Alternatives and Improvements to High-Stakes Testing

An Honors Thesis (HONRS 499)

By

Jennifer Shumaker

Dr. Pamela Hartman

Ball State University
Muncie, Indiana

March 27th, 2008

Expected Date of Graduation: May 3rd, 2008
Abstract

In response to the large number of implemented standardized high-stakes tests in today’s schools, and the pressure attached to them as a result of the No Child Left Behind Act, this study will give a brief analysis of the history of high-stakes testing and its introduction into America’s public schools. This study’s purpose is to discuss the benefits and disadvantages of high-stakes testing, examining various improvements and ways in which the criteria measured by the tests could possibly be met and assessed in other ways.

Acknowledgements

-I want to thank Dr. Pamela Hartman for helping me through this entire project and checking it for grammatical, technical, and organizational errors. She was enormously influential in my decision to focus on this project’s topic and her experience and knowledge with the subject matter was a huge help and encouragement to me.

-I would also like to thank Dr. Joanne Edmonds for advising me in this project from start to finish. Without her help, I would have not known how or where to begin, and her answers to the number of questions I had on the way were also a huge help to me.
Introduction

In response to the large number of implemented standardized tests in today’s schools, and the pressure attached to them as a result of the No Child Left Behind Act, this study will give a brief analysis of the history of the high-stakes testing and its introduction into the curriculum of America’s education system. As a future educator, the issue of accountability tied to high-stakes testing is something I will be faced with on a regular basis. Thus I feel it is my duty as an educator, both for myself and for my future students, to be as aware as possible of all that these tests entail. This study’s purpose is to discuss the benefits and disadvantages of high-stakes testing, examining various improvements and ways in which the criteria measured by the tests could possibly be met and assessed in other ways.

I will begin by first describing the history of high-stakes tests: What exactly are high-stakes tests? Why were they developed? Who developed them? When did this occur? To properly evaluate and assess the impact and need for high-stakes tests, it is important to first know about and understand the beginnings from which they sprung to life. Secondly, I will cover the advantages of high-stakes testing. This is something that, amid the incessant buzz about high-stakes testing drama in American public school, is rarely mentioned. To have a well-rounded perspective of these tests, I thought it was necessary to look at them from both sides of the ongoing debate. Following this perspective, a discussion of the disadvantages of high-stakes testing will serve as a comparative analysis for readers. Last of all, in correlation with the title of this thesis, “Improvements and Alternatives to High-Stakes Testing,” I will discuss improvements
that have been developed by educational researchers, as well as the supplemental or individual alternatives that are currently being suggested and implemented.

While it may seem to some as if those behind the creation of these tests (i.e. legislators vacant of educational experience and thus authentic knowledge of proper assessment techniques) are out to harm education, stripping it of all possible effectiveness, their initial intentions were not of ill design. For education to be effective, goals must be set; for these goals to be reached, standards must be put into place; to ensure these standards are met, some form of assessment is required. Thus, the public schools of America are given high-stakes testing.

Today, schools are constantly challenged and pressured to meet specific academic standards. While it is important to give some merit to these high standards to which all teachers and students should ascribe, when testing these standards is given too much importance, becoming the sole purpose of education, measuring these standards has the propensity to wreak havoc on the educational system if these standards are not met. In many schools across America, Adequate Yearly Progress (AYP), an individual state's measure of progress toward the goal of 100 percent of students achieving a state’s academic standards in at least reading/language arts and math is not being met, and schools are punished as a result.

To measure the academic progress, or lack thereof, within schools, what do assessors use? High-stakes testing. These tests are labeled “high-stakes” because they are used to make decisions that have extremely important consequences for students. A student’s score on this tests determines high school graduation or promotion to the next grade. In addition, teachers are evaluated based on these test scores. If a majority of
student scores happen to be rather low, teachers are often held accountable, placed under
close watch and supervision by administrators and even possibly the threat of losing their
job.

Failing schools that have not met AYP for a number of years can be merged with
nearby schools that have, or a special management team can be assigned to oversee the
operation of the failing school. How did educational assessment come to this point of
severe consequences? Let us take a closer look and examine the beginnings of, and
reasoning behind, the small pamphlets of paper creating such a stir.

The History of High-Stakes Testing

Beginning in the 1970s, testing began to filter into America’s public schools. According to the National Research Council, an organization developed to improve public policy in relation to public education, “the minimum competency testing movement gave large-scale, standardized achievement tests a visible and popular role in holding students (and sometimes schools) accountable” (National Research Council 15). The use of these accountability assessments became quickly widespread and, while thirty-three states had implemented the use of minimum competency testing by the mid-1980s, ten years later, eighteen states had gone even further, using test-based assessments to determine high school graduation (National Research Council).

At this time, little academic improvement had been seen in schools. John F. Jennings, director of the Center on Educational Policy, who advocates for improvements in public schools, comments on what was seen at that time. According to Jennings, SAT I scores had barely increased, college professors found that their students were not
adequately prepared for college-level classes, and business owners claimed the students they employed were not able to successfully take part in a workplace environment. In response to this, in 1990 the administration of President George Bush, Sr. determined a new system of accountability, a “national crusade” to transform the nation’s educational system (Jennings 16). Within this crusade, central focus was placed on national goals, national standards, and a national test that would be used to assess the progress schools had made toward achieving those standards. However, legislators felt more was needed than simply the claim of using a “national crusade” for standards. Specifics needed to be put in place. A concrete form of assessment, an assessment that could be evaluated and accurately measured, was needed. Objective, one-correct-answer tests seemed to be the perfect fit.

A presidential advisory committee composed of business people and educators, recommended to Bush on January 16, 1991, the creation for the first time of national examinations of elementary and secondary education students. Paul O’Neill, chief executive officer of the Aluminum Company of America and chair of the group, asserted that teachers should teach to standards and that having national tests would be the only way to reach the national goals. (Jennings 18)

This development of “teaching to the standards,” used to determine how well teachers were doing their job, was, oddly enough, developed by an executive officer of a huge corporation rather than a teacher. This makes it easy to see why many teachers today, who actually have experience within a classroom setting, can pinpoint the inherent flaws within this standardized system of education that looks at the educational process as more of a cut and dry evaluation of numbers rather than an assessment of educational growth.

By 1994, forty-five states had developed and instituted testing programs for their elementary and secondary schools. However, of these forty-five states, only a few used any form of curriculum that informed teachers and students of what the tests were
specifically focusing on and, thus, the tests were not tied to the schools’ educational curriculums (Jennings). Clearly, this resulted in fairly poor test results for the students, and a consequently poor image of the educational progress of the nation’s schools. From this, the perception of the American public, known to be “edgy and impatient with schools and [to] want results,” (Jennings 3) looked at these poor test results as a direct failure on the part of America’s educators and school administrators. A poll taken in 1989 previously found that 69% of citizens preferred a standardized national curriculum. By 1994, this had risen to 83% (Jennings). This call to action on the part of America’s educational system was clear – something more needed to be done. However, was the use of standards and testing the best way to turn the nation’s schools around? Years later, this question is still asked.

While statewide standards were determined to be the best way to hold each state’s schools accountable, a disconnect existed between these standards and individual school curriculums. This disconnect between curriculum and assessment was deeply rooted:

The U.S. Constitution embodies the idea that government should be limited in its powers and that the closer the government is to the people, the better it will function. In education, this had meant that although states have authority over the schools, the power to determine the content of education has usually been delegated to local school boards. (Jennings 4)

With local school boards determining the curriculum, there was no accurate way to ensure all schools were appropriately meeting the standards. Therefore, to alter and improve the method of statewide accountability, the curriculum at each school needed to become a rather standardized one, appropriate for all schools, and thus fairly assessed by the over-arching statewide standards. Unfortunately, this idealized conclusion was fairly unrealistic in its idea of “fairness,” for how could all schools in each state, consisting of
entirely diverse populations and located in entirely diverse locations, be forced to use the same curriculum and tested on the same standards? More than that, with the development of national standards to meet, schools across the entire nation, let alone the state, began to be assessed by the same criteria.

The National Council on Education Standards and Testing, an organization developed by Congress to address the subject matter, assessment, and standards of performance for the tests, wanted to see the development of national standards and a system of assessments. "The council gave three major reasons for the necessity of this monumental change in American education. National standards were needed to enhance the civic culture." Members of the council also argued that "raising standards would enhance America’s economic competitiveness" (Jennings 22). This standards-based reform movement was based on the idea of setting high and clearly detailed standards for students as well as teachers and holding them accountable for the success or failure to meet those standards. It was believed this was the answer to a much-needed improvement in America’s schools. And what would be a better “system of assessments” for these national standards than a system of nationally standardized tests?

With this system of national standards developed, in 1994 the Goals 2000: Educate America Act became a law, “support[ing] the certification of education standards and national skill standards, and encourag[ing] states to develop their own standards as well” (Jennings 8). The “age of standardized testing” quickly sprang into action, with a number of new tests constantly being developed and implemented for various forms of evaluation. “In addition to mandating that students take more academic classes, states had expanded testing to determine whether to promote students” (Jennings
Thomas Kean, an education group chair and former Republican governor of New Jersey, proposed the mandate that “all high school seniors [be] required to take a national examination of their knowledge and skills” in order to graduate (National Research Council 18). Additionally, a system of examinations for all fourth-, eighth-, and twelfth-grade students was being developed to assess their skills and knowledge in the five core subjects: English, mathematics, science, history, and geography (Jennings 21).

High-stakes testing was in full force and currently still is. The question is, have these tests done all they were supposed to do? Have they improved the educational system of America? Many statistics point to the answer “no.” Large increases seen in educational spending soon after this system was put into place led legislators to voice their complaints and call for yet another change in the original system for which they had formerly voiced their approval. Author and educational columnist Alfie Kohn finds that, today,

some parents and students around the country are simply refusing to be part of the testing programs. In the last year alone (2001), boycotts have sprung up in several states. These acts of civil disobedience – along with less admirable responses like cheating – will probably grow until the message is heard: Raising scores is completely different from helping students to learn. (“Beware” 2)

Finally, this development of national accountability through high-stakes testing has led to one of the most controversial pieces of educational legislation ever put into practice: The No Child Left Behind (NCLB) Act.

NCLB is an act signed into law by President George W. Bush on January 8, 2002. This law is an offshoot of the Elementary and Secondary Education Act (ESEA), which was passed in 1965 by President Lyndon Johnson (U.S. Department of Education). The purpose of ESEA was to equalize education for all students by providing federal funding
to the neediest schools and students. Following in this law’s footsteps, the stated purpose of NCLB was to also equalize education for all students. The initial intent of this law was to improve the education students received by creating equal assessment measures and thus curriculum within every school across the nation. It was thought this would equalize the education all students received, placing all schools at an equal playing field, thereby “leaving no child behind,” as the title so eloquently phrases it.

A large part of this law is based on accountability through test results. Students are not the only ones whose progress is measured based upon the test scores. These scores determine the fate of teachers as well. The purpose of this “accountability” system was to “check up” on those teachers who may not have been teaching the necessary material. While there are teachers who do not fully commit to their profession, lazily taking the easy route of irrelevant worksheets and videos within their classrooms, the majority of America’s educators do take their professions very seriously. However, many of the standardized test scores of these teachers’ students would not say so. This is one of the largest problems with NCLB. A standardized high-stakes test, the sole form of assessment and accountability tied to NCLB, cannot adequately measure whether or not a teacher is doing her job any more than it adequately measures a student’s knowledge.

Presently, a “bipartisan consensus that the law is hurting more than helping efforts to close achievement gaps” pervades the educational field (Jehlen “Rating NCLB” 24). Educators argue that the heavy focus on accountability through testing crowds out the option for authentic learning. Students who are poor test-takers or who suffer from test anxiety are virtually placed on the education chopping block, their intelligence being evaluated and based on their test performance.
In the political race to presidency for the fall of 2008, the topic of the NCLB act is one of the biggest issues presidential candidates face. Voters want to know how candidates will handle this huge problem, what they will change and improve. All candidates, regardless of their party affiliation, proclaim to be in favor of alterations, at the least, to the unpopular law. All know this is necessary to be elected by the American people. Issues they intend to change include the treatment of those schools who do not meet the necessary Adequate Yearly Progress (AYP) required by the law; the skills that are measured on the tests; and the consequences and strings that politicians, desperate to see an improvement in public education, attach so fervently to these tests.

The Advantages of High-Stakes Testing

While many would never claim that high-stakes testing has benefited education, there are two sides to every story or to every controversy. In fact, a report by the Business Roundtable, an organization that advocates policies promoting economic growth (2001), found that much of the anti-testing sentiment in the populace has been “seriously exaggerated.” Citing a recent survey of 1,023 parents of school-age children, which found that 83% of respondents believe tests provide important information about their children’s educational progress, 9 out of 10 wanted comparative (i.e., test) data about their children and the schools they attend (Cizek 29).

As seen in the previous section about the history of high-stakes testing, these tests were initially implemented to benefit the educational process. Unfortunately, over the course of time, the initial reasoning behind the tests’ development has been clouded by various other requirements and mandates by those who were not necessarily the most adequately suited to make those decisions, such as the chief executive officer of the
Aluminum Company of America and chair of the Presidential Advisory Committee, Paul O'Neill, who mandated teachers teach to standards and believed having national tests would be the only way to reach national goals (Jennings 18).

However, despite the influence of sources outside of the education field on high-stakes tests, positive aspects of the test have been found. In 1995, Richard Phelps of the US General Accounting Office conducted a study that estimated the extent and cost of high-stakes testing in the US. To his surprise, he concluded that the benefits of these high-stakes tests far exceeded the cost and recorded his findings in the book *Kill the Messenger: The War on Standardized Testing*. Of these benefits he discovered, the most favorable included organizational clarity and efficiency for needed improvements in school curriculum plans, allowing the school and its district to develop a more comprehensive and appropriate action plan of the necessary steps to take to improve the educational curriculum. He also found it propelled an increase in student motivation and provided a check on grade inflation and more public confidence in schools ("Schools/Assessment"). Unlike many educators and administrators who find themselves exasperated by the limitations and demands they feel high-stakes tests create, Phelps found high-stakes testing to be a beneficial venture to improve the quality of America's public education and recommended that the testing continue.

The primary goal, however, that initially propelled the implementation of high-stakes testing in American schools was not just an improvement in the quality of education students generally received, but, more importantly, in the equality of education throughout this highly diverse nation. Educators and legislators asked themselves, "How can we ensure all students are equally prepared to enter the collegiate or work
environment?" Those students who were not adequately prepared in their elementary and secondary school settings would be at a great disadvantage later in life and would likely not receive the same number of benefits as those who did receive the necessary components of a quality education.

Of course, the word "standardized" definitely holds a negative connotation and many feel it should never be used to describe education, something that necessitates diverse methods and content from area to area within the United States. However, within diversity, equality must somehow find a stronghold. If this does not occur, the term "diversity" could become tainted by the inequality that could ensue.

With this intent of equality in education, one of the largest concerns educators, administrators, parents, and community members have with the standardized high-stakes tests is that the tests are not so "standardized." The problem of test bias presents itself, leading many to feel the test questions and content favors and is geared toward middle-class, Caucasian students much more than students of minority cultures and lower socioeconomic statuses. According to research conducted by Kurt F. Geisenger of The University of St. Thomas, numerous methods and models have been used to discern whether tests are biased. The testing industry has always "attempt[ed] to establish whether test scores are equally valid for members of underserved groups relative to the majority group" (Geisinger 189).

Yet, many argue, with equality must not come a lowering of academic standards. Rather, the educational bar must be raised for the majority, kept the same for some, and hopefully lowered for none. This leads to the highly controversial No Child Left Behind Act (NCLB) of the George W. Bush administration. The idea behind this finely worded
act was to ensure high standards for all children and send the message that no one should give up on students who have difficulty and, subsequently, do not do well in school.

One of the most widely used and winced-at buzz words in education today is “accountability.” Holding educators accountable seems to be the current ticket to success and, while educators certainly should take pride and ownership for the success of their students, they should not be made into scapegoats, responsible for all that is wrong. However, does high-stakes testing adequately measure this? While there are obviously many who would dissent, the standardized, standard-driven high-stakes tests gather the support of many who feel today’s educational system, and thus today’s educators, have severely dumbed down the students of America, leading to its much lower academic rigor than countries such as Japan and China. In fact, “most advanced nations have administered large-scale national- or state-level high-stakes tests for decades, and most other countries are increasing their use of standardized tests” (“Schools/Assessment” 21).

However, at what cost does this occur? Has the quality of education improved with the implementation of high-stakes tests? A big part of the argument to use high-stakes testing to improve education is that we must use hard data rather than subjective determinants to assess the quality of education. Although the use of data obviously stifles the creative part of the educational process, focusing on the “black or white” of things rather than the colorful, imaginative opportunity to express oneself, it provides specific factual information administrators and educators can refer to to note any needed changes or improvements in the curriculum. While this focus on the data-driven standards and high-stakes tests has been criticized, others, such as senior writer and editor of NEA
Today, claim, “[Although] too much testing can get in the way of learning, good tests, well used, can be powerful tools” (“Using Data” 26).

Many administrators, teachers, parents, and students believe high-stakes standardized testing has cast a dark cloud over the entire educational arena. However, through their research, Drs. Goodman and Hamilton, professors of Educational Psychology at the University of Massachusetts at Amherst, wish to dispel a number of, in their opinion, entirely incorrect assumptions they feel have propelled this belief. One large concern people have is that state assessments do not measure what is important in K-12 education. The argue, however, “The most recent survey of state assessment programs by Education Week (2003) reveals an increasing commitment by states to administer criterion-referenced assessments that were aligned with state standards” (qtd. in Education Week, 1997, Goodman and Hambleton 93).

Another positive effect of high-stakes tests may be the resulting opportunity to pinpoint problems within the educational curriculum. “Many NEA (National Education Association) members say the separate reporting is helpful when it’s used to point up problems, not to punish educators” (“Pinpointing Problems” 26). The detailed results and scores of high-stakes tests, itemized by subject area and the topics within each subject, allow educators and administrators to prioritize the needs of students. They are able to view areas that need the most improvement, and the areas in which students have been successful. Teachers are given the opportunity and ability to help craft academic programs to meet the needs of students who score low. One such teacher interviewed in a qualitative study of high-stakes tests conducted by Richard J. Cizek of the University of North Carolina at Chapel Hill has claimed, “My appreciation of having had the privilege
of introducing standardized tests in my school cannot be too strongly emphasized. No school can accurately determine the progress of its pupils, either as a group or individually, without using these tests” (E.M.W. 23).

A problem many educators face today is motivating the unmotivated. Somehow, the process of learning becomes a chore rather than an exciting opportunity for students. But it does not begin in this way. For many, the memory of that first day of school is a conglomeration of both fear and excitement, propelled by anticipation. The yellow bus, the brown, squishy seats (without seat belts!), the “school smell,” the new back pack that matches the new coat (the coat that will be ruined by mud at recess by the end of the week), the glimmer of the newly waxed tile floors, the posters on the classroom wall, the smell of chalk dust, and, of course, that welcoming teacher’s voice that somehow seems to be the one source of solace amongst a whirlwind of uncomfortable newness. In all of this is one important element that seems to be lost over the course of the thirteen-year educational journey: eager anticipation. This eager anticipation is what leads to the desire to learn.

How can educators motivate jaded secondary students? This is something about which secondary educators across the nation are passionate. Should education be fun? Of course. Should it be challenging? Undoubtedly. Should it create fear in students? What is the proper response to this question? On the surface, of course, education should never incite fear within students, but is testing what is needed to motivate them? One student interviewed seems to think so: “I think the tests have helped me in many ways. They have helped me to work both faster and better, and I have more interest in getting ahead
in my studies” (24). The challenge to meet the standards in the form of passing the dreaded high-stakes test is what motivates the unmotivated in the eyes of many.

The Disadvantages of High-Stakes Testing

The title of this thesis, “Improvements and Alternatives to High-Stakes Testing,” clearly shows that there are some disadvantages to high-stakes testing. While it is not a completely horrible practice and, as seen in the previous section and advantages do exist, a great number of things have hindered the testing process from being all that legislators and designers of the tests thought it could be and believed it would provide, turning it into, in the words of Alfie Kohn, prominent educational columnist, “the use of crude rewards and punishments to make people ratchet up the scores” (“Beware of the Standards” 4).

When high-stakes tests were initially proposed, it was thought they would increase the academic rigor of American schools, “catching America up” to the education level of other nations. It was also thought to be a way to hold teachers accountable, thus also providing the American public with possible scapegoats for the deficiencies in public education. Perhaps testing has provided the latter scapegoat opportunity, pushing many educators out of their profession as they tire of the constant badgering and finger-pointing by legislators, administrators, and parents of students. However, little research has shown it to increase the quality of education American students receive. Nor have these tests been successfully proven to equalize education for all students, “leaving no one behind.”

Many education researchers and experts, teachers included, would argue that the disadvantages of high-stakes testing far outweigh the proposed benefits. A large body of
research has found high-stakes tests to have a negative impact on both the classroom and the overall school environment. One finding is that teachers are often affected, pressured to meet the standards and avoid the terrifying gauntlet of standard-driven accusations. Many teachers, as a result of high-stakes testing, have these standards as a constant reminder and, therefore unfortunately, may feel pressure to teach to the test.

However, "noninstructional factors explain most of the variance among test scores when schools or districts are compared." In the 1992 National Assessment of Educational Progress, the four variables of number of parents living at home, parents' educational background, type of community, and poverty rate accounted for 89 percent of the differences in state scores (Kohn, "Standardized Testing" 1).

Of course, all educators and students should want education to meet academic standards, for these standards have been carefully designed to allow for the most beneficial and appropriately challenging education possible. But "the more comprehensive and detailed a list of standards, the more students (and even teachers) are excluded from this process, and the more teaching becomes a race to cover a huge amount of material" (Kohn, "Beware of the Standards" 2). In addition, when these standards were designed, it was highly unlikely the authors agreed to a data-driven assessment of all standards, fueled by the threat of punishment in the form of decreased funding for the school if they were not appropriately met. Additionally, these standardized, high-stakes tests were never intended to measure the quality of learning or teaching; rather, their main purpose was to rank, not rate (Kohn, "Standardized Testing").

Of course, one cannot discuss the topic of high-stakes testing without mentioning the No Child Left Behind Act (NCLB). "Some 45 million tests required by NCLB will be
administered in one school year. The testing industry rakes in more than half a billion dollars a year for these tests” (Jehlen, “Testing How the Sausage is Made” 29). NCLB, designed to hold teachers and schools accountable for the success or failure of the education process, the decisions made are based upon student scores on the high-stakes tests. Caught in the momentum of these decisions that partly fall out of their control, many educators feel their voices do not matter. Kimberly Meigs, an English educator from Tahlequah, Oklahoma, recounts her experience in this testing aftermath:

In the last four years, because of student scores, our department has had to totally revamp our English II curriculum...because we were under great pressure from our district administrators to prove we were doing our jobs. In standardizing our curriculum, we had to eliminate most creative student projects, in-depth class discussions, thematic research opportunities, and group work. (Jehlen, “NCLB” 31)

The environment of the classroom and school before high-stakes testing time surely becomes a tense one. Teachers write “Two weeks until the test” on the board, supposedly in an attempt to motivate their students and, perhaps, themselves. Test-prep materials are rampant in classrooms: practice essays, practice test questions, timed test-taking skills; the list is endless. In fact, a 2005 Education Week study reported that “15 states with 42 percent of the nation’s students had chosen to make their reading and math tests entirely multiple-choice” (Jehlen, “Testing How the Sausage is Made” 31). This is merely a part of the educational process; students in today’s public education system have never known any differently. Many are unsure as to why “the big test” is so important, but they sure know they need to pass.

Yet with all of the hullabaloo preceding this test, little is done relative to the test after the big extravaganza has passed through the school system...until next year when
educators across the nation receive the privilege of repeating the entire process. Though some studies have shown “a relatively short period of introducing students to the content and format of the tests is sufficient to produce scores equivalent to those obtained by students who have spent the entire year in test-prep mode,” (Kohn, “Fighting the Tests”), each educator becomes a facilitator of test preparation rather than a partner in the educational process. Administrators become nagging bosses, who want to prepare the educators as much as possible to prepare the students as much as possible to do as well as they possibly can on the high-stakes test to make the school look as good as possible. It is a far-stretching feeding chain that affects all involved.

With all of this preparation, the big question is, has this relentless focus on tests produced students who are better educated or more competitive in the world economy? Educational researcher Keith Baker would answer, probably not. “High national test scores don’t correlate with healthy economies, and all the intense, high-stakes testing hasn’t had any visible impact on national test scores anyway” (Jehlen, “Testing How the Sausage is Made” 29). International comparisons not only bring forth silly statements about test scores and the economy, they evoke howls of woe and outrage, claiming the studies show that the American educational system has failed.

While some argue high-stakes testing increases motivation to learn in students, others have found it creates a very cynical, laissez faire attitude in many students. John F. Jennings, author of Why National Standards and Tests?, has found a flaw in the evaluative feedback students receive, stating, “Students are not told exactly what they are going to be tested on; nor are they told afterward which questions they answered wrong so that they can learn the material for another time. Is it any wonder, then, that high
school students are not motivated to learn?” (2). Indeed, the lack of student self-assessment and improvement is a large flaw of the high-stakes testing system. Educators are taught the importance of prompt feedback to students about their work through assessment and evaluation. By doing this, students can view what they need to improve on and, thus, do better the next time. How can this occur for those who fail a high-stakes test if they are unable to view the questions they answered incorrectly?

As noted in the previous section, administrators are able to view the accumulative results of the high-stakes tests for their schools, and can then, based on the results, develop a school improvement plan, which will hopefully involve the educators of the building. But where is the student involvement in this? Walk into any classroom, of any age group, and you will see how intrigued and attentive most students become when their tests or other homework is passed back to them. They enjoy seeing how well they did, but, more importantly, they enjoy evaluating themselves. Responses such as, “I thought I answered this question correctly! Why is it wrong?” or “Wow! I did much better than I thought I would! It looks like I’m not as horrible in this subject as I perceived I was,” can be heard. How can students take pride and ownership of their accomplishments on the tests if they are unable to view how they did, assessing which questions they answered correctly and which they did not?

A separate concern of educators, administrators, and students about the highly influential high-stakes tests is how to help those students who suffer from severe test anxiety, “a type of performance anxiety that results from the fear of being judged by teachers, parents, and classmates” (Black 42). In a school setting where so much is at stake to do well on a test, those who already fear that intimidating bubble sheet of
unknowns become even more perplexed and anxiety-ridden, inevitably resulting in low test scores. Worrying about and doing poorly on these tests creates a vicious cycle, hindering students’ performance and self-efficacy. In fact, “high-stakes testing leads to increased grade retention, which has repeatedly been proved to be counterproductive in terms of its effects on students” (Neill 10).

With all of the negative consequences of high-stakes tests, the big question may be, are they here to stay? Many sources would agree that, yes, they will remain in America’s public education system for quite some time. This, then, brings about another important question. How can students, educators, and administrators effectively handle what is forced upon them? Perhaps there is no one way to deal with the stress and anxiety that seems to be inherently tied to high-stakes testing. However, knowing as much as possible about how to prepare for (and not cater to) the tests, as well as deal with the subsequent stress that precedes and follows, will make the incessantly spinning merry-go-round slow down a bit, allowing for a much clearer and more creative frame of mind.

**Improvements to High-Stakes Tests**

High-stakes tests. Upon reading this, words likely to come to the mind of readers include accountability, standards, punishment, No Child Left Behind (NCLB), Adequate Yearly Progress (AYP), test preparation, and test anxiety. Do any positive connotations surround this topic, or has its reputation become so utterly tarnished that a makeover is beyond possibility? Are high-stakes tests are not doomed to be the evil villain lurking in the hallways, classrooms, and offices of America’s schools? Must they be the demise of
every educator in America? Or, perhaps, is there hope for these ill-fated pseudo-assessments?

Some would argue it is not the tests that must be bludgeoned and conquered, but the strings that are attached to them, such as misdirected accountability, high stakes, and punishment though the process of reduced funding and forced curriculum revision. For instance, opponents of these strangling “attached strings” argue that accountability should no longer be based on test scores alone. Rather, it should include classroom and school-based evidence of achievement and regular detailed investigations by trained observers as well. In addition, “the data [from these tests] should be used to help educators do their work better, not to punish them for low scores” (Jehlen, “Don’t Be Left” 19).

The punishment that follows high-stakes tests is a sad aspect of education today. This is accountability gone wrong. However, it is not accountability that is irreversibly poor. Rather, it is accountability that is compacted into mere test scores and punishments, making them “far too narrow to provide sufficient information for any fair or helpful educational decisions” (Neill 9).

Another concern critics hold about high-stakes tests is pinpointing exactly what the tests measure. Do they measure critical thinking skills? Do they measure the intelligence of students, assessing how advanced they are in comparison to others of their grade level? Do they have the ability to do this? Much research points to a negative response to these questions. As Kohn points out,

“In a study published in the Journal of Educational Psychology, it turned out that high scores on both the CTBS and the MAT were more likely to be found among students who exhibited the superficial approach to learning. Similar finding have emerged from studies of middle school students and high school students.” (Kohn “Victims” 1)
This type of lower-order thinking and feigned academic growth is not what the educators nor administrators of America’s schools should want. Rather, students should be pushed both in the classroom and on tests and other evaluative projects to demonstrate their critical thinking skills, applying what they have learned to multiple forms of assessment such as writing rather than merely selecting the best answer among a list of options.

What must be changed in these tests to allow for creativity and expression at an advanced level? The use of open-ended questions would be a very effective way of allowing students to use critical, higher-order thinking skills. It would also allow them to work at their own level, rather than fitting themselves into the tiny Scantron bubble, displaying no more of themselves than the simple No. 2 lead of their pencils. Unfortunately, this open-ended approach does not lend itself well to the data-driven assessment used in high-stakes tests.

In turn, to prepare for high-stakes tests, the classroom is turned into a data-driven, high-stakes world rather than an environment that fosters creativity and higher-order thinking. In preparation for these tests, students are almost encouraged to conform to the superficial thinking required on these high-stakes tests. Monty Neill, Director of the National Center for Fair & Open Testing, examines this: “If instruction focuses on the test, students have few opportunities in which to display the attributes of higher-order thinking, such as analysis, synthesis, evaluation, and creativity, which are needed in order for them to achieve success in school, in college, and in life” (9).

To limit this “teaching to the test” virus to which so many educators find themselves susceptible, tests should be written in a way that is more realistic and not quite as predictable, making them harder to coach for. If the topics found on the test
changed annually, schools would be less likely to narrow their curriculum and would, instead, continue the natural progression within the curriculum devoted to student growth, which is matched to standards rather than test prep packets of multiple-choice, lower-order questions. Harvard University testing expert Daniel Koretz counters, “But that would produce small gains—more meaningful, but smaller. And nobody wants that” (qtd. in Jehlen, “Testing How the Sausage is Made” 33).

Undoubtedly the classroom is not an environment isolated from the national education dilemmas surrounding it; as long as high-stakes tests are used, and educators are held accountable for the performance of their students on these tests, the classroom will, unfortunately, likely remain a test-prep workshop for the weeks or even months before that big, looming test. One of the most disturbing factors related to this is the fact that high-stakes testing implementation knows no boundaries; elementary students are in no different position than are secondary students and “virtually all specialists condemn the practice of giving standardized tests to children younger than 8 or 9 years old” (Kohn “Victims” 1). Seven- and eight-year-old students are held victims to the test-prep packets and the bubble sheets. Assessments such as performance tasks or portfolios would likely be much more appropriate to evaluate students at this age level.

The all-powerful high-stakes tests wield such an immense power over the education world that the educators, administrators, and students involved become mere peons serving a ruthless dictator. High-stakes tests seem to determine all: a school’s image and funding, department curriculum, and administrator and educator careers; even graduation or promotion to the next grade level is based upon the results of a single test. Organizations such as the National Research Council, the American Educational
Research Association, the American Psychological Association, and even companies that manufacture and sell the exams purport to be against this practice. Unfortunately, "just such high-stakes testing is currently taking place, or scheduled to be introduced soon, in more than half the states" (Kohn "Victims" 2).

The resulting "teaching to the test" that creates the environment of high-stakes testing significantly narrows the curriculum. This narrowing ultimately hinders the educational process and opportunity for growth of all students, especially in reading and math. A narrowed curriculum may allow American schools to look good on paper (if that paper is a Scantron form given to an NCLB official), but it certainly does not serve to progress the academic growth of students by challenging them to express themselves in a meaningful and creative way. Despite this curriculum pressure inflicted upon educators by high-stakes tests, administrators and educators need to review and enforce existing curriculum regulations that ensure all students access to a wide variety of subjects.

Gordon Cawelti, Senior Research Associate with the Center on Innovation and Improvement and former Executive Director of ASCD, claims, "Any balanced curriculum, regardless of its approach, should highlight the interconnectedness of various fields of knowledge, expose students to a wide variety of experiences that can help them clarify their interests and talents, and incorporate appropriate ongoing assessments to gauge student mastery" (Cawelti 67). Various educational researchers and curriculum experts have developed a number of approaches to ensure a narrowed curriculum is not the unfortunate, and common, result in classrooms across America.

One such approach, developed by Broudy, Smith, and Burnett before the push for high-stakes testing (1964), is entitled the Broad Fields Approach. In this, traditional
subjects are used to place focus on higher-order thinking skills, ensuring a much more advanced, well-rounded curriculum. The five categories upon which students will focus include symbolics of information (featuring the subjects of English, foreign language, and mathematics); basic sciences; studies of cultures and social institutions; exemplars in art, music, drama, and literature; and large-scale social problems (Cawelti).

Another curriculum-enhancing approach developed later by E.D. Hirsch (1987) is the Crucial Issues Approach. Its purpose is to prepare students to efficiently handle important issues that threaten their well-being and also to care about crucial issues, such as the democratic process; competing economic systems; population control, health and famine, and family issues; conflicting cultures and religions; and deteriorating environmental conditions (Cawelti).

Throughout college, future educators are taught the importance of providing students with the quickest possible feedback of students’ work to ensure that they will be recognizant of the material and will thus have the greatest opportunity to grow from their mistakes. Unfortunately, high-stakes tests do not follow this motto, for students must wait a great length of time before they can view their scores on the test. “The lengthy turnaround time which is involved with regard to scoring most standardized tests makes them nearly useless for helping a particular individual” (Neill 9). At this point, when students finally do receive their scores, it is highly unlikely that they will remember any of the questions, which they would need to do for their scores to be meaningful. However, students are not allowed to view any of the questions. Rather, students are given a sheet of categorized scores, viewing how well they do in specific subjects, but not which questions they missed.
Not only is the process of these delayed scoring methods criticized, but the way in which student responses are assessed is also questionable. Administrators should not rely solely on the summative assessment of tests to assess whether or not students are meeting AYP and growing academically. Rather, these assessments should be mixed with the formative assessment of ongoing classroom-based evidence of student learning, such as the use of portfolios displaying growth in student writing over the course of a semester, or a comparison of evaluative rubrics, used by both the student and the teacher, to evaluate a student’s performance on day-to-day work and classroom involvement.

Many educational specialists agree that multiple assessments “can make a powerful contribution to student achievement and provide useful information for school improvement when combined with other information, such as grade-progression rates, teacher quality, and school climate” (Jehlen “Don’t Be Left” 19). As noted earlier, higher-order thinking skills should make up a large portion of the questions on high-stakes tests. However, for those who want a quick, cut-and-dry measurement of student work, open-ended, critical thinking questions are not the answer.

If students are evaluated using multiple assessments, their strengths will become as apparent as their weaknesses, something the one-size-fits-all assessment of high-stakes tests does not accomplish. These other measures of assessment could include district-level or school assessments and performance assessments, which can provide “richer, more consistent information that enhances validity, diagnostic capacity, and the ability to assess progress” (Neill 10). However, the use of multiple assessment should not only be used for students; schools should also be measured based on multiple forms of achievement when officials determine whether or not the school has met AYP.
Alternatives to High-Stakes Testing

With the many suggested improvements to high-stakes testing come many alternatives as well. Amidst the furor over high-stakes testing, many alternatives have been created and suggested. Some are partial alternatives, not completely replacing the test, while others are completely new options. The term “options” may be rather ironic, considering the one-size-fits-all test students would be mandated to take. However, they are options that allow students to truly show both their strong and weak areas. Is this not what the No Child Left Behind (NCLB) act wishes to assess? As NCLB continues to dominate America’s schools, driven by the need for standard-driven accountability, perhaps more and more alternatives should become available.

One such consideration educational researchers have suggested is the use of portfolios. Used as part of a multiple, formative assessment, portfolios would allow others to view the growth a student has made through a creative outlet. Most important, in these portfolios students could take ownership of what they are accomplishing rather than apathetically responding to lower-order questions by filling in circles with a pencil. Assessment does not have to be such a torturous experience and, although there is definitely a time and place for tests, other evaluative options must be given a position in the process of high-stakes assessment as well.

Not only would portfolios allow for multiple and formative assessments, they also are much more appropriate for and beneficial to ESL (English as a Second Language) students. They “demonstrate growth over time in content areas and English-language proficiency” (“TESOL’s Position” 12). ESL students are placed at a great disadvantage
when taking high-stakes tests. Students are first placed under an immense amount of pressure when made aware of the tests' connection to NCLB and thus the success of the school. Then they are given a time limit in which to answer all questions. Any student would suffer some degree of test anxiety upon hearing these things. But for the ESL students, this is beyond difficult. Imagine having to read question after question and respond promptly, doing all in a language you are not entirely familiar with. Yes, students receive instruction to prepare for this, but no preparation can replace the benefits of receiving an education and being tested in a language one has been speaking her entire life.

Another possible alternative and improvement is a test that provides immediate feedback. The fact that students must wait for a significant length of time before receiving any evaluative feedback has been a prominent criticism of high-stakes testing. A relatively new test format called the Immediate Feedback Assessment Technique (IFAT) would allow students to immediately see if they have answered each question correctly as they take the test. Studies have shown that, when students are given immediate feedback from their tests, the level of their test anxiety is likely to decrease (Black). Immediately showing students if they have answered incorrectly would also give students the opportunity to learn more than they would if they were to take the test from a traditional test form. However, before implementing it in the classroom, it is important to remember that certain students may struggle with this format and prefer something a bit more traditional.

Psychologist Michael Epstein assessed IFAT's contribution to learning in multiple experiments. He and his colleagues “found that participants consistently learned
more when they used the IFAT rather than a traditional answer form that provided no feedback, such as the more widely used, computer-scored, fill-in-the-blank Scantron form" (DiBattista and Gosse 313). A second advantage of the IFAT testing system is that students readily accept its use and this promotes learning.

“Overall, students’ reactions to the IFAT are positive across students with a broad range of characteristics, suggesting that its use may contribute to the important but usually neglected goal of creating a more positive reaction to testing among students” (DiBattista and Gosse 313). Is this possible? Can testing really hold positive connotations for students? If tests are seen as an opportunity to grow rather than an opportunity to be punished, students might drag their feet less on that fated day of the question-and-answer period.

One disadvantage of the IFAT testing system is that students suffering from severe test anxiety may receive more negative immediate feedback, thereby intensifying feelings of anxiety and impairing test performance. Overall, it is important to remember that no quick remedy exists to fix test anxiety. With testing, anxiety is never far behind and, to an extent, is a positive emotion to feel, for it facilitates growth. Often, the causes are unknown and are individually specific to each student. Therefore, no one method or form of assessment can determine or help solve test anxiety. Perhaps the most important thing for teachers do, regarding test anxiety in their classrooms, is realize that the most important part of education is learning itself and, because of this, other forms of activities and assessment should be implemented as well.

Like the computerized IFAT, which allows students to immediately view their incorrect and correct answers, other computerized testing software has sparked interest in
recent years. Programs such as Criterion and Vantage are used to rate millions of essays per year in thousands of districts (Jehlen, "Testing How the Sausage is Made"). However, “their main use is not for making high-stakes decisions, but to help teachers teach” (Jehlen, “Testing How the Sausage is Made” 33). However, while this program is primarily used as a teacher aide, its essay-rating capabilities are sometimes used to score high-stakes tests. The Vantage program evaluates and scores the West Virginia seventh-grade writing assessment on its high-stakes test. It is also used to assess the essay portion of the Graduate Management Admission Test (GMAT) used by business schools. However, in both of these cases, a human scorer is also used.

So, the question is, should computers be given jobs and responsibilities consistently given to humans in the past? Vantage argues its program is more accurate than a person. “When a group of expert human scorers and the Vantage program score the same essays, Vantage consistently comes closer to the group’s average score than any human” (Jehlen, “Testing How the Sausage is Made” 34). Technology is hourly becoming more and more advanced, simplifying the lives of people all over the world. Should this progressive, fast-paced world be given a part in the seemingly archaic world of high-stakes testing? Perhaps a breath of fresh air would do the musty archives which house the Scantron pages some good. Indeed, the students would be open to this new change.

Discussion of Findings

Through the process of researching for “Improvements and Alternatives to High-Stakes Testing,” I examined multiple resources from differing viewpoints. I chose the
topic for this thesis because high-stakes testing is so predominant in the education field today, something, as an educator, I will be faced with daily and reminded of as I plan my instruction and meet with administrators and colleagues to discuss curriculum. I often ask myself the question, “Will high-stakes testing continue? Will NCLB affect me years from now as an educator?” These are questions, I am sure, all educators ask themselves, questions I wanted to attempt to come to some conclusions about through my research. And a number of conclusions I did reach.

First of all, examining the advantages and disadvantages of testing was rather eye-opening for me. Because I have always heard about the negative consequences of the tests, I was surprised to find research and studies that showed some of the advantages of high-stakes tests. For example, I was surprised to find that, in a recent survey of 1,023 parents of school-age children, 83% felt tests provided important information about their children’s educational progress (Cizek). I had always thought the majority of parents were against high-stakes tests due to the strain in places on the students and the classroom environment.

Most amazing to me is the large amount of research on both sides of the spectrum to either promote or argue against high-stakes testing. I suppose such a highly controversial topic will always have an abundant amount written about it; like moths to a bright light come the researchers and critics to the topic of high-stakes testing. I am grateful that such a large amount of research existed to represent both sides of the story because it really opened my eyes to opinions I had harbored before beginning my research.
Another very important result of this research were the comparisons and conclusions I derived from writing about the advantages and disadvantages of high-stakes tests. Although the process did not cause me to be in favor of these tests, it did open my eyes, allowing me to see that they and their developers are not the complete embodiment of evil and destruction to education that I initially thought them to be. At a time when little academic improvement was “seen” in schools by college professors and business owners, the first Bush administration determined a new system of accountability focused on national goals, national standards, and a national test to assess the progress schools made toward achieving those standards (Jennings). Unfortunately, in response to a desperate attempt to please the American public (at least a portion), quick decisions were made that looked good on paper. What could be better than easily measurable data that “matched” the standards?

A fourth conclusion I came to through my research was that the decision to implement high-stakes testing was a hasty one, not clearly thought through, an attempt to correct the perceived shortcomings of the education system of today. Instead, to correct the problems of high-stakes tests, a committee of educators, not a Presidential Advisory Committee made of business executives, would need to go back to ground zero, completely revamping not only the format of the tests, but their overall purpose. Tests should be made to assess growth, but as any educator knows, they are not and should not be the end-all be-all of assessments.

Yet another conclusion I have reached is quite simple: multiple improvements need to be made to high-stakes tests. As noted above, their initial purpose needs to be reconsidered. Yes, they should serve as somewhat of an equalizer of academic standards
from state to state, but punishment should not be inflicted if these standards are not met.

If a student fails an exam, do you take away his resources to help him prepare for the next exam? No. You work with him, helping him, encouraging him to do his best and providing him with all of the help you find within you. Educators know this; educators are told this daily in their methods courses. So what happens when they enter the workforce? Suddenly, their curriculum and goals are controlled by those who believe punishment will enforce academic growth, going against all that they have ever known.

Finally, from my research I have gathered that educators and curriculum experts are continuously developing and suggesting improvements and alternatives; they are not simply calmly sitting at their desks while the education of American students is turned upside down. Using an additional form of assessment such as a portfolio as recommended by TESOL ("TESOL’s Position") and other educational experts would greatly increase students’ ability to showcase their strengths rather than squeeze all that they know into the tiny bubbles of the Scantron sheets. Other alternatives, such as the Immediate Feedback Assessment Technique (IFAT), are being developed to aid those students who suffer from severe test anxiety. I feel this is something that would be an advantageous approach to take; however, as noted previously, the immediate feedback does tend to create even more tension in some students. Perhaps, as is done with many tests such as the Praxis I teacher licensure test, students taking high-stakes exams should be presented with the option of either taking a computerized, immediate feedback test or a regular paper test.

When considering the numerous academic sources I have consulted, comparing and contrasting the points set forth in each, I have surmised that, overall, while high-stakes
testing may not be something that will very soon, if ever, take a bow and exit the stage of No Child Left Behind. I do feel the implementation of various improvements and possible supplements or alternatives could make the present torturous drudgery for students and teachers alike much more tolerable, perhaps even academically beneficial.

As a future educator, I know I will soon enter the game of high-stakes charades, and I want to be well-prepared to not only prepare the students and myself as much as possible, but also to develop my curriculum in such a way that it does not cater to the almighty test. Rather, I want my curriculum to provide numerous outlets for creativity, active inquiry, and academic growth while still aligning with academic standards. All in all, I think students, administrators, and educators need to remember one thing above all: simply because it is a test does not mean it is the only, nor the best, way to meet and assess academic standards.
Works Cited

2007.

wSDF19Wi&p_lva=6&p_li=&p_faqid=6&p_created=1095256734&p_sp=cF9zc
mNoPSZwX3NvcnRfYnk9JnBfZ3JpZHNvcnQ9JnBfcm93X2NudD0xMTUmcF9
wcm9kczoFc9jYXRePSZwX3B2PSZwX2N2PSZwX3BhZ2U9MQ**>


Kohn, Alfie. “Standardized Testing and Its Victims.” Education Week 27 September
2000.


