though the students were learning what they were supposed to be learning, was it really what they needed? The second teacher shared her frustration, "We did all this so they could pass one test? Just sticking them in a small group was a waste of time. It didn't make them a better reader; it just helped them pass the test." The inner turmoil of the teachers was apparent. On one hand, they knew what they were expected to teach

and they knew how they were going to be judged but on the other hand, they were not sure what was best for students in their classroom.

No sooner did they express their concerns; the teachers went right back to talking about the skills they now needed to teach to ensure their students were ready for the assessments. Specifically the fifth grade teachers felt they had to teach more problem solving. One teacher said simply, "I do a lot more problem solving then I would have done before." She believed that since the new assessment included more word problems than the previous assessment she was obligated to spend class time on complex word problems. In a separate focus group, another teacher said almost the identical thing, "We definitely need tons more word problems." Neither teacher seemed to be necessarily complaining about having to do more word problems, but seemed more resigned to the fact that if it was on the assessment then obviously that is what they would do. Clearly, they were teaching to the test, not out of a desire to cheat or create an unfair advantage for their children, but because that is what teaching has become in their district: find out what is on the test and then adjust accordingly.

One math teacher did express concern that the problem solving on these new tests was too difficult and, no matter how much time is spent trying to prepare the students, it was not always going to work. She lamented,

> It is hard on the kids because we moved away from drill and kill to more in depth problems with multiple steps. But try as you may to walk kids through those sorts of problems over and over…They are with you when they are with you, but when they are on their own, they are just gone.

She expressed that the problems were so complex that when she walked students through each problem step-by-step they could follow along and complete the computation. However, when

they were on their own and faced a multi-step problem without a teacher to guide them, they were lost. They did not know what steps to take or in which order to do them. She felt the problems on the new assessment were overly complex and were not a true reflection of the student's grade level ability.

Teachers in the lower grades focused on the increased amount of time spent on teaching computation in order to prepare their students for the state assessment. One explained, "I am trying to have one assignment a week from now on, directly on computation, straight computation. Some they work on in class and then some they take home to work on as well." This teacher was spending class time, as well as homework time on test preparation. She had seen the standards and determined that her students' computation skills were not high enough for the new assessment. Another third grade teacher said it was clear that "We have to get their computation accuracy higher." She had screened her students and realized their skills were not up to the new rigorous standards. She did not dispute the standards or choose to focus on what she thought was important. Rather, she adjusted what she did to help her students prepare for the test. This resignation to teaching to the test was most obvious in writing. One teacher stated,

> The one that is probably the biggest thing that they don't experience is the writing assessment. The on-demand writing assessment is not how we teach writing so trying to introduce them to some of the ways they are going to have to take that test beforehand, so they are not completely floored. That is just not how we teach writing, at least in our district.

As a result they were considering revamping their entire writing curriculum to align with the test even though they believed they were teaching writing well before. Again, there was not anger or resentment in these comments but a resignation that obviously they would have to change their

writing curriculum again because the assessment had changed. For them it seemed to be a part of the natural order: the assessment changes, your curriculum changes. Not the other way around.

The most interesting conversation surrounding teaching to the test focused on the need to review material before the test. It was clear that district administration did not want teachers teaching to the test and so they had issued a directive that teachers were to focus on the standards and not take time away from instruction to focus on specific review. One teacher said,

> We used to spend weeks test prepping. We would set a date then we would spend two or three weeks before the test getting them ready to take the test, then we would take the test. Now we just teach all the time.

While this teacher was relieved by the change and enjoyed not teaching to the test in that way, other teachers were troubled. One commented she was happy about more instructional time but said, "That is what is so hard not to do, is not to do the review." Another teacher disagreed with the new directive completely. She said, "I still think that review is important…. I know that I would hate to be thrown into a test and be told here you go and be tested over something I learned in August, clear in March or April." She believed review was necessary to maximize the students' test scores. For her, the test was not about know what information her students had retained over the course of the year, but about reviewing the testable indicators and maximizing the score on that given day.

After spending so much time hearing how they were adjusting their instructional time to focus on problem solving, increased computation, and increased writing, it was interesting to hear that the district was taking a stand against teaching to the test. In the teachers' mind, they were adjusting to the new standards, not teaching to the test. They, or at least the district, only

saw the three weeks blocked out prior to the test as teaching to the test. The teachers saw a natural order in subjugating their professional expertise to teaching to the test. It has occurred so many times over the course of their career that they did not even recognize it for what it was.

Teaching test-taking skills. The majority of teachers stated they took time out of class to teach test-taking skills to help their students do well on the assessments. One third grade teacher commented, "In recent years the students have not gotten all the testing taking skills in previous grades, so it has been up to us to teach them how to take the test." When the old state assessments were in full swing, even first and second grade teachers were taking time out of class to teach test-taking skills. However as soon as they were able, they dropped that practice and now the third grade teachers felt they had to take even more time to prepare their students for the new state assessments. The clear indication by the teacher was that she wished the younger grades would return to teaching test-taking skills so that her students would be more prepared. Another added, "As a teacher when you know there is something you can teach your kids that is going to help them on the test, you start to do it." Again, she illustrated a pragmatic approach to the new assessments. It was obvious to the teacher that she was almost obligated to teach test-taking skills. These tests are so important that naturally a teacher would do everything she could to help her students succeed. She did not even question the underlying implications to what she was saying. Still another had a more long-range view of test taking skills being a necessity for students at different stages of their educational careers.

> They are going to take tests for eight years of their life. Even if state assessments all of a sudden were not a thing, they would still need to know how to take a test because the further along they go, a test is more and more the yardstick the teacher is going to use to measure whether or not they have learned. Beyond that they need to be able to take the

ACT; that is an important test. So I think they need test taking skills. Now when you get out that is another question. When was the last time I took a multiple-choice test?

Two teachers disagreed with their colleagues about the importance of explicitly teaching test tasking skills. One said, "Honestly, they don't really need the help. They pretty much figure it out." She was an upper grade teacher and seemed to believe her students had already learned all the necessary test-taking skills. It was not that she did not think students needed the skills; it is just that they already had them so she was not going to take additional time out of her class to review them. The other commented, "There is no set aside time to review. There is no time set aside for test prep. We just teach all the way through and we take the test and we keep going like it is a normal day." This teacher seemed to take the district's directive on no set aside time for test prep and apply it to time taken to teach test-taking skills.

Nonetheless, the rest stated the need to teach test takings skills was magnified in several ways by the new assessments. First, students were assessed on curriculum content they had not been taught since the district had not yet fully converted to the new standards. For example, one teacher observed, "I think the new test was harder for the kids because they didn't have the opportunity to prepare for the new items." For another it was the fear of the unknown, she said, "What are we going to be tested over? What is it going to look like?"

Second, teachers believed students were unfamiliar with the technology used to complete the assessments. The same teacher went on to point how students "didn't know how to do all the enhanced items. They didn't know how to click and drag or where they were supposed to respond." Another teacher agreed, "I can't say the students necessarily knew how to use the computerized tools on the math test so we had to practice some of those things." Another teacher worried about the students' technology abilities,

> Can I actually use the computer, and use the different aspects of the computer to answer the questions, the click and drags and the highlights? Sometimes they can and sometimes they can't. So it has been a learning curve the last couple of years.

Whether they were worried about the content or the technology, most teachers commented that teaching test taking skills was best accomplished by having students take practice tests throughout the year. One teacher justified this practice by noting that all the new items "were on the practice tests, so making sure the students see those tests and making sure they know how to answer the questions correctly because you can't really help them once they are in the actual test." This indicated the primary use of technology within the buildings was test prep, rather than increased learning. For example, one teacher shared her strategy for preparing students for the content via use of the technology,

> We give weekly reading assessments on the computer and I talk a lot about test taking skills like you need to read all the questions before you read the passage, how can you count up answers that you know aren't right. These are the types of things you will see on the state assessments and they are the types of things you can do, like go back and highlight the text or make sure you can find the answer in the text, and things like that.

The teachers almost uniformly believed that teaching test taking skills, like teaching to the test, was a natural part of teaching. It was a necessity for children today because they were going to have tests at all levels of their education. It was a necessity for the teachers because only by taking time to teach test-taking skills could they ensure the students' true knowledge shine through on the test.

Loss of instructional time. In addition to taking time to teach to the test and to teach test-taking skills, the teachers gave several other examples of how the assessments resulted in a

loss of instructional time. The first was assessments themselves and how much instructional time has been converted to giving assessments. One teacher was exasperated by this new reality, "I feel like over time when you add up how much you assess it would be a staggering amount. Wow, I don't even know how many hours or days of instructional time that you lose." He went on to outline the new state assessments include four sessions for math, four sessions for reading, two sessions for writing, and two sessions for either science or social studies. Later, the teacher added, "You can't completely get rid of it, but you have to have the happy medium and I don't think we have it." So even in his complaint he was not ready to say the tests were unnecessary, only that he wished they would take less class time. A second teacher concurred, "Yes, there needs to be a balance and I don't think we have hit that." The teachers also complained that the new assessments were supposed to be shorter but in actuality, they take more time. One teacher explained that even though the individual test sessions are shorter "they are spread out over more days so it makes it seem longer." And when you put them all together it is even longer. Another teacher explained, "In order to get all of the assessments and all of the parts taken we have three, four weeks, where we are testing. That is quite a bit."

The veteran teachers feared a return to the assessment mania that encapsulated the height of NCLB. One teacher reminisced, "We used to do pep rallies and the 5th graders would go and do cheers for the 3rd graders. They would write them letters about how to prepare." To the teachers, these activities were nice, but a waste of instructional time. Another added, "Kids from the lower grades would form two lines and all the 4th and 5th graders would run through. Kids would yell and scream. There were posters all over the school. They really built it up." The teachers believed that all this hype was necessary at the time because the assessments were so important and the stakes were high, but they did not want to return to those days. One teacher

even remembered the parental involvement fondly, “There was more community involvement with assessments. The parents would bring snacks.” But she believed parent involvement did not outweigh the loss of instructional time. A fourth summarized it this way, “I mean the whole school basically felt like it shut down and everybody contributed to the state assessments and supporting those kids and telling them how well they could do and telling them they could do this.” Another added, “Nowadays I try to play the ‘Eye of the Tiger’ and get them dancing, but thank goodness it is not like it used to be.” They all pointed out that they were thankful that they were no longer doing those sort of activities, but, as one teacher put it, “I would like to say those days are gone, but I could see us going back to them.”

Finally, the teachers believed the loss of instructional time was hurting their students’ learning. One teacher said there are units she would like to teach but,

> I just don’t have the time. Sometimes I think, if I had more freedom I would do this unit and I could teach the same things they are asking us to teach and I think I could get more out of my students.

This loss of instructional time was another example of the teachers being resigned to the fact that the assessments drove the instructional time, instead of the other way around. They believe they lack the professional discretion to determine what is best for their students.

Educational Triage

Educational triage occurs when schools and teachers allocate scarce resources to students who are closest to advancing an accountability level, often called bubble students, at the expense of low achieving students who have no hope of passing and high achieving students who are guaranteed to score exemplary (Jennings & Sohn, 2014). As with medical triage, schools make decisions about which students they can most easily help and which students should be addressed

only after all the borderline students are served. In both cases, it is a matter of allocating scarce resources to individuals where the organization can get the greatest impact. Some researchers have studied schools where this was an explicit strategy of the school (Booher-Jennings, 2006; White & Rosenbaum, 2013). However, in most cases educational triage is a by-product of strategies implemented by the school to improve overall test scores (Neal & Schanzenbach, 2010). No teachers in Appraisal Public Schools talked specifically about focusing on bubble students and definitely not at the expense of other students. Teachers did, however, talk about the use of formative assessments, RTI tier time, remediation, and special education as tools the schools have utilized to improve learning.

Formative assessments. As described above, the teachers reported that Appraisal school district has implemented an RTI system based on a series of benchmarking tests, tiered instruction, and progress monitoring in response to the new standards and assessments. In addition, a teacher explained that with the new reading series, "we added weekly assessments as well." These formative assessments are used to identify which students need additional help and which students are on track. One teacher explained the process,

> After we use the screener and progress monitoring to identify the students who need help, we set up tier time to give them extra help. And if they don't improve, they have to do it again. A couple times a week at least.

In other words, it helps the school triage, that is, determine which students are in need of immediate help and which are not. While this is good educational practice and does support student learning, at least one teacher saw a clear connection to the state assessment, "I think it was to kind of help us with thinking about the state assessment."

There were several examples of how the teachers saw the formative assessments as benefiting state assessment scores, without necessarily benefitting overall student learning, because placement was based on how the student did on a weekly test rather than how they did on their classwork. One teacher noted that when a student's placement within the tier structure changed based on a weekly assessment, rather than in response to what the student demonstrated through their classwork, it was not time well spent. She said, "Weekly testing does give us some indication of where the students are, but it is not the only indication. I would say every other week or month would be ok and you could still judge where they are at based on their classwork." She did not believe that weekly reading assessments, coupled with weekly progress monitoring for Tier II and III students, was absolutely necessary for increased learning. Another teacher questioned the validity of the formative assessments,

> You sit a kid down and they read to you but what if they have an off day and they don't meet the target? I had one kid that progress monitored clear above the target the whole time but then he sat down for the assessment and doesn't make the target.

She wondered why one summative assessment outweighed the weekly snapshots of student learning. This teacher felt she knew the student was making gains, and was disturbed that because one assessment indicated failure the student was required to remain in tiered instruction. She also wondered if the weekly progress monitoring was not going to count, then was it worth all the class time taken to administer them. It is interesting that her thought process focused on how to allocate limited time to ensure her students got credit on the summative test for learning, rather than just for learning. A third teacher did not think the formative assessments aligned clearly with the new textbooks so a student could be doing great on their classwork but still end up in tier time because of a weekly assessment that tested a different topic. She explained, "On

the benchmarking test there are questions early on that they don't learn until later in the school year so they already look like they are behind because of those questions." Therefore, students end up in tier time partially because of things they have not learned yet.

However, it was clear that, despite their misgivings about weekly assessments versus classwork, the teachers were following district policy and assigning the students to tier time, even if they questioned the results. This seemed to be another example of the teachers acquiescing to what they were told to do because it was going to help the students on the test. They knew the educational triage was in place to help the students do well on the assessment and they wanted the students to do well on the assessments, so they were going to let the system win out over their professional judgment.

Tier time. Despite the reservations of some teachers, the district utilizes these benchmarking tests and progress monitoring to tier instruction so they can focus on students who need the most help. One teacher reported,

> We benchmark three times a year and we sort the data according to kids. We then further test the kids who are below benchmark and decide where they need to be…. Once we get them into the tiers, we have a pretty step-by-step protocol.

Another teacher added, "This protocol helps prepare them for the state assessments." As with the formative assessments, the instruction during tier time is educationally sound and helpful to students, but most teachers saw it as built around preparing them for the state assessments. They also acknowledged that implementing the prescribed protocol faithfully was difficult with their limited amount of time. One teacher commented, "When we do have time for small [Tier III] groups we get the kids who are really struggling, which is hard to do every day." This difficulty

was exacerbated by the fact they were not sold on the fact that preparing for the state assessments should always be their end goal.

One 5th grade math teacher did not agree with her colleagues, she believed the RTI protocol really did focus on the students' needs. She said,

> I don't know if I feel that the screener and tier time has a super correlation to the state assessments as much as it has to do with overall learning and hitting the skills when we have a struggling student. It helps fill the holes for them.

This comment did come from a teacher who only taught math and had just begun the RTI protocol with her students earlier in the year and therefore had not yet seen the impact of tier time on the state assessment results. Her fellow teachers conveyed they wished RTI was solely for increased student learning but felt the fact the beginnings of RTI were a reaction to low state assessment scores belied that.

Other teachers assumed tier time was designed to prepare students to pass the state assessments and were, in fact, concerned that it was not. One teacher explained, "The growth we believe our students are having on a daily basis during tier time isn't even monitored on our benchmarking test because the benchmarking test doesn't align with the state assessment." She did not believe they should even be utilizing tier time if it was not going to help them on the assessments. She was not questioning the validity of tier time based on its impact on student learning, but only on its impact on the state assessments. Teachers were frustrated by the fact that to many, successful student learning is synonymous with a successful outcome on the assessment. They do not believe that one assessment should be the only determinant. One teacher put it this way, "I think that is when teacher judgment comes into play….If you feel like

a student didn't perform as well on that assessment as what they would do in the classroom that is also taken into consideration." But too often, it is not.

Remediation. Research on educational triage indicates that many schools use retention, mandatory after school remediation, or mandatory summer school as part of their protocols to help all students succeed on the high stakes assessments (Booher-Jennings, 2006). This was not the case at Appraisal school district. While they do offer summer school, teachers indicated the school district did not require remediation based solely on failure to pass the state assessments. One teacher explained, "I know one of the qualifications for going to summer school was state assessments….but it wasn't like if you didn't pass you had to go." Several teachers did mention extra tutoring for students who did not pass the assessment was an option, but not required. One teacher said, "I was paid to tutor after school those kids who were struggling with the old state assessment. I would watch and monitor them and we would spend three or four days a week after school but they are not pushing that anymore." A second teacher reported, "I spend a lot of time working with my lower kids on fact fluency after school because they are just not fluent in their multiplication facts, or even their addition and subtraction facts but it is not required." Therefore, Appraisal school district did not demonstrate all aspects of the educational triage found in many other schools, however, this may be because they had not consistently struggled to meet AYP under No Child Left Behind. Many of the schools that took the next step in educational triage by mandating remediation and/or retention did so because they were a chronically failing school. None of the schools in Appraisal were chronically failing. Although there were years when they did not pass the assessment in one subject or another it was not persistent.

Special education. The final piece of educational triage is the use of special education. Historically, the way in which schools dealt with special education students was often a clear indication of whether they were practicing educational triage to improve their scores on the test or practicing it to improve overall learning. That is because some schools would move students to special education so they would not be counted in their results (Nichols & Berliner, 2007). To mitigate this problem, the new state assessments allow only the lowest one percent of students statewide to take an alternate assessment. All other students must now take the regular state assessments. Several teachers indicated it was unrealistic to think special education students would be able to pass the regular state assessments. One teacher explained, "They took out that middle modified assessment. There is now only the general assessment or the alternate assessment. For students that is an extreme gap." A second teacher agreed, "It seems like if they have an Individual Education Plan (IEP) they should have a different test. They shouldn't be expected to take the regular test."

This change in the state assessments did change how the schools served special education students. One teacher exclaimed, "The new assessments have made a difference in how we prepare special education students for the test." Because the district's goal was to do well on the assessments, they changed how they serviced special education students. A teacher provided detail, "We have moved to more inclusion and less resource room time so that those kids are getting the same core instruction that their peers are getting as opposed to a totally different curriculum." This is exactly what the law was designed to do and the teachers did not seem concerned with this change in service because they seemed to be equating success on the assessments with success in overall learning. Overall, this is a benefit to the special education student because they are not being marginalized to avoid the testing and they are receiving better

instruction. If they are doing better on the assessments, then obviously it is a positive change for the student.

However, as soon as the changes do not benefit the student on the assessments, the teachers become frustrated. One teacher lamented the lack of accommodations available to special education students on the test. She said,

> I gave my student a multiplication table, a simple tool, and his scores shot up. He knows how to use the tool. He doesn't have the multiplication tables memorized but he knows how to use it well enough. He is now getting 100% on his assignments. I am thinking, "How am I going to help this kid?" Because when he gets to the assessment he is not going to be able to use that tool. It is hard.

So the teachers are fine with changes to the services as long as it increases assessment scores, but they are uncomfortable with changes that do not result in improved scores. This indicates teachers are resigned to the fact that assessments are driving their classroom and school and not the other way around.

Pressure

The utilization of instructional and educational triage in districts and schools is a direct result of the pressure administration and teachers feel to make sure their students succeed on high stake assessments (Lauen & Gaddis, 2012a). While state sanctions for schools who fail to meet AYP was always a concern for the lowest achieving schools, for most schools the true pressure comes first from the publication of scores by the media and second by parents threatening to move schools based on the results (Nichols et al., 2012). This pressure then moves to the teachers who often feel that their job depends on how their students perform on this one assessment (Reback et al., 2011). Finally, the pressure is passed on to the students which

often inhibit their academic success (von der Embse & Hasson, 2012). Some of this pressure has decreased in the last several years because of the transition to new assessments and the fact that the Race To The Top (RTTT) and ESEA waivers have eliminated AYP in many states, but teachers clearly still remember the pressure on the old assessments and most anticipate a return to that pressure as the new tests get fully implemented.

Public scrutiny. Like most media outlets across the nation, the Appraisal newspaper regularly published the state assessment scores from the local schools. One teacher commented that the paper often compared their scores to nearby school districts. She said, “Everyone could find our school’s scores and scores for schools in the surrounding area. They were all in the paper. There are some parents that don’t pay any attention to it, but most do.” Even though last year’s test was supposed to be a baseline year and many newspapers around the state chose not to publish the scores, the paper in Appraisal did. One teacher revealed, “I was surprised when they published the scores from last year’s test. I thought it was way too early to do that.” The teachers were appalled, but not surprised, that the local newspaper would report baseline scores.

In general, the teachers agreed that parents were influenced by the state assessment scores being published in the newspaper. One teacher reported, “I have heard parents talking about scores. More in the past, but maybe that is because I don’t have kids in school any more. People would judge schools based on test scores.” A second teacher concurred, “Parents would see the scores in the paper and complain about different schools. It was unfortunate." But the teachers accepted this scrutiny was an inevitable result of NCLB.

Despite this acceptance, the teachers were offended that people would use those scores to determine the quality of a school. One teacher was frustrated by the parents’ choices. She said,

> I have friends who would decide where to live based on test results. I would try to explain that those tests are rigged, maybe that's not the right word, but maybe test scores should not be the only deciding factor, but they didn't understand.

She was implying the tests were not a true reflection of the students' ability or the quality of the school. Another was exasperated because parents did not understand how a school district worked and why one test score should not tell them anything about the quality of an individual school. She stated,

> We have five elementary schools and you would hear, "You don't want to send your kids to that school because they didn't do as well on the state assessment." Even though they were the exact same teachers with the exact same curriculum and we all lesson planned together and got together for collaboration meetings, it was just very political. The parents thought they knew which schools had done what, but here we are all working in the same district, teaching the same things.

The teachers were also troubled by the fact that parents did not understand the role socio-economic status played on test scores. Teachers reasoned if parents understood how far behind some of these students were when they came to school, they might not be so quick to judge the quality of the school. One teacher pointed out, "You can't start a kid off at a third grade level and expect them to be at a fifth grade level within a few months. It just doesn't happen that way." If parents looked at the quality of the teacher and how much hard work they put into their jobs, it would be a better indicator of the quality of the school. One teacher spoke of her time working in a high poverty school. She said,

> I know for a fact how hard those teachers worked. They worked just as hard as any teacher in a high achieving school did. They were just as good, if not better, than any

> teacher. They just had a lot more issues to deal with. Most of those kids made tremendous growth from the beginning of the year to the end of the year, so it wasn't fair that parents didn't think the school was good.

A second teacher was equally frustrated that newspapers never took into account how poverty impacted test scores. She commented,

> Ideally, you would take the low socio-economic group and spread them out evenly over all the schools, but it just doesn't always work out that way. So those kids at some schools are at a lower socio-economical level and have less privilege so they are going to struggle a bit more and their scores are going to be a little lower. But they started at a lower level. It doesn't mean they are not making significant gains. They definitely are, but it looks different in the newspaper.

Rather than judging a school based on the one score the newspaper decided to publish, the teachers wished that parents would look at the individual gains of students, where they started from, and where they are now. They felt it was disturbing that the media and the parents were using the scores for the wrong reason. They were intended to measure student progress, not to compare schools, but that seems to be one of their primary uses.

Teachers were equally troubled by the fact that the increased public scrutiny only made their already difficult job even harder and turned the focus away from students. One teacher from a low socio-economic school complained, "Here we are working with the hardest kids and there were always people checking in to see if we were teaching what we were supposed to be teaching because our kids weren't doing as well on one test." They felt they were not being treated as professionals based on the result of one test, rather than the scope of their work. But even more so, they were offended that these tests were taking the emphasis away from the

children. One teacher reported, "It becomes about how my school looks compared to that school across town. It becomes about how this district looks compared to another district of the same size. It should be about the kids sitting in my classroom." The teachers believe the tests are a distraction that took away from what was best for kids, but at the same time, teachers know they cannot ignore the test. One teacher explained the dilemma this way,

> I think the emphasis has switched from the students to everybody else. I think we need to get that emphasis back on the students, what is best for the students. We need to think about how the assessments are going to help them, rather than how is the school doing. However, this is my job so I have to get these students to achieve because I want to keep my job.

The increased public scrutiny brought on by the newspapers and parental pressure has had a significant impact on teachers.

Pressure on teachers. As increasing amount of pressure is put on districts and schools to perform well on high stakes assessments, the next logical step is teachers. President Obama's Race to the Top education initiative required that states receiving waivers from NCLB include student growth measures (SGMs) as part of their teacher evaluations. Since many states are transitioning to new assessments, multiple SGMs were allowed under the law. The Midwestern state where Appraisal school district is located mandated each district develop a teacher evaluation tool where three student growth measures had to comprise at least 20% of the overall evaluation. This immediately puts pressure on the teachers. One teacher said, "I struggle with the fact that my estimation as a teacher is based on someone else's achievement, not my own. I know it is only 20% of the whole thing, but that is still a lot."

Teachers were concerned the state assessments were not a good way to measure teacher effectiveness because one test score does not always capture how much a student has learned. One teacher worried,

> We give them the test and then all we can do is hope that they are in a good mood. Hope that their mom didn't make them wear something they didn't want to wear and it's going to put them in a bad mood. Now they are not going to cooperate with anyone. That is our fear. I sure hope everybody just takes a step back and sees this for what it is, something out of our control.

Teachers were frustrated that their performance evaluation was being at least partially based on things outside their control. One teacher said, "You are doing your best but a certain amount is out of your control. It is on their family situation or other things. The students are doing their best, but it is hard when it is an evaluation year." The teachers would rather see themselves judged on their effort and ability, rather than what one test result indicates. Another teacher agreed, "Anytime they tell you they are going to base your scores on a kid where 50% of the time you have no control over, it is stressful." The teachers did not see this as fair or as treating them as professionals.

The idea behind SGMs is increasing the accountability on teachers will result in increased student learning. Teachers see this as the government assuming they will not do a good job unless they are held accountable or forced to do a good job. However, one teacher commented this attitude is really counterproductive, "For me it is more of a distraction then a good form of accountability." They see it as being forced to jump through hoops particularly when they are being evaluated on assessments that are not yet finalized. One teacher told a story about one of the best teachers in the building. She said, "One of our guru teachers was in her

retirement year but she could only be marked at the lowest level because without state assessment scores they said she couldn't demonstrate how good she was." The teachers considered this to be ridiculous. If there were no state assessment scores available, that should not be held against the teacher. Unfortunately, the policy stated that unless teachers could demonstrate three successful growth measures they could not be ranked highly. Another teacher shared a similar story.

> When I was evaluated last year, one of my SGMs was the state assessment and since there was no data yet, I did not think I should be rated down but I was. Even though in all my observations and walk-throughs I had excellent ratings and remarks, the highest rating I could get as a teacher was "proficient" because my student growth measures weren't ready yet.

The policy was enacted before all the pieces were complete. The teachers saw this as completely unjust but realized it was a federal requirement, enforced by the state, so it was really out of their principal's hands. As a result, they were not necessarily upset with their school or district. Rather, they saw this as another disconnect between policy and the classroom. However, another teacher spoke about a situation that was within the school's control. This issue was being judged based on students who are not even in her classroom. She reported,

> We had some transferring of students that had not been identified correctly so they needed to transfer to a different class where they could get the support they needed. But they are still showing up on my data and I am trying to get them off my data so that my SGMs are accurate because that will play into my evaluation.

Even though the teachers understood that the school was just following federal and state policy, when it began to negatively impact their evaluation, they became concerned. Another teacher

provided a different example. She said, “They say it’s not a big deal but parents can go and see that their child is in a class where the teacher has a basic rating, when that doesn’t really reflect on the type of teacher you are.” Overall, this has led the teachers to being resigned to their fate. Just like with preparing for the tests, they seemed to be saying, “Just tell me what you want me to do and I will do it.” To the laughter of her co-workers, one teacher proclaimed, “My opinion is I am doing the best I can. I don’t think I am the worst teacher in the state, so if you think all my SGMs are terrible then fire all of us.”

While most of the teachers focused on the state assessments, some of the teachers did see the pressure lessening because there are two other SGMs in addition to the state assessment and those SGMs can use other means to demonstrate student growth. One teacher commented, “Even though the test is more rigorous I don’t feel the pressure I felt with NCLB to get a perfect score. I don’t think we are being told every single kid has to score well. We are looking for growth.” Another teacher concurred, “It is nice that there are options. It is not just the state assessment. So if something happens on those scores, you have other things you can talk about, which I think helps spread it out a little bit.” The teachers are still feeling the pressure and would prefer to have no SGMs but they are happy there is some latitude. They are also enjoying the fact the tests and SGMs are in transition. One teacher said, “Maybe I should be afraid, maybe I should be more concerned but for now I am not even sure what we are supposed to be using so how can I be worried?” Another teacher agreed, “Right now, it is a little stressful and it makes me anxious but all I can do is give the test. You should probably ask me in a few months as we are watching the kids take the test.” This comment was met with laughter because all the teachers anticipated the pressure returning.

The teachers acknowledged the reprieve from the extreme pressure of NCLB was probably going to be short lived, especially if their test scores impacted their evaluation and their evaluation impacted their pay. One teacher said that district leaders were downplaying the significance of the SGMs and telling teachers just to do their best, which she appreciated hearing but, in the end, whatever the district did with SGMs would not matter because, "Once they start basing our pay on it, the pressure will definitely come back." Another teacher was a union rep and she commented this might have ramifications across the district. She reported,

> Our principals have been told that they are ranking their teachers too high and they need to lower their rankings because that is how the new evaluation tool is supposed to be used. As the union representative, I have taken that information from the teachers and I have shared that with the union. They are going to share that with the superintendent because that is a major concern. Why are we being rated so low if we are doing the job we are supposed to be doing?

Again, there was a general resignation to the fact the pressure was going to return. One veteran teacher summed it up this way, "I sure hope the pressure doesn't come back but I know it is just the trickle down. The State puts pressure on the superintendent. The superintendent puts pressure on the principal and then it trickles down to the teachers."

The only reprieve from this ongoing pressure was trust in the administrator. In one school, the teachers trusted the principal understood one test score did not define the teacher and that much of the results were outside their control. One teacher said, "We are blessed that we have supportive, understanding administrators. I can't imagine what it would be like with a more negative principal." At this school, the teachers were significantly less concerned about SGMs and their use on the evaluation than teachers at other schools were. They did not feel

blamed or shamed if their scores were not where they wanted them to be. For example, one teacher said,

> When our scores don't look so good they sit down with us and ask why. It is not us; it is just something that needs to be fixed. Maybe it is the student or something else. The pressure has been taken off. You are not a bad teacher. It is just something that needs to be fixed. Our administrators make a big difference. We are lucky.

Another teacher commented the principal understood that growth was more important than an arbitrary cut score. She explained,

> Our principal understands that in 4th grade the students usually have the same issues that they had in 3rd grade. You can raise their fluency by 30 points, but they still might be 20 points from the target. The SGM target is important and we aim for them, but they aren't super important. Our principal is pretty understanding about that.

The teachers at this school were in agreement that the stress a teacher feels on SGMs was directly related to how much they trusted their principal. If the teachers trust their principal then they are confident that the principal knows they are a good teacher even if they do not hit all their targets in a given year. If the teachers do not trust their principal, they are significantly more nervous about the evaluation process when their test scores have not hit the targets.

Student pressure. Teachers believed student feel pressure from several sources. Their parents can convey the seriousness of the test for their future. One teacher reported, "I will never forget when a student came in and said, 'My mom said she would buy me a digital camera if I pass both state assessments.' It was amazing what the parents would say." Or their principal can impress on them the importance of not letting down the school or their classmates. Finally, teachers can inadvertently, or overtly, pass the pressure they feel onto their students. One

teacher explained, “We don’t mean to put that pressure on the students. Just like our principals don’t mean to put that pressure on us. But we all feel it and it trickles down.” Regardless, the teachers in Appraisal school district recalled the pressure the students felt under NCLB, were relieved that the students were not currently feeling that pressure, and were convinced that the pressure on students would inevitably return.

All of the veteran teachers who taught during the height of NCLB vividly remembered the inordinate amount of pressure the students were under and what that did to them. One teacher recalled, “We had kids in tears, especially the kids that couldn’t do it. They would cry for hours because of these tests.” Another teacher said, “The old way we tested put so much pressure on the kids. I had students who got physically sick.” Still another reported, “Kids were saying, ‘Oh my goodness, I have to take the test today? I can’t sleep. I can’t function.” These comments went across schools and across grade levels. There did not seem to be any exceptions. One teacher spoke as a parent, “My own kids would come home really stressed out about the test. I mean really, really worried they were going to flunk if they didn’t do well on the test. I don’t want kids to feel like that.” Several other teachers pointed out that it was not just the struggling students who were afraid of failing. It was the high achieving students as well. One said, “It is not just the pressure of making sure they passed the state assessments. It was also scoring where they needed to score.” Another teacher concurred, “It is really hard for them to face that pressure without shutting down.”

Some teachers remembered fondly the stress free days before NCLB and compared it to the height of NCLB. One teacher recalled, “I remember kids saw tests as an opportunity to do well but there wasn’t this crazy pressure to perform. I think the way they saw it was a reflection

of the way we presented it to them." They accepted their part in the previous hysteria and did not want to be a part of it again. Another teacher lamented,

> Ten years from now, no one is going to say to our kids, "I noticed in 3rd grade you were just proficient in reading." But that is how we are treating it. We are treating these tests like they are a big deal - even then parents started getting worried about it. They are coming up to school making sure their kids are ready. In the old days, I just don't remember parents asking me when I was administering the California Achievement Test (CAT). People cared, but it wasn't at this level of hysteria.

This comment highlights the wide spread nature of this hysteria. This teacher had come from California and she saw the same reaction in both states. Another teacher focused on the impact of public pressure, "I don't remember the newspaper publishing your math scores. The scores weren't paraded down the hallway. There were no banners hung in the lunchroom. It just got out of hand."

The teachers relished the fact that there has been less pressure on the students the last few years. One teacher reported,

> The third and fourth graders have never taken a high stakes test. They have no other measuring stick. To them it has never been a big deal. Even the older kids who are old enough to remember the pep rallies, even for them the assessments are no big deal. And that is fine with me.

Children of all ages are embracing the decreased pressure. Another teacher commented, "Last year's class didn't seem to put any pressure on themselves, which, in a way worried me, because a little bit of pressure is good." Even when the test was a pilot, this particular teacher could not bring herself to relax completely. She still wanted her kids to take it seriously because it

mattered to her. The teachers all agreed that the pressure would return, as one put it, "I think that for some of the kids it will be stressful because they want to do well. It will be interesting to see what happens when we have been giving the same test for a couple of years."

Eliminating High Stakes Testing

The last question in every interview was if you could get rid of the state assessments, would you get rid of them, keep them, or modify them. There was not a single teacher who advocated for keeping it. A few found redeeming qualities, including accountability and data-driven instruction, but the vast majority were in favor of getting rid of high stakes assessments altogether. Most of the teachers' responses fell into one of two categories. First, for a variety of reasons, teachers did not believe that the state assessments were improving instruction. Second, they did not believe that state assessments were a good indication of what their students had learned.

Not improving instruction. The overall consensus among the teachers was the state assessments were not necessary because they did not improve classroom instruction. The teachers were fairly unequivocal in their responses. One teacher said, "I would get rid of state assessments. I think I can show growth in other ways. It doesn't really help me and it takes a week of my teaching away." A second teacher simply stated, "I have always been opposed to state assessments." A third said, "I would rather not give them. Just let me teach more." A fourth teacher said bluntly,

> I would rather not give state assessments for the simple fact that it is not really useful for teaching my kids. That is my job, to teach these kids and help them learn, and I don't feel like the assessment necessarily helps me do that.

The teachers were not opposed because it was difficult for them or made their job harder, they just did not see a benefit to education of their students.

There were a variety of other reasons that teachers were opposed to continuing high stakes testing. A veteran teacher worried about the cyclical nature of education. She said, "You don't run a business the way education is run because we are constantly reinventing the wheel. We have a tendency to swing way to the right and then way to the left, and we are never in the middle." A second teacher concurred with a laugh, "If education was a business, we would have gone bankrupt a long time ago." Another veteran teacher highlighted the fact that the public seemed to care more about assessment results than the schools. She said,

> I can't think of the last time someone discussed the state assessment results with me and how we did. I know what my information is if I look it up. However, I have never had a principal come to me and say your kids did this and this and this is what it shows us about their learning. But everyone else seems to know how our building does and how our district does. It is political, but we get no real information from the state assessments.

She realized public scrutiny was part of the intent of NCLB but that scrutiny was not impacting her on a daily basis because, in her opinion, it was all political. There was further clarification from another teacher who indicated the public scrutiny was limited to one segment of the population. She said, "A lot of it is socio-economic. It is the higher socio-economic parents who are more concerned about the test scores [when picking a school]. The lower socio-economic parents care less about scores." She believed that lower socio-economic families worked with their neighborhood schools while high socio-economic parents were moving around town and building new houses based on the proximity to schools who did well on the state assessment.

Several teachers believed tracking students' assessment results over time would be more beneficial to their classroom instruction. For example, a teacher said,

> I wish that the assessments would follow the students more but I feel so much pressure is put on a teacher for one year and then the next year you get a new set of students. I wish it would follow students throughout rather than one year on this particular teacher. I wish it had that kind of focus.

A second teacher agreed. "I prefer benchmark testing which can be nationally normed or normed by the district. I think that would give me enough of a sense of how my students are doing, rather than at the state level."

Finally, teachers were critical of the idea that teachers would work harder if they were under pressure. One teacher commented, "Those that care, care. Those that don't, don't. And we care." She did not believe they needed a test to make them try harder. Another teacher summed it up this way,

> Whereas now, because I feel all this pressure, "Let's do it for the school. The principal wants you to do your best." We really went above and beyond here, and maybe other schools were doing the same insane, crazy things, but it became this production with pep rallies, parades down the hall, snacks every day, little cards, reminders, and notes. I just thought, "Well, no wonder these kids have anxiety." If they are paying even a little attention to what is going on around them, they don't see it as a chance to do really well and show what they have learned. They see it as they have to preform, like they are on a stage and everybody is expecting something.

Overall, the teachers just felt the tests took up instructional time and caused undue pressure. This was time and energy that could better be used improving instructional time.

Not a good indication of what the students know. The second category of responses as to why teachers would not keep state assessments was the tests were not a good indication of what the students had learned. As one teacher put it, "Most teachers understand where their kids are without having to go through all that. Obviously if they need proof positive, that is one way to give it to them, but it is not that hard to show student growth." A second teacher said it even more succinctly, "I do not think state assessments are a valid picture of what our kids know." A third teacher simply said, "It is not a good indication of what they know." One reason given for the lack of validity is that state assessments represent only one moment in time. A teacher described it this way,

> I am not sure that one test a kids takes on one day throughout a year is ever a true reflection of what they know. It depends on the day, if the student is having a bad day or maybe they didn't take their medication, or they woke up grumpy, or they just don't feel good. It seems like we put a lot of stake in these state tests but everyone has had a bad day before or some kids just don't test well.

Another teacher commented, "Test a student at a different time of day and their score might be different."

Another common complaint was the test is not accurate for all types of students. One teacher commented they were a waste of time for all the above average students. She said, "If they come in here and they are already in the well above average range, do I really need to keep testing them? Do I really need that data to say they are above average?" Another teacher complained the new tests were too difficult for the average student. She said, "I just wish they would write the tests like they wanted kids to succeed. Sometimes it seems they don't even want them to get the questions right, unless the kid is gifted. Why throw the tricky questions in

there?" Still another teacher believed the assessments would be more beneficial if special education students were exempt.

> It is hard for me to give the test to kids who I make accommodations for all year long. I feel we are setting them up for failure. I am trying to help them succeed throughout the year and I know when it comes to the day of the test, they are just going to get slammed. It breaks your heart watching them struggle through it. So if I had the choice I would say special education students do not have to take it.

Finally, teachers disagreed with the assumption that passing a test was the same as student success. One teacher commented,

> I don't feel like it is fair to say that this test in fourth grade is going to tell you whether you are college ready or not. Lots of things can happen between now and then. Lots of successful people are bad test takers but very successful people in the real world. They made it through college just fine.

A second teacher put it this way, "We need the curriculum. We need exemplars. We need guidance on what needs to be done in the classroom." That is what teachers believe impacts success. Not one test score from one week out of the year.

Overwhelmingly the teachers had no desire to keep the state assessments as they were. There were two ways they felt high stakes testing had improved their classroom instruction. Two ways they were willing to possibly keep state assessments if they could modify them around these two improvements. The first reoccurring theme was that testing increased the accountability for teaching the standards. The second theme was that it made their classroom instruction more data driven.

Accountability for the standards. The movement towards standards based teaching preceded the No Child Left Behind Act, but the adequate yearly progress mandate of NCLB forced teachers to increase their focus on the standards, if only the ones that appeared on the test. Even when asked about an ideal world where there were no high stakes tests, these teachers did not back away from the standards. One teacher commented, "I don't want to see so much pressure on the test but I think it is important that we teach to the standards. I appreciate the rigor of the standards." Another teacher agreed,

> We need to have the standards because we need to know where we are going. I think whether we had the state assessments or not, we would still be using the standards and trying to teach what has been asked of us. I don't know that there would be as much pressure, that is for sure, but we would still be teaching the standards.

Clearly, the teachers have accepted standards based education and believe it has had a positive impact on their classroom instruction. One teacher stated, "We feel like if we align what we teach to the standards, and the standards are aligned to the test, then our students will be okay. We just try to focus on the standards and not the test." They just wished the pressure from the tests would decrease. A veteran teacher commented, "I think we need to teach the standards but I think that the pressure just brings up so much concern for the teachers. I agree you need the information but at the same time you need to be able to teach the standards."

Even while they were criticizing the high stakes tests, they were careful not to criticize the standards. One teacher bemoaned the tests but then said,

> I don't know that I feel that way about the math standards necessarily. In fact, I kind of like what we are doing in math now better than the old standards. There is certainly more rigor and writing, but that is not necessarily a bad thing.

Another teacher added she would not mind the tests if they matched the standards, "If they would come talk to us teachers maybe we could give them some guidance on how to make the test more effective and actually measure the standards we are teaching. But nobody seems to bother to ask us." A third teacher put it this way, "It is nice to have a test at the end of the year to know what they have learned all through the year, but it has to match what we are supposed to be teaching. So I would keep the assessment, but only if it really matches the standards."

Likewise, the teachers seemed to have embraced the concept of accountability. They did not like the high stakes nature of the tests, and did not necessarily like the tests as written, but they did not seem to mind being held accountable for student success. One teacher said, "I don't like the tests but I would worry about getting rid of them completely because I think they do hold us accountable as teachers. They give us a goal." A second teacher simply stated, "I do think it is good that we have some accountability. That is why the tests don't really stress me out. I just try to do the best job I can." Therefore, it was not accountability, but the tests themselves that the teachers did not like. One veteran teacher commented, "I think there has to be other options. I don't know what the right answer is, but I don't think it is the state assessment. I do think we need accountability but not the state assessments." Another teacher concurred, "I still think there are benefits for having some sort of assessment system, just not the one that is being used now."

Data-driven instruction. The second result of high stakes testing the teachers seemed to embrace was data driven instruction. No Child Left Behind forced schools away from class averages and overall impressions and toward having data that demonstrates all children are learning. Here again the teachers indicated they wanted data, and they knew they needed assessments to get the data, but they did not think high stakes assessments were the way to go about it.

One set of teachers focused on the benefits of getting data in general. They acknowledged assessments have always been a necessary part of education. These teachers clearly saw the benefits but wished they could change various aspects of the current system. One teacher focused on the amount of time the current assessments take. She stated,

> I am not opposed to testing. I could see some benefits if we modified some things so that they were not so time consuming. But I think there is something to be said for seeing what kids have mastered and how prepared they are for the next grade. I just don't like how the system is set up now.

Another teacher focused on when they got the data. She said, "We have to have data. We have no concerns with that. But we just don't get a lot of valuable information from the state test because we don't get the data until so late in the year." A third teacher focused on the pressure. She said, "It is nice see the numbers go up and say they are improving but the pressure on the state test is just such a distraction." It was evident that the teachers were not opposed to assessments or accountability, just high stake assessments.

Another group of teachers focused on the need for specific data to drive their instruction. They preferred a pre and posttest model, rather than the once a year state assessment. One teacher simply said, "We have to have data, but we need to pre and posttest so we know what the kids need to learn." Another teacher was more critical of the current state assessment process. She said,

> I would like to see the students take a test at the beginning of the year and then at the end. I always find it ironic that the state calls it a growth measure when it is only one point in the year. How is that called growth when there is no comparison? Now it can't take

eight days at the beginning of the year and eight days at the end. My fantasy would be a short pre and posttest.

In addition, several teachers were upset the state assessments did not provide them with enough specific data. One teacher stated, "What we need is some actual feedback about how they did. Not just a score, but a breakdown about specific skills." A second teacher said, "If you could use the data to find the holes and make sure we are hitting all the standards, then I don't think the data hurts." A third teacher lamented, "The data I got from the state assessments didn't help me know how my class did on specific skills. It really wasn't even broken down in the way it used be." After years of testing they want the specific data to help their students succeed. While they are happy to get rid of the tests, they do not want to get rid of the data.

Final thoughts. Throughout the interviews, three quotes stood out to me. Each one encapsulated much of what all the teachers were saying. The first focuses on the fact that high stakes tests should not be the end all of education. A relatively young teacher who just wanted the best for all her students made it. She said,

> I am glad that all the craziness with the pep rallies is slowing down. I have had fewer parents asking about the state assessments. I think this good because they are in a position where they can't really control how their kid does on the test. They can help their kid read every night. They can make sure their kid does their homework, gets to school on time, and is fed. But if all they are going to do is focus on that one test score and that is their only gauge that my kid was a great third grader, then I think that is sad. It is sad because there are so many other victories that third graders have. They are here from August to May and there are so many other things third graders do. That four day

> period where we put them in front of a screen, cross our fingers, and hope for the best should not be the only focus.

The veteran teachers focused on the pendulum of education. They realize that while policy and programs may change around them, the children do not change. She said,

> I want to believe that all the craziness over testing is over and that we will just begin to see it as a tool to measure achievement and observe patterns of learning. I want to believe that, but I don't know. Over the years, I have seen so much ebb and flow on issues and assessments. I have been a teacher for over 20 years but I feel like I have changed jobs seven or eight times. I don't ever feel like I can get too comfortable. The one thing that is consistent for me is that the kids show up every day. In 23 years, that is the one thing that is constant, the kids show up and I just do my job.

The final quote summarizes the feelings of all teachers, on many topics. This was from a teacher was neither inexperienced or a seasoned veteran, but from a teacher in the prime of their career. She said,

> That is the problem with the politicians and in education today. They don't live it. They don't see it. They don't know that the decisions they are making are actually taking the attention away from the students and putting it on something else. That is what I struggle with. I love my job. I love teaching. I love coming to school every day and hanging out with these kids and just learning together. If we were able to take the emphasis off everybody else and put it back on the students I think that would be to their benefit. I don't care about me. I just want what is best for them and I think a lot of teachers would agree.

These three quotes encapsulate the passion the teachers demonstrated throughout the interviews. To a person, the teachers in Appraisal school district loved their students, enjoyed their job, respected their principal, and were loyal to their school district. The concerns they expressed about high stakes testing were not out a selfish desire to work less or a fear that they were going to get fired. Their concerns were expressed out of a genuine desire to help their students and renew respect to their profession.

CHAPTER 5

Conclusions and Implications

The findings presented in chapter 4 provided insight into teachers' perceptions of the influence of high stakes testing on classroom instruction. This chapter presents the conclusions and implications derived from those findings. First, I revisit Campbell's law and use it to help explain the conclusions I drew from my findings. Second, I offer implications for policy makers and researchers.

Campbell's Law

In 1975, Donald Campbell summarized his views on the corrupting effect of quantitative indicators on social programs. He claimed, "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor" (Campbell, 1975, p.34). His contemporaries began referring to this as Campbell's Law.

Campbell was not the only one warning of the deleterious effects of quantitative indicators. In physics, Heisenberg's uncertainty principle stated that trying to measure a microscopic object will alter the conditions surrounding that object and therefore render the measurement inaccurate or at least uncertain (Madaus & Clarke, 2001). English banker Charles Goodhart (1975) postulated when a measure becomes a target, it ceases to be a good measure. Whenever a government attempts to regulate financial assets they become unreliable as indicators of economic success. Similarly in economics, Robert Lucas Jr (1976) argued if economists only rely on historical data to justify their decisions the outcome could be corrupted by that very data. He said just because a bank has never been robbed does not mean it is possible

to safely get rid of all the guards. The guards are the reason the bank has never been robbed, not an indication they are no longer needed.

In education, high stakes tests were created to measure the effectiveness of schools. However, schools that raise their test scores receive praise as good schools without much, if any, scrutiny into the details about how they crossed the accountability threshold (Nichols & Berliner, 2007). Success is its own measure of success. However, because schools operate with finite resources, they have to strategically weigh how they will utilize these resources (Koretz, 2011). It is logical in this era of accountability for schools to concentrate their protocols and procedures around doing well on the test and focusing their resources on students on the verge of passing (Campbell, 1975). This results in educational and instructional triage where students deemed most likely to survive the accountability process receive treatment first.

In medical terms, triage occurs during emergencies where the number of victims overwhelms the ability of the medical professionals to treat all the patients with their normal standard of care. The emergency dictates that doctors make informed decisions about which patients to treat. They do a quick assessment of each patient and determine if they are healthy enough to survive, wounded but treatable, or wounded but would require too much time and effort to save. The slightly wounded and mortally wounded patients who would normally get treatment are pushed aside while the doctors focus on the ones they are most likely to save. As applied to education, the pressure of high stakes testing creates an emergency where teachers focus on the average students to the exclusion of students who are either mortally ill (i.e., low achievers) or healthy enough to survive on their own (high achievers). The idea of educational triage runs counter to the ideology of schooling that purports to help all students because all students are considered valuable to society. It also runs counter to the ideology that schools are

the great equalizer that help the poorest of the poor achieve the American dream (Labaree, 2012). Educational and instructional triage suggests that some students have more worth than other students. Triage on the battlefield is one thing. So is triage in a hospital emergency room after a disaster. But it is another thing entirely when it is used in a school.

Campbell's Law can be used to explain why educational and instructional triage is prevalent in schools but is often not done with malice. Schools and teachers only want the best for their students. However, Campbell's law helps explain how their utilization of educational and instructional triage is inevitable. It is logical for institutions and people to do whatever is necessary to meet the standards set out for them. Campbell's law can be used to explain the following five conclusions derived from my research.

Assessments Drive Classroom Instruction and School Procedures

Campbell's law helps explain why educational and instructional triage might be a natural byproduct of high stakes testing. Throughout the interviews the teachers tended to focus on the limitations and obstacles presented by the high stakes tests. The teachers also spoke at length about how high stakes tests impacted the structure of the school day and the instruction in their classroom. They did not particularly characterize these impacts as negative, but rather as inevitable. The veteran teachers saw high stakes tests as the new normal, and the teachers with less experience have never known any other way. The more emphasis put on the results of the assessments, the more schools and teachers will allow the assessments to drive classroom instruction and school procedures, not the other way around.

Research indicates the high stakes testing has a significant negative impact on teacher morale (Cawelti, 2006; Willis, 2011). This is mainly the result of public pressure created by the publication of results by the media and the use of results on teacher evaluations. Research also

indicates a negative impact on student morale as a result of importance placed on these tests for the future of the student (Dutro & Selland, 2012; Cornell, Krosnick, & Chang, 2006). As a result, instead of assessments being designed to measure the effectiveness of the academic programs and the quality of the teaching, the policies and procedures are altered to fit the assessment. This was certainly the case in Appraisal school district, where the local newspaper routinely published the results of the state assessment and compared schools to each other and adjourning school districts. The teachers perceived that parents put significant weight into these results, often choosing to purchase new houses in proximity to high achieving schools. The teachers also felt pressure because their professional evaluations would be impacted by the results of these tests. Finally, the teachers perceived that students felt extreme pressure, especially on the old state assessments, and anticipated this pressure returning as the results of the new assessments were published. They related numerous stories of students crying and feeling sick because of the pressure of high stakes testing. It seemed only natural to them to focus their time and energy on the state assessment. This is not a criticism of the teachers or the district. In fact, it is the most logical approach for them to take. If there is one primary measure of effectiveness, it only makes sense one would alter what is done to maximize the results of that measure. This is why Campbell's law asserts that the more emphasis put on a measure the more likely it is to distort the very thing it is trying to measure. Once educational and instructional triage takes hold, no longer is high stakes assessments measuring the overall effectiveness of the schools and teachers, rather it is measuring their effectiveness at preparing students for the assessment.

More Instructional Triage than Educational Triage

Campbell's law also helps explain why, as the external pressure decreased on the schools in Appraisal, the amount of educational triage decreased. During the height of NCLB, the teachers recalled the school implementing many activities that focused the students on the importance of the assessment, but took away valuable class time. They mentioned pep rallies, writing notes of encouragement to fellow students, parents bringing in special treats, school wide chants, and high-fives down the hallway. They also mentioned remediation for students who were predicted not to do well on the test as well as making changes to their special education program. All of these school-wide, educational triage techniques dramatically decreased as soon as the assessment changed and the accountability eased.

Because of the transition to the new state assessments, it has been three years since the assessment results have been published or used for state accountability purposes. The teachers noted a corresponding decrease in many of the school wide activities, like pep rallies, notes home, and remediation. As soon as the pressure decreased, the schools backed away from these activities and put their time back into classroom instruction. The teachers also predicted that these activities would increase again as the pressure on the schools increased. They were convinced that the school would resume these educational triage activities if the leadership believed it would increase test scores. Classroom learning is important but if you can get more "bang for your buck" by getting the students excited for the test, then that would take precedent. When results on the test are the only measure, then this is a logical course to take. Without knowing it, they were explaining the effects of Campbell's law.

Over the same time period however, the teachers expressed no decrease in the amount of internal pressure they felt. Because assessment results were part of the evaluation tool and they

were expected to show growth over time, the teachers never reduced the amount of pressure they put on themselves. Even though the assessment results were not being published and the administration was, at least temporarily, backing off on school-wide test prep initiatives, the teachers did not feel they could let up. How their students performed on the test was still a major indicator of success, particularly when their teacher evaluation was tied to assessment scores. As a result, the amount of instructional triage did not change with the new assessments.

As long as teachers perceive the tests to be high stakes they will inevitably take steps to ensure their students do well on them. To do otherwise, would not make any sense. Several teachers described sacrificing instructional time to administer practice tests and review before the state assessments, even though the district leaders discouraged it. The majority of teachers also stated they took extensive time out of their class to specifically teach test taking skills. This included time spent on teaching necessary technological skills and developing simulated questions. In addition, the teachers gave multiple examples of ways they changed what they taught to align with the new assessment. This included adopting new curriculum in math and reading that increased emphasis on teaching higher level thinking skills and complex problems solving and extensive test review. In this way, Campbell's law helps explain why teachers' indicated there was more instructional triage then educational triage currently occurring in Appraisal school district.

Teachers are Resigned to High Stakes Testing

After sixteen years of high stakes testing, the teachers were resigned to the reality of high stakes testing. Throughout the interviews the teachers expressed a sentiment of "just tell me what to do and I will do it." They did not question whether the changes brought about by educational and instructional triage were good for students or not, as long as they were

convinced the changes were good for the assessment results (Booher-Jennings, 2006). Their professional judgment was subjugated to the greater good of higher test scores. The teachers did not seem to believe they had control over the decisions they make at school (Ingersoll & Perda, 2008). For them, the curriculum was set, the instructional strategies were locked in place, the weekly assessments were required, and the tier time was mandatory. All of these changes they approached with an unquestioning attitude. They said things like, "If that is what they want me to do" and "I guess that is the way we do it now." They seemed to accept whatever they were asked to do as long as test scores improved (Nichols & Berliner, 2007). Repeatedly, the teachers implied that if the administration implemented a new curriculum to improve test scores, they would embrace it. They want autonomy in many areas but they realize autonomy in selecting curriculum is no longer an option (Pearson & Moomaw, 2005). There was a definite sense of resignation.

In addition, several veteran teachers indicated they felt as though they had changed jobs two or three times during their teaching career. To them, teaching in the era of high stakes testing was completely different than teaching prior to them (Connell, 2009). These long time teachers remembered fondly the lack of stress and the ability to focus on the student and not on the test score. They recalled having more freedom to teach what they wanted and less intrusion into their instructional time. They noted how they now feel expected to fall in line with whatever practices and procedures are proven to effectively increase test scores without being able to take into account the individual students in their classroom (Koretz, 2011). Despite these fond memories, the teachers did not seem bitter; instead they seemed resigned. They did not question the importance of good test scores. They admitted, "It is a different time" and "We have to demonstrate results." They did not blame the school district or their principals for the change.

To them, the definition of education had changed and they were going to meet that new definition. They acquiesced. "Just tell me what to do, and I will do it."

Redefining Success

This acceptance of high stakes testing seemed to come from a new definition of success. Many people now believe that successful student learning is synonymous with successful test results (Stiggins, 2002). The teachers indicated the effectiveness of programs and policies were determined by their impact on test scores, rather than their impact on overall student learning. State assessment drove curricular decisions and the weekly test-aligned screeners drove instructional decisions (Cawelti, 2006). Students were rewarded based on their assessment scores, more than on their classwork. While some teachers still held out a belief that learning could not be defined by a single number, many teachers succumbed to the belief that increased test scores meant that learning had occurred (Kohn, 2000). When touting the success of their building, they referenced test scores. When explaining the new textbooks, they touted their alignment to the new assessment. When they talked about student growth, they cited the number of students moving out of tier time. Despite their misgivings about the new assessment and their expressed desire to see a decrease in standardized assessments, most teachers still defined success in terms of assessment scores. They did not realize their definition of success was feeding into the very thing they wished would go away (Nichols, Glass, & Berliner, 2006).

A few teachers did identify the hypocrisy. They wished the emphasis would be taken off the school and the teacher and put back where it belongs – on the children. These teachers went into teaching because they loved children and wanted to see them grow into successful young people, not because they wanted to see their test score grow. Many teachers would rather leave the profession then hurt their children (Margolis, Meese, & Doring, 2016) . Children are more

than their test score. They disagreed with their colleagues and did not accept the new normal. Teachers are opposed to the idea that a lack of autonomy was going to lead to better learning (Crocco & Costigan, 2007). They seemed to believe the more the definition of a quality education was distilled down to a single number, the less accurate that definition was (Haertel, 2013). In this way, Campbell's Law helps explain the discrepancy between the two sets of teachers in their definition of success.

Two Positives: Increased Accountability and Data Driven Instruction

While some teachers may believe that high stakes testing has distorted the definition of a quality education, they were still proud of the job they did. All the teachers were proud of their work and extremely proud of their students. They repeatedly commented on the good things happening in their buildings. Despite the near unanimous hope that high stake assessments would disappear, the teachers repeatedly referenced increased accountability and the advent of data driven instruction as two positives that have resulted because of high stake assessments (Casselman, 2015).

While the notion of 100% proficiency under NCLB was an unrealistic goal that created undue pressure, the teachers in Appraisal agreed the idea that no individual child should be left behind was a good one. NCLB helped them mainstream special education students and forced them to focus on all their subgroups (Ross-Hill, 2009). It changed the focus from what teachers teach, to what children learn. They acknowledged that other teachers could no longer say, "I taught it so the students should have learned it." Teachers now know they are responsible for the learning of all children (Darling-Hammond, 2010). This is a significant shift in thinking that will forever change the face of education.

Likewise, NCLB brought about the idea of data driven instruction. Now that teachers feel a greater responsibility for the learning of every child, they began to believe that having data to demonstrate learning is invaluable. Prior to NCLB, they felt confident their children were learning, but they often had no way to demonstrate it in a manner the public would accept. Teachers reported they appreciated the ability to prove they were effective teachers. They also appreciated having detailed data that broke down skills and provided specific information about where the learning process for a specific student needed to be shored up. This ability to pinpoint distinct learning deficiencies was made possible by detailed testing. The teachers expressed that while this sort of testing had been around before NCLB, it was now more accessible and prevalent in the classroom. As a result, they used it much more frequently and thought it a great benefit to the students' learning.

Implications

My study findings and conclusions indicate that educational and instructional triage is occurring as an inevitable response to high stakes testing. As a result, high stakes tests are unable to accurately measure what they purport to measure, namely the effectiveness of schools and teachers. In this section, I offer four suggestions policy makers and researchers.

Low Stakes Tests are a Better Indicator of Student Success then High Stakes Tests

Teachers are not opposed to assessments or accountability. They understand that the public wants standardized tests to demonstrate learning and they are okay with that (Moses & Nanna, 2007). However, they are opposed to the high stakes nature of tests that stress out students and strain already limited instructional time (Musoleno & White, 2010). Likewise, Campbell's law demonstrates that the more high stakes the test, the less accurate it is. Therefore,

if policymakers continue to ask for annual assessments they should be purposely designed to be low stakes.

Teachers agree that assessments are necessary to gather data on how the students are doing. However, the teachers are clear that assessment should not be high stakes. The results should not be used to rank schools because they are not a good reflection of the teacher's ability to teach or how much the student has learned (Vernaza, 2012). The scores should not be released to the press because the multiple purposes of schools ensures that, from someone's point of view, they are always doing a poor job and those pushing for reforms can always find a negative comparison to prove their point (Reese, 2007). They should not be used as part of teacher evaluations because there is little reliability and validity between test scores and other measures of effective teaching (Haertel, 2013). Finally, they should not be used as a sole basis for remediation of a child because the pressure often has a debilitating effect on the student's performance (Watson et al., 2014).

However, the teachers are not opposed low stakes testing. In fact, they welcome the opportunity to demonstrate about how well their students were performing on low stake assessment (Winters, Trivitt, & Greene, 2010). Teachers believe that they want to help all students learn, but when they are forced to teach to the middle they know they are not meeting the needs of all learners (Willis, 2011). They expressed the focus of education should be on the students, not on the school or teachers. Focusing on what the principal and the teacher should be doing, only takes away instructional time for the students. They want data to demonstrate their success. They merely want it in a low stakes environment.

Administrators Need to Build Trust

If political realities prevent high stakes testing from being eliminated completely, then administrators need to find a way to decrease the pressure teachers feel. The less pressure teachers feel the more effective they are at their job because they are able to focus on student learning, rather than student testing (Tschannen-Moran, 2009). Administrators may not be able to control the pressure that the media creates by reporting the results. Likewise, policy prevents administrators from removing scores from their teacher evaluations completely. However, administrators can minimize the stress created by high stakes tests by acknowledging one test score does not always reflect everything that is going on in the classroom (Feuerstein, 2011).

Administrators often feel caught between the policy requirements and doing what is best for their teachers (Walker, Kutsyuruba, & Noonan, 2011). The more they are able to articulate that the policy requirements are not going to supersede personal observations or what they know is going on in the classroom, the less stress the assessments will invoke. Otherwise these sorts of mandates can drive a wedge between principals and teachers (Kutsyuruba, Walker, & Noonan, 2011). The most effective way to mitigate the impact of assessment mandates is building trust between the teacher and the principal (Moye, Henkin, & Egley, 2005).

Teachers who trust their administrator are more effective in the classroom (Hoy & Tschannen-Moran, 1999). Teachers need to be able to trust that their administrator understands a teacher might have a particularly hard class one year, or that test anxiety might lower the scores, or that kids could have a bad day, or that some test questions could be poorly written. Overall, if teachers do not trust their administrators it can have a profoundly negative impact on the classroom (Tschannen-Moran & Hoy, 1998). In schools where there is a high level of trust between teachers and principals, there is generally higher achievement (Rothstein, 2004).

Loss of Professionalism

There is ample research on the deprofessionalization of teaching that has been occurring sense the advent of the standards movement and NCLB. The standards movement has made teachers feel their opinion is not valued and they have little autonomy in their decision-making (Hodges, Tippins, & Oliver, 2013). This is because the standards movement implicitly states that teachers cannot be trusted to determine what children most need to learn, but need a set of standards to ensure that they teach what the standards writers believe is important. Secondly, teachers believe the public does not view them as highly trained professionals, but rather as doing a job anyone could do (Pike, 2014). The negative publicity surrounding test results and the constant barrage of stories of failing schools only reinforce the lack of respect. The devaluing of the teaching profession is exacerbated when high stakes testing is part of their evaluation system because it implies they are not working hard enough and if there was a consequence they will try harder (Cochran-Smith, Piazza, & Power, 2013). This is quite demeaning to a profession filled with people who love children and give their all to help them.

The deprofessionalization of teaching is causing teachers to leave the profession and others not to choose the profession in the first place (Goldstein, 2011). This is especially true in schools serving the most challenging students (Crocco & Costigan, 2007). Aspiring teachers realize they are not going to be rewarded monetarily but often go into the profession because they care deeply for the success of the children and want to feel like they are making a difference (Wellman, 2007). When they do not feel appreciated or empowered they are going to leave the profession, which is devastating when research shows teachers have the most significant impact on student learning (Hodges et al., 2013).

More importantly, if teachers were empowered in their classroom and less beholden to the results of a test, they are more effective (Willis, 2011). When they do not have to subjugate their professional opinion in the name of standards and testing, they are more effective at teaching higher order thinking (Cawelti, 2006). Teachers are more effective in an environment where they feel trusted and treated as a professional (Johnson, Kraft, & Papay, 2012). High stake tests have contributed to the undermining of teaching as a profession. More can be accomplished by trusting them to do their best as professionals, rather than trying to catch them doing wrong on a high stakes test.

Impact of ESSA

In December 2015, Congress finally replaced the No Child Left Behind Act with the Every Student Succeeds Act (ESSA). This act restored some autonomy to the states and allowed them more control over their assessments and accompanying sanctions. With the current backlash in some areas against high stakes testing, especially as related to the Common Core state standards, it will be interesting to see if any states back off on their sanctions and make their assessments less high stakes. If this does occur, it will provide a rich opportunity for researchers to continue to exam the influence of high stakes test on elementary classroom instruction. Studying schools in states who reduce pressure, or comparing schools in states with different levels of sanctions, provides rich opportunity to further research educational and instructional triage, especially as it relates to Campbell's law.

REFERENCES

REFERENCES

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

Ambrose, D. (2012). Battling creaticide: An interview with David C. Berliner. *Roeper Review, 34*, 77-80. doi: 10.1080/02783193.2012.660680

American Recovery and Reinvestment Act of 2009, U.S.C. 111-5 (2009).

Apple, M. W. (2006). *Educating the" right" way: Markets, standards, God, and inequality*: Taylor & Francis.

Apple, M. W. (2011). Democratic education in neoliberal and neoconservative times. *International Studies in Sociology of Education, 21*(1), 21-31.

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher, 36*(5), 258-267.

Au, W. (2009). Social studies, social justice: W(h)ither the social studies in high-stakes testing? *Teacher Education Quarterly, 36*(1), 43-58.

Aydeniz, M., & Southerland, S. A. (2012). A national survey of middle and high school science teachers' responses to standardized testing: Is science being devalued in schools? *Journal of Science Teacher Education, 23*(3), 233-257.

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., . . . Shepard, L. A. (2010). Problems with the use of student test scores to evaluate teachers *EPI Briefing Paper* (Vol. 278). Washington D.C.: Economic Policy Institute.

Baker, G., Gibbons, R., & Murphy, K. J. (1994). Subjective performance measures in optimal incentive contracts. *Quarterly Journal of Economics, 109*(4), 1125-1156.

Baker, M., & Johnston, P. (2010). The impact of socioeconomic status on high stakes testing reexamined. *Journal of Instructional Psychology, 37*(3), 193-199.

Berliner, D. (2011). Rational responses to high stakes testing: The case of curriculum narrowing and the harm that follows. *Cambridge Journal of Education, 41*(3), 287-302.

Biklen, S. K., & Bogdan, R. (2006). *Qualitative research for education: An introduction to theory and methods* (5th ed.). Boston, MA: Allyn & Bacon.

Blau, P. M. (1963). *The dynamics of bureaucracy: study of interpersonal relations in two government agencies* Chicago, IL: University of Chicago Press.

Booher-Jennings, J. (2005). Below the bubble:"Educational triage" and the Texas accountability system. *American Educational Research Journal, 42*(2), 231-268.

Booher-Jennings, J. (2006). Rationing education in an era of accountability. *Phi Delta Kappan, 87*(10), 756-761.

Borg, M. O. M., Plumlee, J. P., & Stranahan, H. A. (2007). Plenty of children left behind: High-stakes testing and graduation rates in Duval County, Florida. *Educational Policy, 21*(5), 695-716.

Bracey, G. W. (1991). Why can't they be like we were? *Phi Delta Kappan, 73*(2), 104-117.

Brown, C. P. (2007). Examining the streams of a retention policy to understand the politics of high-stakes reform. *Education Policy Analysis Archives, 15*(9), 1-28.

Campbell, D. T. (1975). Assessing the impact of planned social change. *Evaluation and program planning, 2*(1), 67-90.

Campbell, D. T., Stanley, J. C., & Gage, N. L. (1963). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin.

Casey, L. M. (2013). The will to quantify: The "bottom line" in the market model of education reform. *Teachers College Record, 115*(9), 1-7.

Casselman, B. (2015). No Child Left Behind Worked. Retrieved 4/11/2106

Cawelti, G. (2006). The side effects of NCLB. *Educational Leadership, 64*(3), 64-68.

Chamberlin, J. L. (2007). Poverty, school size and charter designation as predictors of student achievement on a statewide high-stakes testing program. *AASA Journal of Scholarship & Practice, 4*(1), 21-26.

Cherry, K. (2014). History of intelligence testing: The history and development of modern IQ testing http://psychology.about.com/od/psychologicaltesting/a/int-history.htm. Retrieved 05.01.2014

Chiseri-Strater, E. (1996). Turning in upon ourselves: Positionality, subjectivity, and reflexivity in case study and ethnographic research. In P. Mortensen & G. Kirsch (Eds.), *Ethics and representation in qualitative studies of literacy* (pp. 115-133). Urbana, IL: National Council of Teachers of English.

Chudowsky, N., & Chudowsky, V. (2007). No child left behind at five: A review of changes to state accountability plans. Washington, D.C.: Center on Education Policy.

Chudowsky, N., Chudowsky, V., & Kober, N. (2007). Answering the question that matters most: Has student achievement increased since no child left behind? Washington, DC: Center on Education Policy.

Cochran-Smith, M., Piazza, P., & Power, C. (2013). *The politics of accountability: Assessing teacher education in the United States.* Paper presented at the The Educational Forum.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. (1966). Equality of educational opportunity (pp. 1-32). Washington, DC: US Government printing press.

Connell, R. (2009). Good teachers on dangerous ground: Towards a new view of teacher quality and professionalism. *Critical studies in education, 50*(3), 213-229.

Cornell, D. G., Krosnick, J. A., & Chang, L. (2006). Student reactions to being wrongly informed of failing a high-stakes test: The case of the Minnesota basic standards test. *Educational Policy, 20*(5), 718-751.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: Sage.

Creswell, J. W. (2015). *Educational research : Planning, conducting, and evaluating quantitative and qualitative research* (5th ed.). Upper Saddle River, N.J.: Peasron Education.

Crocco, M. S., & Costigan, A. T. (2007). The narrowing of curriculum and pedagogy in the age of accountability: Urban educators speak out. *Urban Education, 42*(6), 512-535.

Cuban, L. (2007). Hugging the middle teaching in an era of testing and accountability, 1980-2005. *Education Policy Analysis Archives, 15*(1).

Darling-Hammond, L. (2010). Performance counts: Assessment systems that support high-quality learning. Washington, DC: Council of Chief State School Officers and Stanford Center for Opportunity Policy in Education.

Darling-Hammond, L., & Adamson, F. (2013). Developing assessments of deeper learning: The costs and benefits of using tests that help students learn. (pp. 1-28). Stanford, CA: Stanford Center for Opportunity Policy in Education.

Davis, W. A., & Havighurst, R. J. (1948). The measurement of mental systems: Can intelligence be measured? *The Scientific Monthly, 66*(4), 301.

Dee, T., & Jacob, B. (2010). Evaluating NCLB. *Education Next, 10*(3), 54-61.

DeFauw, D. L. (2013). 10 writing opportunities to 'teach to the test'. *Reading Teacher, 66*(7), 569-573. doi: 10.1002/TRTR.1161

Dutro, E., & Selland, M. (2012). 'I like to read, but I know I'm not good at it': Children's perspectives on high-stakes testing in a high-poverty school. *Curriculum Inquiry, 42*(3), 340-367. doi: 10.1111/j.1467-873X.2012.00597

Feuerstein, A. (2011). The politics of accountability and teacher preparation. *Action in Teacher Education, 33*(1), 3-23. doi: 10.1080/01626620.2011.559421

Forgione, P. D., Jr. (2012). The "race" is underway: Be vigilant & prepared for our children's sake! *State Education Standard, 12*(2), 13-21.

Freeman, M., Preissle, J., Roulston, K., & Pierre, E. A. S. (2007). Standards of evidence in qualitative research: An incitement to discourse. *Educational Researcher, 36*(1), 25-32.

Garrison, M. J. (2009). *A measure of failure: The political origins of standardized testing*. Albany, NY: Suny Press.

Gay, L. R., Mills, G. E., & Airasian, P. W. (2012). *Educational research: Competencies for analysis and application*. Boston, MA: Pearson Education.

Gillborn, D., & Youdell, D. (1999). *Rationing education: Policy, practice, reform, and equity*. Philadelphia, PA: McGraw-Hill International.

Goldstein, D. (2011). The test generation. *American Prospect, 22*(4), 14-19.

Goodhart, C. (1975). Problems of monetary management: The UK experience. In A. Courakis (Ed.), *Inflation, depression and economic policy in the west: Lessons from the 1970s*. Oxford: Basil Blackwell.

Goodwin, B. (2014). Better tests don't guarantee better instruction. *Educational Leadership, 71*(6), 78-80.

Groen, M. (2012). NCLB--The educational accountability paradigm in historical perspective. *American Educational History Journal, 39*(1/2), 1-14.

Guilfoyle, C. (2006). NCLB: Is there life beyond testing? *Educational Leadership, 64*(3), 8.

Guisbond, L., Neill, M., & Schaeffer, B. (2012). NCLB's lost decade for educational progress: What can we learn from this policy failure? Jamaica Plain, MA: FairTest: The national center for fair and open testing.

Gunzenhauser, M. G. (2003). High-stakes testing and the default philosophy of education. *Theory into Practice, 42*(1), 51.

Guthrie, J. W., & Springer, M. G. (2004). A nation at risk revisited: Did "wrong" reasoning result in "right" results? At what cost? *Peabody Journal of Education, 79*(1), 7-35.

Haertel, E. H. (2013). Reliability and validity of inferences about teachers based on student test scores *The William H. Angoff Memorial Lecture*. Princeton, NJ: ETS.

Hargrove, T. Y., Jones, M. G., Jones, B. D., Hardin, B., Chapman, L., & Davis, M. (2000). Unintended consequences of high-stakes testing in North Carolina: Teacher perceptions. *ERS Spectrum, 18*(4), 21-25.

Heritage, M. (2010). Formative assessment and next-generation assessment systems: Are we losing an opportunity *National Center for Research on Evaluation, Standards, and Student Testing (CRESST) and the Council of Chief State School Officers (CCSSO)*. Washington, DC.

Herman, J., & Linn, R. (2013). On the road to assessing deeper learning: The status of Smarter Balanced and PARCC assessment consortia. *National Center for Research on Evaluation, Standards, and Student Testing (CRESST)*. Washington, DC.

Herman, J., & Linn, R. (2014). New assessments, new rigor. *Educational Leadership, 71*(6), 34-37.

Hernandez, J. (2009). New education secretary visits Brooklyn school. *Retrieved from New York Times at http://cityroom.blogs.nytimes.com/2009/02/19/new-education-secretary-visits-brooklyn-school*. Retrieved 12/15/2014

Hodges, G. W., Tippins, D., & Oliver, J. S. (2013). A Study of Highly Qualified Science Teachers' Career Trajectory in the Deep, Rural South: Examining a Link Between Deprofessionalization and Teacher Dissatisfaction. *School Science and Mathematics, 113*(6), 263-274.

Hong, W.-P., & Youngs, P. (2008). Does high-stakes testing increase cultural capital among low-income and racial minority students? *Education Policy Analysis Archives, 16*(6), 1-21.

Hoover, H., Dunbar, S., Frisbie, D., Oberley, K., Bray, G., Naylor, R., . . . Qualls, A. (2003) The Iowa test of basic skills: Interpretive guide for school administrators. Chicago, IL: Riverside Publishing.

Hout, M. S. S. (2012). Do high-stakes tests improve learning? *Issues in Science & Technology, 29*(1), 33-38.

Howell, W. G., West, M. R., & Peterson, P. E. (2007). What Americans think about their schools: The 2007 "education next" PEPGY survey. *Education Next, 7*(4), 12-26.

Hoy, W. K., & Tschannen-Moran, M. (1999). Five faces of trust: An empirical confirmation in urban elementary schools. *Journal of School Leadership, 9*, 184-208.

Huddleston, A. P. (2014). Achievement at whose expense? A literature review of test-based grade retention policies in U.S. schools. *Education Policy Analysis Archives, 22*(18), 1-31. doi: 10.14507/epaa.v22n18.2014

Hursh, D. (2007). Assessing no child left behind and the rise of neoliberal education policies. *American Educational Research Journal, 44*(3), 493-518.

Ingersoll, R., & Perda, D. (2008). The status of teaching as a profession. *Schools and society: A sociological approach to education*, 107-118.

Jeffrey, A. J., Auger, R. W., Pepperell, J. L., Sleegers, P., den Brok, P., Verbiest, E., . . . Daly, A. J. (2013). “If we're ever in trouble they're always there”: A qualitative study of teacher-student caring. *The Elementary School Journal, 114*(1).

Jennings, J., & Sohn, H. (2014). Measure for measure: How proficiency-based accountability systems affect inequality in academic achievement. *Sociology of Education*(2), 125. doi: 10.1177/0038040714525787

Johnson, S. M., Kraft, M. A., & Papay, J. P. (2012). How context matters in high-need schools: The effects of teachers’ working conditions on their professional satisfaction and their students’ achievement. *Teachers College Record, 114*(10), 1-39.

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

Jonsson, P. (2011, July 5). America's biggest teacher and principal cheating scandal unfolds in Atlanta. *Christian Science Monitor*.

Journell, W. (2010). The influence of high-stakes testing on high school teachers' willingness to incorporate current political events into the curriculum. *The High School Journal, 93*(3), 111-125.

Kansas State Department of Education. (2015a). Building Report Card Retrieved March 13, 2015, from http://online.ksde.org/rcard/district.aspx?org_no=D0410

Kansas State Department of Education. (2015b). Kansas K-12 Reports. Retrieved March 13, 2015, from http://online.ksde.org/k12/CountyStatics.aspx?org_no=D0410

Kearns, L.-L. (2011). High-stakes standardized testing & marginalized youth: An examination of the impact on those who fail. *Canadian Journal of Education, 34*(2), 112-130.

Kleinsasser, A. M. (2000). Researchers, reflexivity, and good data: Writing to unlearn. *Theory into Practice, 39*(3), 155-162.

Kohn, A. (2000). *The case against standardized testing: Raising the scores, ruining the schools*. Portsmouth, NH: Heinemann.

Koretz, D. M. (2008). *Measuring up*. Cambridge, MA: Harvard University Press.

Koretz, D. M. (2011). Lessons from test-based education reform in the US. *American Educator, 14*, 9-23.

Krieg, J. M. (2008). Are students left behind? The distributional effects of the No Child Left Behind Act. *Education, 3*(2), 250-281.

Kutsyuruba, B., Walker, K., & Noonan, B. (2011). Restoring broken trust in the work of school principals. *International studies in educational administration, 39*(2), 81-95.

Kvale, S. (1995). The social construction of validity. *Qualitative inquiry, 1*(1), 19-40.

Labaree, D. F. (2012). *Someone has to fail*. Cambridge, MA: Harvard University Press.

Ladd, H. F., & Lauen, D. L. (2010). Status versus growth: The distributional effects of school accountability policies. *Journal of Policy Analysis and Management, 29*(3), 426-450.

Lai, E. R. (2011). Performance-based assessment: Some new thoughts on an old idea. *Always Learning Bulletin, 20*, 1-4.

Lapan, S. D., Quartaroli, M. T., & Riemer, F. J. (2011). *Qualitative research: An introduction to methods and designs* (Vol. 37): John Wiley & Sons.

Lauen, & Gaddis. (2012a). Accountability pressure, academic standards, and educational triage. Washington, DC: Society for Research on Educational Effectiveness.

Lauen, & Gaddis. (2012b). Shining a light or fumbling in the dark? The effects of NCLB's subgroup-specific accountability on student achievement. *Educational Evaluation and Policy Analysis, 34*(2), 185-208.

Leistyna, P. (2007). Corporate testing: Standards, profits, and the demise of the public sphere. *Teacher Education Quarterly, 34*(2), 59-84.

Leman, N. (1999). *The Big Test*. New York, NY: Straus and Giroux.

Lincoln, Y. S., & Guba, E. G. (1986). But is it rigorous? Turstworthiness and authenticity in naturalistic evaluation. In D. D. Williams (Ed.), *Naturalistic evaluation* (pp. 15-25). San Francisco, CA: Jossey-Boss.

Llopis-Jepsen, C. (2013a). Kansas opts to create its own Common Core tests. *Topeka Capital-Journal*. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=edsnbk&AN=14AA133510D2E280&site=eds-live

Llopis-Jepsen, C. (2013b). KU assures high quality for Common Core tests. *Topeka Capital-Journal*. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=edsnbk&AN=14AAB87EE098D558&site=eds-live

Lucas Jr, R. E. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester conference series on public policy*, *1*, 19-46.

Lucido, H. (2010). *Educational Genocide: A plague on our children*. Letham, MD: R&L Education.

Madaus, G. (1988). The distortion of teaching and testing: High-stakes testing and instruction. *Peabody Journal of Education, 65*(3), 29-46.

Madaus, G., & Clarke, M. (2001). The adverse impact of high stakes testing on minority students: Evidence from 100 years of test data. In G. Orfield & M. Kornhaber (Eds.), *Raising standards or raising barriers? Inequality and high stakes tests in public education*. New York: The Centruy Foundation.

Madaus, G., & Russell, M. (2010). Paradoxes of high-stakes testing. *Journal of Education, 190*(1/2), 21-30.

Maltese, A. V., & Hochbein, C. D. (2012). The consequences of 'school improvement': Examining the association between two standardized assessments measuring school improvement and student science achievement. *Journal of Research in Science Teaching, 49*(6), 804.

Marchant, G. J., Paulson, S. E., & Shunk, A. (2006). Relationships between high-stakes testing policies and student achievement after controlling for demographic factors in aggregated data. *Education Policy Analysis Archives, 14*(30), 1-34.

Margolis, J., Meese, A. A., & Doring, A. (2016). Do Teachers Need Structure or Freedom to Effectively Teach Urban Students? A Review of the Educational Debate. *Education and Urban Society, 1*, 24.

Marshall, C., & Rossman, G. B. (2010). *Designing qualitative research*. Thousand Oaks, CA: Sage.

Martin, P. C. (2012). Misuse of high-stakes test scores for evaluative purposes: Neglecting the reality of schools and students. *Current Issues in Education, 15*(3), 1-10.

Mason, E. J. (2007). Measurement issues in high stakes testing: Validity and reliability. *Journal of Applied School Psychology, 23*(2), 27-46.

Mauthner, N. S., & Doucet, A. (2003). Reflexive accounts and accounts of reflexivity in qualitative data analysis. *Sociology, 37*(3), 413-431.

Maxwell, J. A. (2012). *Qualitative research design: An interactive approach*. Thousand Oaks, CA: Sage.

McNeil, L. M. (2000). Creating new inequalities: Contradictions of reform accountability system in Texas. In D. Flinders & S. Thorton (Eds.), *The Curriculum Studies Reader* (Vol. 81, pp. 728-734): Pyschology Press.

McNeil, M. (2013). Rifts deepen over direction of education policy in U.S. *Education Week, 32*(30), 1-16.

Mehta, J. (2008). How did we get here? Paradigms, fields, and professions in education policy. *Politics of Education Association Bulletin, 33*(1), 3-7.

Merriam, S. B. (1995). What Can You Tell From An N ofl?: Issues of validity and reliability in qualitative research. *PAACE Journal of Lifelong Learning, 4*, 50-60.

Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. San Francisco, CA: Jossey-Bass.

Mora, R. (2011). "School is so boring": High-stakes testing and boredom at an urban middle school. *Perspectives on Urban Education, 9*(1).

Morgan, D. L. (1996). Focus groups. *Annual review of sociology*, 129-152.

Moses, M. S., & Nanna, M. J. (2007). The testing culture and the persistence of high stakes testing reforms *Education and Culture* (Vol. 23, pp. 55-72). Prude: Prude University Press.

Moye, M. J., Henkin, A. B., & Egley, R. J. (2005). Teacher-principal relationships: Exploring linkages between empowerment and interpersonal trust. *Journal of Educational Administration, 43*(3), 260-277.

Musoleno, R. R., & White, G. P. (2010). Influences of high-stakes testing on middle school mission and practice. *RMLE Online: Research in Middle Level Education, 34*(3), 1-10.

National Center for Ecuation Statistics. (2015). NCES Local codes. Retrieved 10/29/2015

National Commission on Excellence in Education. (1983). *A Nation at Risk : The imperative for educational reform*. Washington, D.C.: National Commission on Excellence in Education.

Neal, D., & Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. *The Review of Economics and Statistics, 92*(2), 263-283.

Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, MA: Harvard Education Press.

Nichols, S. L., Glass, G. V., & Berliner, D. C. (2006). High-stakes testing and student achievement: Does accountability pressure increase student learning? *Education Policy Analysis Archives, 14*(1), 1-172.

Nichols, S. L., Glass, G. V., & Berliner, D. C. (2012). High-stakes testing and student achievement: Updated analyses with naep data. *Education Policy Analysis Archives, 20*(20).

No Child Left Behind. (2002). Act of 2001, Pub. L. No. 107-110, § 115 *Stat* (Vol. 1425, pp. 107-110). Washington, DC.

Noblit, G. W. (1999). *Particularities: Collected essays on ethnography and education* (Vol. 44). New York: Peter Lang

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237-257.

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

Parkison, P. (2009). Political economy and the NCLB regime: Accountability, standards, and high-stakes testing. *Educational Forum, 73*(1), 44-57.

Patton, M. Q. (2002). *Qualitative evaluation and research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Pearson, L. C., & Moomaw, W. (2005). The relationship between teacher autonomy and stress, work satisfaction, empowerment, and professionalism. *Educational research quarterly, 29*(1), 37.

Pecheone, R., Kahl, S., Hamma, J., & Jaquith, A. (2010). Through a looking glass: Lessons learned and future directions for performance assessment. Stanford, CA: Standford Center for Opportunity Policy in Education.

Pershey, M. G. (2010). A comparison of African American students' self-perceptions of school competence with their performance on state-mandated achievement tests and normed tests of oral and written language and reading. *Preventing School Failure, 55*(1), 53.

Peshkin, A. (1988). In search of subjectivity—one's own. *Educational Researcher, 17*(7), 17-21.

Peshkin, A. (1993). The goodness of qualitative research. *Educational Researcher, 22*(2), 23-29.

Peshkin, A. (2000). The nature of interpretation in qualitative research. *Educational Researcher*(9), 5. doi: 10.2307/1177087

Pike, M. (2014). *The Perceptions of Teacher Status and Impact on the American Educational System.* (Doctoral dissertation), Baylor University. Retrieved from http://hdl.handle.net/2104/9009

Plank, S. B., & Condliffe, B. F. (2013). Pressures of the season: An examination of classroom quality and high-stakes accountability. *American Educational Research Journal, 50*(5), 1152-1182. doi: 10.3102/0002831213500691

Porter, A., & Chester, M. (2002). Building a high-quality assessment and accountability program: The philadelphia example. *Brookings Papers on Education Policy*, 285-337.

Reback, R., Rockoff, J., & Schwartz, H. L. (2011). Under pressure: Job security, resource allocation, and productivity in schools under NCLB. Washington, DC: National Bureau of Economic Research.

Reese, W. J. (2007). Why Americans love to reform the public schools. *Educational Horizons, 85*(4), 217-231.

Reich, G. A., & Bally, D. (2010). Get smart: Facing high-stakes testing together. *Social Studies, 101*(4), 179-184.

Rentner, D. S., Scott, C., Kober, N., Chudowsky, N., Chudowsky, V., Joftus, S., & Zabala, D. (2006). From the capital to the classroom: Year 4 of the No Child Left Behind Act. Washington, DC: Center on Education Policy.

Rice, J. M. (1913). *Scientific management in education*. New York: Hinds, Noble and Eldredge.

Rose, L. C. (2004). No child left behind: The mathematics of guaranteed failure. *Educational Horizons, 82*(2), 121-130.

Ross-Hill, R. (2009). Teacher attitude towards inclusion practices and special needs students. *Journal of Research in Special Educational Needs, 9*(3), 188-198.

Rothman, R. (1995). *Measuring up: Standards, assessment, and school reform*. San Fransico, CA: Jossey-Bass.

Rothstein, R. (2004). Class and schools: Using social, economic, and educational reform to close the achievement gap. Washington, DC: Economic Policy Institute.

Ryan, A. M., & Stoskopf, A. (2008). Public and catholic school responses to IQ testing in the early 20th century. *Teachers College Record, 110*(4), 894-922.

Ryan, G. W., & Bernard, H. R. (2003). Techniques to identify themes. *Field methods, 15*(1), 85-109.

Sahlberg, P. (2010). Rethinking accountability in a knowledge society. *Journal of Educational Change, 11*(1), 45.

Schaeffer, B. (2012). Resistance to high-stakes testing spreads. *District Administration, 48*(8), 34-36.

Shadish, W. R., & Luellen, J. K. (2004). Donald Campbell: The accidental evaluator. In M. Akin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (pp. 80-87). Thousand Oaks, CA: Sage

Springer, M. G. (2008). Accountability incentives: Do schools practice educational triage? *Education Next, 8*(1), 74-79.

Stake, R. E. (1971). Testing hazards in performance contracting. *Phi Delta Kappan*, 583-589.

Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan, 83*(10), 758.

Stiglitz, J. E., Sen, A., & Fitoussi, J.-P. (2010). Report by the commission on the measurement of economic performance and social progress. Paris.

Supovitz, J. (2009). Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change, 10*(2-3), 211-227.

Taylor, F. W. (1914). *The principles of scientific management*. New York: Harper.

Tienken, C. (2012). Neoliberalism, social darwinism, and consumerism masquerading as school reform. *Interchange, 43*(4), 295-316. doi: 10.1007/s10780-013-9178-y

Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative inquiry, 16*(10), 837-851.

Tschannen-Moran, M. (2009). Fostering teacher professionalism in schools the role of leadership orientation and trust. *Educational Administration Quarterly, 45*(2), 217-247.

Tschannen-Moran, M., & Hoy, W. (1998). Trust in schools: A conceptual and empirical analysis. *Journal of Educational Administration, 36*(4), 334-352.

U.S. Department of Education. (2002). *No child left behind act*.

Vernaza, N. A. (2012). Teachers' perceptions of high-stakes accountability in Florida's Title I elementary schools *Current Issues in Education* (Vol. 15). Phonix, AZ: Arizona State University.

von der Embse, N., & Hasson, R. (2012). Test anxiety and high-stakes test performance between school settings: Implications for educators. *Preventing School Failure, 56*(3), 180-187.

Walker, K., Kutsyuruba, B., & Noonan, B. (2011). The fragility of trust in the world of school principals. *Journal of Educational Administration, 49*(5), 471-494.

Watson, C. E., Johanson, M., Loder, M., & Dankiw, J. (2014). Effects of high-stakes testing on third through fifth grade students: Student voices and concerns for educational leaders. *E Journal of Organizational Learning & Leadership, 12*(1).

Wellman, N. (2007). Teacher voices: The impact of high-stakes testing on teacher caring. *Teacher Education and Practice, 20*(2), 204-216.

White, K., & Rosenbaum, J. (2013). Inside the blackbox of accountability: How high-stakes accountability alters school culture and teh classification and treament of students and teachers. In A. R. Sadovnik, J. A. O'Day, G. W. Bohrnstedt, & K. M. Borman (Eds.), *No Child Left Behind and the Reduction of the Achievement Gap: Sociological Perspectives on Federal Educational Policy*. New York, NY: Routledge.

Williams, T., Kirst, M., & Haertel, E. (2005). Similar students, different results: Why do some schools do better? A large-scale survey of california elementary schools serving low-income students. Washington, DC: American Institutes for Research.

Willis, C. (2011). High-stakes testing and the moral decisions of leaders. *Journal of Cases in Educational Leadership, 14*(4), 47-53.

Winters, M. A., Trivitt, J. R., & Greene, J. P. (2010). The impact of high-stakes testing on student proficiency in low-stakes subjects: Evidence from florida's elementary science exam. *Economics of Education Review, 29*(1), 138-146.

Yeh, S. S. (2005). Limiting the unintended consequences of high-stakes testing. *Education Policy Analysis Archives, 13*(43).

APPENDICES

APPENDIX A

INTERVIEW CONSENT FORM

Department of Counseling, Educational Leadership, Educational and School Psychology
Box 142, Wichita, KS 67260-0142

Focus Group Interview Consent Form

Purpose: You are invited to participate in a study that will examine how the new Kansas State Assessments have influenced elementary teachers' approach to preparing students for standardized tests.

Participant Selection: You were selected for participation in this study based on your position at the elementary school and length of employment with respect to having given the state assessment for several years. Approximately 24 individuals have been invited to participate in a focus group interview.

Explanation of Procedures: As a participant, you will be asked to be involved in a four to six member focus group interview conducted by me. The interviews will consist of 8-10 open-ended questions to seek your perception of how your teaching practices has changed. For example one question might be, "What school wide structures does your school have in place to ensure all your students do well on the assessment?" Or another, "What are the most effective strategies you use in your classroom to prepare students for the state assessment?" The focus group interviews will last approximately 60 minutes and will take place at your school. With your permission, I would like to audio record the interview so an accurate transcript can be created which will facilitate data analysis and assist me in reporting accurate findings.

Discomfort/Risks: There are no risks, discomforts, or inconveniences expected from your participation in this study. However, if a question makes you uncomfortable, you are under no obligation to respond. You also can skip any question you do not wish to answer.

Benefits: The purpose of this study is to examine how the new Kansas State Assessments have impacted elementary teachers' approach to preparing students for standardized tests. The results of this study will inform educators on how the Kansas State assessments have changed how teachers organize their classroom and instruct their students. This study hopes to add to the body of knowledge about educational and instructional triage. Results may be published in journals and presented at conferences so I can share with others what is learned from the study.

Confidentiality: Every effort will be made to keep your study-related information confidential. However, in order to make sure the study is done properly and safely there may be circumstances

where this information must be released. By signing this form, you are giving the research team permission to share information about you with the following groups:

- Office for Human Research Protections or other federal, state, or international regulatory agencies;
- The Wichita State University Institutional Review Board;
- Dr. Joel Abaya, my dissertation Chair at Wichita State University

Recordings and transcriptions of the interviews will be stored in a secure, online, password-protected program. At the conclusion of the study, transcripts, and recordings will be stored in a single password protected file and maintained indefinitely at WSU by my dissertation committee chair. Transcripts and recordings will not be labeled with identifiable information.

Focus group discussions are confidential to the group and all participants are asked and encouraged to please not share what was discussed outside of the focus group.

I may publish the results of this study. If I do, I will only discuss group results. Your name will not be used in any publication or presentation about the study.

Refusal/Withdrawal: Participation in this study is entirely voluntary. Your decision whether or not to participate will not affect your future relations with Wichita State University and/or myself. If you agree to participate in this study, you are free to withdraw at any time.

Contact: If you have any questions about this research, you can contact me at Jamie Finkeldei, (316) 393-7230 (cell phone) or my advisor Dr. Joel Abaya, (316) 978-6392 (office phone). If you have questions pertaining to your rights as a research subject you can contact the Office of Research and Technology Transfer at Wichita State University, 1845 Fairmount Street, Wichita, KS 67260-0007, telephone (316) 978-3285.

You are under no obligation to participate in this study. Your signature below indicates that:

- You have read (or someone has read to you) the information provided above,
- You are aware that this is a research study,
- You have had the opportunity to ask questions and have had them answered to your satisfaction, and
- You have voluntarily decided to participate.

You are not giving up any legal rights by signing this form. You will be given a copy of this consent form to keep.

_____

Printed Name of Subject

_____ _____

Signature of Subject Date

_____

Printed Name of Witness

_____ _____

Witness Signature Date

APPENDIX B

FOCUS GROUP INTERVIEW PROTOCOL

Hello, my name is Jamie Finkeldei. I am a doctoral student in the Educational Leadership program at Wichita State University. I appreciate your willingness to participate in my dissertation research entitled, "The Influence of High Stakes Testing on Elementary Classroom Instruction." This study will examine how the new Kansas State Assessments have changed how teachers teach. You were selected for participation in this study based because you began your teaching career before these new assessments so you can offer some comparison.

To begin, I would like to share a few procedures for our conversation. With your permission, I would like to audio record our conversations for response clarity and accurate analysis of data when reporting the findings of this study. Please give your first name before continuing with your response to a question. This will assist in following the flow of the conversation during the transcription process. After the initial transcription, the names will be replaced with pseudonyms and subsequent analysis will result in text without identifiers. Although we will be on a first name basis, no names or identifying comments will be used when I report the results of this session. You can be assured of complete confidentiality. This focus group session will last approximately 45-60 minutes. Again, thank you for your participation.

Faculty Focus Group Interview Questions

2. Please introduce yourself by stating your name, the grade level you teach, and the number of years you have taught in the classroom.
3. What are your thoughts on the new Kansas state assessment that you gave last spring?

4. What school wide structures does your school have in place to ensure all your students do well on the assessment,
    a. for example MTSS, tier groups, tutoring, summer school, etc?
5. What other school wide strategies has your school implemented to ensure all your students do well on the assessment,
    a. for example pep rallies, incentives, posters, cheers?
6. What do you think works best in your classroom to prepare students for the state assessments?
7. How has preparing for the state assessment changed your classroom instruction?
8. What are your thoughts on the new Student Growth Measures (SGM)?
    a. Does that change how you view the assessments?
    b. If we got rid of state assessments do you think students would learn more or less? Why?