

**THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE**

DEPARTMENT OF CURRICULUM AND INSTRUCTION

**“HIGH STAKES STANDARDIZED TESTING IN THE UNITED STATES:
TRENDS, CONSEQUENCES, AND ISSUES”**

**HEATHER C. SMITH
Spring 2011**

**A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Elementary and Kindergarten Education
with honors in Curriculum and Instruction**

Reviewed and approved* by the following:

**Dr. Jacqueline Edmondson
Associate Dean for Undergraduate and Graduate
Studies
Thesis Supervisor and Honors Advisor**

**John Tippeconnic
Professor of Education and Director of the American
Indian Leadership Program
Thesis Second Reader**

***Signatures are on file in the Schreyer Honors College.**

Abstract

Education is ever changing to provide future generations the best possible education. In spite of past educational reforms, there still remain gaps in student achievement that occur in predictable ways on socioeconomic status and race. Policymakers have turned to standardized testing to resolve this apparent problem. The implementation of high-stakes standardized testing in public school has created clear positive and negative consequences. There exists a set of very long-term historical testing data. The Chinese government first administered written exams, known as the Keju, to candidates working in the Civil Service, a highly respected occupation. As the system with the longest and best-preserved records, the Keju system exhibited the same positive and negative consequences seen in today's public schools as a result of high-stakes testing. In the mid-1800's, testing began on American soil in the form of entrance exams. Since then, pressure for high scores on standardized tests has increased from diagnosis testing, to evaluation testing, to test-based accountability. The pressure for good test scores, high stakes, reached an all-time high after No Child Left Behind (NCLB) was signed into law in 2001. Public schools are now rewarded or punished based on their student population test scores. Punishments range from closing schools to losing government funding. With severe punishments lurking on test results, standardized testing has created a very negative environment in public school systems. Teaching to the test, test cheating, and unhealthy competition between schools are a few of the issues surfacing in public school systems as a result of high-stakes testing. It is

important to note that standardized testing is a good indicator of a student's performance, but not an exact measure of a student's academic ability.

Standardized testing is very practical in the sense that they are easy to grade, easy to track progress, fast, and relatively inexpensive. Standardized tests only have negative consequences when they are used as the sole measurement of a student or school district's educational quality. In light that past reforms have not fixed all gaps in student achievement, are policymakers correct in turning to standardized testing to solve this apparent problem?

TABLE OF CONTENTS

Introduction	1
History and Various Uses of Testing	3
Reliability and Validity	13
Testing Bias	18
The Positive Aspects of Standardized Tests When Used As Intended	22
With High-Stakes Consequences, Standardized Tests Have Negative Impacts	26
Is NCLB-Mandated High Stakes Standardized Testing Even Working?	35
Conclusions	37
Works Cited	40

Introduction

The purpose of this thesis is to better understand the history of standardized testing and also to explore the issues related to it. I wanted to explore both the advocates and opponent's claims for standardized testing in public schools. This thesis will give educators and other people with interests in education the opportunity to consider testing from different perspectives and gain insight into a current important issue. People who read this paper will become better informed as an advocate or opponent of standardized testing.

To explore this issue, I conducted an extensive literature review. I read various books and journals both supporting and opposing standardized testing in public schools. I read publications from the biggest names in standardized testing such as Alfie Kohn, Nicholas Lemann, Lucy Calkins, and Diane Ravitch. Diane Ravitch's book was the most current, which explains her involvement with the testing movement in 2010. I purposely chose to read publications both praising and opposing standardized testing to collect data and consider all sides of the issues around standardized testing. I researched publications through various online catalogs to ensure that I was reading the most relevant material.

As I read through research, I purposely sought out materials that included the voices and perspectives of teachers. I also spoke informally with several teachers from different school districts to see if their perspectives coincided with what I was reading. I learned that it is difficult to find support for high stakes standardized testing from classroom teachers.

This paper will begin with the history and evolution of standardized testing and its uses throughout the world. I will explain key features of standardized tests, including reliability and validity, and then I will then identify the positive and negative consequences impacting education as a result of high-stakes testing in K-12 public schools. The paper will then conclude by suggesting ways for the reader to speak out in his or her community about the dangers of high-stakes testing in public schools today.

Current Contexts

Today's society is a "testing culture." There are tests in elementary school, tests to go to college, tests to get a driver's license, and tests to get a job in some professions (Koretz, 2008). The APGAR test is even administered within seconds of a baby's birth. Why does society, and even school systems, turn to testing? Standardized tests are considered to indicate a person's mastery of a large range of skills and knowledge, and the government, employers, parents, teachers, and principals rely on these tests to draw conclusions about the test-taker's competencies (Young, 2005).

Tests come in many forms, and public school systems in the United States rely on many types of formative and summative assessment, such as midterm and final exams. Other tests that schools face today are reading tests, such as running records, dibbles, and the vocabulary STAR test. College-bound students graduating high school also take the Standardized Achievement Test or American College Test. Public school children may also take achievement tests, such as the Iowa Test of Basic Skills, which measure specific knowledge and

skills in particular content areas such as science, math, English, and social studies. In this paper, I am focusing exclusively on the use of standardized testing. A standardized test is a test designed in such a way that the questions, conditions for administering, scoring procedures, and interpretations are intended to be as consistent as possible (Popham, 1999). Standardized testing is an effort to make all assessments measuring academic achievement across the nation as equal as possible so that comparisons can be made across groups.

In order to better understand standardized testing in schools, I will examine their increased use over the years and also their consequences, both positive and negative. This provides one the opportunity to look at testing from different perspectives. With an estimated 140-400 million standardized tests given each year in the United States, one may also gain insight into an important current issue facing schools (Young, 2005). Readers may also speak out in future school board meetings and inform others regarding their feelings about high stakes standardized testing in schools.

History and the Various Uses of Testing

There is quite a long history related to standardized tests. The Keju examination system is perhaps the first documented historical use of testing in the World. The Keju examination system originated in the year 606 and officially ended in 1905, which is a total span of 1,298 years (Suen, 2006). The Chinese government administered written exams, known as the Keju, to

candidates working in the Civil Service, a highly respected occupation (Young, 2005). In addition to high stakes employment testing, the government also orchestrated educational testing and a test-driven education system. These tests were considered high stakes because they determined who would eventually serve as grand councilors and other high-power positions. In the hierarchical Confucian society, success on these tests also decided overall class, power, status, and prestige (Suen, 2006).

The next recorded use of testing was throughout the Middle Ages. In the 15th century, a scientist from Germany, named Gutenberg, invented the printing press, which made books and tests relatively inexpensive to produce. By the 16th century, many Europeans could read and attend school, which is when European students began taking written entrance exams. European universities were the first universities in the world to use tests as part of their entrance requirements (Young, 2005).

Standardized testing was first used in the United States in the mid-1800's. Boston introduced written exams for the government-funded schools in 1845. Harvard University followed shortly after with the first American college entrance exams in 1851 (Young, 2005). It wasn't until later in the eighteenth century that a standardized entrance exam was developed for colleges across the United States.

In 1909, public schools in America began using the Thorndike Handwriting Scale. Schools chose to use this form of a test to measure the quality of a student's handwriting. This was the first use of testing in the public school

system. It is important to note that this test was administered to every student in the school. The Thorndike Handwriting Scale was also presented to children in grades 5, 6, 7, and 8.

Although Edward Thorndike (Thorndike Handwriting Scale) developed an early multiple-choice test, Frederick J. Kelly was the first to use his multiple-choice test as part of a statewide assessment. Kelly's multiple-choice test was created in 1915 and was called the Kansas Silent Reading Test. Multiple choice offered a quick way to test and score ability of a large group (Merritt, 2006).

Standardized testing was first administered in a large-scale format during World War I (Young, 2005). These tests were standardized multiple-choice IQ tests given by a specialist. The purpose of these tests was to assess the intelligence of the World War I military recruit and do so in the most efficient way. The better a recruit scored on the multiple-choice test, the higher their rank would be.

In 1916, Alfred Binet developed the intelligence test. He was commissioned by the French government to find a method to differentiate between children who were intellectually 'normal' and those who were inferior. Binet developed an original assessment that he tested on Paris school children. The assessment consisted of tasks such as following commands, copying patterns, naming objects, and putting items in order. From that data, he developed the intelligence quotient (IQ). According to Binet, a purpose for designing the intelligence test was to "identify students who could benefit from

extra help in school” (Learninginfo.org, 2010). Binet’s intelligence test has been the model for every IQ test ever since (Young, 2005).

Following Binet, other tests for specific aptitudes were developed, such as Rorschach’s famous inkblot test. Although the development of the inkblot test isn’t easily traceable, it is said that the use of inkblots to stimulate imagination can be traced back to the 15th century. Rorschach’s inkblot test was used for a diagnostic purpose within the psychological realm. For example, data from the inkblot test could diagnose a thought disorder.

Throughout the early 1900’s, entrance exams were implemented as a requirement for most colleges and universities. The College Entrance Exam Board developed a series of written exams that would test a potential student’s knowledge of math, science, literature, and Latin. This exam was graded by professors and eventually led to the development of the Scholastic Aptitude Test. The SAT, was developed in 1921. This was a multiple-choice exam that made grading easier and faster. The SAT was adapted from the Army Alpha test previously mentioned in WWI. The SAT was thought to have measured pure intelligence, regardless of the quality of the test taker’s high school education and was only used for scholarship applicants when applying to college. Much later in 1942, the SAT became a test for all college applicants. In 1948, the Educational Testing Service was chartered and the SAT was on its way to becoming the basic college admissions device for millions (Frontline, 2010).

From 1930-1960, the low stakes testing era, standardized tests were used as “global assessments of a large range of educational skills” (Ward, 1999, p.48).

This era was considered “low stakes” because there were no dire consequences or huge rewards for student performance on a standardized test. There were also very few testing programs *imposed* on schools; however, many school districts purchased and administered achievement tests annually (Koretz, 2008).

One example is the Iowa Test of Basic Skills (ITBS), which was one of the five commercially published low-stakes tests published in 1995 (Langerfeld, K.L., Thurlow, M.L., & Scott, D.L., 1996). The IBTS was originally used as a tool to improve instruction. There are various content sections that measure specific skills, such as vocabulary, word analysis, reading comprehension, listening, language, math, social studies, and science. This standardized test solely indicates a single student’s strengths and weaknesses in a particular content area. It does not determine their full educational ability in relation to their grade level curriculum.

The Elementary and Secondary Education Act (ESEA) was enacted in 1965 as part of President Lyndon B. Johnson’s War on Poverty. ESEA required all schools receiving federal money to prove that they were accomplishing educational goals through the results of standardized tests (Young, 2005). Then Attorney General Robert F. Kennedy suggested that standardized tests be required to demonstrate that the money spent on public education was helping students to learn. These federal funds were allocated to high-poverty schools. It is important to note that these funds were not given as a result of high or low test scores.

Shortly after ESEA was implemented, the National Assessment of Educational Progress (NAEP) was launched in 1969. This testing program was established to assess educational progress of the United States as a whole. The focus of this test was not on the individual student, but on what America's students know and can do in various subject areas as a whole. NAEP, which is administered by the National Assessment Governing Board (NAGB), tested students of the ages 9, 13, and 17 in areas such as writing, science, citizenship, and computers (Ward, 1999). Since 1969, the test has grown to assess more academic subject areas such as the arts, economics, geography, mathematics, reading, science, U.S. history, and writing. These assessments are given by pencil and paper once every two years. The initial results are released about one year after administered. This federally funded program was one of the few testing programs of this era to have no consequences for poor results.

NAEP can be contrasted to required state tests, which measure student performance on the state's own curriculum standards. State tests allow comparisons of results over time within the state. State mandated tests also have the ability to let parents know how their child is performing in relation to other subgroups within the state. On the contrary, NAEP is the only assessment that allows comparison of test results from one state with another, or with results for the rest of the nation. NAEP does not tell parents how their child is performing in relation to their state's grade level curriculum standards.

Two years after NAEP was established, the first minimum competency-testing program was created. These tests were designed to ensure that all

students could reach an acceptable minimal level of mastery of basic skills (Koretz, 2008).

Other minimum competency tests focused on individual students and not groups of students. These tests were often used as “promotional gates” to either move on to the next grade or determine that a student was ready to graduate from high school. However, students were well aware that these tests were typically easy and had a low minimum score (Koretz, 2008).

One example of a promotional gate exam is the Regents Exit Exam used in New York, one of the most established high school graduation exams. The first Regents examinations were administered in November of 1866. Originally, there were only five exams that assessed algebra, Latin, American history, natural philosophy, and natural geography. In 1897, the New York State Education Department expanded the Regents Exams to forty-two different tests. The Regents Exams increased the number of states with mandated statewide testing programs (Koretz, 2008). By the end of the 1970’s, 60% of the states in the United States were using mandated testing programs.

Another federally-mandated testing program was authorized in 1974, the Title 1 Evaluation and Reporting System (TIERS). This testing program required evaluation of Title 1 fund-receiving schools through student scores on standardized achievement tests (Koretz, 2008). At this time, schools were not held accountable for their student’s performance, but just had to demonstrate that their students were taking them.

As scores were collected and compared from one year to the next, education critics began to comment on the perceived nationwide decline of performance on these tests. This led to demands for higher education standards and the advancement in the accountability movement (Koretz, 2008). Throughout President Ronald Reagan's administration, there was a shift of public opinion toward a belief that every student should attend college (Levine, 1993). International comparisons on similar tests to NAEP, such as the Program for International Student Assessment (PISA), an international evaluation of fifteen-year-old school pupil's scholastic performance, indicated that US students' cores were lower than many other countries in reading and mathematics. These international comparisons led to the publication of *A Nation At Risk* in 1983 by the Congressional Research Service. This report claimed that the declining test scores indicated a "rising tide of mediocrity" in US schools (Fowler, 2010). In addition to criticizing American public schools, the authors of this report hoped to motivate educators to make the education system better for *all* students (Ravitch, 2010). This report instigated educational reforms marked by an increase in the reliance on standardized testing and accountability (Koretz, 2008).

During President George H. W. Bush's administration, an education bill for national tests and standards was proposed, but the legislation died in Congress. The Senate passed a resolution condemning the standards by a vote of 99 to 1. All national education standards were condemned as unlawful and deleterious federal dabbling in local and state affairs. National standards would

have held all states accountable for teaching students the same content in every subject.

In 1994, ESEA was reauthorized. When Clinton was running for office, he promised assessments for accountability. He created the “Goals 2000” program for states to write and pick their own standards, tests, and curriculums (Ravitch, 2010). Clinton’s administration spent more time on other aspects of educational reform such as increasing school choice, making college more affordable with work study grants and scholarships, hiring more quality teachers, reducing class sizes, and increasing the education budget. Although Clinton didn’t build his education platform around standardized tests, on February 5, 1997 in his State of the Union address, President Clinton argued that “good tests will show us who needs help, what changes in teaching to make, and which schools to improve. They can help us end social promotion. For no child should move from grade school to junior high, or junior high to high school until he or she is ready.” Although Clinton didn’t create new standardized testing requirements or legislation, he believed in assessments for accountability in the public school system.

Clinton pushed the standards movement out and replaced it with the “accountability movement”, which took off full force in the twenty-first century. Accountability was for *every* student, including those with disabilities and those who had English as a second language. During George W Bush’s administration, the consequences of standardized testing intensified with the reauthorization of the ESEA as No Child Left Behind. Under NCLB, *every* student had to be

“proficient” in reading and math by 2014. NCLB required schools to test its students annually in grades three thru eight, and also in one secondary grade (Koretz, 2008).

There were severe sanctions for low and poor performing schools that did not meet Adequate Yearly Progress (AYP). AYP determines whether a public school or school district is making annual progress towards the academic goals established by each state. Each state’s goals call for improvement to eventually meeting the NCLB goal in 2014. The sanctions for not meeting AYP could consist of a warning, loss of federal funding, or the closing of that school. These high stakes consequences often times can produce a fearful school environment, rubbing off on the administrators, teachers, and eventually students. Although all students in the same state took the same standardized test, the results were disaggregated for different racial groups, English language learners, disability students, and low SES students (Koretz, 2008). With such pressure on schools for good test results, standardized test-based accountability, not standards, became the national education policy (Ravitch, 2010). When test scores became the nation’s education policy, standardized testing received its negative stereotype. Standardized testing has many practical uses, but the sole measurement of a school, student, or teacher is not one of them.

When President Barack Obama took office in 2008, he promised to fix what he considered to be NCLB’s demanding requirements.. In March 2010, he published a blueprint of his NCLB reauthorization plan. The reauthorization retains the structure and spirit of the Bush-era NCLB law, such as annual testing

and data-driven accountability, but it also adds resources and flexibility to meet new goals (Chaddock, 2010). Obama's new goal for his revision is that by 2020 all students' will graduate from high school prepared for college and a career (Quiad, 2010). Obama's plan is intended to relieve some of the pressure schools face in relation to AYP, measuring drop out rates, and every school having 100% proficiency on standardized tests. Instead of punishing failing schools, Obama's new plan provides resources to help them. Five percent of the lowest performing schools in each state will be identified and then strong measures to upgrade those schools will follow (Quiad, 2010). Obama also plans to reward schools for progress, particularly with poor and minority students. The NCLB revision also includes testing other subjects besides reading and math to avoid narrowing the curriculum in schools (Chaddock, 2010).

With all of these goals and revisions in mind, the NCLB reauthorization could be the next big step in the history of standardized testing. Whatever the outcome, testing is a part of the education system in the United States, and it will continue to be in our future. Because of this, every citizen should speak out about the correct and incorrect uses of standardized testing in schools. In order to do so, a basic understanding of the key features of standardized tests is needed, particularly in relation to the reliability and validity of the tests.

Reliability and Validity

Aside from all of the negative aspects of high-stakes standardized testing in schools, tests simply cannot be trusted because they are error-prone. When

tests lose their full reliability, they should not be used in schools for high-stakes purposes. The NCLB law mandates that testing “shall be used for purposes for which such assessments are valid and reliable” (Koretz, 2008).

In the field of educational assessment, the “validity of assessment” can be defined as the degree to which evidence and theory support the interpretations of test scores that are entailed by proposed uses of the test (American Educational Research Association, 1999). Generally, most large-scale state and commercial standardized tests are achievement tests that measure the knowledge, skills, and abilities (Zucker, 2003). One of the primary purposes of the Pennsylvania State School Assessment (PSSA) is to give schools in Pennsylvania an accurate assessment on how well their students are performing at the time that the test is administered. Anytime that the test does not indicate this, the test becomes invalid.

“Construct irrelevance” refers to a situation in which a test score is indicative of some object other than the intended construct being tested (Suen, 2006). For example, a geography test should examine how much knowledge a student holds about geography. When the geography test score indicates the student’s family income rather than knowledge, the test has no validity. The most frequent discussed sources of “construct irrelevance” are teaching to the test, narrowing the curriculum, test cheating, and mastering test-taking skills (Suen, 2006).

Standardized tests are error-prone, which effect both validity and reliability. Test error occurs when a test score indicates something false about

the student taking the test. Whether a question is poorly worded, a student didn't get sleep the night before, or the grading machine missed a question, a test error takes place (Young, 2005). The resulting test score may give inaccurate information about the student who took the test, therefore making the score unreliable.

Many students take the SAT numerous times before settling with a final score. When students re-take the SAT, their test score is often slightly different every time (Koretz, 2008). The student never takes the same SAT twice. There are many different versions of the test, all which are said to test the same skills at the same levels. Different variations of the same test may prove to be unreliable because a student often receives a different score each time the test is taken. Ravitch (2010) states that some tests may even be worded poorly, have ambiguous answers, or have two answers that are equally correct. An error in mechanics may cause a student to receive a misleading score. In these cases, the reliability of the test score is significantly lowered.

Another entity that a standardized test score does not reflect is the 'outside conditions' when a student is taking the test (Ravitch, 2010). Particular room circumstances such as temperature, noise level, and who is sitting next to whom may effect the way a student performs on a test. Student performance may also be affected by their state of mind or distractions outside of the classroom. A student may be fighting with his or her best friend, only slept for 4 hours the previous night, or missed breakfast that morning. All of these scenarios would cause a student to perform at a lower level than typical of that

student. The way a student handles standardized test taking may also skew the scores. One student may care so much about their score that they are overly nervous and perform poorly on the test. When a student has test anxiety, they will perform significantly lower than what a teacher knows that student can do. The higher the stakes, the higher the anxiety, and the lower the reliability of the test scores (Kohn, 2000). Timed tests also increase nervousness in anxious students. On the flip side of test anxiety, students may care less about the test because the score is not reflected in their final grade. Students may 'blow off' the test by coloring in random ovals to complete the test quickly. All of these outside factors are not taken into consideration when examining the validity of a test score.

Another unreliable source of error found within standardized testing is an inconsistency in scoring. For multiple choice "bubble" questions, a machine will scan the answer sheet and provide immediate results. These technological machines are great for saving time and effort, but their accuracy may be questionable at times. A well-publicized case in October of 2005 made known of possible machine grading errors. Over four thousand student's SAT scores were lowered about 130 points due to a machine glitch when scanning their answer sheets (Pope, 2006). A case of moisture made the machine shift student answer sheets so that they didn't line up correctly with the sensor. Some students may not have been accepted into their college of choice due to that mistake. Therefore, one should watch how much trust is being placed on these machine-graded exams. It may not happen often, but machine errors can occur. Other

errors in standardized test grading are found in graders for the writing section. There are multiple graders across the nation for the writing and essay portion of the test. Although trained together, some graders may grade more harshly than others (Koretz, 2008). Some graders may even change the way they grade to be like other graders. A high-level student may receive a low writing score because he or she had a harsh grader and a lower-level student may receive a high writing score because he or she had a light grader. The perception of these two students in writing may be skewed due to the grading process. Any source of error in the process of standardized testing lowers the reliability of the test score.

Another element to consider with test score reliability is the proficiency score of each state. NCLB demands that by 2014, every child in every state must score proficient on the standardized test of its state. However, there is no set score that determines a national proficiency. Proficiency may vary from the 6th to the 77th percentile (Ravitch, 2010). This is such a big difference that if a student is proficient in Colorado, they could be failing in Massachusetts.

The system with the longest and best-preserved records, the Keju examination system revealed these same problems with reliability and validity throughout history. The fact that these problems persisted in the Keju system, despite repeated attempts to remove them, suggests that standardized tests are not the most valid and reliable test. When measures are taken to remove these problems, they simply mutate to new forms (Suen, 2006). In high stakes testing, success is defined by a high score on the test. The factors of test-taking skills, test

cheating, and curriculum narrowing appear to be so inherently chronic under high stakes conditions that they defy preventative measures (Suen, 2006).

Since NCLB law mandates that tests should be used for assessments when “valid and reliable”, a standardized test should not be used as the sole measurement of a student's academic ability or proficiency. The National Research Council agrees in stating, “tests are not perfect, an educational decision that will have a major impact on a test taker should not solely be made on the basis of a single test score” (National Research Council, 1999, p.276).

Testing Bias

Another characteristic to take into consideration about standardized testing is test bias. Standardized test questions may be biased against the students taking the test. This is another source of test error that may lower reliability. A test bias is a systematic distortion of scores that undermines the validity of a particular student or subgroup taking the test (Young, 2005). As previously mentioned, teaching to the test and inflation from testing materials are ultimately forms of test bias. One of the biggest faults of standardized tests is that they often do not take diversity into account (Ravitch, 2010). Student test anxiety, disabilities, cultural differences, language differences, gender, and socioeconomic status are all diversities amongst our students today. Test questions do not take these diversities into account and therefore may undermine the validity of a particular student score.

Gender is not so much an inherent bias within the test. However, there are different expectations of males performing well on the math section and females performing well on the reading section. This may influence a student's performance going into the testing atmosphere with these stereotypes in mind (Young, 2005). A student's socioeconomic status, however, is an inherent bias within the test. Tests require a set of knowledge and skills more likely of a privileged background (Kohn, 2000). Students who have attended preschool, own computers, overheard conversations of current events from parents, and have been on interesting trips are considered to have a privileged upbringing. When a student is lacking one or few of these experiences, they may be placed at a disadvantage on the test. Content or wording on the test may distort the performance of a low socioeconomic student. For example, there may be a math word problem on the test about 'yachting'. A poor student most likely will not know what that means and may answer the question incorrectly solely on the misunderstanding of the word (Koretz, 2008). A student's location could also double as a similar test bias. If the same question about yachting was administered to a student in a western landlocked area, they may fail to know the meaning of a lanyard, which would obscure the competence of the Midwest in math.

A student's gender, socioeconomic status, and location are all subject to test bias. However, the two biggest subgroups impacted by test bias are children with special needs and English language learners (Koretz, 2008).

The 1990's increased the participation of special education students in large-scale testing programs to ensure that they were assessed in the same fashion as mainstream students. NCLB requires that 95% of students with disabilities *must* be included in state and district-wide testing programs with "appropriate accommodations where necessary" (Thurlow, 2001). Due to the indistinct nature of accommodations in the law, different states interpret the meaning inconstantly. One state may even permit an accommodation that another state prohibits. The purpose of these accommodations is not to raise the test scores of special needs children, but rather to help them score as well as their proficiency warrants them to score (Koretz, 2008).

Although accommodations are a violation of standardization, they increase the validity of test scores for students who have disabilities. An example of an accommodation for a blind student would be having a Braille version of the test. An example of an accommodation for an orthopedic student would be having that student dictate her answers rather than writing them down. Once again the purpose of accommodations is not to raise their test score, but rather help each student score in a way that reflects his or her abilities. Providing insufficient accommodations for a student may cause him or her to score lower than his or her knowledge or ability level. Providing too many accommodations for a particular student may cause him or her to score higher than his or her knowledge or ability level. Both of these scenarios indicate a test bias because the accommodations, whether too little or too many, undermine the validity of a student's test scores.

Learning disabilities are the largest classification of a disability and they are also the hardest to assign a fair accommodation to for a student (Koretz, 2008). For example, a dyslexic student may need help reading the questions on the test, but not drawing an inference to answer. Reading the test to him or her would fundamentally change what the test measures: to first decode, then infer meaning (Koretz, 2008). Not reading the test to the dyslexic student would significantly decrease the child's score, and reading the test to the dyslexic student may increase a child's score. Either way, there would be a test bias for this student. Under NCLB, there are serious consequences for special needs students not meeting AYP but only a few risks in providing excessive accommodations. Therefore, many schools are over-accommodating and providing an unfair advantage to these students (Koretz, 2008).

The second biggest test bias subgroup is students who are English language learners (ELL). The biases for these students are similar to assessing students with disabilities, even though having English as a second language is not a disability. ELL students come from different cultures. Some questions, even if read correctly, could be confusing to some students due to the different cultural traditions (Koretz, 2008). An intelligent Israeli student may have cultural or linguistic factors that undermine his or her aptitude from poor test results. The most common ELL test bias is from linguistic barriers. A lot of American words or phrases have polysemy, words that have multiple unrelated meanings. For example, "cutting a price" means reducing the price, not cutting it with a knife (Koretz, 2008). Non-native English speakers may only know the

more common meaning of any word. The most obvious ELL test bias is when a student cannot speak or reading English fluently. If a student cannot read the question, they cannot give an appropriate answer. A test translator would be the most apparent accommodation given to an ELL student, however; there are so many different languages in schools today that a translator in every school for every language is seemingly unrealistic (Koretz, 2008). A translator reading the test to an ELL student may also fundamentally change what the test measures, as in the case of the dyslexic student. There are also advantages with certain languages compared to others. German and Spanish have a similar structure to English, whereas Mandarin is completely different. All of these situations pose a testing bias that undermines the validity of a particular student or subgroup taking the test.

The Positive Aspects of Standardized Tests When Used As Intended

With all of the negative publicity standardized testing has received from educational activists, it is a common misconception to think that these tests are “bad”. The negative impact of standardized testing in schools comes not from the tests themselves, but rather from the consequences of their misuse. Allocation of funds, meeting AYP, and possible closing of failing schools are just a few examples of the consequences of high stakes standardized tests. In these cases, the students’ test results are used as the sole measurement of the education they have received. It has been stated that test scores are only an *indicator*, not the

definition of the education a student has received (Young, 2005). The information taken from tests can be extremely valuable for certain uses if the tests themselves are valid and reliable (Ravitch, 2010).

Test results can show students and parents what the student has learned, what they have not learned, and what they need to improve. To some extent, these tests can be used to make comparisons across groups of students. A correct use of a test would be to assess a student's comprehension of a particular unit taught in class. Teachers may use the test results to see how much of a particular lesson was truly taken in by the students. The correct use would be to test what is taught, not to teach what is tested (Ravitch, 2010). Such comprehensive exams are valid and reliable, because they reflect the information taught in class.

Standardized tests are a very practical assessment. These tests are fast, easy to administer, and relatively inexpensive compared to performance assessment (Kohn, 2000). While standardized tests quickly determine a student's knowledge, performance assessment determines a student's use of knowledge and problem solving skills. Performance assessment takes a long time to administer and grade for each student. So why wouldn't standardized tests be used? Standardized tests are less time consuming than more complicated assessments that require personal time with each and every student. No school district has the time during the school year to fully evaluate each student's performance individually for every single subject. A standardized test is a quick measurement with quick results. The standardized test also is a

snapshot overview of a person's mastery of a large range of skills and knowledge (Young, 2005). As previously mentioned, this overview should not be the sole measurement of a student's academic achievement.

The standardization of each test is a step towards equality for all test-takers. This holds all students, across the nation, accountable for the same material (Young, 2005). Standardization attempts to level the playing field with many diverse school systems. Each student is exposed to the same content, as well as the same format of the test. These tests are also easy to administer because there are explicit directions that come with each test packet. All students are delivered the same directions in the same way (Young, 2005).

Standardized tests also save time with easy grading. A teacher or administrator doesn't have to go through each individual test and grade written work. With standardized testing, a machine does the grading work for the evaluator, which saves time. The machine grading is also consistent and non-biased about the student answer sheet being graded (Young, 2005). The objectiveness of the grading machine is another positive aspect of using this form of assessment. Both the test and grading machine lower the risk of human error. For example, a teacher who may be a bit harsher when grading tests, or a teacher who is in a hurry may make errors or overlook problems when grading a test. A grading machine will mark an incorrect response as wrong and a correct response as right. A grading machine does not have any emotional ties to a student to judge their test differently. When a teacher is assessing their students, their emotions and prior knowledge of the student may become a bias and affect

how they score the student (Young, 2005). Therefore, objectivity is a positive aspect of standardized testing.

Another aspect of standardized testing highly looked upon is the ease of tracking a student's progress from year to year. With mountains of data stored on electronic files, computers can easily track the progress of an individual student over many years. Standardized tests also group students together by race, gender, and SES, which can allow schools to track the achievement of these different performance groups over time to make improvements in instruction (Young, 2005). Tracking progress from school year to school year goes hand in hand with accountability. As noted earlier, under NCLB, schools must show an adequate yearly progress. With the help of standardized testing, tracking the data of these groups of students becomes very easy. Each school can set a cut score for students and hold them accountable to it. Standardized testing can also hold schools across the nation accountable for the same cut score (Young, 2005). Tracking progress and setting high expectations for students on standardized assessments may force students to take their education more seriously (Nathan, 2002).

Standardized tests are very practical because they are fast, easy to administer, and relatively inexpensive. Computers storing data make it easy for schools to track the progress of their students and student sub-groups from school year to school year. Standardized tests are also very objective and are non-biased when grading each answer sheet. Lastly, the standardization of each test holds each student across the nation accountable for the same content

standards. With no dire consequences or rewards pending on the scores of a test, standardized testing is a positive tool in assessment for schools to use freely.

Very few teachers cited in my research had anything positive to say about the standardized tests they are required to administer to their students. It is also important to note that little to nothing is documented in literature where teachers support high stakes standardized testing.

With High-Stakes Consequences, Standardized Tests Have Negative Impacts

A current belief is, “testing ruins schools”. In an interview regarding his book, *The Case Against Standardized Testing*, Alfie Kohn stated, “tests damage low-income students by reducing their education to test-prep”. In concurrence with Kelly Gallagher’s book, *Readicide*, Kohn also states, “children are not coming home chatting excitedly about school or wanting to read a book for fun” (Kohn, 2000). Additionally, Kohn states that tests are ruining school spirit. The question of “how well are kids learning” has been changed to “who is beating whom”. There are other professional opponents who are in agreement with both Kohn and Gallagher’s books and beliefs. On the other hand, the advocates of standardized testing claim that tests are very practical, objective, and correct as an indicator of student achievement. Standardized testing becomes negative when it has high stakes and is used as the sole measurement of a student’s academic achievement.

As previously listed, standardized tests could have many positive uses in schools if used appropriately. Most recently, however, students, teachers, and parents view standardized tests in an increasingly negative way. In 1975, Donald Campbell, a well-known sociologist, stated, “the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption”(Patton, 2008, p.258). This became known as ‘Campbell’s Law’. After NCLB was passed in 2001, standardized test results were used to label schools as passing or failing, yet this was not the purpose of the tests when they were designed. The tests were designed to be indicators of what students learned in particular subject areas, not indicators of what they were taught or the quality of their school. The increased pressure resulted in the corruption of schools to meet AYP any way they could. This pressure, also known as high stakes, created negative consequences associated with standardized testing in schools. In high stakes testing, success is defined by a high score on a test (Suen, 2006).

Bess Keller wrote in the highly respected weekly education journal, *Education Week*, “no single measure should determine a student’s fate”(Keller, 1998, p.1). Under NCLB, students, teachers, administrators, and schools are judged solely on the measure of standardized test scores. Ravitch (2010) states, “not everything that matters in education can be measured; spirit of innovation, inquiry, imagination, and dissent are lost.” However, NCLB continues to use math and reading test results as the primary measure of students and schools. From the consequential pressure and high stakes associated with student test scores, standardized testing has had a severe negative impact on students,

teachers, and the K-12 public school systems over the past decade. Standardized testing does have many practical uses, but the sole measurement of a school, student, or teacher is not one of them.

One of the most widely talked about issues with high-stakes standardized testing is the narrowing of classroom curriculum. It is a common misconception that what is taught in the classroom and what is actually tested are the same thing. However, what students are tested on doesn't always match up with classroom objectives (Ormrod, 2003). Therefore, to improve student test results, teachers sometimes discard other topics planned in order to spend more time on concepts they know will be on the test. This is often referred to as "teaching to the test". When standardized tests only measure reading and math, other important subjects such as history, civics, literature, science, arts, and geography may be left out of the classroom agenda (Ravitch, 2010). A test should always follow curriculum, not replace it. However, NCLB only mandates testing, not a curriculum. Not only do children lose out on important subject matter, but they also lose social and moral development. With a class driven by reading and math, there is no time for class meetings, building a sense of community, creative play, or working on conflict-resolution skills (Kohn, 2000). These skills are important for creating a well-rounded, functioning citizen in society.

Robert Tobias, the New York City public schools testing director for thirteen years, has stated, "Much of test preparation is not designed to increase student learning, but rather to try to beat the test". For example, when a teacher is teaching his or her students about the Pythagorean theorem, one might say

“two of the most common ratios are 3:4:5 and 5:12:13”. What the teacher actually means is that those two ratios will most likely be used on the particular test they will take. The real world will not help a person out with right triangles legs only in the ratio of 3:4:5 and 5:12:13 (Rubinstein, 2000). Another example of teaching to the test took place in Montgomery County, Maryland in 2001. A teacher was reviewing for the county’s math exam with her students. A question on the review sheet read, “how much should each band member raise if b = number of band members, when $f(b)=\$1200/b$?” The question on the actual county test read, “how much money should each cheerleader raise when n = number of cheerleaders, and $f(n)=\$420/n$ ” (Strauss, 2001). With questions almost identical, a student will temporarily memorize the procedure to pass the test, not commit the procedure to memory. Another form of ‘teaching to the test’ is when a teacher teaches “testwiseness” (Burley, 2002). Testwiseness is teaching valuable test taking skills, such as how to guess on a multiple-choice question. When the curriculum content is narrowed and test-taking skills are taught, students potentially lose out on a rich and full education (Young, 2005).

Most all standardized, high-stakes tests are multiple-choice. The multiple-choice format limits teaching and learning to *knowledge*, at the expense of *skills and abilities*, such as critical thinking, creative thinking, and problem solving (Haladyna, 2002). Students unable to generate a response will just have to recognize or guess at one for a correct answer. Roger Farr, professor and standardized test writer from Indiana University, has stated, “I don’t think that there’s any way to build a multiple choice test that allows a student to show

what they can do with what they know” (Checkley, 1997, p.5). Children will never go beyond Piaget’s Concrete Operational Thought into Higher-Level Thinking on a multiple-choice exam. Having a student graduate with the ability to correctly chose one bubble out of four on a multiple-choice exam is not as important as a graduating a student well-educated and equipped to live a gratifying life.

The negative impact that high-stakes standardized testing has on students goes beyond the loss of history, civics, the arts, science, and literature. Students are also bribed with food, sports tickets, and exemption from finals and sometimes punished with no diploma for their test scores (Keller, 1998). These reinforcement tactics teach students that test scores are the only things important in education. With this lesson taught, testing undermines the love for learning and the desire for students to acquire knowledge. Both of these are ingredients of intrinsic motivation (Ravitch, 2010). Because a test score has no correlation with reading at home, students may have no interest in reading a book for their own enlightenment. It is difficult to maintain the same level of student motivation because teachers and students know that the tests are the primary measure of their success or failure (Ravitch, 2010).

Another negative impact on students from high stakes testing is through unhealthy competition. Before the implementation of high-stakes standardized tests, healthy competition was present in schools. Students would compete between schools in science fairs, essay contests, debates, and athletic events. Today, there is exceptionally unhealthy competition driven between schools,

such as receiving the highest test scores, not sharing resources, and teaching to the test (Ravitch, 2010). There is also unhealthy competition between individual students in schools. Students may now worry how their test scores compare to other students test scores instead of worrying about how well they are doing for the sake of wanting to do well. Those students whose scores do not compare well to those of high performing students may believe they are failing, even when they are not (Kohn, 2000). This leads to a loss of self-esteem and decreased motivation in school. This downward spiral of poor performance and negative views toward education are extremely difficult to break.

Students are not the only ones negatively affected by high-stakes standardized testing in K-12 public schools. As previously mentioned, teachers are pressured into narrowing the curriculum in which they teach to improve his or her class test scores. The accountability pressures corrupt the purpose of schooling by causing teachers to focus on the measure, rather than the goal of education.

The standardized testing movement most often affects teachers emotionally. Educators feel as though they need to defend themselves and what they teach. Under such pressure for their students to perform well, teachers may feel they need to teach to the test or cheat. This then may cause good teachers to lose their love for teaching and leave the schools (Kohn, 2000). Teachers may also feel emotionally biased against low-performing students. A superintendent in Florida observed, “when a low-performing child walked into the room, instead of seeing a challenge or opportunity for improvement, teachers saw the student

as a liability” (Wilgoren, 2001). Best-selling author and writer for the *New York Times* stated, “not everything of value that a teacher imparts on his or her students can be captured on a test” (Gladwell, 2008). However, teachers are feeling solely responsible for the scores his or her students receive. Teachers also feel helpless when students’ scores are publicized because they are not the only person who influenced that score.

When high-stakes standardized testing has a negative impact on the students and a negative impact on the teachers, the K-12 public school system as a whole is impacted negatively as well. Story lines and newspaper headlines every year read, “performance is getting better, scores increasing dramatically”, but this is only true at face value (Fuller, 2006). Test scores, indeed, may be increasing, but an actual increase in student learning and achievement is not. This is often called test inflation. Test scores may appear to be increasing because teachers are teaching to the test. The students lose out other valuable subjects, but their test scores in reading and math increase. Schools spending millions of dollars for test preparation materials also go hand in hand with teachers teaching to the test. Another form of test inflation is when schools use incentives, such as food and prizes, to boost student performance on test. Ravitch (2010) states that results from incentives are meaningless because gains that are purchased with cash are short-lived and have nothing to do with a real education. “Buying” good test scores with incentives and test preparation materials does not create an actual increase in student learning, which is the goal of education. Relocation of good teachers to failing schools and test cheating

are also forms of test inflation used today (Koretz, 2008). Ways to make *real gains* in education are by working more effectively, teaching more content, and working harder. These are hardly possible in a curriculum driven by high stakes standardized tests.

Test inflation is not the only corruption in K-12 public school systems. A problem on the rise is simple test cheating. Test results are often used to “determine funding, salary increases, and other administrative aspects” (Ormrod, 2003, p.538). With such an unwarranted amount of pressure on teachers and principals, scores are often raised by hook or by crook. This dishonesty of school systems takes four different forms (Ravitch, 2010). The first is old-fashioned cheating. Schools may correct student answers or even leak questions to the students before the test. A second form of test cheating is called “skimming”. Skimming is when schools restrict admission of low-performing students into their district (Ravitch, 2010). Schools may do this by requiring interviews with parents they know won’t show up, have students write essays, compare attendance records, or ask for letters of recommendation (Carnoy, Jacobsen, Mishel, & Rothstein, 2005). A third way of gaming the system is by reducing the participation of low-performing students on state tests. Schools may encourage those students to stay home that day or suspend them. School administrators may also shift students to other categories, such as special education or English language learners. The last form of test cheating is when schools make the test less challenging or lower the cut score. For example, the test scores in New York jumped from 65.8 proficiency in 2006 to 86.5 % in 2009.

To the public, those schools were improving, but in actuality, the cut scores were lowered to increase proficiency. It was also noted in 2008, that sometimes students in New York could receive enough correct answers to pass by randomly guessing (Senechal, 2009). All in all, the colossal pressure of standardized testing has forced school systems to turn to deceit. This dishonesty in schools today steers them away from the original goals of education.

In Ondine Rarey's production of *Testing Mrs. Grube*, a failing school in Central Harlem begins preparation for state-mandated tests in September that are administered in April. Even with a scripted three-hour daily test prep, Marie Grube's fifth grade students continue to fail every practice test, until she deviates from policy (Ondine Rarey Productions, 2007). In this documentary, teachers became managers and instruction was clearly teacher-centered, expository, and not responsive to individual developmental levels. The second part of the documentary, *A Different Standard*, acts as a corrective to the school in *Testing Mrs. Grube*. This documentary explores a different New York school in Central Park. At this school, learning is inquiry-based and children are clearly engaged and learning through hands-on discovery approach instruction. Like the Central Harlem school in *Testing Mrs. Grube*, this school is also required to take the standardized tests. The innovative school struggles to merge its instructional approaches with the reality of the tests. The teachers of the Central Park school defend their pedagogical positions in the face of the standards testing movement. These two documentaries, viewed as a set, show that there are two

clear paths schools can follow under NCLB; the path of teaching to the test, or the path of inquiry and innovative learning.

Throughout the research, teachers seem to agree that high stakes standardized testing negatively impacts public schools, administrators, teachers, and children. Lucy Calkins agrees with the theme in this movie by stating that a concern for high scores makes the curriculum match the test (Calkins, Montgomery, & Santman, 1998). Other teachers agree that narrowly focused curriculum can be harmful. Teaching to the tests may achieve higher scores in the short run, but these scores are *not* evidence that the students are strong in that subject (Calkins et al, 1998). Interviewed teachers have stated that the learning environment for children is less innovative and that teachers, too, have a rough work environment. Another common issue mentioned amongst teachers, literature and movies was the concern that not every student performs his or her best when taking a standardized test. Calkins mentions that test anxiety is *very* common (Calkins et al, 1998). Test anxiety affects the overall trustworthiness of the test results. Reliability and validity are two key factors in standardized testing.

Is NCLB-Mandated High Stakes Standardized Testing Even Working?

It is safe to say that if improvement of *educational achievement* is the end goal, testing is not working (Peterson, 2003). Despite twenty years of agitation and reform, mostly started by A Nation At Risk and NCLB 2001, student

achievement has idled, if not declined. In 2006-2007, 25,000 public schools didn't make AYP and in the 2007-2008 school year, 30,000 public schools didn't make AYP (Ravitch, 2010). Thirty thousand schools are around 30% of all public schools in America, therefore 30% of American public schools are failing to make AYP. As previously noted, the number of failing schools keeps increasing. Often times, after reviewing test results from various schools, the government spends more time punishing failing schools that don't do well rather than spending time helping those students and schools improve (Young, 2005). Federally taking over schools, removing resources, and cutting off funding are a few of the many ways failing schools are punished. These punishments do not raise test scores or improve student education.

In some school districts across the country, test scores have improved. Some may credit these scores to increased learning, but in most cases, test inflation or test cheating forces higher scores under high stakes. Teaching to the test will definitely improve test scores. Ravitch (2010) states that data reflects mainly the amount of time invested in test preparation activities. In these cases, the *education* of a student isn't improving, just the test score on a standardized reading, writing, and math test. Tests fail to measure "carrying out extended analyses, solving open ended problems, and displaying command of complex relationships" (Ayers, 1993, p.118). Even the reading questions rarely examine how students interrelate parts of the text and do not require justification or evidence of an answer. Tests are "a quick finding of answers rather than a reflective interpretation" (Resnick & Resnick, 1990, p.71). Because a

standardized test score is not a full measurement of a student's academic ability, requiring high test scores from schools are not going to improve education. Well-educated teachers, small class size, beautiful facilities, and curriculums rich in arts and science are what will improve education as a whole. As of today, nationwide mandated standardized high-stakes testing is *not* improving educational achievement in schools.

Conclusions

There are a few positive and several negative aspects of high stakes standardized testing in K-12 public schools. Standardized testing has been around for a long time and will continue to be used in schools in the future. The pressure and consequences for which the scores are used have increased over the years to present day public schools, due to the standards-based accountability mandated by NCLB in 2001 (Zucker, 2003). Pressure, or "high stakes", comes from important decision-making based solely on a test score. These important decisions may determine school funding, school resources, or closing the doors of failing schools. Standardized testing is used for a good *indicator* of a student's performance, not a *full measurement* of their education and capabilities. Therefore, schools should not be solely judged based on their passing or failing test scores. Some schools are more affected by low socioeconomic status and an abundance of ELL students than other schools. These test-biased circumstances may cause a significantly lower performance on

tests. Struggling schools need help, not a loss of funds, resources, or operation rights. In addition to a potential test bias, standardized tests are error prone and incapable of measuring every aspect in education. Therefore, it is learned that standardized testing should never be the sole measure determining a student, teacher, or school's fate.

In addition to being a constructive indicator of a student's performance level, standardized testing also equalizes the playing field by ensuring that every student takes the same test. Standardized tests are also seen in a positive light because they are quick, easy to grade, and are inexpensive. Additionally, machine graders grade each test objectively. Standardized testing receives its negative connotation when it is implemented under 'high stakes'. This is currently happening under NCLB. The high pressure placed on test scores has forced many negative outcomes in public schools. The curriculum is narrowing to solely math and reading, teachers are cheating to achieve optimum results, and the moral and love for learning in schools is declining.

In 2001, the reauthorization of the Elementary and Secondary Education Act under No Child Left Behind set the stakes for standardized testing the highest in history. Today in 2010, Obama's plan to change NCLB lowers the stakes by eliminating AYP, measurement of dropout rates, and punishments for low scoring schools. Also, schools will not be required to reach 100% proficiency by 2014. The possible elimination of these requirements would significantly lower the pressure on schools and allow for a healthy, public education system. Teachers will feel they have time to teach other subjects such as, social studies,

science, and art. Hopefully current “test score” atmosphere in schools will return to a general love for learning. Honesty will also be restored in schools.

Standardized testing will still be used in schools, however; without pressure and high stakes, standardized testing can be used sensibly. A sensible use of the test would be in supplementation of other information, such as portfolios, evaluations, and experiments (Koretz, 2008). Summative tests are also a good evaluation of content learned in a particular subject. Instead of using standardized test scores to evaluate the effectiveness of a school, inspection teams could do the job. Physical conditions, curriculum plans, and the quality of teaching and classroom learning are far greater measurements of a school than test scores (Ravitch, 2010).

Standardized testing is not the “cure-all” in education. To improve the *education* in schools, there should be an established curriculum, standards for proficiency in that curriculum, matching textbooks, and tests based on the curriculum taught (Ravitch, 2010). The school system should be aligned with curricular goals, not with random skills testing. Standardized tests under low stakes or no stakes could have a positive outcome in our schools today. According to Nicholas Lemann, “We have a moral obligation to be precise about what the problems in American education are-like subpar schools for poor and minority children-and to resist heroic ideas about what would solve them, if those ideas don’t demonstrably do that” (Lemann, 2010).

Until changes are made, teachers, students, administrators, parents, and community members can all work to change the use of standardized testing in

our schools today. Many professionals believe high-stakes testing is corrupting our schools year by year, and that people in the community should speak out to help their school and the K-6 public education system. An advocate for or an opponent against could speak out at a school board meeting, form a group of supporters, write a letter to a school administrator, write a letter to a state legislator, or write to the local newspaper. With help from the community, the use of standardized testing in schools can be changed and K-6 public schools can be restored to a healthy, honest environment for students to become well-educated, productive members in society.

Works Cited

American Educational Research Association. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Ayers, W. (1993). *To Teach: The Journey of a Teacher*. New York: Teachers College Press.

Burley, H. (2002). A Measure of Knowledge. *American School Board Journal*, 189 (2), 23-27.

Calkins, L., Montgomery, K., & Santman, D. (1998). *A Teacher's Guide to Standardized Reading Tests*. Portsmouth, NH: Heinemann.

Carnoy, M., Jacobsen, R., Mishel, L., & Rothstein, R. (2005). *The Charter School Dust-Up: Examining the Evidence on Enrollment and Achievement*. New York: Economic Policy Institute and Teachers College Press.

Chaddock, G. R. (2010, March 15). *Obama's No Child Left Behind Revise: a little more flexibility*. Retrieved Aug 20, 2010 from The Christian Science Monitor: <http://www.csmonitor.com/USA/Politics/2010/0315/Obama-s-No-Child-Left-Behind-revise-a-little-more-flexibility>

Checkley, K. (1997, June). Assessment That Serves Instruction. *ASCD Education Update*, 1, 4-6.

Fowler, F. C. (2010, August). *Policy Studies for Educational Leaders*. Retrieved August 30, 2010 from http://www.angelfire.com/theforce/shu_cohort_viii/images/fowlernotes6to8.pdf

Fox, D. (2001). Radical Dilemmas in the Anti-High-Stakes-Testing Movement. *Radical Teacher*, 61, 28-35.

Frontline. (2010). *Secrets of the SAT*. Retrieved December 12, 2010 from Frontline PBS Online: <http://www.pbs.org/wgbh/pages/frontline/shows/sats/where/history.html>

Fuller, B. (2006). *Is the No Child Left Behind Act Working? The Reliability of How States Track Achievement*. Policy Analysis for California Education. Berkley: University of California.

Gladwell, M. (2008, December 15). *Most Likely to Succeed*. From The New Yorker: http://www.newyorker.com/reporting/2008/12/15/081215fa_fact_gladwell?currentPage=all

Haladyna, T. M. (2002). *Essentials of Standardized Achievement Testing. Validity and Accountability*. Boston: Allyn and Bacon.

Holloway, J. (2001). The Use and Misuses of Standardized Tests. *Educational Leadership*, 59 (1), 77-78.

Hoover, H. (2003). *Iowa Tests of Basic Skills: Interpretive Guide for School Administrators*. Chicago: Riverside Publishing.

Keller, B. (1998, May). In Age of Accountability, Principals Feel the Heat. *Education Week*, 1-16.

Kohn, A. (2000). *The Case Against Standardized Testing*. Portsmouth, NH: Heineman.

Koretz, D. (2008). *Measuring Up: What Educational Testing Really Tells Us*. Cambridge, MA: Harvard University Press.

Langerfeld, K. T. (1996). *High Stakes Testing for Students: Unanswered Questions and Implications for Students with Disabilities*. Retrieved December 20, 2010, from National Center on Educational Outcomes: <http://education.umn.edu/NCEO/onlinepubs/synthesis26.htm>

Learninginfo.org. (2010). *History of IQ Tests*. Retrieved December 8, 2010 from A to Z of Brain, Mind, and Learning: <http://iq-test.learninginfo.org/iq01.htm>

- Lemann, N. (2010, September 27). Comment Schoolwork. *The New Yorker* .
- Levine, A. (1993). *Higher Learning in America. 1980-200*. Baltimore: The Johns Hopkins University Press.
- Lin, Q. (2002). Beyond Standardization: Testing and Assessment in Standards-Based Reform. *Action in Teacher Education* , 23 (4), 43-49.
- Madaus, G. F., & O'Dwyer, L. M. (1999, May). A Short History of Performance Assessment. *Phi Delta Kappan* , 688-695.
- Merritt, S. (2006). *Mastering Multiple Choice*. Canada: The Brain Ranch.
- Nathan, L. (2002, April). The Human Face of the High-Stakes Testing Story. *Phi Delta Kappan* , 595-600.
- National Research Council. (1999). *High Stakes: Testing for Tracking, Promotion, and Graduation*. Washington, DC: National Academy Press.
- Ondine Rarey Productions. (2007). *Testing Mrs. Grube and A Different Standard*. From Two Documentaries by Ondine Rarey: www.ondinerareyproductions.com
- Ormrod, J. (2003). *Educational Psychology. Developing Learners*. Columbus, OH: Merrill Prentice Hall.
- Patton, M. Q. (2008). *Utilization-Focused Evaluation*. Thousand Oaks, CA: SAGE Publications, Inc.
- Peterson, P. E. (2003, Spring). A Ticket To Nowhere. *Education Next* , 39-46.
- Pope, J. (2006, May 14). *SAT Error Lowers Scores for 4,000 Students: problem lay with scanning of their answer sheets*. Retrieved 2010 from Seattle Pi: http://www.seattlepi.com/national/262827_satmistake14.html
- Popham, W. J. (1999). Why Standardized Tests Don't Measure Educational Quality. *Educational Leadership* , 56 (6), 8-15.
- Quiad, L. (2010, Feb 1). *'No Child Left Behind': Obama To See Sweeping Changes in Law*. Retrieved Aug 20, 2010 from The Huffington Post: http://www.huffingtonpost.com/2010/01/31/no-child-left-behind-obam_n_443924.html
- Ravitch, D. (2010). *The Death and Life of the Great American School System*. New York, NY: Basic Books.

- Resnick, L. B., & Nolan, K. J. (1995). *Standards for Education*. Washington DC: Bookings Institution.
- Resnick, L. B., & Resnick, D. P. (1990). *Tests as Standards of Achievement in Schools*. Princeton, NJ: Educational Testing Service.
- Rubinstein, J. (2000). Cracking the MCAS Grade 10 Mathematics. *Princeton Review* , p. 31.
- Senechal, D. (August, 17 2009). *Guessing My Way to Promotion*. Retrieved 2010 from Gotham Schools: <http://gothamschools.org/2009/08/17/guessing-my-way-to-promotion/>
- Sizer, T. R. (1992). *Horace's School: Redesigning the American High School*. Boston: Houghton Mifflin.
- Strauss, V. (2001, July 10). Review Tests Go Too Far, Critics Say. *Washington Post* , p. A09.
- Suen, H. (2006). Chronic Consequences of High-Stakes Testing? Lessons from the Chinese Civil Service Exam. *Comparative and International Education Society* , 50 (1), 46-65.
- Thurlow, M. L. (2001). *2001 State Policies on Assessment Participation and Accommodations*. Retrieved January 2010 from Synthesis Report 46: <http://education.umn.edu/NCEO/OnlinePubs/Synthesis46.html>
- Ward, A. W. (1999). *Assessment in the Classroom*. Albany, NY: Wadsworth Publishing Company.
- Wilgoren, J. (2001, March 14). Florida's Vouchers a Spur to Two Schools Left Behind. *New York Times* , pp. A1, A18.
- Young, K. (2005, May). *Testing*. Retrieved January 2010 from Standardized Testing: <https://www.msu.edu/~youngka7/testing.html>
- Zucker, S. (2003, December). *Fundamentals of Standardized Testing*. Retrieved November 21, 2010 from Pearson Education: http://www.pearsonassessments.com/NR/rdonlyres/20DB5E75-059E-4EB7-BFB1-F9C171ABE0C3/0/Fundamentals_of_Standardized_Testing_Final.pdf

Heather Smith
324 Wister Circle
Mechanicsburg, PA 17055
(717) 574-2481 | hcs5010@gmail.com

EDUCATION	BS in Elementary Education (K-6), 2011 The Pennsylvania State University Minor: Education Policy Studies Honors: Curriculum & Instruction Qualified: Middle School Math K-8 Dean's List, Schreyer's Honors Society, Pi Lambda Theta Education Honors Society, Penn State's Presidential Freshman Award 2008, Student PSEA
TEACHING EXPERIENCE	Student Teaching, 4th Grade (Jan. 10-April 15, 2011) Willow Dale Elementary School, Mr. Markowski Cooperating Teacher, Centennial School District <ul style="list-style-type: none">• used centers for guided reading, grammar, geography, vocabulary, and spelling to complement the child-based hands-on curriculum to allow students to explore, internalize, and make deeper connections to the skill being practiced• implemented a positive discipline plan which promoted student responsibility, problem solving skills, and student accountability to become a self-manager• completed an inquiry into the teaching and learning project where I taught a 4 week unit on African culture and customs• taught with an inquiry based approach to promote student learning through questioning and finding evidence to support their explanations Preservice Teaching, 3rd Grade (Sep. 28-Dec. 10, 2010) West Beaver Elementary School, Mrs. Bishop Cooperating Teacher, Midd West School District <ul style="list-style-type: none">• Taught science unit on reflections & mirrors within the FOSS Science Curriculum• Taught a social studies unit on landforms within the Harcourt Trophies Curriculum• Taught math meeting, fact practice, and math lesson everyday within the Saxon Math Curriculum• Taught Writer's Workshop (Lucy Calkins), a newly implemented writing curriculum Practicum Observed classroom and taught lessons one-on-one, to small group and whole class. Kindergarten at Logan Elementary, Mrs. Kelley Coordinating Teacher; Altoona Area SD <ul style="list-style-type: none">• Completed lessons on various children's books, poems, and themes• Lessons involved cooperative learning, teacher modeling, hands-on approach, and interdisciplinary teaching
VOLUNTEER WORK	Volunteered as Movie Elective Teacher, State College Friends School 2010; Preschool Teacher, Gloria Gates Memorial Foundation 2007-2009; 2 nd Grade Sunday School Teacher, First Church of God Mechanicsburg 2007-2009; Volunteer Reader, Altoona Public Library 2007-2009
CONTINUING PROFESSIONAL DEVELOPMENT	National Science Teachers Association , current member PSEA , current member Teacher Survival Skills Series , October 2010 Responsive Classroom Seminar , October 21, 2010 EdLion Webinar Series: Managing Your Classroom Effectively , March 2010 Improving Parent Teacher Communications , October 29, 2007
PROFESSIONAL GOALS	<ul style="list-style-type: none">• I have a serious commitment to be the best teacher for each group of students I have the opportunity to teach.• To keep up with the most effective teaching methods, I wish to attend workshops, obtain a master's degree, and interact with other teachers for advice and to learn new ideas.• I plan to begin work on my Master's in Educational Leadership through Penn State.